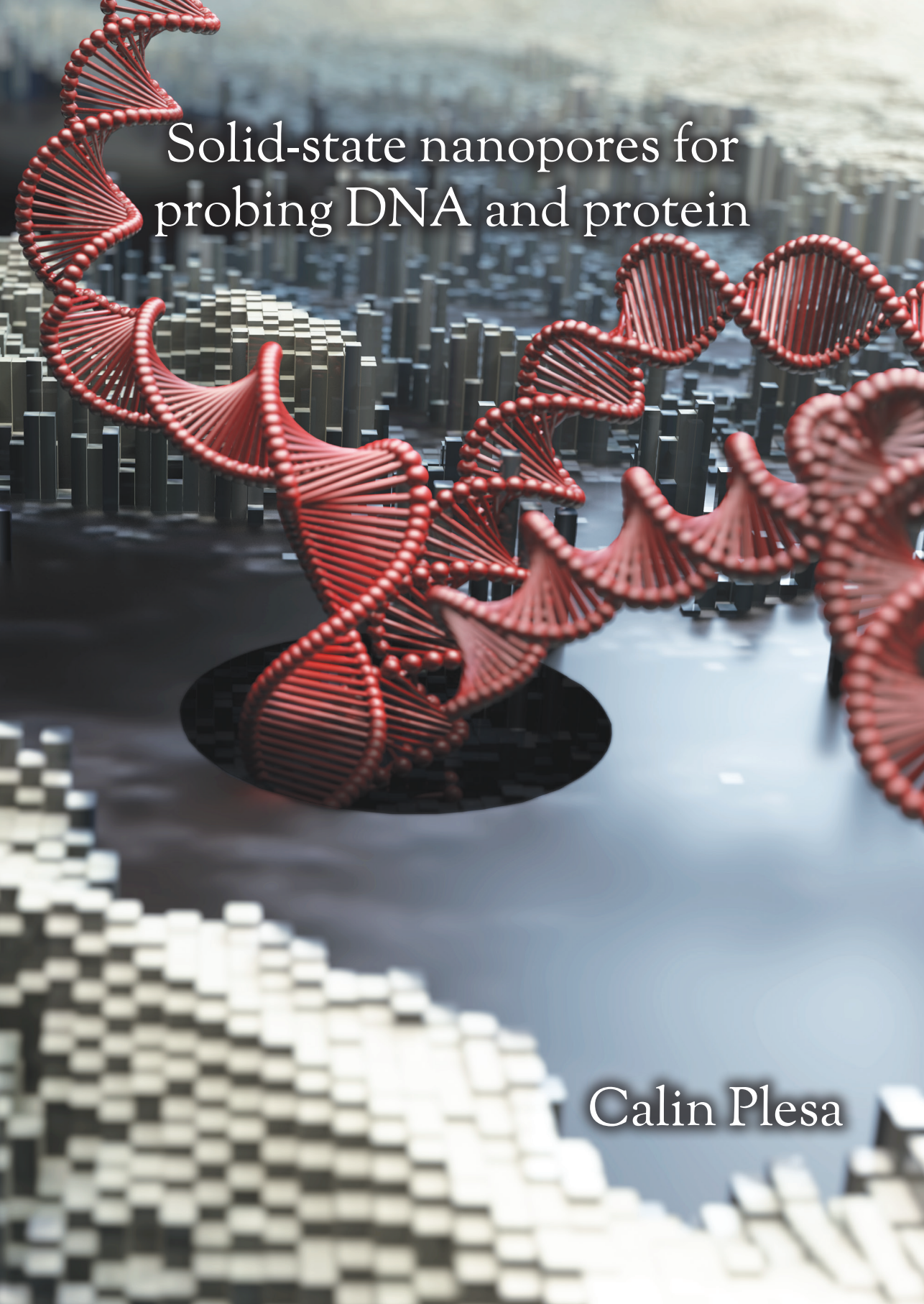
Solid-state nanopores for probing DNA and protein



Calin Plesa

Solid-state nanopores for probing DNA and protein

Solid-state nanopores for probing DNA and protein

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K. C. A. M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op vrijdag 30 januari 2015 om 10:00 uur

door

Calin PLESA

Master of Science in Nanoscience
geboren te Boekarest, Roemenië.

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. C. Dekker

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. C. Dekker	Technische Universiteit Delft, promotor
Prof. dr. H. W. Zandbergen	Technische Universiteit Delft
Prof. dr. N. H. Dekker	Technische Universiteit Delft
Prof. dr. H. Dietz	Technische Universität München
Prof. dr. U. F. Keyser	University of Cambridge
dr. G. Maglia	Katholieke Universiteit Leuven
dr. C. Storm	Technische Universiteit Eindhoven
Prof. dr. M. Dogterom	Technische Universiteit Delft, reservelid

Keywords: nanopores, DNA, protein, knots, DNA origami, translocation

Printed by: Gildeprint

Front & Back: Erik Major

Casimir PhD Series, Delft-Leiden 2014-36

ISBN 978-90-8593-208-6

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Contents

1
Introduction

In this introductory chapter, I provide a brief background context for the topics within this thesis, starting with the discoveries that created the field of molecular biology. At the heart of this field lie several polymers, DNA, RNA, and polypeptides which store information, regulate processes, and catalyse reactions. The central dogma describes the flow of information among these three and establishes the basic processes behind life. Nanotechnology is a relatively new field of research which focuses on manipulating matter on scales below 100 nm. This field has increasingly focused on studying biology using a number of tools including nanopore sensing, which is the subject of this thesis. This technique uses a nanoscale hole in a thin membrane to probe biomolecules such as DNA, RNA, and polypeptides, and brings together aspects from both molecular biology and nanotechnology. I briefly describe the history and the basic approach, and I outline the primary challenges in this field. At the end of this chapter, I provide an outline of this thesis.

1

1.1. Introduction

Molecular biology has revolutionized our view of life. This field was created by a confluence of different discoveries in genetics and chemistry, starting 150 years ago when Gregor Mendel first proposed his laws of biological inheritance to describe how biological traits were passed from parents to their offspring. Each biological trait is stored as a unit of information which was subsequently termed a *gene* by Hugo de Vries and Wilhelm Johannsen. A critical step linking genetics and chemistry came when Thomas Hunt Morgan showed that these genes could be physically found on chromosomes inside cells. Chromosomes are large structures which can be seen inside the nucleus of cells when stained with certain dyes, as shown in Figure 1.1. Several experiments then established that chromosomes were made from DNA and proteins and that DNA was the carrier of genetic information within cells. These discoveries gave birth to molecular biology. After the 1950s, this field experienced phenomenal growth and was further driven by the major medical treatments which were enabled by this new knowledge.

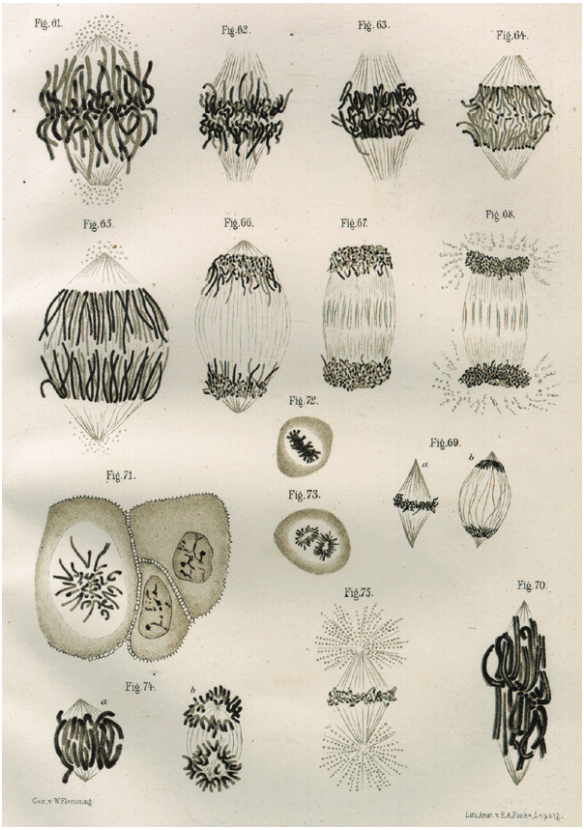


Figure 1.1: An illustration of mitosis by Walther Flemming in 1882[1]. Identical pairs of chromosomes are pulled to opposite side of the cell in preparation for cell division.

1.2. Molecular biology

1.2.1. Deoxyribonucleic acid (DNA)

A critical point in our present day understanding of the molecular mechanisms behind life was the discovery of the structure of DNA, shown in Figure 1.2, by Watson and Crick in 1953[2] using the x-ray diffraction images of Rosalind Franklin. Biochemically, it had been known that DNA was composed of four bases (adenine, guanine, cytosine, thymine), a sugar group, and a phosphate backbone. Watson and Crick discovered a double helix structure of DNA which has two intertwined strands with paired bases on the interior. Bases on one strand are complementary to bases on the other strand, such that the biological information is encoded into each strand and can be easily replicated. Each adenine base pairs with thymine through two hydrogen bonds, while cytosine pairs with guanine through three hydrogen bonds.

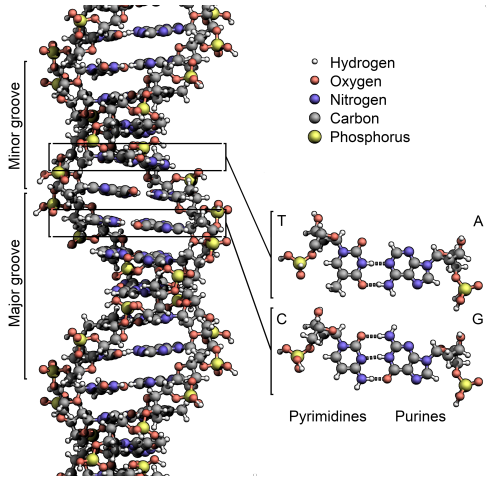


Figure 1.2: A schematic illustration of the structure of a DNA molecule (credit: Richard Wheeler).

1.2.2. Central dogma

After the structure of DNA was discovered, Francis Crick proposed[3] the 'central dogma in molecular biology', which states that information is transferred in-between nucleic acids and from nucleic acids to protein, but never from protein to nucleic acids, as shown in Figure 1.3. In general the sequence of the bases in DNA is the primary long-term information storage medium in the cell. DNA is *replicated* such that each of the daughter cells has the same DNA sequence after dividing. DNA is also *transcribed* into ribonucleic acid (RNA) which acts as both a medium for short-term information storage as well as having some enzymatic capabilities which can catalyse some biochemical reactions. Although RNA can adopt a two-strand double-helix structure similar to DNA, it is often found in single-stranded form inside the cell. Its high flexibility allows it to easily pair with itself, forming complex secondary

1

structures that are defined by its base sequence. The information within some RNA molecules is *translated* into protein. Proteins are long chains of amino acids which fold into well-defined structures based on their sequence and are responsible for most of the enzymatic activity within the cell. Proteins called polymerases replicate DNA and transcribe DNA into RNA. A large complex called the ribosome, made primarily from RNA, is responsible for translating the nucleotide sequence into an amino acid sequence. Since there are twenty amino acids and only four DNA bases, each amino acid is encoded within DNA as a sequence of three bases called a codon. Different codons direct the ribosome to either **i)** start protein sequence, **ii)** add a certain amino acid to the protein chain, **iii)** or stop the protein synthesis. In addition to enzymatic activity, proteins also help to regulate transcription and package DNA, among many other functions. This complex list of roles allows life to emerge from its molecular constituents.

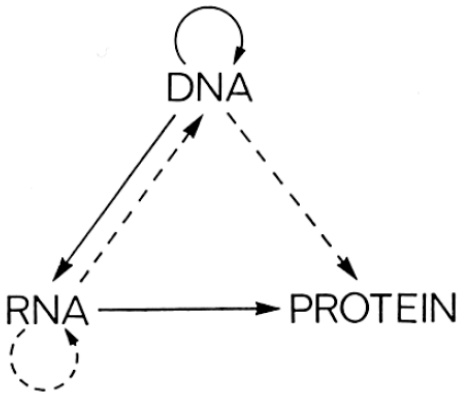


Figure 1.3: The central dogma in molecular biology states that information only flows from nucleic acids to protein and not in reverse as shown in this figure from Francis Crick's 1970 publication[4]. Solid lines denote the major transfers while dashed lines represent special transfers which only occur in certain circumstances.

1.3. Nanotechnology

Nanotechnology is our ability to manipulate single atoms and molecules on the nanometer scale. It is difficult for the human mind to comprehend just how small a nanometer is. To get an impression, think of the map of Europe on a 10 euro cent coin. Now imagine you could zoom in to this map on the surface of the coin and see very fine details. You zoom in to the country of the Netherlands, to the city of Delft, to a building on the TUDelft campus, and to an office in the building. In this scenario the size of a typical computer monitor would correspond to about one nanometer in size on the face of the coin, which is equivalent to the length of about four copper atoms in a row or the radius of a DNA molecule.

In addition to being incomprehensibly small, the nanoscale world is unlike the one we are used to. At the nanoscale, the force of gravity is insignificant while elec-

tromagnetic (van der Waals) forces dominate. Furthermore, quantum and surface effects become increasingly important. Liquids behave like sluggish molasses as objects move at low Reynolds numbers. All of this leads to a number of interesting and often counter-intuitive phenomena.

1.3.1. Tools for nanotechnology

The capacity to understand nanoscale phenomena and engineer nano-devices is dependent on our ability to observe and manipulate matter at this scale. Techniques capable of imaging at small scales have always been far ahead of our ability to manipulate at these scales. Light-based microscopy, first developed in the seventeenth century, enabled many of the major discoveries in microbiology. This technique, which is based on visible light, has a diffraction-limited resolution of about 200 nm. This resolution limit was surpassed by electron microscopes in the twentieth century, with modern transmission electron microscopes (TEM) capable of achieving resolutions below 0.1 nm. This was accomplished by replacing light with an electron beam in vacuum.

A more recent branch of microscopy is scanning probe microscopy. In this approach an atomically sharp tip is scanned over a surface to determine its topology and other properties. In its first incarnation, named scanning tunnelling microscopy (STM), a tunnelling current was measured between a metallic tip and a conductive surface. This allowed lateral resolutions of 0.1 nm and even higher resolution in height. Subsequently the requirement of a conductive surface was removed with the development of atomic force microscopy (AFM), which is used in several chapters of this thesis. In this approach, a variety of forces (most often van der Waals forces) between the tip and surface are measured as a function of the distance between them. When a cantilever, with a tip on the free end, is brought close to the surface, the interaction between the tip and the surface leads to a deflection in the cantilever. This deflection is amplified and measured using an optical lever based on a laser reflected from the back of the cantilever onto a set of photodiodes, as shown in Figure 1.4. This provides a simple and powerful way to image nanostructures on top of a flat surface, such as mica.

Although scanning probe methods are capable of some limited manipulation on the nanoscale, the bulk of our current nanofabrication abilities are based on lithography. In its most simple form, this approach is akin to making an old-fashioned photograph. First a photosensitive material, termed a resist, is exposed to light or electrons with an exposure pattern designed to make the desired structure, as shown in Figure 1.5. By focusing the pattern through optical elements, the exposed pattern can be made much smaller than the mask used to define the pattern. Subsequently the resist is developed, a process which removes either the exposed or unexposed areas of the resist material depending on the requirement. This developed resist often acts as a subsequent mask for another type of material underneath, which is selectively etched away in the regions where it is not covered in resist. Finally, the resist is removed and we are left with a planar nanostructure. Improvements to this basic approach have fuelled the increases in computational power seen since the invention of the integrated circuit. Lithography processes are

1

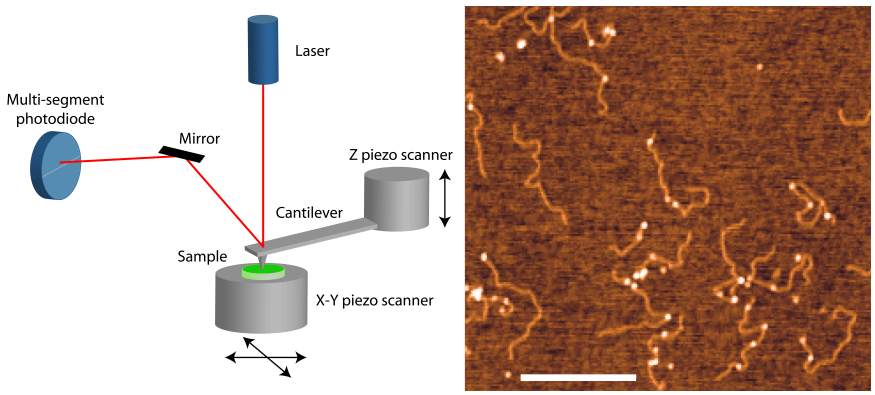


Figure 1.4: Left: Schematic illustration of an atomic force microscope. Right: AFM scan of some 2.2 kbp DNA molecules with bound protein. The scale bar is 500 nm. (Credit: Calin Plesa)

typically referred to as top-down techniques since they start with a large piece of material and remove material to form increasingly smaller structures.

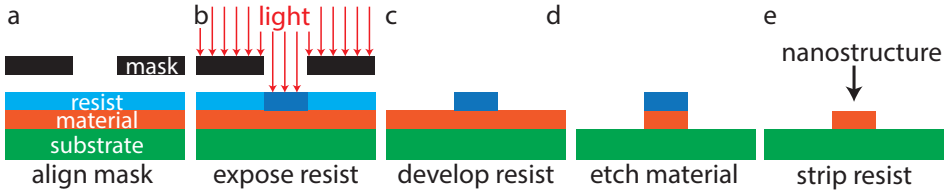


Figure 1.5: Schematic side-view illustration of a photolithography process.

1.3.2. DNA origami

An alternative to the top-down approaches are bottom-up techniques were small parts are combined to make a larger structure, much like Lego. Since it would be prohibitively time consuming to assemble nanostructures by adding pieces one by one, most bottom-up techniques rely on self-assembly. The nature of the complementary interactions in nucleic acids makes them ideal building materials for this approach. Ned Seeman first outlined many of the principles necessary to create structures using DNA in the 1980s[5] in an attempt to improve crystallography techniques. More recently, a technique called DNA origami has been introduced which uses many short single-stranded DNA oligonucleotides (staples) to fold a much longer single-stranded DNA scaffold[6]. Through the sequence design of the staples, this approach can be used to create well-defined two-dimensional and three-dimensional nanostructures with high yield. DNA origami is used in two of the chapters in this thesis to create designed nanostructures.

1.4. Nanopores

A nanopore can be broadly defined as a small hole with diameter < 100 nm in a thin membrane with a thickness < 100 nm. Nanopores are versatile nanoscale sensors capable of probing a wide variety of molecules and phenomena. Many of the principles underlying nanopores can be traced back to the development of the Coulter counter by Wallace H Coulter starting in the 1950s[7]. In this approach, cells in a solution could be quickly counted and sized by passing them through a small hole between two chambers containing solution, while simultaneously measuring an electrical current through the hole. The passage of a cell through the orifice results in a drop in the current signal proportional to the cell's volume. This technique is now generally known as resistive pulse sensing. Nanopores are scaled down versions of this approach which enable detection of much smaller objects such as individual biomolecules.

The initial breakthroughs in this field came from experiments with α-hemolysin, a biological nanopore produced by *Staphylococcus aureus*. This pore can insert itself into lipid bilayers, resulting in the formation of a 1.5 nm pore in the membrane. Addition of single-stranded DNA molecules to one side of the membrane and application of an electric field results in detectable translocation events caused by the molecules passing through the pore[8]. Much of the progress in this field was pushed by the realization in the 1990s that this approach may be used for low-cost long-read-length sequencing of DNA[9]. In this approach, a DNA strand would translocate through the pore, with each base producing a unique signature in the ionic current, allowing a long linear readout of the molecule's sequence. Follow-up experimental studies revealed that translocation through α-hemolysin could discriminate between different homopolymers[10, 11], taking the first step towards sequencing.

Solid-state nanopores were developed to address some of the limitations of biological nanopores. These include the small fixed size of the pore, the stability of the pore, and the limited range of conditions in which the pore and lipid bilayer are functional. Solid-state nanopores are made by creating a hole in a thin insulating membrane. A number of fabrication techniques exist, but the most popular methods use either a focused beam of electrons (TEM) or ions (FIB) to create the nanopore[13, 14]. Silicon nitride (SiN) has been the material of choice for the membrane, although pores have also been created in SiO_2, Al_2O_3, HfO_2, graphene, MoS_2, and other materials. Pores created with these approaches can be made with almost any diameter. The membrane themselves are made with the photolithography processes described earlier, which create a free-standing membrane in the center of a silicon chip.

After fabrication, the membrane is placed into a flowcell such that the nanopore is the only ionic connection between two aqueous chambers. These are filled with a high-concentration salt solution, typically 1M KCl. If an electric field is applied across the membrane using silver/silver-chloride electrodes, a steady ionic current can be measured corresponding to the passage of ions through the open pore. When DNA is added to one of the solutions, transient blockades are observed in the ionic current due to the passage of DNA molecules through the pore (Figure 1.6). These events

1

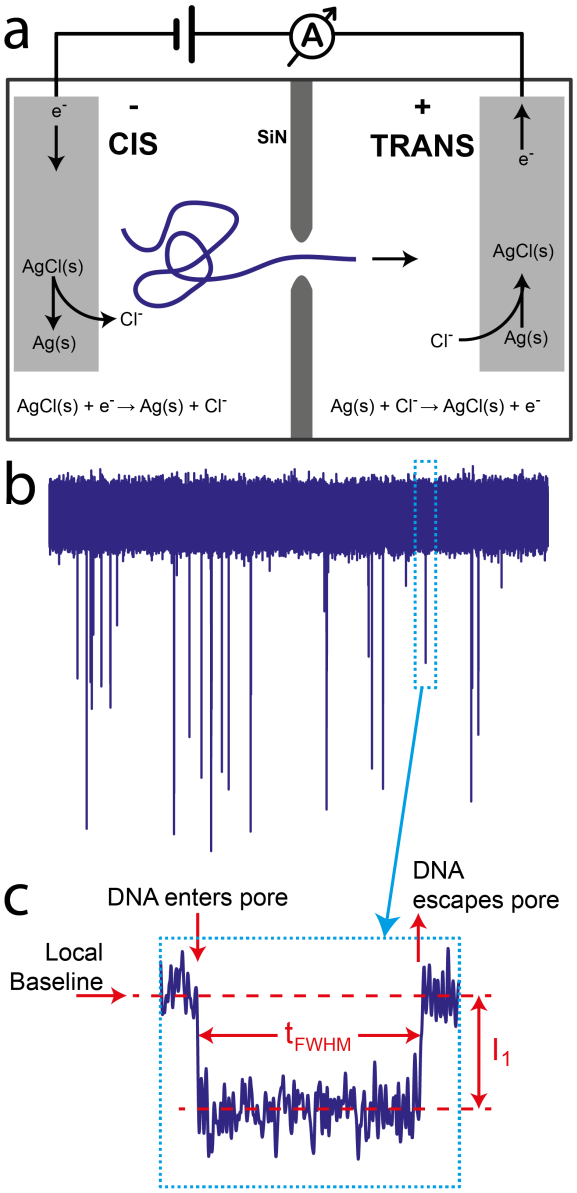


Figure 1.6: a) Schematic side-view illustration of DNA translocating through a nanopore. b) Typical ionic current trace showing DNA molecules translocating through a nanopore. c) A close-up of one DNA translocation event[12].

are only observed in one polarity, were the electrophoretic force on the negatively charged DNA causes it to translocate through the pore. These blocakde events are digitized and recorded. Their amplitude carries information about the volume of the molecule, while the duration can be related to the molecule's length. This is the basic approach used in this thesis, aspects of which are further described in detail in each subsequent chapter.

1.5. Why study nanopore transport?

The field of bionanoscience lies at the intersection of physics, biology, chemistry, and engineering. Broadly speaking, it aims to use the tools provided by nanotechnology to study biologically relevant phenomena at their native length scales. Solid-state nanopores are a perfect tool for this purpose since they can probe at length scales 1 - 100 nm in various aqueous media. Furthermore, biological nanopores themselves, are universally found in living organisms, where they are used to control transport both within cells and across the cell boundary. Given their importance, understanding transport through nanopores is an important goal in biology, but this represents only one of multiple aspects which motivate this work, as described next.

DNA sequencing was the initial application envisioned for nanopores and is just as relevant today as when it was first proposed. Recent developments in biological nanopores have demonstrated that they can be used in conjunction with enzymes to sequence DNA, with the first commercial product already on the market. Solid-state nanopores could provide many advantages, but a number of major issues still remain. The DNA molecules translocate through wide pores far too quickly. New methods to slow down the translocation velocity are required. Additionally, solid-state nanopores still suffer from much higher noise levels compared to biological nanopores. Using thick layers of high dielectric constant materials is one approach to improve this situation. Finally, the resolution of the technique must be improved. A number of possible approaches are being investigated including the use of thinner membranes such as graphene and MoS_2 as well as different non-ionic sensing modalities such as tunnelling current and plasmonics.

Biosensing represents a major potential application for nanopores, where they could be used to detect medically relevant biomarkers or study DNA-protein complexes. In addition to the issues already pointed out, an open question in this area is how to carry out high-resolution measurements in physiological conditions, where the ionic current provides a much lower signal.

From a more fundamental level, much remains unknown about the translocation process and its dynamics. We currently do not understand how the polymer blob outside the pore unfolds and how this affects the translocation process. Typical experiments show a wide spread of translocation times for a single length polymer and it has been hypothesized that the width of this distribution could be reduced by controlling the initial conformation of the molecule at the start of the translocation process. Finally, we have only just begun exploring how knotting phenomena and non-equilibrium behavior affect the translocation process.

An important development in solid-state nanopores would be the ability to control the geometry and chemistry of the pore itself. This would reduce variability

and allow biomolecules to be attached inside or close to the sensing region. DNA origami nanostructures, discussed in Chapter 5, could provide a solution.

1.6. This thesis

This thesis describes a number of fundamental studies and techniques using solid-state nanopores that address some of the topics described in section 1.5.

In **Chapter 2** we investigate what happens if a single individual molecule is recaptured many times through a nanopore by quickly switching the polarity of the electric field. We use this approach to study the non-equilibrium folding of different DNA molecules after translocation, which is linked to the molecule's characteristic Zimm relaxation time.

Chapter 3 focuses on the translocation dynamics of a DNA molecule translocating through a nanopore. A synthetic DNA-origami construct is used to determine the translocation velocity of different parts of a DNA molecule. We observe significant intramolecular, intermolecular, and pore-to-pore velocity fluctuations. This knowledge is used to estimate the uncertainty in determining the spatial position of DNA bound structures.

Knots are an ever-present phenomenon in long linear polymers such as DNA. In **Chapter 4**, we use solid-state nanopores to detect knots in long DNA molecules. The knotting occurrence is determined for different length DNA molecules and the size of the knots is observed to be very small. Furthermore, we see that knots are capable of slipping out of linear molecules during the translocation process.

DNA origami nanostructures can significantly enhance the abilities of solid-state nanopores. In **Chapter 5** we provide a fundamental characterization of DNA nanostructures docked onto solid-state nanopores. We observe large ionic permeabilities for different nanostructure designs, mechanical bending, and mechanical buckling.

Many proposed biosensor applications require the detection of proteins translocating through a nanopore. In **Chapter 6**, we show that detection of proteins is extremely difficult due to the resolution limitations of the system. We conclude that the majority of proteins translocate faster than can be observed.

In **Chapter 7** we demonstrate that it is possible to detect single proteins bound to a DNA strand in typical experimental conditions. A new model system based on anti-DNA antibodies is introduced and characterized.

Previous studies with nanopores all use silver/silver-chloride based electrochemistry. In **Chapter 8** we introduce a new set of electrochemical reactions based on silver/silver-glutamate. We investigate the use of high-concentration glutamate solutions as redox-capable thickening agents and demonstrate the translocation of DNA using this approach.

Data analysis in an important part of each nanopore experiment. In **Chapter 9** we describe a number of new techniques and algorithms for the analysis of nanopore data.

References

[1] W. Flemming, *Zellsubstanz, kern und zelltheilung* (Vogel, 1882).

[2] J. D. Watson and F. H. C. Crick, *Molecular structure of nucleic acids,* Nature **171**, 737 (1953).

[3] F. H. C. Crick, *The biological replication of macromolecules,* in *Symp. Soc. Exp. Biol*, Vol. 12 (1958) pp. 138–163.

[4] F. H. C. Crick, *Central dogma of molecular biology,* Nature **227**, 561 (1970).

[5] N. C. Seeman, *Nucleic acid junctions and lattices,* Journal of Theoretical Biology **99**, 237 (1982).

[6] P. W. K. Rothemund, *Folding DNA to create nanoscale shapes and patterns,* Nature **440**, 297 (2006).

[7] W. H. Coulter, *Means for counting particles suspended in a fluid,* (1953), USPTO Patent 2656508.

[8] J. J. Kasianowicz, E. Brandin, D. Branton, and D. W. Deamer, *Characterization of individual polynucleotide molecules using a membrane channel,* Proceedings of the National Academy of Sciences **93**, 13770 (1996).

[9] D. W. Deamer and M. Akeson, *Nanopores and nucleic acids: prospects for ultrarapid sequencing,* Trends in Biotechnology **18**, 147 (2000).

[10] M. Akeson, D. Branton, J. J. Kasianowicz, E. Brandin, and D. W. Deamer, *Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules,* Biophysical Journal **77**, 3227 (1999).

[11] A. Meller, L. Nivon, E. Brandin, J. Golovchenko, and D. Branton, *Rapid nanopore discrimination between single polynucleotide molecules,* Proceedings of the National Academy of Sciences **97**, 1079 (2000).

[12] C. Plesa and C. Dekker, *Data analysis methods for solid-state nanopores; submitted,* (2015).

[13] J. Li, D. Stein, C. McMullan, D. Branton, M. J. Aziz, and J. A. Golovchenko, *Ion-beam sculpting at nanometre length scales,* Nature **412**, 166 (2001).

[14] A. J. Storm, J. H. Chen, X. S. Ling, H. W. Zandbergen, and C. Dekker, *Fabrication of solid-state nanopores with single-nanometre precision,* Nat Mater **2**, 537 (2003).

Part I

DNA

2

Non–equilibrium folding of individual DNA molecules recaptured up to 1000 times in a solid state nanopore

We investigate translocation of linear and circular double–stranded DNA molecules through solid state nanopores where each molecule is recaptured and re–translocated many times. Single molecules can be recaptured by switching voltage polarity for hundreds or even thousands of times. The large number of recapture events allows statistics on the translocation of individual molecules. Surprisingly, we observe that recaptured DNA molecules do not translocate in a linear head–to–tail fashion, but instead translocate as a folded blob where multiple parts of the DNA molecule simultaneously translocate through the pore in parallel. This folding is observed through the presence of up to 13 DNA double strands from the same molecule simultaneously inside the pore, as well as many smaller fold numbers occurring during the course of a translocation event. The strong folding is particularly prominent when the molecule is recaptured at short time scales, i.e. shorter than its characteristic time to relax to its equilibrium configuration. At longer recapture times, both the amount of folding and the mean duration of translocation approach the values observed in non–recapture experiments. The data shows that the translocation time of a molecule depends on the molecule's conformation at the start of the translocation process, with extended molecules having a longer translocation time. The observations can be attributed to a high–density non–equilibrium DNA configuration that arises in the close vicinity of the nanopore immediately after translocation, which dissipates on a time scale given by the Zimm relaxation time.

This chapter has been published as: C. Plesa, L. Cornelissen, M. W. Tuijtel, and C. Dekker, *Non-equilibrium folding of individual DNA molecules recaptured up to 1000 times in a solid state nanopore*, Nanotechnology **24**, 47, 475101 (2013).

2.1. Introduction

Solid-state nanopores are a versatile tool for biophysics with a broad range of applications[1, 2]. In this technique biomolecules are placed into one of two chambers separated by a membrane containing a nanometer-scale pore. Upon application of an electric field across this membrane, charged molecules experience an electrophoretic force pulling them through the pore, as shown in Figure 2.1a. Experimental research using nanopores has been able to provide a number of insights into the relevant polymer physics. These include, for example, the observations that long polymers translocating through large pores have a power-law length dependence of their translocation time[3], a diffusion-limited capture process[4], and a strong preference for end capture of molecules which has been attributed to configurational entropy[5].

Despite this body of work, a number of important open questions remain. For example, (1) how does the local translocation velocity vary as the molecule passes through the pore? Or (2) why are the distributions in the translocation times for one particular length so disperse? The former question is critical for many nanopore applications, for example DNA analysis, where one aims to map the measured temporal signals into the spatial domain, while understanding the latter question is key to determining if the differences observed between two translocation events are due to the physics of the translocation process or actual physical differences between two molecules. Both questions relate to the conformational state and behavior of the polymer coil outside the pore as the translocation occurs.

Storm et al.[3] showed that the observed non-linear dependence of translocation time on the molecule's length could be explained by the hydrodynamic drag of the polymer coil outside the pore balancing the applied electrophoretic force pulling the polymer through. Due to the large polymer coil outside the pore, we expect the velocity at the start of the translocation process to be lower, because the large coil needs to unravel and move towards the pore, compared to the velocity at the end of the translocation process where the coil has shrunk, providing less resistive force and increasing the translocation velocity significantly. Indeed, modeling simulations carried out by Lu et al. predicted this type of behavior[6]. Their work suggests that such behavior should be present in all translocations. Additionally, they attributed the large distributions observed in translocation times (the second question) of equal length molecules to the large variety of molecular configurations possible at the start of the translocation process, with the translocation time increasing as the distance from the center-of-mass of the molecule to the pore entrance (at the start of the translocation process) becomes larger.

In this work we apply a technique which allows us to systematically alter the molecular configuration of a single molecule, switching between the normal relaxed, extended form and a much more condensed conformation. We carry out many repeated measurements on the same molecule in different configurations, in order to observe how this affects the translocation time and other parameters.

In 2007, Gershow and Golovchenko demonstrated that it was possible to recapture a translocating dsDNA molecule by quickly reversing the electric field after a molecule had passed through the nanopore[7], as illustrated in Figure 2.1b. By con-

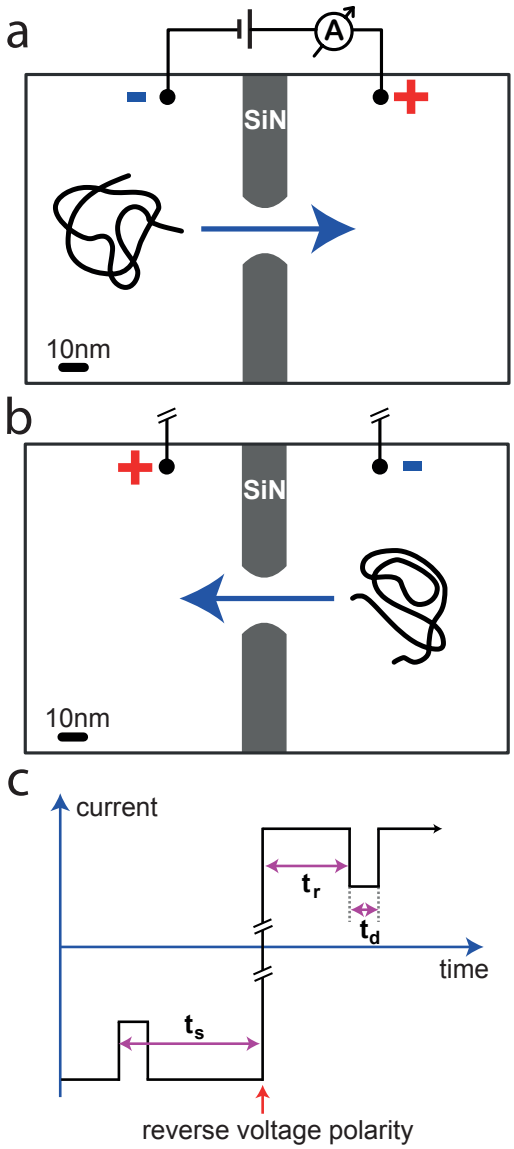


Figure 2.1: a) Schematic representation of the nanopore setup, with two aqueous reservoirs separated by a SiN membrane containing a pore. b) Once the passage of a DNA molecule has been detected, the electric field is switched and the same molecule is recaptured to translocate once more through the pore. c) An idealized current trace showing a recapture event and the various timings.

trolling the duration between a translocation event and the electric field switching time, they showed that the process of capturing the DNA into the nanopore fits a drift-diffusion transport model. This allowed them to probe the capture length, the point beyond which the electrophoretic force could no longer overcome diffusion due to Brownian motion. They were able to recapture 5 kbp DNA molecules up to 22 times.

Here we show that it is possible to obtain much higher numbers of recaptures (hundreds, or even thousands) of long double-strand DNA molecules. This enables us to obtain large-number statistics of the translocation characteristics of individual molecules. Our data shows that both the translocation time and the amount of folds observed are highly dependent on parameters of the recapture process which can be experimentally controlled. This behavior is found to differ if the molecule is in a non-equilibrium state versus in a relaxed state.

2.2. Recapture experiments

We use 20 nm diameter pores in 20 nm thick SiN membranes, fabricated as previously described[8]. After drilling the pores with a TEM, the membrane's surface was manually painted with a layer of PDMS in order to reduce the membrane capacitance, thus lowering the noise and increasing the signal-to-noise ratio[9]. Membranes were mounted into a PMMA flowcell where they separated two aqueous chambers each containing an Ag/AgCl electrode. Double-stranded DNA molecules were added to one reservoir and translocation events were detected as a drop in the nanopore ionic conductance. After detection of a translocation event, the electric field's polarity was reversed, leading to the recapture of the molecule that had just translocated through the nanopore. The nanopore's ionic current was recorded with an Axopatch 200B amplifier and digitized with a National Instruments 6251 DAQ card controlled by a custom LabView script. Recorded current data were filtered with a 10 kHz Gaussian low pass filter and analysis was carried out in Matlab with level fitting done in OpenNanopore[10].

We automated the translocation event detection in order to perform multiple recaptures. Briefly, our script scans the previous 1 ms of 3kHz-filtered current data for events where the current deviates more than 5σ away from the baseline, where σ is the standard deviation of the filtered baseline current. Upon detection of an event, an interval of time (t_s) was allowed to pass before the polarity of the electric field was switched (Figure 2.1c). This switching interval was varied during the course of multiple recaptures, in order to probe the dependence of various parameters on the recapture time. The control script attempted to recapture the molecule indefinitely until the molecule escaped.

We investigated four different types of molecules: linear and circular dsDNA with a contour length of 48.5 kbp or 97 kbp. Lambda dsDNA (48.5 kbp) was purchased from New England Biolabs and diluted to a concentration of 3.7 pM in 1M KCl, 10 mM Tris, 1 mM EDTA at pH 8. The lambda DNA solution was left at room temperature and not heated, resulting in a fraction of molecules in a circular configuration[11], as well as higher order structures such as dimers[12]. In order to ensure that only a single molecule was being studied during the recapture process, we measured

the rate of translocation events at 50 mV applied voltage before attempting any recaptures. The rate of events was determined to be very low: below 0.0014 Hz, or approximately 1 event every 10 minutes.

Using PDMS-coated nanopore membranes, we observed that each time the voltage polarity was reversed, the current baseline required around 2.5 ms to settle. Due to this effect, events recaptured less than 3 ms after switching were, although still detected by the recapture script, not included in the data set since their dwell times and amplitudes could not be properly determined. For each recapture event, the switching time (t_s), recapture time (t_r), and dwell (translocation) time (t_d) was recorded, as shown in Figure 2.1c. The switching time (t_s) is the time between an event and the subsequent voltage reversal. The recapture time (t_r) is the time between a voltage reversal and the next event. The dwell time (t_d) is defined as the time in between the leading and trailing edges of an event.

2.3. Results

Over the course of these recapture experiments, many (>30) molecules were recaptured, each producing current traces such as shown in Figure 2.2a. For the analysis, we selected 8 molecules that were each recaptured at least 50 times. Four different types of dsDNA molecules were analyzed: 2 linear lambda molecules, 4 circular lambda molecules, 1 linear lambda dimer molecule, and 1 circular lambda dimer molecule. The average number of recaptures for each type of molecule was 64, 99, 1067, and 561 respectively. An example of a 97 kbp linear molecule translocation event is shown in Figure 2.2b, and an example for the 97 kbp circular molecule is given in Figure 2.2c. More event examples, for each type of molecule, are provided in Supplementary Section 2.5.4. The presence of the 97 kbp lambda DNA dimers within the population was not surprising as these have been observed in the past[3] in translocation experiments where the lambda DNA solution, like in the current preparation, was not heat treated but kept at room temperature. Additionally, these longer molecules are selected out in experiments where we screen for high number of recaptures, since they diffuse slower and are thus less likely to escape.

We investigated the amount of folding for each molecule using two different approaches: (1) The large number of recaptures for each molecule meant that current blockade histograms could be made for each individual molecule, as shown in Figure 2.3. These histograms display the blockade levels of the current relative to the local baseline level. Since recapture data contains events at both polarities, we take the absolute value of the current in order to use both positive and negative events. These histograms contain a large baseline peak at zero, corresponding to the portions of the current trace where there was no translocation occurring, as well as smaller peaks corresponding to the current blockade produced by an integer number of DNA strands present in the nanopore. Interestingly, the frequency of occurrence for each peak is found to decrease exponentially with the peak number (i.e. number of dsDNA strands that is inserted), as is evident from the straight lines in Fig. 2.3e. This plot was made by determining the height of each (non-baseline) peak in the current histograms and then normalizing the values using the

height of the first non-baseline peak. The steepness of the slopes in this semilog plot are a measure for the difficulty of inserting a new fold into the pore, and we thus observe that it is harder to insert new DNA into the pore for circular dsDNA as compared to linear dsDNA, and that it is easier to obtain a higher number of folds for longer molecules compared to shorter molecules. The maximum number of dsDNA strands simultaneously present in the pore can be determined from the left-most current-blockade peak in the current histograms of Fig. 2.3a-d. This approach thus

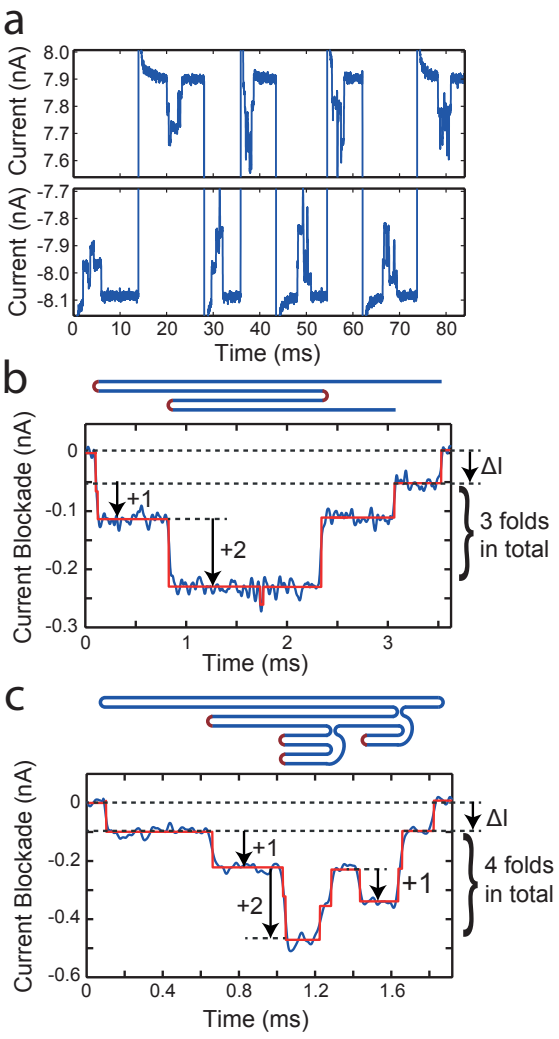


Figure 2.2: a) A typical current trace showing a 97 kbp linear molecule being recaptured 7 times through a nanopore. b) An example current trace of an event for the 97 kbp linear molecule. A possible molecule conformation is shown above in blue. c) An example event for the 97 kbp circular molecule. A possible molecule conformation is shown above in blue.

provides a direct measure of the maximum numbers of parallel dsDNA in the pore. It, however, does not indicate how often a certain molecule was folded along its length. For example, a molecule having many spatially separated short folds would produce a similar histogram as a molecule having one single long fold. (2) A more proper representation of the latter can be produced by counting the number of folds present in each event. This is accomplished by counting the number of rising edges, above the initial dsDNA blockade ΔI present in each event, as shown in Figure 2.2bc. Since an unfolded event caused by a simple head-to-tail translocation would only produce a single blockade level with value ΔI, any larger current values correspond to folds in the molecule. The magnitude of ΔI is determined using the single-dsDNA peak in the current histogram of that particular molecule. The magnitude of the rising edge divided by ΔI then provides the number of folds present at that point in the translocation process. Circular molecules have a baseline twice as large as that of linear molecules (i.e., $\Delta I_{circ}=2\Delta I_{lin}$) since there are always 2 dsDNA segments present in the circular topology and each fold brings 2 more dsDNA segments into the pore. In our analysis, we determined the total number of folds n present in each event and calculated the average number of folds, henceforth called the 'fold count', using all events in a given population in order to quantify the amount of folding present. Although the number of folds in any single event is an integer value, the fold count is a non-integer real number. The standard error of the mean was used as a measure of the uncertainty in these values.

We analyze current blockades seen in the current histograms for each type of molecule. Linear and circular molecules could easily be distinguished from their current histograms, as shown by comparison of Figure 2.3ab and 2.3cd respectively. The insertion of linear DNA molecules produced a current blockade of 0.056 nA, i.e. very comparable to the blockade of 0.061 nA seen in non-recapture control experiments performed on a pore of the same size (Supplementary Section 2.5.2). Circular molecules, however, exhibited a blockade of 0.111 nA, i.e., as expected, twice as large, since there are always at least two dsDNA strands present inside the pore due to the circular topology. The current histograms reveal that the longer molecules tend to have a significantly larger number of parallel DNA strands simultaneously inside the pore. For the linear 48.5 kbp, the current histogram peaks indicate the presence of events where there are up to 5 DNA segments in the pore simultaneously, while this number increases to up to 13 for the dimer which is twice as long. In the case of the circular molecules, we see a similar trend, with the 48.5 kbp circular molecule having up to 6 dsDNA segments simultaneously and the 97 kbp molecule up to 8. Current histograms from non-recapture control experiments on 48.5 kbp linear dsDNA typically show the presence of up to 5 DNA strands in the pore at 50 mV.

What is causing these remarkably high numbers of parallel DNA strands in the pore that we observe in the recapture experiments? We suggest that this relates to the DNA configuration directly after translocation. Since the strength of the electric field quickly drops off with increasing distance away from the nanopore, a large amount of DNA is densely piled up in the close vicinity of the pore when a DNA molecule translocates through the nanopore. The molecule then relaxes to

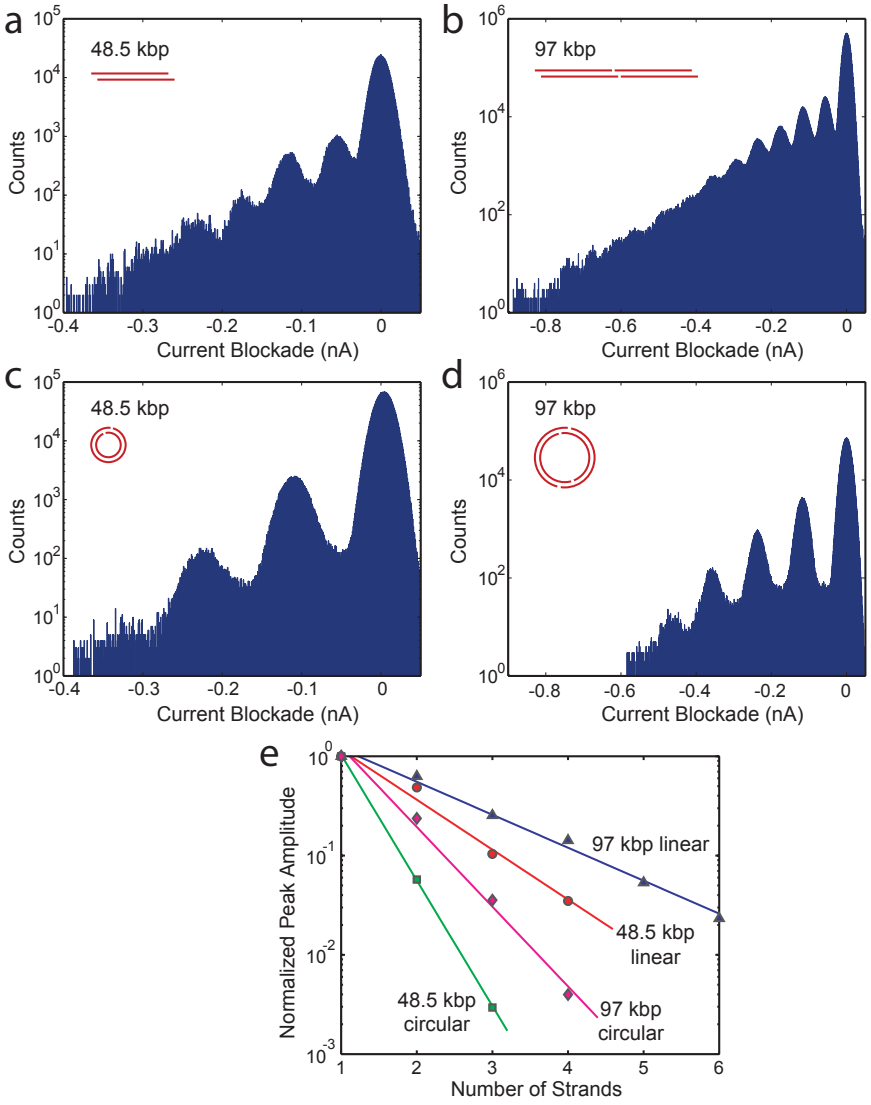


Figure 2.3: Current histograms for four types of molecules: a) Linear 48.5 kbp lambda dsDNA from a molecule recaptured 67 times. b) Linear 97 kbp dsDNA from a molecule recaptured 1067 times, showing much higher folding levels. c) Circular 48.5 kbp lambda dsDNA from a molecule recaptured 203 times. d) Circular 97 kbp dsDNA from a molecule recaptured 561 times. e) The frequency of occurrence for each peak on the current histograms. Peak amplitudes have been normalized using the amplitude of the largest (non-baseline) peak. The slope is a measure of the difficulty of inserting a new fold into the pore. The exponential decay constants of the fits are 2.91, 1.85, 1.16, and 0.76 for the 48.5 kbp circular, 97 kbp circular, 48.5 kbp linear, and 97 kbp linear molecules respectively.

equilibrium on a time scale given by the Zimm relaxation time. The Zimm relaxation time[13] can be estimated as

$$\tau_{Zimm} = \frac{\alpha M \eta_s [\eta]_0}{RT} \tag{2.1}$$

where M is the molecular weight, R is Boltzmann's constant, T is temperature, η_s is the solvent viscosity (0.89 cP for 1 M KCl), $[\eta]_0$ is the intrinsic viscosity, and α is a coefficient which depends on the type of molecule and experimental conditions. In a solution containing a polymer, the intrinsic viscosity is a measure of the polymer's contribution to the measured viscosity of the solution. The intrinsic viscosity of a particular polymer can be determined by measuring the viscosity of the solution at different concentrations of polymer and extrapolating the resulting trend to zero polymer concentration. For 48.5 kbp linear, 48.5 kbp circular, 97 kbp linear, and 97 kbp circular molecules, we find relaxation times of 67 ms, 15 ms, 205 ms, and 46 ms respectively. Details of this calculation have been provided in Supplementary Section 2.5.1. A different DNA configuration, and subsequently different translocation characteristics, can be expected when the recapture is smaller or larger than these Zimm times. In the following analysis these relaxation times mark the crossover points between equilibrium and non-equilibrium behavior.

We experimentally examined the amount of folding and the dwell time for each event as well as how these quantities vary as the molecule approaches equilibrium. Figure 2.4 and Figure 2.5 show a strong dependence for both the fold counts and dwell times on the recapture time. For the 48.5 kbp circular molecule (Figure 2.4a), we see the mean dwell time systematically increase from 1.4 ms at small recapture times to 1.85 ms at the largest recapture times probed. Correspondingly the fold count, shown in Figure 2.4b, reduces from about 1.3 at small recapture times to 0.4 at large recapture times. Interestingly, in the population distribution of folding, Figure 2.4c, we see the number of unfolded events reduced by a factor of 2 when comparing events occurring above the Zimm time (60% of events unfolded) to those occurring below (31% unfolded). Detailed folding statistics for each molecule can be found in Supplementary Section 2.5.3. Analyzing the data from the 97 kbp molecules, we see the dwell time increase from 3 ms to 6 ms, and 2 ms to 3 ms for the linear and circular molecules, respectively, as shown in Figure 2.5a. Looking at the fold counts for these molecules, Figure 2.5b, we see a decrease from 9.5 to 5.5 folds on average for the linear molecule and from 2.5 down to below 1 fold for the circular molecule. The distribution of folds for the 97 kbp circular molecule, shown in Figure 2.5c resembles that seen for the 48.5 kbp circular molecule (Figure 2.4c), although with higher fold counts both above and below the Zimm time.

Control experiments for DNA translocation without recapturing show that the most probable dwell time for 48.5 kbp linear molecules at 50 mV was around 3.3 ms. Using the previously observed power-law length dependence[14] for the dwell time with an exponent of 1.26, we would expect a 97 kbp linear molecule to translocate with a dwell time of about 8 ms. Indeed, Figure 2.5a shows the dwell time converging to a value of about 6 ms as the recapture time increases, though this value should be consider a lower limit in view of the limited statistics at the higher

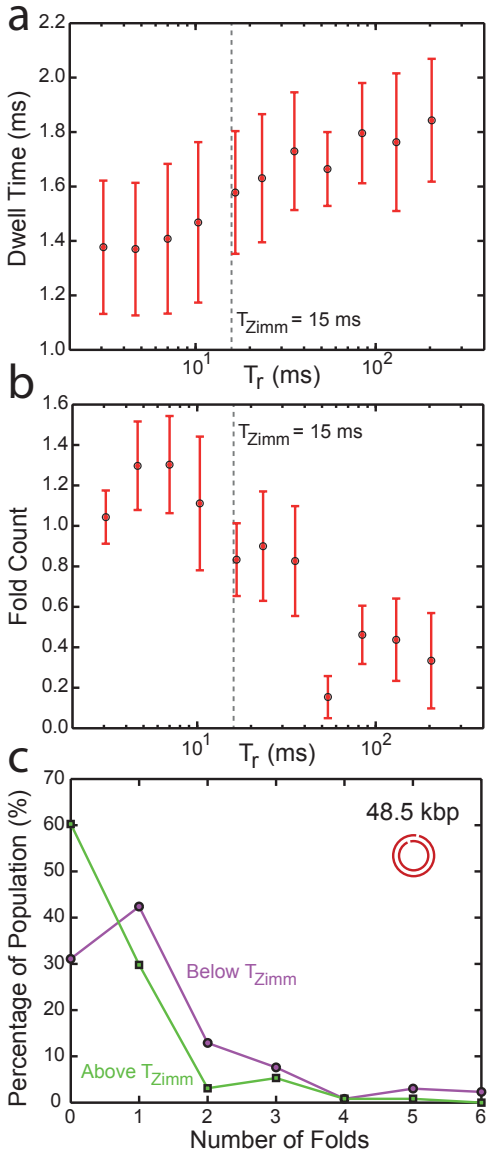


Figure 2.4: a) The mean dwell time as a function of the recapture time for circular 45.8 kbp dsDNA molecules. At very large recapture times, we expect the dwell times to converge to about 1.7 ms, the most probable dwell time observed in non-recapture experiments. b) The fold counts for events at a given recapture time. The error bars shown represent the standard error of the mean, determined using the dwell time (a) or fold count (b) of the events in each bin. c) The percentage of molecules with a given number of folds, for recaptures occurring below the Zimm time (green squares) and above the Zimm time (magenta circles). Above the Zimm time 60% of events are unfolded while 30% have one fold. At times below the Zimm time, only 31% of events are unfolded and 42% of events have one fold.

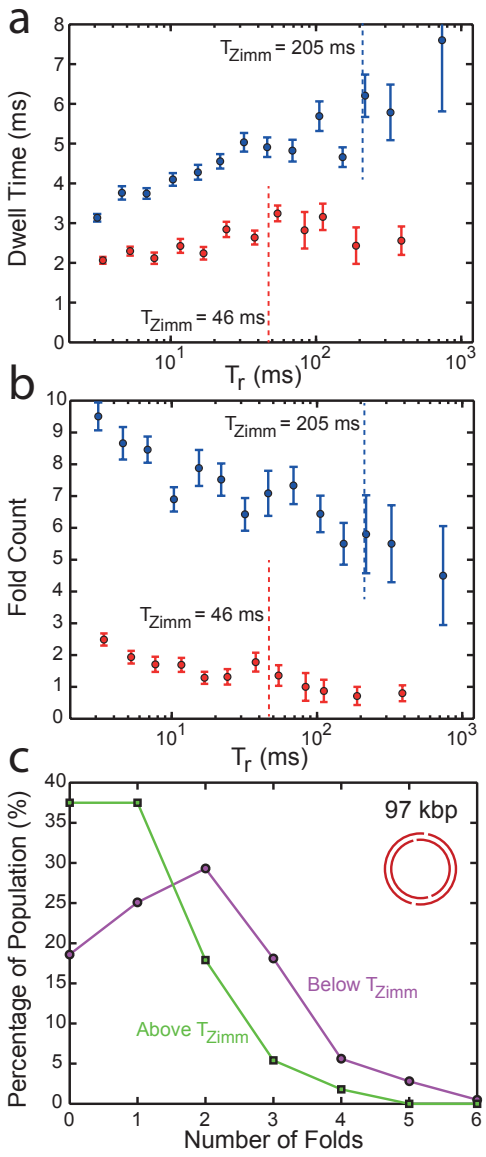


Figure 2.5: a) The dwell time as a function of the recapture time for circular (bottom-red) and linear (top-blue) 97 kbp dsDNA molecules. b) The fold counts for events at a given recapture time. For very short recapture times, the DNA density is very high in the close vicinity of the pore, which leads to very high fold counts. The error bars shown represent the standard error of the mean, determined using the dwell time (a) or fold count (b) of the events in each bin. c) The percentage of molecules with a given number of folds for a 97 kbp circular molecule, for recaptures occurring below the Zimm time (green squares) and above the Zimm time (magenta circles).

recapture times. The most-probable dwell time for 48.5 kbp circular molecules was estimated to be around 1.7 ms at 50 mV, using the events from the control experiments which only contained the two-dsDNA current level. This estimate is in good agreement with the limiting value observed in Figure 2.4a.

We can also examine the effect of Zimm relaxation in the regime where folding plays no role. This allows us to analyze how the initial molecular configuration affects the translocation time of the molecule. For this, we selected only unfolded events and observe how their dwell time changes above and below the Zimm relaxation time. Due to the high fold counts in our recapture experiments, only the circular molecules had a sufficient number of unfolded events for this analysis. Unfolded events were separated into those occurring above and those below the Zimm relaxation time. We expect the molecules recaptured below the Zimm time to have, on average, a smaller coil radius then the molecules that were allowed sufficient time to relax. The mean dwell times and standard errors for these populations are shown in Table 2.1. Indeed, in both cases we see the dwell time increase significantly when the molecule is recaptured only at times above its relaxation time. This observation agrees with the notion that a more spatially extended molecule takes a longer time to translocate.

Table 2.1: The mean dwell time and the standard error for unfolded events in circular molecules, shown for events recaptured below and above the Zimm time.

Type	48.5 kbp circular		97 kbp circular	
Regime	$t_r < \tau_Z$	$t_r > \tau_Z$	$t_r < \tau_Z$	$t_r > \tau_Z$
t_d (ms)	1.68	1.78	1.67	2.41
S.E. (ms)	0.03	0.02	0.18	0.34

2.4. Discussion and Conclusion

These nanopore experiments provide a way to probe the non-equilibrium configurations of DNA and its approach to equilibrium. The exceptionally high fold counts that we observe in these experiments can be explained by looking at the recapture process. In the translocation process, the DNA molecule is transported through the nanopore, driven by the electric field which is very high right at the pore, but drops very quickly (with $1/r^2$) away from the pore. This causes a fast pile up of DNA material, leading to a large amount of DNA on the trans side of the pore. This effect is very prominent, since we are dealing with very long (16 or 32 μm) DNA molecules and applying very low voltages (50 mV). Note that translocation occurs fast, i.e., the DNA is piled up in a few ms, which is much faster than the Zimm equilibrium time for these long DNA polymers. As a result, the translocated DNA does not yet have time to equilibrate to its equilibrium configuration. This high-density pile up of DNA leads to multiple points along the molecule that can subsequently be recaptured into the pore, resulting in events with a large number of folds. Alternatively, at very long recapture times, the molecule is allowed to relax and the DNA blob will

return to its characteristic large Flory radius on a timescale given by τ_{Zimm}, and the DNA densities and thus fold counts will approach those seen in non-recapture experiments where the molecules are in equilibrium when captured. Gathering sufficient statistics at very long recapture times is difficult, however, since longer switching times increase the probability of the molecule being lost. In practice, we were able to obtain data for recapture times between about 3 and 300 ms, with better statistics at the shorter t_r end.

Gershow and Golovchenko's original paper proposed the repeated interrogation of single molecules as a way to improve the accuracy of the measurement and as a way to probe time-dependent processes affecting the molecule. The data in this study shows that this approach may be viable in the future, but care must be taken to prevent unwanted effects such as folding that lead to complications. The use of short DNA molecules should reduce or eliminate the amount of folding observed, but will require the use of integrated nanopore amplifiers in order to achieve the fast switching speeds required[15–17]. Smaller-diameter pores introduce the complication of an entropic barrier present in the capture process[2], but may be useful to eliminate the folding observed, albeit with a higher risk of clogging and DNA-pore interactions. For probing protein-DNA structures[18], which require the use of large pores, analysis algorithms could take advantage of the fact that the location of folds as well as noise will change with each subsequent recapture event while the position of DNA-bound protein remains the same.

What do these results teach us about the influence of the molecular configuration on the translocation process? Our observations show that it is important to consider the timescales of the experimental process relative to the Zimm relaxation time, since the behavior of a polymer will be influenced by its state of (non)equilibrium. Additionally, the capture of a denser non-equilibrium DNA blob yields high fold counts and short translocation times. The observation that the mean dwell time of unfolded events increases with recapture time supports the simulations of Lu et al[6]. It is clear that reducing the large translocation time variation seen in nanopore experiments will require an approach which can either control the initial molecular conformation or be able to apply a strong controlling force to the molecule. In the former approach, integration of a entry nanochannel with a nanopore[19, 20], or combining a nanopore with a thin film gel[21], could force a DNA molecule into an extended conformation before the start of the translocation process. Such a method should provide multiple advantages including longer translocation times, improved reproducibility from event to event, and better discrimination between molecules with different physical properties such as length.

We have shown in this work that it is possible to carry out large number of recaptures on a single DNA molecule and use this to gain insight into both the properties of the molecule as well as the physics governing the translocation process. While these results show that a number of issues such as high fold counts will need to be addressed, the high-number recapture technique promises to open a number of interesting avenues for future research and extend the versatility of nanopores as a biophysical technique.

Acknowledgement

We thank Derek Stein, Johan Dubbeldam, and Rutger Hermsen for discussions. This work was supported by the European Research Council under research grant NanoforBio (no. 247072) and the Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) Academy Assistants Program.

2.5. Supplementary Info

2.5.1. Calculation of Zimm Relaxation Time

We aim to quantitatively estimate the Zimm time for long linear and circular dsDNA. The Zimm time equation 2.1 given in the main manuscript is a condensed form of the full expression

$$\tau_1 = \frac{M\eta[\eta]_0}{RT(1+1.662h)F(h)} \tag{2.2}$$

This equation, given by Hearst[22], contains a parameter h, which describes the hydrodynamic interaction between the polymer and the solvent. Within equation 2.2, R is the ideal gas constant, T is the temperature in K, M is the molecular weight, η is the solvent viscosity, and $[\eta]_0$ is the intrinsic viscosity. The function $F(h)$ is a sum of inverse eigenvalues and is introduced later in equation 2.5. We use the method of Bloomfield and Zimm[23] to determine the value of the hydrodynamic interaction parameter, h,

$$h(\varepsilon) = \frac{2^{\frac{\varepsilon}{2}} N^{\frac{1-\varepsilon}{2}} \rho}{(12\pi^3)^{\frac{1}{2}} b\eta} \tag{2.3}$$

where the friction coefficient for one Kuhn segment is taken to be $\rho = 2.69 \cdot 10^{-10}$ Pa·m as determined by Pluen et al[24], the Kuhn length is taken to be $b = 100$ nm, the viscosity of 1M KCl is $\eta = 8.9 \cdot 10^{-4}$ Pa·s, and the number of Kuhn segments is $N = 165$ and $N = 330$ for the 48.5 kbp and 97 kbp molecules respectively. Additionally, the excluded volume expansion parameter, ε, is found using the experimental relation to salt concentration as determined by Douthart and Bloomfield[25] for lambda DNA molecules

$$\varepsilon = 0.05 - 0.11 \cdot log(I) \tag{2.4}$$

where I is the ionic concentration of the solution expressed in Molar. Since our measurements take place in 1M KCl, the log term goes to zero and we take the value of the expansion parameter to be $\varepsilon = 0.05$. Using Eq. 2.3 we find values of $h = 1.80$ and $h = 2.51$ for the two molecules. This places us in the so-called "partial draining" regime, in between the free-draining ($h << 1$) and non-draining ($h >> 1$) cases. We use to approach of Hearst[22] to determine the intrinsic viscosity of the linear molecules, by finding the sum of inverse eigenvalues, $F(h)$, using

$$F(h(\varepsilon)) = \sum_{k=1}^{\infty} \left(k^2 + 4h \frac{\lambda_k'}{\pi^2} \left(1 - \frac{1}{2\pi k} \right) \right)^{-1} \tag{2.5}$$

This is further split into two parts, as Hearst has done,

$$F(h(\varepsilon)) = \sum_{k=1}^{8} \left(k^2 + \frac{4h \cdot (\pi k)^{\delta}}{\pi^2 \cdot 2^{\delta}} \left(\pi k \int_0^{\pi k} t^{-\delta} \cos(t) dt - \delta \int_0^{\pi k} t^{-\delta} \sin(t) dt \right) \right)^{-1} + \sum_{k=9}^{\infty} \left(k^2 + 2hk^{\frac{3}{2}} \left(1 - \frac{1}{2\pi k} \right) \right)^{-1} \tag{2.6}$$

where δ is given by

$$\delta = \frac{1 + \varepsilon}{2} \tag{2.7}$$

In the first sum of Eq. 2.6, we have substituted the integral for the diagonal elements of the matrix G, while in the second term we use $k^{\frac{3}{2}}$ and keep the correction factor that Hearst dropped. This equation was solved in Maple (using the Euler-Maclaurin summation formula for the second term), and gave values of $F(2h) = 0.4575$ and $F(2h) = 0.3613$ for the 48.5 kbp and 97 kbp linear molecules respectively. The intrinsic viscosity was then determined using

$$[\eta]_0 = \frac{N_a b^3 N^{\frac{3}{2}}}{\sqrt{12\pi} M} [2h \cdot F(h)] \tag{2.8}$$

where $M = 3.15 \cdot 10^7$ g/mol and $M = 6.30 \cdot 10^7$ g/mol for the 48.5 kbp and 97 kbp molecules respectively. This produced values of $[\eta]_0 = 109$ deciLiter/gram for the 48.5 kbp DNA and $[\eta]_0 = 169$ dL/g for the 97 kbp molecule. The former value compares quite well to experimentally determined values for linear lambda DNA of 112 dL/g[25], and 132 dL/g[26], as well as 119 dL/g using the estimation formula of Tsortos et al[27].

Placing all of these values back into Eq. 2.2 gives us Zimm relaxation time values of of $\tau_1 = 67$ ms for the 48.5 kbp linear DNA and $\tau_1 = 205$ ms for the 97 kbp linear molecule.

We proceed with the same calculation for the circular molecules by first noting that the ratio of the intrinsic viscosity of linear to circular molecules has been predicted and experimentally measured[25] (in high salt conditions) to be

$$\frac{[\eta]_L}{[\eta]_C} = 1.6 \tag{2.9}$$

We use Eq. 2.9 to estimate the values of the intrinsic viscosity to be $[\eta]_0 = 68$ dL/g for the 48.5 kbp circular DNA and $[\eta]_0 = 106$ dL/g for the 97 kbp molecule.

The former value compares favorably to experimentally measured values for circular lambda DNA of 65 dL/g[25] and 90 dL/g[26]. We compute the eigenvalues using the approach of Bloomfield and Zimm[23]

$$\sum_{1}^{\infty}(\lambda_k^{'})^{-1} = \sum_{k=1}^{10}\left(2\pi^2k^2\int_0^1\left[q\left(1-\frac{q}{2}\right)\right]^{-\delta}\left[\left(\frac{q}{2}\right)^{2\delta}+\left(1-\frac{q}{2}\right)^{2\delta}\right]^{\frac{1}{2}}\cos(\pi kq)dq\right)^{-1} + \sum_{k=11}^{\infty}\left(\frac{\pi^{2+\delta}k^{1+\delta}}{\Gamma(\delta)\cos(\frac{\pi\delta}{2})}\left[1-\frac{1}{2}\left(\frac{1}{2\pi k}\right)^{2\delta}\frac{\Gamma(1+\delta)}{\Gamma(1-\delta)}\right]\right)^{-1} \tag{2.10}$$

The first ten eigenvalues, needed for the first sum in Eq. 2.10, are 15.088, 44.348, 82.842, 128.867, 181.435, 239.879, 303.701, 372.522, 446.032, and 523.974 for $\varepsilon = 0.05$ while the second term is an approximation[23] which is valid for large values of k. The sum of inverses, given in Eq. 2.10, has a value of 0.3479 (factoring in the double degeneracy of the eigenfunctions) and can then be used to determine the Zimm relaxation time using a modified version of Eq. 2.2

$$\tau_1 = \frac{M\eta[\eta]_0}{2RT\lambda_k^{'}\sum_k\frac{1}{\lambda}} = \frac{M\eta[\eta]_0}{RT(15.088)(0.3479)} = \frac{0.1905M\eta[\eta]_0}{RT} \tag{2.11}$$

This gives Zimm relaxation time values of of $\tau_1 = 15$ ms for the 48.5 kbp circular DNA and $\tau_1 = 46$ ms for the 97 kbp circular molecule, as summarized in Table 2.2.

How do these calculated relaxation times compare to measurements by others? Many measurements of the relaxation time exist for linear lambda DNA in more viscous conditions[28–34]. Liu et al determined the relation $\tau_1 = 0.0947\eta$, relating the relaxation time of linear lambda DNA to the viscosity of the solution. Extrapolating to our conditions gives a value of 84 ms, very close to our calculated value of 67 ms.

Table 2.2: The Zimm relaxation times for the four different molecules.

Size	Zimm Time (ms)
48.5 kbp linear	67
97 kbp linear	205
48.5 kbp circular	15
97 kbp circular	46

2.5.2. Control Experiments

Non-recapture control experiments were carried out using 48.5 kbp lambda DNA (heated to 65°C for 10 min) at a concentration of 1.56 nM in a 20 nm PDMS coated pore, at 50 mV and 100 mV applied voltage as shown in Figure 2.6 and 2.7.

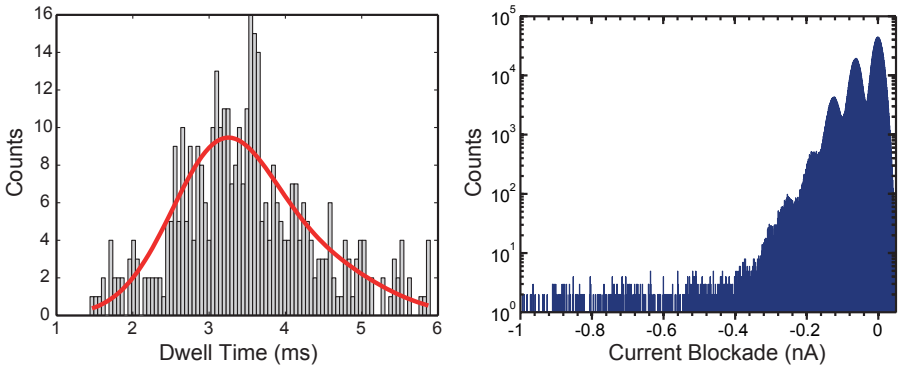


Figure 2.6: (Left) Dwell time histogram for unfolded events translocated at 50 mV. The most probable dwell time is 3.3 ms. (Right) Current histogram for all 50 mV events. The first DNA peak occurs at 0.061 nA (1.22 nS).

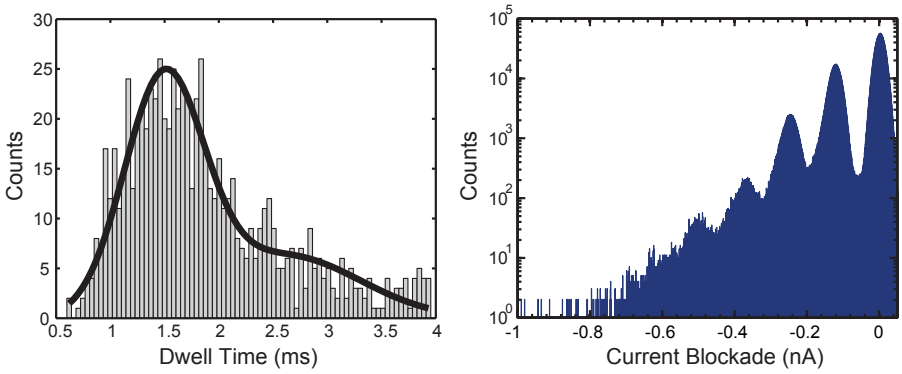


Figure 2.7: (Left) Dwell time histogram for unfolded events translocated at 100 mV. The most probable dwell time is 1.5 ms. (Right) Current histogram for all 100 mV events. The first DNA peak occurs at 0.123 nA (1.23 nS).

2.5.3. Observed Fold Counts

Fold counts were determined for all experiments using a Matlab script coupled with the level fitting event output from OpenNanopore[10]. Briefly, this script looked at the magnitude of the current decreases for each successive level, relative to the blockade for an unfolded molecule (as determined using first peak in the current histogram). Events with a recapture time below 3 ms were removed since the baseline did not have sufficient time to settle.

Table 2.3: The observed fold counts for linear molecules in recapture experiments, given as the percentage of events with a given number of folds, relative to the total population of events in that data set.

Size	**48.5 kbp**			**97 kbp**		
Type	linear			linear		
Data Set	$t_r < \tau_Z$	$t_r > \tau_Z$	all t_r	$t_r < \tau_Z$	$t_r > \tau_Z$	all t_r
0 folds	15.0	33.3	20.2	0.2	0.0	0.2
1 fold	33.3	33.3	33.3	2.4	14.8	3.0
2 folds	10.0	16.7	11.9	2.6	0.0	2.4
3 folds	11.7	8.3	10.7	6.9	14.8	7.3
4 folds	13.3	0.0	9.5	8.4	18.5	8.9
5 folds	10.0	8.3	9.5	9.6	7.4	9.5
6 folds	6.7	0.0	4.8	9.2	7.4	9.1
7 folds	0.0	0.0	0.0	11.3	3.7	10.9
8 folds	0.0	0.0	0.0	12.0	14.8	12.1
9 folds	0.0	0.0	0.0	9.6	7.4	9.5
10 folds	0.0	0.0	0.0	7.9	0.0	7.5
11 folds	0.0	0.0	0.0	6.0	3.7	5.9
12 folds	0.0	0.0	0.0	4.9	0.0	4.7
13 folds	0.0	0.0	0.0	2.6	0.0	2.4
14 folds	0.0	0.0	0.0	2.4	7.4	2.6
15 folds	0.0	0.0	0.0	1.7	0.0	1.6
16 folds	0.0	0.0	0.0	0.4	0.0	0.4
17 folds	0.0	0.0	0.0	0.6	0.0	0.6
18 folds	0.0	0.0	0.0	0.2	0.0	0.2
19 folds	0.0	0.0	0.0	0.6	0.0	0.6
20 folds	0.0	0.0	0.0	0.2	0.0	0.2
21 folds	0.0	0.0	0.0	0.0	0.0	0.0
22 folds	0.0	0.0	0.0	0.0	0.0	0.0
23 folds	0.0	0.0	0.0	0.2	0.0	0.2
24 folds	0.0	0.0	0.0	0.0	0.0	0.0
25 folds	0.0	0.0	0.0	0.0	0.0	0.0
Mean Fold Count	2.32	1.33	2.04	7.66	5.7	7.56

Table 2.4: The observed fold counts for circular molecules in recapture experiments, given as the percentage of events with a given number of folds, relative to the total population of events in that data set.

Size	**48.5 kbp**			**97 kbp**		
Type	circular			circular		
Data Set	$t_r < \tau_Z$	$t_r > \tau_Z$	all t_r	$t_r < \tau_Z$	$t_r > \tau_Z$	all t_r
0 folds	31.1	60.3	45.6	18.6	37.5	22.5
1 fold	42.4	29.8	36.1	25.1	37.5	27.7
2 folds	12.9	3.1	8.0	29.3	17.9	26.9
3 folds	7.6	5.3	6.5	18.1	5.4	15.5
4 folds	0.8	0.8	0.8	5.6	1.8	4.8
5 folds	3.0	0.8	1.9	2.8	0	2.2
6 folds	2.3	0	1.1	0.5	0	0.4
7 folds	0	0	0	0	0	0
8 folds	0	0	0	0	0	0
9 folds	0	0	0	0	0	0
10 folds	0	0	0	0	0	0
Mean Fold Count	1.23	0.59	0.91	1.77	0.96	1.61

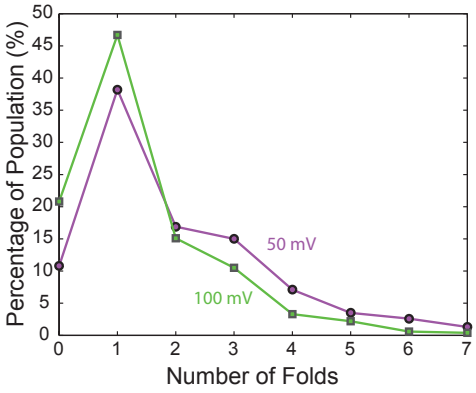


Figure 2.8: The percentage of molecules with a given number of folds, at 100 mV (green squares) and 50 mV (magenta circles). Data from non-recapture control experiments using 48.5 kbp lambda DNA.

Table 2.5: The observed folding counts for the linear 48.5 kbp molecules in non-recapture control experiments, at applied voltages of 50 mV and 100 mV.

Size	**48.5 kbp**	**48.5 kbp**
Type	linear	linear
Data Set	50 mV control	100 mV control
0 folds	10.8	20.8
1 fold	38.2	46.7
2 folds	16.9	15.1
3 folds	15.0	10.5
4 folds	7.1	3.3
5 folds	3.5	2.2
6 folds	2.6	0.6
7 folds	1.3	0.4
8 folds	1.4	0.0
9 folds	0.7	0.0
10 folds	0.3	0.0
11 folds	0.3	0.1
12 folds	0.2	0.0
13 folds	0.0	0.2
14 folds	0.2	0.0
15 folds	0.3	0.1
16 folds	0.3	0.0
17 folds	0.2	0.1
18 folds	0.0	0.0
19 folds	0.1	0.0
20 folds	0.2	0.1
21 folds	0.0	0.0
22 folds	0.0	0.0
23 folds	0.1	0.0
Mean Fold Count	2.45	1.47

2.5.4. Sample Events from Recapture Experiments

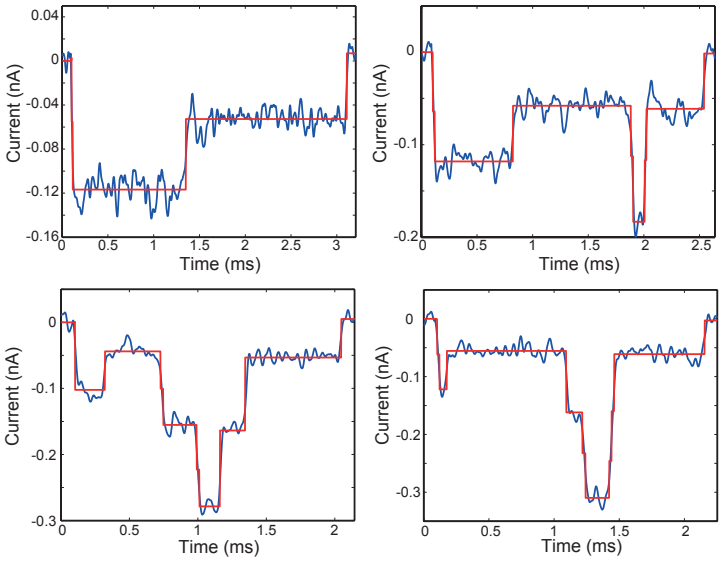


Figure 2.9: Example events from the recapture of a 48.5 kbp linear DNA molecule.

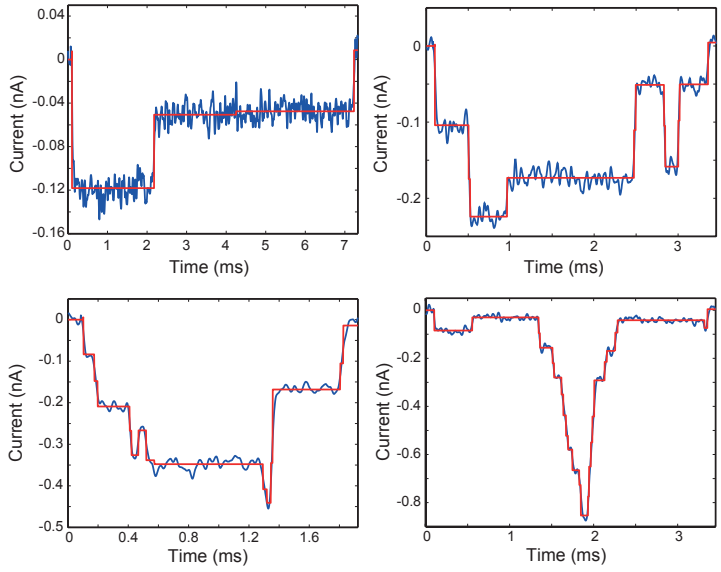


Figure 2.10: Example events from the recapture of a 97 kbp linear DNA molecule.

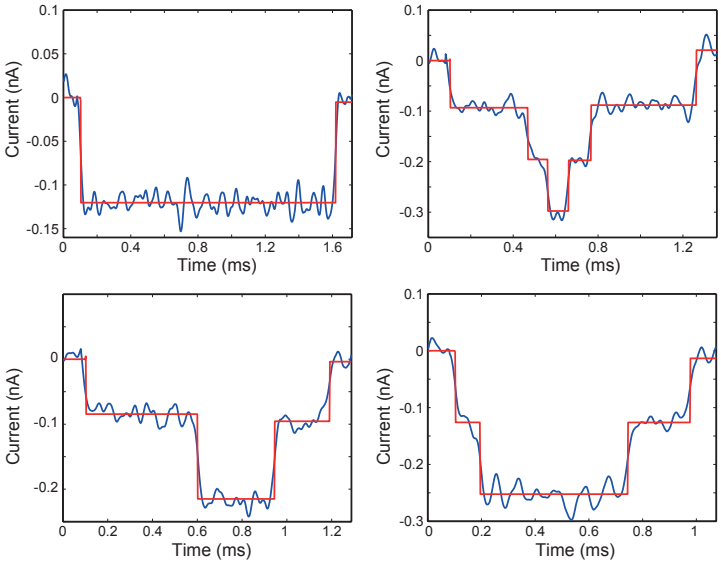


Figure 2.11: Example events from the recapture of a 48.5 kbp circular DNA molecule.

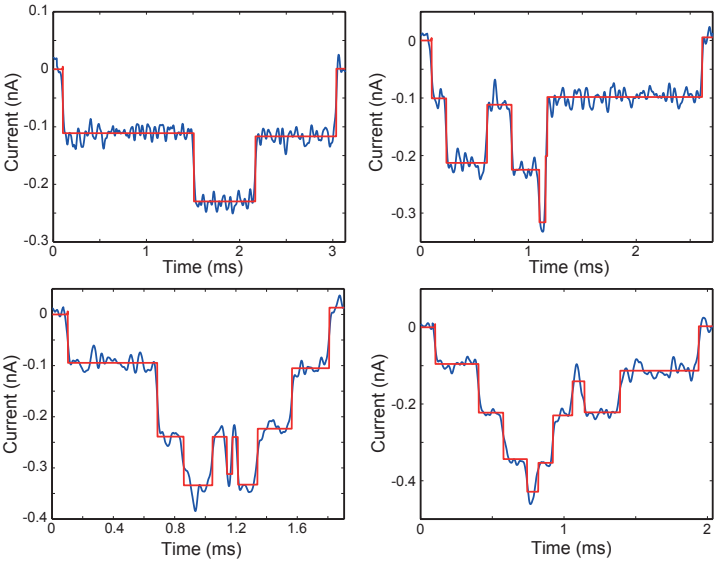


Figure 2.12: Example events from the recapture of a 97 kbp circular DNA molecule.

References

[1] C. Dekker, *Solid-state nanopores,* Nat Nano **2**, 209 (2007).

[2] M. Wanunu, *Nanopores: A journey towards DNA sequencing,* Physics of Life Reviews **9**, 125 (2012).

[3] A. J. Storm, C. Storm, J. Chen, H. Zandbergen, J.-F. Joanny, and C. Dekker, *Fast DNA translocation through a solid-state nanopore,* Nano Letters **5**, 1193 (2005).

[4] P. Chen, J. Gu, E. Brandin, Y.-R. Kim, Q. Wang, and D. Branton, *Probing single DNA molecule transport using fabricated nanopores,* Nano Letters **4**, 2293 (2004).

[5] M. Mihovilovic, N. Hagerty, and D. Stein, *Statistics of DNA capture by a solid-state nanopore,* Physical Review Letters **110**, 028102 (2013).

[6] B. Lu, F. Albertorio, D. P. Hoogerheide, and J. A. Golovchenko, *Origins and consequences of velocity fluctuations during DNA passage through a nanopore,* Biophysical Journal **101**, 70 (2011).

[7] M. Gershow and J. A. Golovchenko, *Recapturing and trapping single molecules with a solid-state nanopore,* Nat Nano **2**, 775 (2007).

[8] X. J. A. Janssen, M. P. Jonsson, C. Plesa, G. V. Soni, C. Dekker, and N. H. Dekker, *Rapid manufacturing of low-noise membranes for nanopore sensors by trans -chip illumination lithography,* Nanotechnology **23**, 475302 (2012).

[9] V. Tabard-Cossa, D. Trivedi, M. Wiggin, N. Jetha, and A. Marziali, *Noise analysis and reduction in solid-state nanopores,* Nanotechnology **18**, 305505 (2007).

[10] C. Raillon, P. Granjon, M. Graf, L. J. Steinbock, and A. Radenovic, *Fast and automatic processing of multi-level events in nanopore translocation experiments,* Nanoscale **4**, 4916 (2012).

[11] J. C. Wang and N. Davidson, *Thermodynamic and kinetic studies on the interconversion between the linear and circular forms of phage lambda DNA,* Journal of Molecular Biology **15**, 111 (1966).

[12] A. D. Hershey, E. Burgi, and L. Ingraham, *Cohesion of DNA molecules isolated from phage lambda,* Proceedings of the National Academy of Sciences of the United States of America **49**, 748 (1963).

[13] L. C. Klotz and B. H. Zimm, *Retardation times of deoxyribonucleic acid solutions. ii. improvements in apparatus and theory,* Macromolecules **5**, 471 (1972).

[14] A. J. Storm, J. H. Chen, H. W. Zandbergen, and C. Dekker, *Translocation of double-strand DNA through a silicon oxide nanopore,* Physical Review E: Statistical, Nonlinear, and Soft Matter Physics **71**, 051903 (2005).

[15] M. Carminati, G. Ferrari, D. Bianchi, and M. Sampietro, *Femtoampere integrated current preamplifier for low noise and wide bandwidth electrochemistry with nanoelectrodes,* Electrochimica Acta (2013).

[16] J. K. Rosenstein, M. Wanunu, C. A. Merchant, M. Drndic, and K. L. Shepard, *Integrated nanopore sensing platform with sub-microsecond temporal resolution,* Nat Meth **9**, 487 (2012).

[17] A. Uddin, S. Yemenicioglu, C.-H. Chen, E. Corigliano, K. Milaninia, and L. Theogarajan, *Integration of solid-state nanopores in a 0.5 um CMOS foundry process,* Nanotechnology **24**, 155501 (2013).

[18] S. W. Kowalczyk, A. R. Hall, and C. Dekker, *Detection of local protein structures along DNA using solid-state nanopores,* Nano Letters **10**, 324 (2009).

[19] Y. Zhang, *Combined Nanochannel-Nanopore Device for Single-Molecule DNA Analysis and Manipulation*, Ph.D. thesis (2012), mcGill University.

[20] Y. Zhang and W. Reisner, *Nanochannel device with embedded nanopore: a new approach for single-molecule DNA analysis and manipulation,* (2013).

[21] D. Sean, H. W. de Haan, and G. W. Slater, *Polymer translocation through a nanopore from a crosslinked gel to free solution,* in *Bulletin of the American Physical Society*, Vol. 58 (APS, 2013).

[22] J. E. Hearst, *Effect of partial draining on the intrinsic viscosity of flexible macromolecules,* The Journal of Chemical Physics **37**, 2547 (1962).

[23] V. Bloomfield and B. H. Zimm, *Viscosity, sedimentation, et cetera, of ring-and straight-chain polymers in dilute solution,* Journal of Chemical Physics **44**, 315 (1966).

[24] A. Pluen, P. A. Netti, R. K. Jain, and D. A. Berk, *Diffusion of macromolecules in agarose gels: comparison of linear and globular configurations,* Biophysical journal **77**, 542 (1999).

[25] R. J. Douthart and V. A. Bloomfield, *Intrinsic viscosities of cyclic and linear λ DNA,* Biopolymers **6**, 1297 (1968).

[26] J. R. Dawson and J. A. Harpst, *Light scattering and hydrodynamic properties of linear and circular bacteriophage lambda DNA,* Biopolymers **10**, 2499 (1971).

[27] A. Tsortos, G. Papadakis, and E. Gizeli, *The intrinsic viscosity of linear DNA,* Biopolymers **95**, 824

(2011).
[28] H. P. Babcock, D. E. Smith, J. S. Hur, E. S. G. Shaqfeh, and S. Chu, *Relating the microscopic and macroscopic response of a polymeric fluid in a shearing flow,* Physical Review Letters **85**, 2018 (2000).
[29] H. P. Babcock, R. E. Teixeira, J. S. Hur, E. S. G. Shaqfeh, and S. Chu, *Visualization of molecular fluctuations near the critical point of the coil–stretch transition in polymer elongation,* Macromolecules **36**, 4544 (2003).
[30] Y. Liu, Y. Jun, and V. Steinberg, *Longest relaxation times of double-stranded and single-stranded DNA,* Macromolecules **40**, 2172 (2007).
[31] T. T. Perkins, D. E. Smith, and S. Chu, *Single polymer dynamics in an elongational flow,* Science **276**, 2016 (1997).
[32] D. E. Smith, H. P. Babcock, and S. Chu, *Single-polymer dynamics in steady shear flow,* Science **283**, 1724 (1999).
[33] D. E. Smith and S. Chu, *Response of flexible polymers to a sudden elongational flow,* Science **281**, 1335 (1998).
[34] R. E. Teixeira, H. P. Babcock, E. S. G. Shaqfeh, and S. Chu, *Shear thinning and tumbling dynamics of single polymers in the flow-gradient plane,* Macromolecules **38**, 581 (2005).

3

Velocity of DNA during translocation through a solid state nanopore

While understanding translocation of DNA through a solid-state nanopore is vital for exploiting its potential for sensing and sequencing at the single-molecule level, surprisingly little is known about the dynamics of the propagation of DNA through the nanopore. Here we use linear double-stranded DNA molecules, assembled by the DNA origami technique, with markers at known positions in order to determine, for the first time, the local velocity of different segments along the length of the molecule. We observe large intramolecular velocity fluctuations, likely related to changes in the drag force as the DNA blob unfolds. Furthermore we observe an increase in the local translocation velocity towards the end of the translocation process, consistent with a speeding up due to unfolding of the last part of the DNA blob. We use the velocity profile to estimate the uncertainty in determining the position of a feature along the DNA given its temporal location, and demonstrate the error introduced by assuming a constant translocation velocity.

C. Plesa, N. van Loo, P. Ketterer, H. Dietz, and C. Dekker, *Velocity of DNA during translocation through a solid state nanopore*, Nano Letters (2014), in press.

3.1. Introduction

The number of solid-state nanopore applications has exploded[1, 2] since their inception over a decade ago[3, 4]. Nanopores have been used to study and detect various biopolymers (particularly DNA[5–9], free protein[10, 11], DNA-origami[12–14], and DNA-protein complexes[15–17]. Despite this advance, some basic properties of the translocation process remain poorly understood. One such fundamental question is how the velocity of a long DNA molecule changes as the molecule translocates through the pore. Previous simulations from the Golovchenko group suggested that the wide distribution observed for the translocation velocities for molecules of equal length can be attributed to drag-induced velocity fluctuations and predicted that the DNA would strongly speed up at the end of the translocation process[18], as described in detail later. Currently, no experimental measurement of the velocity profile has been carried out. Since knowledge of this velocity profile is required to convert temporal signals into positional information, crucial for obtaining biological information from a translocating DNA molecule, this represents a serious gap in our current understanding of the translocation process.

In this study we introduce a new technique to determine the velocity of different segments along a DNA molecule, and we present the first experimental data on how the velocity changes during the course of a translocation event. We systematically observe a higher local velocity at the end of the translocation event, which can be attributed to the unfolding of the DNA blob outside the pore. Additionally, we observe a wide distribution in the intramolecular, intermolecular, and pore-to-pore translocation velocities. We utilize this information to estimate how well we can determine the spatial position from a measured nanopore signal.

Before we turn to the experimental findings, we briefly recapitulate the technique of solid-state nanopores and the DNA translocation process. Nanofabrication[19] is used to create chips with 20 nm thick free-standing SiN membranes. A transmission electron microscope (TEM) is subsequently used to focus an electron beam onto the membrane and create a nanopore of any diameter desired. This membrane is then placed in a flowcell such that it separates two small reservoirs containing 4M LiCl salt solution. Chlorinated silver electrodes are placed into each chamber and an electric field is applied across the membrane. The ionic current passing through the nanopore is recorded by means of a low-noise amplifier. Any biomolecule passing through the pore causes a temporary drop in the measured ionic current for the duration of its translocation through the pore constriction. DNA molecules, which are highly negatively charged, are electrophoretically driven through the nanopore, as shown in Figure 3.1a. Long DNA molecules typically translocate in a head-to-tail fashion[20], but can also exhibit varying degrees of folding, when the molecule is captured along its length rather than at an end. While a piece of the DNA traverses the nanopore, the rest of the DNA molecule remains as a large blob outside the pore[21]. A large variation is typically observed for the total translocation time, even among molecules of the same type. These differences in the intermolecular velocity, as well as intramolecular velocity fluctuations, have been suggested to be caused by variations in the drag force due to the unraveling of this random blob outside the pore[18], which becomes increasingly

smaller as the translocation progresses. The translocation velocity of a molecule is dependent on its particular initial conformation at the moment of capture, and this is believed to be the cause of the wide distribution of intermolecular translocation times typically observed. Both simulations[18] and recapture experiments[22] suggest that extended conformations lead to slower translocation durations, presumably because the mean drag force is larger. Near the end of the translocation process, the drag force due to the DNA blob before the pore decreases quickly, resulting in a significant non-linear increase in the translocation velocity. This effect has been predicted by simulations[18] but never experimentally measured. A recent experimental study by Singer et al examined the translocation time between two DNA-bound PNA probes as a function of distance, for relatively short (<3.4 kbp) DNA molecules translocating through a small (3.7 nm) pore in asymmetric salt conditions[23]. They found a power law dependence of this time with distance between the probes, similar to what is observed for the total translocation time as a function of total DNA length. In contrast, we investigate long DNA molecules, with a specific focus on determining the velocity profile along the molecule and we attempt to shed some experimental light on the open questions that we outlined in this section.

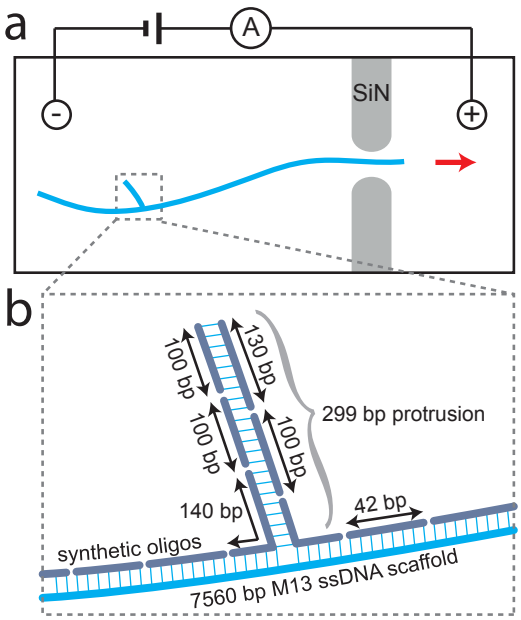


Figure 3.1: a) Schematic illustration of a synthetic DNA construct containing a protrusion translocating through a solid-state nanopore. b) Close up schematic of the protrusion showing it is assembled from multiple individual DNA oligomers.

We designed synthetic DNA constructs with markers at known positions in order to measure the local velocity over different segments along the molecule. DNA nanotechnology has grown significantly over the last few years, particularly due to the

drop in DNA synthesis costs as well as the introduction of DNA origami[24], a simple technique to create nanoscale shapes out of DNA. In the DNA origami approach, a long single-stranded DNA scaffold is folded into any shape desired, through the addition of short oligonucleotides complementary to multiple sections of the scaffold. In this study, we used a similar strategy to create linear long DNA molecules with a protrusion at a defined position along the molecule (Fig. 3.1). Since the position of the protrusion is controlled by design and we can measure the time required to traverse the DNA between say the start of the molecule and the protrusion, we are able to determine the mean translocation velocity along different segments of the molecule. A 7560 base M13 ssDNA scaffold was hybridized with 42 base oligos everywhere along its length except where the protrusion is attached (Fig. 3.1 and Supplementary Section 3.6.5). A protrusion is separately assembled and hybridized to the partially hybridized M13 backbone, as shown in Figure 3.1b. The 299 bp protrusion is created from the hybridization of six complementary oligos, ranging in length from 94 to 140 bases, as detailed in Supplementary Section 3.6.5. Three constructs were created, **i)** a symmetric construct with the protrusion at the exact center of the linear molecule (Figure 3.2a), **ii)** an asymmetric construct with the protrusion positioned 1571 bp (20.8%) from the closest end (Figure 3.2b), **iii)** and a control construct with no protrusion. We chose to use a dsDNA protrusion since this will have a well-defined blockade level that can be easily distinguished from the current signals produced by knots[25], which are at least twice as high in amplitude (as measured from the single dsDNA blockade level). For the asymmetric construct we chose to attach the protrusion 1571 bp away from the end in order to be able to distinguish the protrusion from small folds which can occur at the start or the end of the translocation process. By measuring the time τ_p from the start of the translocation to the start of the protrusion, we are able to determine the mean velocity over the first part of the molecule. Similarly, we can determine the velocity for the segment in which the protrusion itself resides inside the pore, as well as the last segment of the molecule. Given the already large distribution in translocation times τ_{DNA} for DNA molecules of equal length, we use the normalized temporal position (τ_p/τ_{DNA}) for comparisons between different molecules and pores. The presence of nicks along the DNA strand can be expected to reduce the persistence length of these DNA constructs relative to dsDNA. In order to characterize this effect, we carried out gels and determined the persistence length using AFM measurements (Supplementary Section 3.6.1). We found that the persistence length of these constructs is only slightly reduced and that these DNA constructs still behave like worm-like chains as opposed to freely-jointed chains. These synthetic DNA constructs allowed us to probe the velocity of the translocation process in a way not previously possible.

Translocation of constructs containing a protrusion results in events with a characteristic current spike within the DNA blockade. As expected, when the constructs are translocated through a nanopore we observe that a significant fraction (Supplementary Section 3.6.2) of unfolded events contain current spikes caused by the presence of the protrusions. Typical examples are shown in Figure 3.2, where we see (i) the DNA entering the pore, causing a current drop I_1, then (ii) during the

translocation, we observe a sharp spike with an additional amplitude I_1 due to the protrusion, and (iii) eventually the current level returns to the open-pore level when the translocation is complete. These types of events are not observed in the control measurements on the construct without any protrusion. The fraction of events containing protrusions is set by a number of factors as discussed in Supplementary Section 3.6.2. The presence of the protrusion in these DNA constructs is clearly visible as a current spike present within a large fraction of the DNA events observed and allows us to determine the mean translocation velocity for any segment desired.

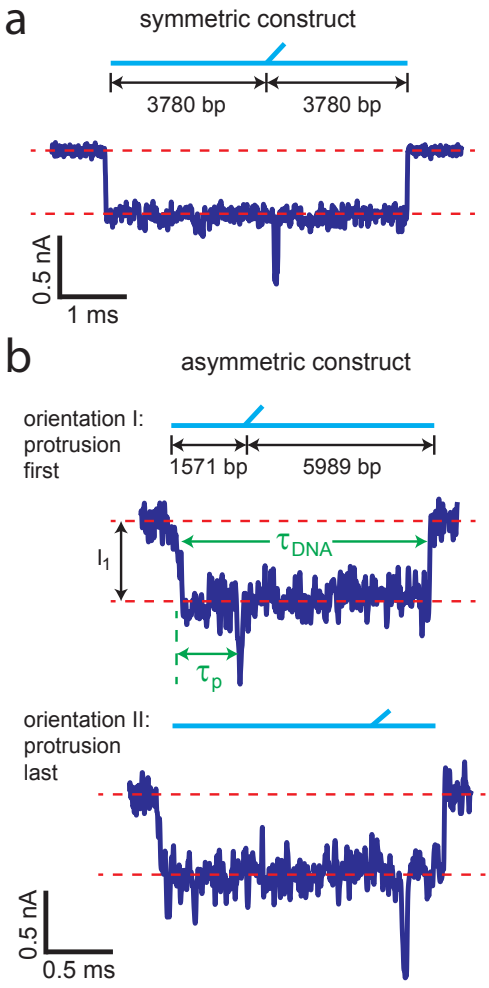


Figure 3.2: a) Example current trace of the symmetric DNA construct. b) Typical current traces produced by the asymmetric DNA construct. Two orientations are possible as depicted in the molecular configurations shown above each current trace.

3.2. Results

We first discuss the results based on the symmetric construct with the protrusion at the center of the DNA molecule. Figure 3.3 shows the distribution of the center protrusion as a function of the normalized temporal position (τ_p/τ_{DNA}). A Gaussian fit of this 296 event distribution has a mean of 0.528 and a standard deviation of 0.137. This distribution is extremely wide considering the well-defined position of the protrusion at the center of the molecule. If the molecules were travelling at a constant velocity we would expect a peak centered at 0.500. The slight shift observed in the mean of the distribution suggests that the molecule, on average, travelled slower in the first half relative to the second half. We also plot the difference between the average velocity over the last segment and the first segment in the DNA construct, to examine the typical velocity fluctuations, as shown in Figure 3.5a. Positive differences occur if the molecule sped up in the latter half while negative differences occurred if the molecule slowed down. We can see that the Gaussian distribution (with several positive outliers cut off in the figure), has a mean value of 0.015 bp/μs and STD of 0.79. Nearly all values fall within the $\pm$2 bp/μs range, with some outliers as far out as +17 bp/μs which can be attributed to DNA-pore interactions. These velocity fluctuations are quite large in comparison with the mean velocity of 2.0 bp/μs for these molecules, calculated using the most probable translocation time. This very sizable variation indicates that the intramolecular speed distribution is very broad, similar to the inter-molecule variation.

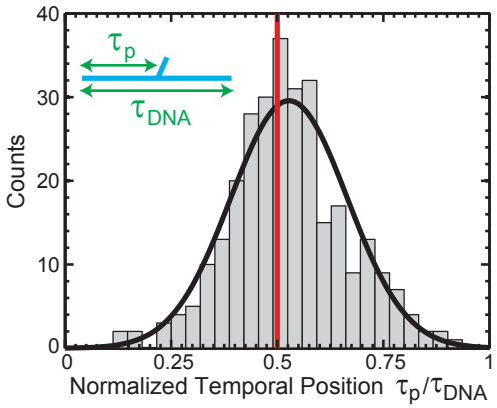


Figure 3.3: Histogram of the normalized temporal position observed for the symmetric construct with a protrusion. The red line is the expected position of the protrusion assuming a constant translocation velocity. The solid black line is a Gaussian fit to the distribution. The distribution has a mean of 0.528 and a standard deviation of 0.137.

Analysis of the translocations for the asymmetric DNA construct reveal that the translocation velocity of the molecule increases as it approaches the end. Figure 3.4a shows the normalized temporal positions observed for the position of the protrusion. The 'protrusion first' orientation (I) has a 114 event distribution with a Gaussian mean of 0.238 and a standard deviation of 0.103, while the 'protrusion

last' orientation (II) has a more narrow distribution of 98 events with a mean at 0.830 and standard deviation of 0.054. If these constructs would translocate with a constant velocity, we would expect the distributions to be centered at 0.208 and 0.792 respectively, which are not within the standard error for the measured means. The small shift observed in the mean values again indicates that the molecules move slower at the start compared to the end. Figure 3.4b shows the average translocation velocity obtained for different segments based on both orientations. We observe that the last 21% of the molecule translocates 19% faster, on average, than the first 21%. The slightly higher mean velocity observed for the first 21% of the molecule (short red segment) compared to the first 79% of the molecule (long blue segment) may suggest that there is a slightly higher velocity right at the start of the translocation process, which was also observed in the simulations of Lu et al[18] and could be due to the low drag at the very start of the translocation process Supplementary Figure 3.13 zooms in on the data for each orientation and shows the average translocation velocity determined for the three different segments of the molecule for each orientation. These results further demonstrate that significant fluctuations in the translocation velocity are present and show that the DNA speed increases during translocation with a strong increase in the local velocity at the end of the translocation process.

Next to the intra- and intermolecular fluctuations in the velocity, we also observe significant variations in the mean translocation velocity between different pores. Although the experiments all concern 10 nm pores that probe the same 7560 bp construct, we observe most-probable translocation times varying from 1.9 ms to 4.8 ms corresponding to velocities of 3.98 to 1.58 bp/μs respectively, as shown in Figure 3.5c. This significant variation can be attributed to DNA-pore interactions, which lead to a characteristic long tail in the translocation time distribution. We also note that pores fabricated in older membranes typically show slower translocation velocities. This effect could be due to more oxides present in the membrane material and requires further investigation. As we have shown, it is however still well possible to determine and directly compare the positions of local protrusions between different experiments by using the normalized temporal position from ensemble distributions. This approach is supported by our finding that the time required to reach the protrusion τ_p scales linearly with the total translocation time of the event τ_{DNA}, as shown in Fig. 3.15.

Table 3.1: The value of the measured normalized temporal position, its standard deviation for three spatial positions, and its standard error of the mean on a 7560 bp DNA molecule.

Spatial Position (bp)	Normalized Temporal Position	STD	SE
1571	0.238	0.137	0.01
3780	0.528	0.103	0.006
5989	0.830	0.054	0.006

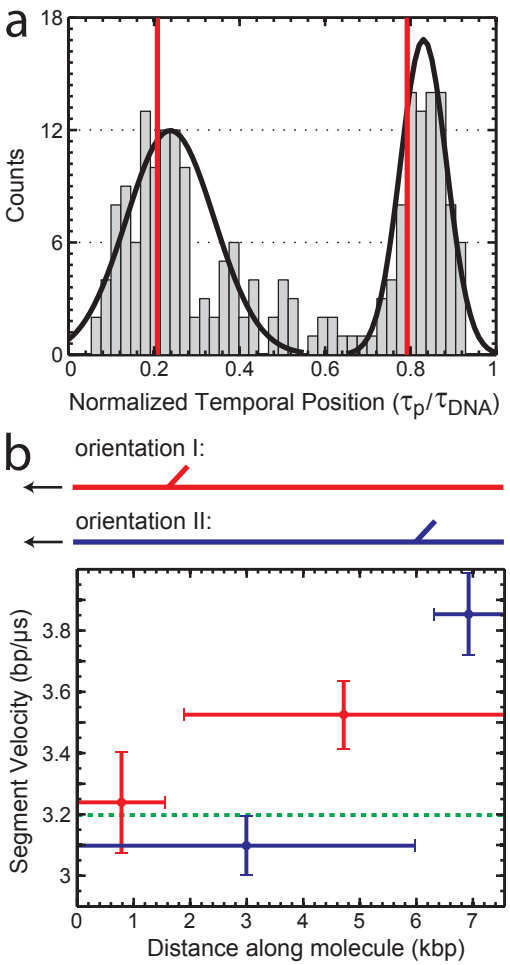


Figure 3.4: a) Histogram of the normalized temporal position observed for the asymmetric construct. We took a cutoff point at 0.6 to separate the two orientations. The distribution of orientation I has a mean of 0.238 and STD of 0.103, while orientation II has a mean of 0.830 and STD of 0.054. Both peaks occur later than the position expected based on a constant velocity (0.208 and 0.792, respectively), which implies the first part of the molecule goes slower while the last part goes faster. b) The mean local translocation velocity over various segments of the asymmetric DNA construct. The horizontal line indicates the length of the segment that was used to elucidate its average speed while the vertical line indicates the standard error. The dashed green line is the mean translocation velocity of the dataset. Red points correspond to orientation I while blue points correspond to orientation II.

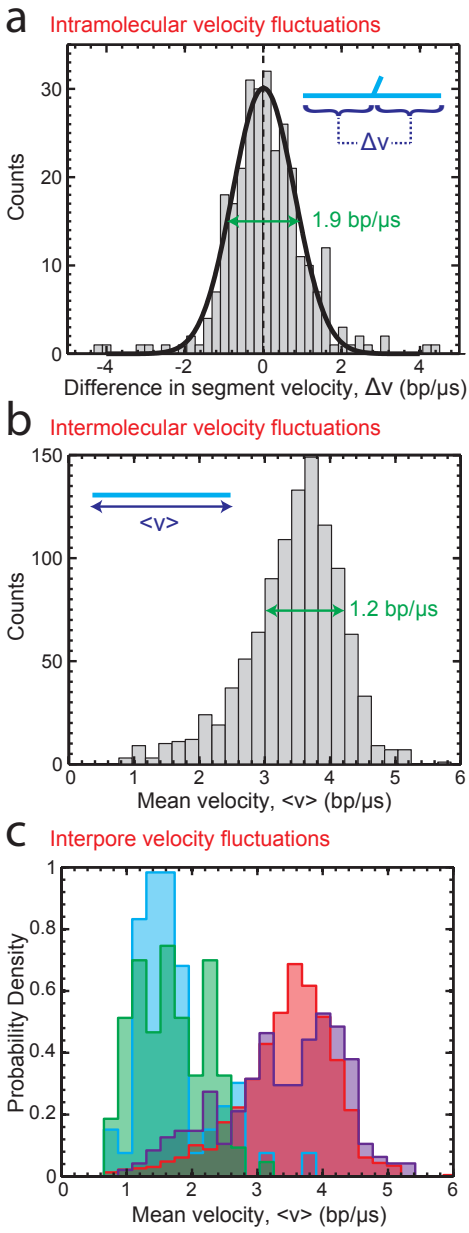


Figure 3.5: Significant velocity fluctuations occur both within molecules (a), among molecules (b), and among different nanopores (c). a) Histogram of the difference in translocation velocity between the last half of the molecule and the first half. Positive values indicate the velocity was faster in the second half of the molecule. The distribution has a mean of 0.015 bp/μs and STD of 0.79. The solid green line shows the FWHM. b) Histogram of the differences observed in the mean velocity for equal length molecules during the same experiment with the control construct containing no protrusion. c) Histograms showing the variation in the mean velocity between different 10 nm pores, *ceteris paribus*. We observe significant differences in the mean velocities for different pores.

3.3. Discussion

How accurately can we determine the spatial position of a local structure (bound protein, side group,...) along the DNA, if we know the temporal profile of the DNA molecule? This question is central to many nanopore applications. The measurements carried out in our study allow us to estimate how accurately position can be determined from either a single measurement or an ensemble of measurements. Figure 3.6 shows the measured normalized temporal position as a function of the spatial position, from the data points of Fig. 3.3 and 3.4a (cf. Table 3.1). We use a cubic spline fit in order to interpolate values and the shaded area shows the range of the standard deviation observed. The solid blue line is the velocity profile that would apply in the case of a constant velocity. We observe that the actual position, is always less than the position estimated assuming a constant velocity, with a statistically significant difference. This approach can be used to quantify the uncertainties in determining spatial position for various measurements, as discussed next.

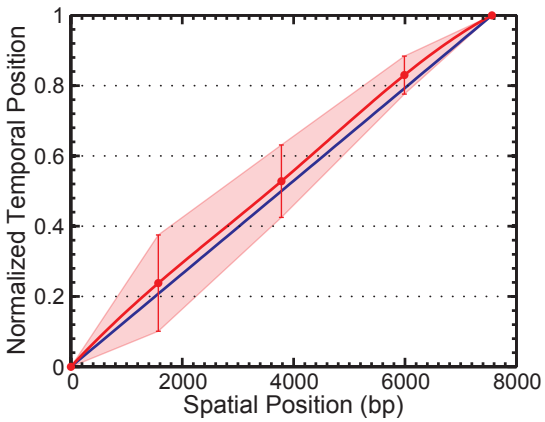


Figure 3.6: The relationship between the normalized temporal position and the spatial position determined for our 7560 bp construct translocating through a 10 nm pore in 4M LiCl. Red points were experimentally measured, while the red line is a cubic spline interpolation, and the shaded area is the interpolated standard deviation. The blue line represents the case of constant velocity. The actual spatial position typically lags behind the position estimated assuming a constant velocity.

Determining the spatial position from a single temporal measurement leads to very large uncertainties, which can be reduced significantly using ensemble measurements. We expect the spatial position corresponding to a single normalized temporal point to fall anywhere within the horizontal intercept of that normalized time point and the shaded red area. For example, if a normalized temporal position of 0.2 is measured, one would estimate that this is caused by a feature at a position along the molecule somewhere between 800 bp to 2200 bp. This result emphasizes the fact that single measurements produce very inaccurate results because the stochastic fluctuations are large. The solution to improving the accuracy thus is to carry out ensemble measurements and fit the resulting distribution of normal-

ized time points, which results in much smaller uncertainties. Using the standard error of the mean for the Gaussian fits (Figure 3.3 and 3.4a), we estimate that the uncertainty in the spatial position (from an ensemble-measured normalized temporal position) to be about 200 bp over the first part of the molecule, which reduces to about 90 bp after the midpoint. The larger uncertainty in the first part of the molecule is due to the wider distribution observed in this region. If the velocity profile is not known or cannot be measured, one could assume a constant velocity (blue line Figure 3.6). What would be the error associated with estimating the spatial position from ensemble-measured temporal data, assuming a constant velocity? Since the value of the mean velocity is slightly higher than the actual local velocity for most of the molecule, the estimated spatial position determined assuming a constant velocity will be between 130 and 330 bp further along the molecule than the true position, at all positions except at the start and the very end of the translocation process. The results demonstrate the importance of knowing the actual velocity profile and provide the first numbers for estimating the uncertainty in the spatial position of a feature along a molecule.

Although these estimates are specific for a 7560 bp long molecule in 4M LiCl, the results apply more generally. In the case of shorter molecules, the speed up at the end of the translocation will concern a larger fraction of the total translocation time, and accordingly there will be a larger difference between the mean velocity and the actual velocity over most of the molecule, and hence larger overestimates of the position if a constant velocity is assumed. In the case of longer molecules the effect will be opposite since the increase in velocity at the end will be an increasingly smaller fraction of the total translocation time and the mean velocity will converge towards the actual velocity over most of the molecule. The effect that the local velocity is smaller than the mean velocity may be even larger than our experiments reveal. In most nanopore measurements on dsDNA, the translocation time is much smaller than the polymer's Zimm relaxation time[21]. Due to the presence of nicks every 42 bp along our DNA-origami construct and the use of very high-salt solutions, the relaxation time of the DNA constructs in these experiments, however, is likely smaller than or similar to their translocation time, which may reduce the magnitude of some of the effects observed.

3.4. Conclusion

This study has introduced a novel method for probing the local velocities of DNA molecules translocating through solid state nanopores using synthetic DNA constructs. This was used to measure the mean velocity over several segments of a 7560 bp DNA molecule. Significant fluctuations are observed in both the intramolecular and intermolecular translocation velocity, as well as between different nanopores of the same diameter. The size of the intramolecular velocity fluctuations is surprisingly large, and they are apparent even when averaged over length scales corresponding to half the total length of the molecule. We also systematically observe an increase in the velocity at the end of the translocation process, an effect attributed to the reduced drag force as the last of the DNA translocates through. We have used the measured velocity profile to estimate the error in determining the

spatial position both if the velocity profile is known or if a constant velocity profile is assumed. These results demonstrate the utility of this approach for illuminating the biophysics of the translocation process.

3.5. Methods

DNA construct self-assembly. The DNA construct was assembled using a circular strand of 7560 bases length derived from the genome of bacteriophage M13[26]. Staple oligonucleotide strands were prepared by solid-phase chemical synthesis (Eurofins MWG, Ebersberg, Germany, HPSF grade). Production of the full DNA construct was accomplished by two separate reactions for assembling the long line backbone and the protrusion. The backbone was assembled in a reaction mixtures containing M13 phage DNA at a concentration of 50 nM, 42 base complementary DNA oligonucleotides at 200 nM each, and 5 mM TRIS, 1 mM EDTA, 20 mM $MgCl_2$ and 5 mM NaCl (pH 8). The reaction mixtures were subjected to a thermal annealing protocol using TETRAD (Biorad) thermal cycling devices. The mixtures were first incubated at 65°C for 15 min and then annealed from 60 to 40°C in steps of 1°C per hour. The protrusions were separately pre-annealed according to the same protocol above with 10mM $MgCl_2$. After separate assembly, the protrusion and backbone were incubated at RT for 12 hours in a ratio of 1.2:1. After the assembly all objects were purified using 100kDa Amicon filters to separate the construct from the excess staple strands. The filter purification was carried out 4 times with a buffer containing 5 mM TRIS, 1 mM EDTA, 5 mM $MgCl_2$ and 5 mM NaCl (pH 8) centrifuging at 2000 rcf for 30 min. The constructs were then linearized with 10U HincII in a total reaction volume of 56.4 μL (for 3 hr at 37°C) and either used as is or purified with an phenol/chloroform extraction and ethanol precipitation.

Nanopore measurements. The synthetic DNA constructs were diluted into solutions of 4M LiCl TE pH8, which facilitates high-resolution nanopore measurements[27]. (Attempts to measure in 1M KCl were hampered by the limited resolution.) In this study, the DNA constructs were translocated through 10 nm nanopores at 100 mV applied voltage. Current traces were digitized at 500 kHz, low-pass Gaussian filtered at 40 kHz, and analyzed with the Transalyzer Matlab package[28]. We selected only non-folded DNA translocation events with a single spike of amplitude I_1 for further analysis.

Acknowledgement

The authors would like to acknowledge Meng-Yue Wu for TEM nanopore drilling, Jaco van der Torre for advice on purification protocols, and Katharina Häußermann for helpful discussions about the assembly strategy. This work was supported by the Netherlands Organisation for Scientific Research (NWO/OCW), as part of the Frontiers of Nanoscience program, as well as the European Research Council under research grant NanoforBio (no. 247072).

3.6. Supplementary Info

3.6.1. Construct Characterization

Our synthetic DNA constructs were characterized using gel electrophoresis, atomic force microscopy (AFM), and nanopore measurements. Figure 3.7 shows the assembly of a DNA construct and the resulting shift in mobility after the attachment of the protrusion. Figure 3.8 shows that the migration speed of a linearized DNA origami construct is similar to a dsDNA molecule of equal length. This indicates no significant difference in the persistence length due to the presence of the nicks that are present every 42 bp in the origami constructs, as any significant reduction in the persistence length would cause the construct to migrate faster in a gel.

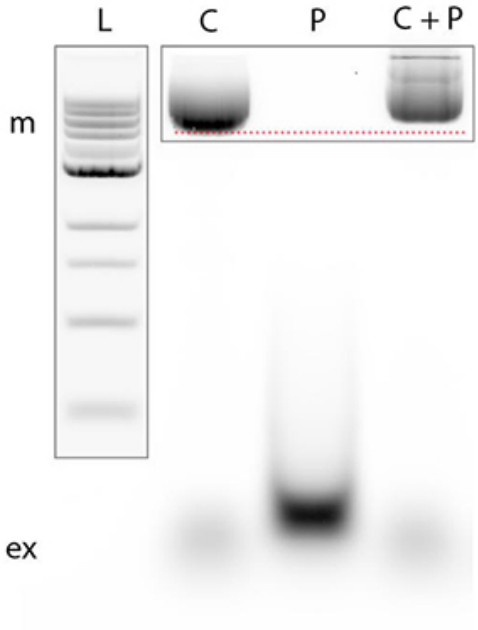


Figure 3.7: Assembly of circular DNA construct with asymmetric protrusion. Labels: **L**, New England Biolabs 1kB DNA ladder; **C**, folded circular DNA object without protrusion; **P**, pre-annealed protrusion; **P+C**, mixture of P and C in 1:1 after incubation at RT for 1 day; **m**, folded objects; **ex**, non-integrated excess staple strands. Marked regions of interest were auto-leveled. The dashed red line assists to note the mobility shift after addition of the protrusion.

AFM was used to image the DNA origami constructs as shown in Figure 3.9 and 3.10. Protrusions are marked with a red arrow. Samples with 10 mM Mg^{2+} were incubated on freshly cleaved mica for several min, after which they were rinsed with water and dried with nitrogen gas. Imaging was carried out with a Nanoscope IV in tapping mode with Olympus AC160TS tips.

In order to investigate the persistence length of the DNA origami constructs, we measured the end-to-end distances of 58 molecules, as shown in Figure 3.11. It has previously been shown that this approach can be used to estimate the persistence

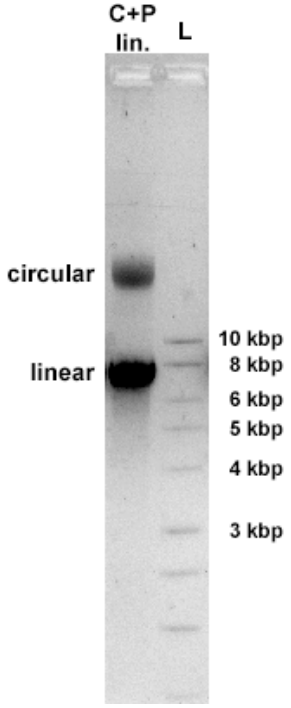


Figure 3.8: The 7560 bp asymmetric DNA construct after linearization (lane **C+P lin.**) alongside a Promega BenchTop 1 kbp DNA ladder (lane **L**). The linear DNA origami construct migrates accurately at the expected size for equivalent-length dsDNA molecules, indicating no significant reduction in the persistence length compared to dsDNA.

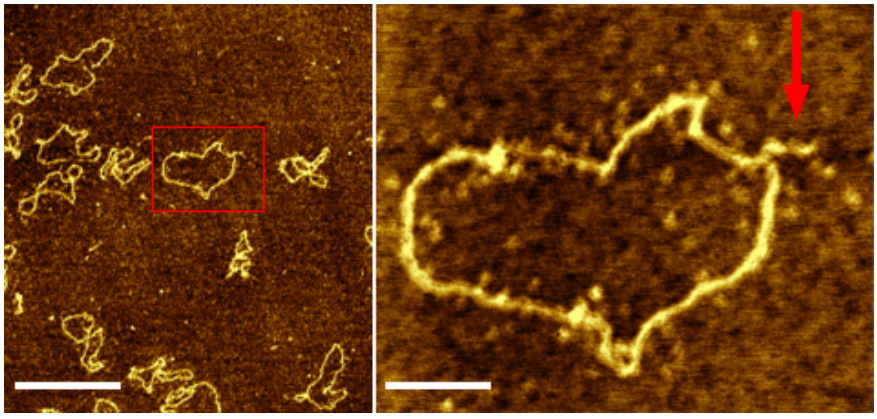


Figure 3.9: AFM scans of some DNA origami constructs before linearization. Arrow indicates the protrusion. Scale bar denote 1 µm (left) and 200 nm (right).

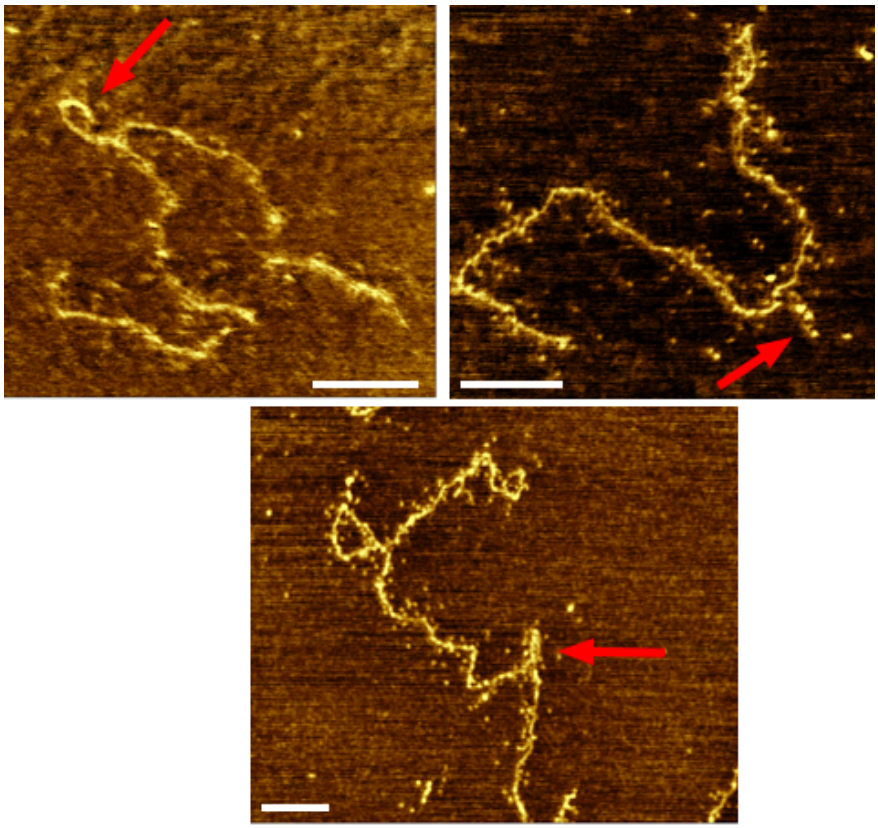


Figure 3.10: AFM scans of some DNA origami constructs after linearization. Arrows indicate the protrusion. Scale bars denote 200 nm.

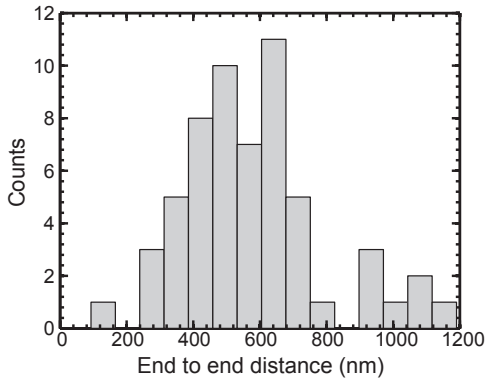


Figure 3.11: The end-to-end distance measured for 58 DNA construct molecules after linearization. We find a root-mean-square end-to-end distance of 612 nm.

length of dsDNA on a 2D surface, given that DNA molecules have been found to re-equilibrate on the surface in AFM measurements in the presence of Mg^{2+}.[29] We find a root-mean-square end-to-end distance (RMSD) of 612 nm. We compare this with several values given by the worm-like chain (WLC) and freely-jointed chain models, as shown in Table 3.2. We can determine the 2D end-to-end distance of a WLC polymer on a 2D surface using

$$< R^2 >_{2D} \cong 4L_pL\left(1 - \frac{2L_p}{L}\right) \quad \text{for} \quad L >> L_p \tag{3.1}$$

where L is the contour length and L_p is the persistence length of the molecule[29]. Assuming WLC behavior and using Eq. 3.1, we thus find a persistence length of 38 nm for our DNA constructs. This value is not significantly lower than the value expected for dsDNA (50 nm), while it is very much larger than values that can be expected for a freely-jointed chain model, both if one would insert a persistence length of ~1nm as expected for ssDNA, or 14 nm for a freely hinging chain of 42bp segments. Indeed, our DNA constructs do not behave like freely-jointed chains that would result in much smaller values for the RMSD (Table 3.2). These experimental data clearly indicate that the synthetic DNA constructs used should recapture the same translocation behavior as regular dsDNA molecules.

Table 3.2: A comparison of the measured root-mean-square end-to-end distance (RMSD) with the values predicted by several models for L = 2570 nm.

AFM measured RMSD for DNA constructs on a mica surface	612 nm
Worm-like chain 2D RMSD using Eq. S1 with L_p = 50 nm	703 nm
Worm-like chain 3D RMSD with L_p = 50 nm	866 nm
Freely-jointed chain 3D RMSD with segment length of 50 nm	507 nm
Freely-jointed chain 3D RMSD with segment length of 14 nm (42 bp)	192 nm

3.6.2. Yield

The gel shown in Figure 3.7, shows that the initial assembly reaction went to completion. We quantified the final yield using nanopore measurements, as shown in Table 3.3. The value of the yield is given as the percentage of unfolded translocation events with a protrusion out of all unfolded translocation events. We observe yields ranging from 29% to 85%. Several factors affect the final yield. **i)** The protrusion is first assembled through the hybridization of six oligomers. Any excess oligomers from this mixture may interfere with the binding of the protrusion if they hybridize to the M13 scaffold strand before a fully assembled protrusion does. **ii)** Since the protrusion is mixed with the scaffold strand at a 1:1 ratio, the yield may be affected by uncertainties in the stoichiometry of the oligomers and the presence of some protrusions which are not fully assembled. **iii)** We observe that purifications, which are required to remove excess oligomers and restriction enzymes, reduce the final yield (Table 3.3). These factors all contribute to the final observed yield.

Table 3.3: The final yield of the linear DNA construct with a side protrusion, as determined using nanopore measurements.

Assembly #	Construct	Yield (%)	Number of purifications after assembly
1	asymmetric	38%	2
		29%	
2	asymmetric	53%	1
3	symmetric	84%	1
		85%	

3.6.3. Velocity Data

Figure 3.12 shows histograms of the translocation velocities measured for each of the three segments of the symmetric construct. The most probable translocation velocity is similar for all three with a value of around 2 bp/μs. A small shift towards higher velocities is seen in the last segment relative to the first segment which can be attributed to the speed up at the end of the translocation process. Figure 3.13 shows the mean local translocation velocity over various segments of the asymmetric DNA construct, including the protrusion. Figure 3.14 shows scatter plots, with the current blockade versus total translocation time, for the asymmetric construct for both events with and without the protrusion.

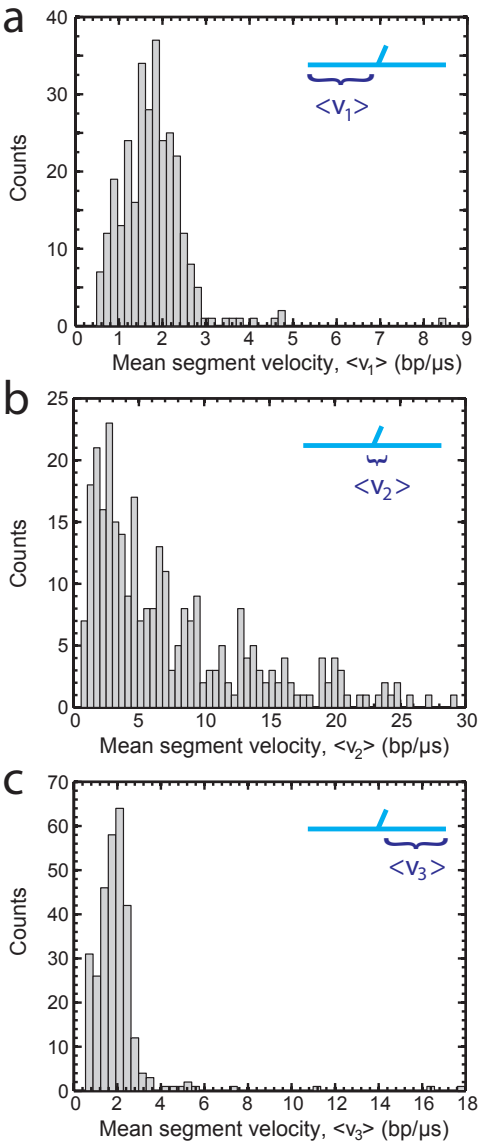


Figure 3.12: The segment velocities observed for each of the three segments of the symmetric construct.

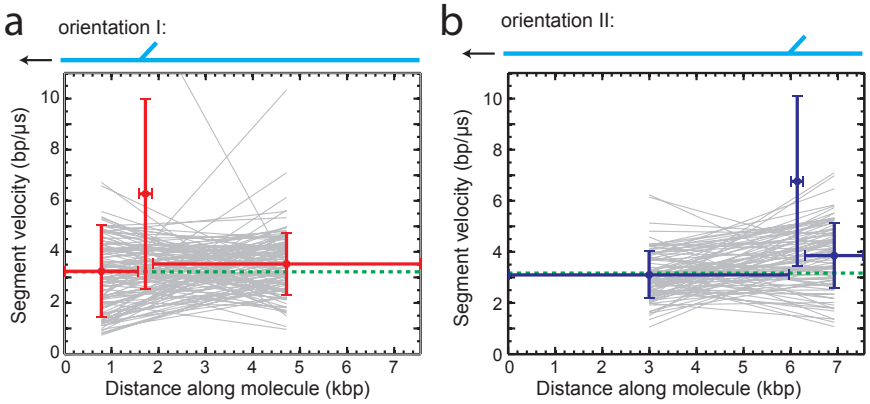


Figure 3.13: The mean translocation velocity over various segments of the asymmetric DNA construct, for each orientation. Each horizontal line indicates the length of the segment that was used to elucidate its average speed, while the vertical line indicates the standard deviation for that segment. The dashed green line is the mean translocation velocity of all data in the dataset. Changes in the local velocity in the first and last segment are overlaid as grey lines where each line is the data for each molecule.

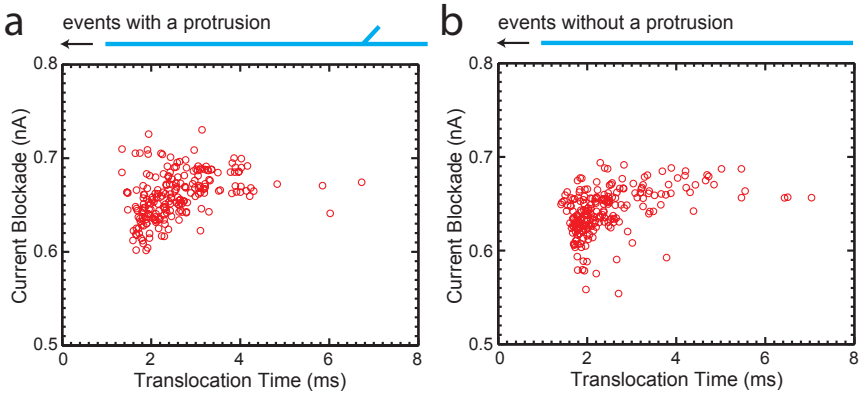


Figure 3.14: Scatter plots showing the current blockade and total translocation time for the asymmetric DNA construct, showing both events with and without a protrusion.

3.6.4. Scaling of τ_p of with τ_{DNA}

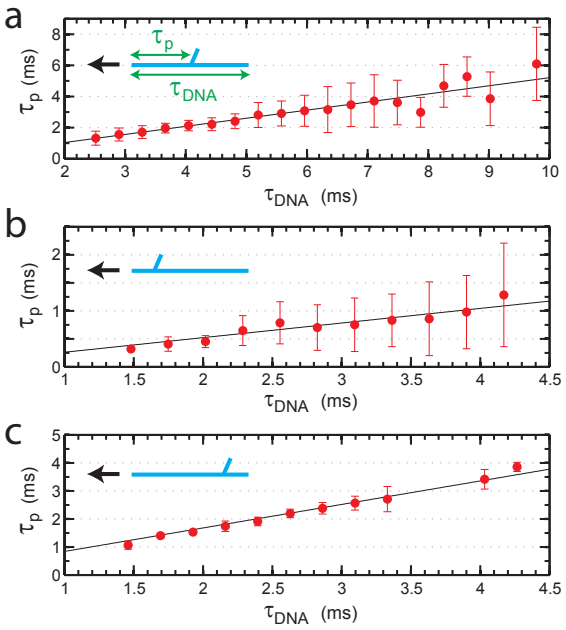


Figure 3.15: The linear scaling observed between the time required to reach the protrusion τ_p and the total translocation time of the event τ_{dna} for the **a)** symmetric construct **b)** asymmetric construct in orientation I, and **c)** asymmetric construct in orientation II. Solid lines are linear fits with an intercept of zero and slopes of **a)** 0.521, **b)** 0.261, and **c)** 0.839.

3.6.5. Protrusion and M13 Backbone Oligomer Sequences

Oligomer sequences are available in the online supplementary information for the publication Plesa et al.[30].

References

[1] M. Wanunu, *Nanopores: A journey towards DNA sequencing,* Physics of Life Reviews **9**, 125 (2012).

[2] F. Haque, J. Li, H.-C. Wu, X.-J. Liang, and P. Guo, *Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA,* Nano Today **8**, 56 (2013).

[3] J. Li, D. Stein, C. McMullan, D. Branton, M. J. Aziz, and J. A. Golovchenko, *Ion-beam sculpting at nanometre length scales,* Nature **412**, 166 (2001).

[4] A. J. Storm, J. H. Chen, X. S. Ling, H. W. Zandbergen, and C. Dekker, *Fabrication of solid-state nanopores with single-nanometre precision,* Nat Mater **2**, 537 (2003).

[5] A. T. Carlsen, O. K. Zahid, J. Ruzicka, E. W. Taylor, and A. R. Hall, *Interpreting the conductance blockades of DNA translocations through solid-state nanopores,* ACS Nano **8**, 4754 (2014).

[6] G. Ando, C. Hyun, J. Li, and T. Mitsui, *Directly observing the motion of DNA molecules near solid-state nanopores,* ACS Nano **6**, 10090 (2012).

[7] D. Fologea, J. Uplinger, B. Thomas, D. S. McNabb, and J. Li, *Slowing DNA translocation in a solid-state nanopore.* Nano Letters **5**, 1734 (2005).

[8] S. W. Kowalczyk, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Modeling the conductance and DNA blockade of solid-state nanopores,* Nanotechnology **22**, 315101 (2011).

[9] R. M. M. Smeets, U. F. Keyser, D. Krapf, M.-Y. Wu, N. H. Dekker, and C. Dekker, *Salt dependence of ion transport and DNA translocation through solid-state nanopores,* Nano Letters **6**, 89 (2006).

[10] C. Plesa, S. W. Kowalczyk, R. Zinsmeester, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Fast translocation of proteins through solid state nanopores,* Nano Letters **13**, 658 (2013).

[11] J. Larkin, R. Y. Henley, M. Muthukumar, J. K. Rosenstein, and M. Wanunu, *High-bandwidth protein analysis using solid-state nanopores,* Biophysical Journal **106**, 696 (2014).

[12] C. Plesa, A. N. Ananth, V. Linko, C. Gülcher, A. J. Katan, H. Dietz, and C. Dekker, *Ionic permeability and mechanical properties of DNA origami nanoplates on solid-state nanopores,* ACS Nano **8**, 35 (2013).

[13] R. Wei, T. G. Martin, U. Rant, and H. Dietz, *DNA origami gatekeepers for solid-state nanopores,* Angewandte Chemie, International Edition **51**, 4864 (2012).

[14] N. A. W. Bell, C. R. Engst, M. Ablay, G. Divitini, C. Ducati, T. Liedl, and U. F. Keyser, *DNA origami nanopores,* Nano Letters **12**, 512 (2012).

[15] A. T. Carlsen, O. K. Zahid, J. A. Ruzicka, E. W. Taylor, and A. R. Hall, *Selective detection and quantification of modified DNA with solid-state nanopores,* Nano Letters **14**, 5488 (2014).

[16] A. Ivankin, S. Carson, S. R. M. Kinney, and M. Wanunu, *Fast, label-free force spectroscopy of histone–DNA interactions in individual nucleosomes using nanopores,* Journal of the American Chemical Society **135**, 15350 (2013).

[17] G. V. Soni and C. Dekker, *Detection of nucleosomal substructures using solid-state nanopores,* Nano Letters **12**, 3180 (2012).

[18] B. Lu, F. Albertorio, D. P. Hoogerheide, and J. A. Golovchenko, *Origins and consequences of velocity fluctuations during DNA passage through a nanopore,* Biophysical Journal **101**, 70 (2011).

[19] X. J. A. Janssen, M. P. Jonsson, C. Plesa, G. V. Soni, C. Dekker, and N. H. Dekker, *Rapid manufacturing of low-noise membranes for nanopore sensors by trans -chip illumination lithography,* Nanotechnology **23**, 475302 (2012).

[20] M. Mihovilovic, N. Hagerty, and D. Stein, *Statistics of DNA capture by a solid-state nanopore,* Physical Review Letters **110**, 028102 (2013).

[21] A. J. Storm, C. Storm, J. Chen, H. Zandbergen, J.-F. Joanny, and C. Dekker, *Fast DNA translocation through a solid-state nanopore,* Nano Letters **5**, 1193 (2005).

[22] C. Plesa, L. Cornelissen, M. W. Tuijtel, and C. Dekker, *Non-equilibrium folding of individual DNA molecules recaptured up to 1000 times in a solid state nanopore,* Nanotechnology **24**, 475101 (2013).

[23] A. Singer, S. Rapireddy, D. H. Ly, and A. Meller, *Electronic barcoding of a viral gene at the single-molecule level,* Nano Letters **12**, 1722 (2012).

[24] P. W. K. Rothemund, *Folding DNA to create nanoscale shapes and patterns,* Nature **440**, 297 (2006).

[25] C. Plesa, D. Verschueren, J. W. Ruitenberg, M. J. Witteveen, M. P. Jonsson, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Direct observation of DNA knots using solid state nanopores; submitted,* (2015).

[26] S. M. Douglas, H. Dietz, T. Liedl, B. Hogberg, F. Graf, and W. M. Shih, *Self-assembly of DNA into*

nanoscale three-dimensional shapes, Nature **459**, 414 (2009).
[27] S. W. Kowalczyk, D. B. Wells, A. Aksimentiev, and C. Dekker, *Slowing down DNA translocation through a nanopore in lithium chloride,* Nano Letters **12**, 1038 (2012).
[28] C. Plesa and C. Dekker, *Data analysis methods for solid-state nanopores; submitted,* (2015).
[29] C. Rivetti, M. Guthold, and C. Bustamante, *Scanning force microscopy of DNA deposited onto mica: Equilibration versus kinetic trapping studied by statistical polymer chain analysis,* Journal of Molecular Biology **264**, 919 (1996).
[30] C. Plesa, N. van Loo, P. Ketterer, H. Dietz, and C. Dekker, *Velocity of DNA during translocation through a solid state nanopore; in press,* Nano Letters (2014).

4

Direct observation of DNA knots using solid state nanopores

Long DNA molecules can self-entangle into knots. Experimental techniques to observe such DNA knots (primarily gel electrophoresis) are limited to bulk methods and circular molecules below 10 kbp in length. Here we show that solid-state nanopores can be used to directly observe individual DNA knots in both linear and circular single molecules of arbitrary length. DNA knots are observed as short spikes in the nanopore current traces of traversing DNA molecules. The observation of knots is dependent on sufficiently high measurement resolution, which can be achieved using high-concentration LiCl buffers. We study the percentage of DNA molecules with knots for different DNA molecules, up to 166 kbp in length. We find that the knotting probability rises strongly with length, and compare our experimental data to simulation-based predictions for long polymers. From the translocation time of the knot through the nanopore, we estimate that the majority of the DNA knots are tight, with small sizes below 100 nm. In the case of linear molecules, we observe that knots are able to slide out upon applying high driving forces (voltage). Our results demonstrate that the solid-state nanopore technique can provide a wealth of information about the position and the size of knots, including the number of DNA strands inside DNA knots.

C. Plesa, D. Verschueren, J. W. Ruitenberg, M. J. Witteveen, M. P. Jonsson, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Direct observation of DNA knots using solid state nanopores*, submitted.

4.1. Introduction

It is well established that long polymers are subject to increasing self-entanglement as their length becomes larger. DNA knots have been the focus of significant study in polymer physics[1, 2] and, although they are ubiquitous in nature[3], their exact role in biological processes is still under investigation[4–6]. Despite their interest, knotting remains among the least understood properties of polymers due to a lack of both experimental techniques to observe them as well as rigorous theoretical approaches to describe and characterize them. Many open questions remain[7], such as what determines the characteristic chain length beyond which knots become prevalent upon cyclization, knot localization[8, 9], the existence of metastable tight knots[10], and many others.

A number of experimental techniques have been developed to study knots, particularly in DNA. Knots have been induced with high-electric fields[11], optical tweezers[12, 13], topoisomerase enzymes[14–16], DNA recombinases[17, 18], and through the cyclization of linear DNA molecules[19, 20]. Electron microscopy[14] (EM) and atomic force microscopy[21] (AFM) have been used to image knots with excellent resolution but these techniques are limited to small molecules and low statistics. Optical techniques[11, 12] have been used to introduce and study the behavior of knots in DNA strands. Gel electrophoresis[15–17, 19, 20], the dominant tool used in knot studies, is a bulk technique where knots are trapped in circular molecules which are limited to lengths below 10 kbp. Here we introduce a new single-molecule technique capable of directly observing individual knots, in both linear and circular molecules with arbitrarily long lengths.

Solid-state nanopores have emerged as an important tool with a large number of potential applications[22, 23] at the crossroads of physics, biology and chemistry. Nanopores have provided a method to investigate various concepts in polymer physics such as polymer translocation through pores[24, 25], the Zimm relaxation time[25, 26], and the polymer capture process[27, 28]. In nanopore sensing, biomolecules in an aqueous solution are placed into one of two reservoirs separated by a membrane containing a nanometer scale pore, as shown in Figure 4.1. Subsequent application of voltage using electrodes bathed in each reservoir produces an electric field and a resulting electrophoretic force on charged molecules such as DNA, causing them to be pulled through the nanopore. The presence of a molecule in the pore causes a blockade in the ionic current that (in high salt conditions) is proportional to the volume of the segment of the molecule in the pore. For long polymers, the translocation process is much faster than the typical relaxation time[29], which allows us to investigate the polymers before they have a chance to relax. This enables us to probe long DNA molecules for the presence of topological structures such as knots. Several theoretical papers have addressed some of the issues arising from the presence of knots in molecules translocating through nanopores[30, 31]. Here we report the first direct experimental observation of knots in measurements on long dsDNA molecules that translocate through solid-state nanopores.

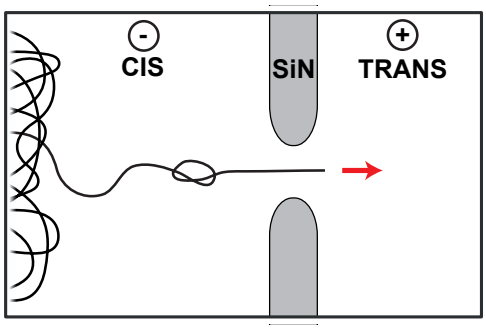


Figure 4.1: A schematic illustration of a DNA molecule with a trefoil (3_1) knot translocating through a solid-state nanopore.

4.2. Results and Discussion

Knots produce distinctive signatures in the current blockade of translocating DNA molecules. When a DNA molecule passes through the pore, it causes a blockade with magnitude I_1 in the current level as shown in Figure 4.2a. If the molecule enters the pore in a folded configuration this produces a blockade level twice as high ($2I_1$) for the duration of the fold, as shown in Figure 4.2b. These folds occur primarily at the start of the translocation process and their probability of occurrence has been shown to increase as the capture point gets closer to the end of the molecule[28]. High-resolution investigation of DNA translocation events reveals the presence of additional sharp blockage spikes with a high amplitude and very short duration occurring within a fraction of the events. Examples of such events are shown in Figure 4.2c-d,f. We observe such events in both linear DNA (Fig 4.2cd) and circular dsDNA (Fig 4.2f). These spikes occur only within the DNA translocation events and have an amplitude (on top of the DNA blockade level I_1) which is an integer multiple of $2I_1$. Such spikes can only be associated to two types of molecular configurations: an internal fold within the molecule, or a topologically constrained DNA knot. The former configuration of local folds has extremely low rates of occurrence due to the nature of the translocation process, which can be seen as follows. The nanopore can be thought of as applying a point force to the polymer at the location of the nanopore, since the electric field strength is highest in the pore and strongly decreases as $1/r^2$ away from the pore[32]. A long DNA molecule coils up into a polymer blob that has a large size (many hundreds of nm) compared to the nm-size nanopore. For translocation, a DNA end is pulled though the pore, thus disentangling the blob. The tension in the DNA strand resulting from the pulling action towards the pore propagates along the strand outwards from the nanopore, which pulls out any internal folds (since the velocity of the leading (captured) DNA is much higher than that of the lagging DNA (behind the fold)). Knots, on the other hand, cannot unfold due to the topological constraints imposed and are pulled towards and translocated through the nanopore. This argument that the observed blockades are due to knots as opposed to folds is further supported by the small size observed for the majority of knots (see below), whereas local folds

would have a wide distribution of larger sizes.

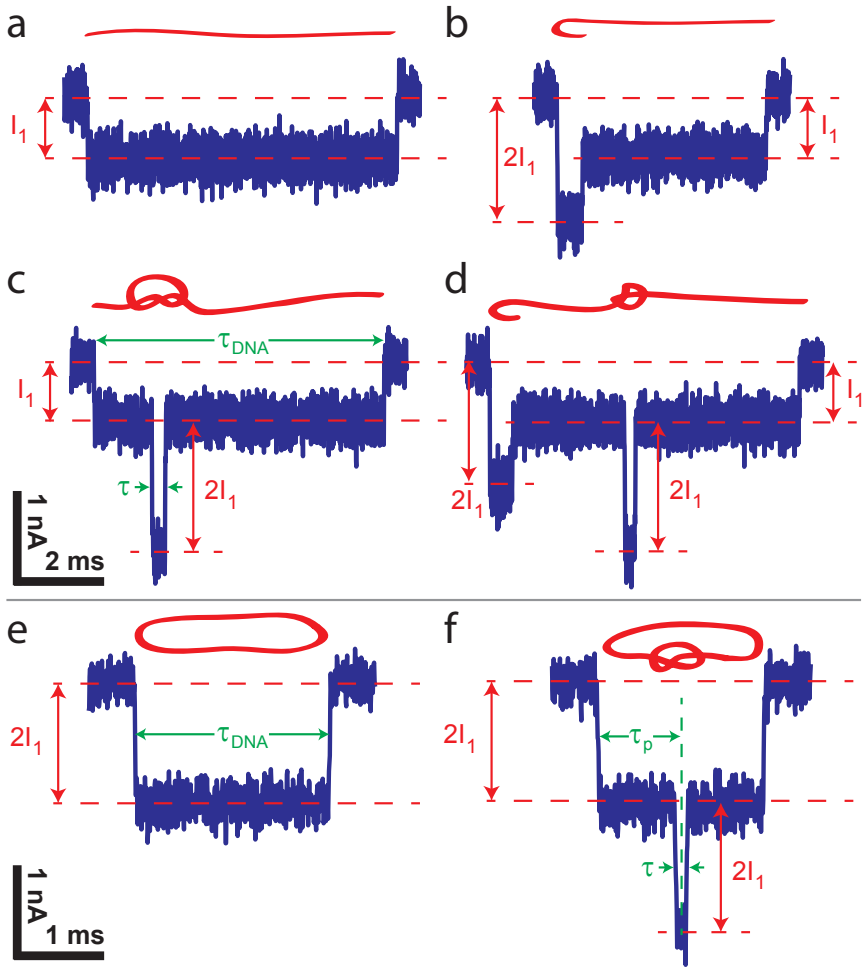


Figure 4.2: Six example events for lambda DNA translocating through a 10 nm pore at 200 mV in 2M LiCl at 30kHz bandwidth. The molecular configuration attributed to each type of event are shown above each current trace in red. a) Current trace of an unfolded event. b) Current trace of an event with a fold at the start. c) Current trace of an event with an internal blockade which can be associated with a knot. d) Another event similar to c but with an additional fold at the start. e) Current trace of an unfolded circular molecule translocating. f) Current trace of a circular molecule containing a knot.

The observed current signatures allow us to probe the knotting probability, knot size, and knot position. The $2I_1$ magnitude of the spikes provides strong evidence for the occurrence of DNA knots, indicating the presence of 3 double-stranded DNA segments. Although the knot amplitude can be used to determine the number of DNA strands inside the pore, the current signatures for the knots do not provide information about the crossing number of the knot. Different knot types (trefoil, figure-eight,...) have the same number of DNA strands when linearly stretched out

(without changing the topology of the knot) and accordingly can have the same amount of DNA strands inside the pore during translocation. As a consequence, our experimental results lead to statements encompassing many knot types. Most spikes that we observe consistently have an amplitude of $2I_1$ beyond the normal DNA blockage level. Occasionally, however, we observe events during which the additional current blockade has an amplitude of only I_1 (beyond the normal DNA blockage level), which may be due to the presence of replication forks, that occur due to the nature of the plasmid replication process, or due to simultaneous co-translocation of two different molecules. This latter phenomenon is only significant if either a high concentration of DNA is used or if there is a significant amount of smaller DNA fragments due to handling. For circular DNA molecules, DNA catenanes may be possible as well.

In common nanopore experiments, knots are hard to observe since their small size (< 100 nm; see below) makes their observation very sensitive to the resolution of the solid-state nanopore measurement. Most nanopore experiments are carried out in 1M KCl[25–28], where most observed knots have a duration (τ) of 15 μs at 100 mV, indicating that their detection is at the edge of what is resolvable (Supplementary Section 4.5.2). For this reason the majority of the experiments presented in this study are carried out in solutions of 2M or 4M LiCl, which as we have previously shown, can increase the translocation time of DNA by a factor of 7x or 10x respectively, relative to 1M KCl[33]. Furthermore, the conductance blockades are also higher in 2M and 4M LiCl, which leads to a higher signal-to-noise-ratio (SNR) and allows for higher measurement bandwidths. Indeed, for 48.5 kbp DNA molecules at the same conditions, we observe fewer than 50% of the knots in 1M KCl relative to measurements in 2M LiCl solution, as further discussed below.

We measured molecules of different length to study how the knotting probability scales as a function of DNA length, and we compare the observed rates of knot occurrence to previous simulation-based predictions. Four different DNA lengths were used: 2686 bp pUC19 linearized with XmnI, a 20678 bp linearized plasmid, 48502 bp lambda phage DNA, and 165648 bp phage T4 GT7 DNA. Figure 4.3 shows the measured knotting probabilities and their standard deviation for linear molecules as a function of DNA molecule length (Supplementary Section 4.5.1). A strong non-linear increase is observed. For 2686 bp molecules in 4M LiCl we find a knotting probability of 1.8%, while for the 20678 bp molecules we observe 13.8$\pm$1.0% of molecules with knots. For 48502 bp molecules we find 26.8$\pm$3.4% of molecules to contain knots, while similar measurements carried out in 2M LiCl find a 24.4$\pm$4.6% knotting occurrence (Supplementary Section 4.5.6). As mentioned above, data taken in 1M KCl showed significantly lower knotting levels, an effect attributed to the limited resolution in these measurements. Nevertheless, it can provide a lower bound for the knot probability. For the 165648 bp DNA this lower bound was measured to be 51.8$\pm$1.9%. These results provide the first experimental validation of the knotting abundances in long polymers.

The observed rates of knotting occurrence are co-plotted in Figure 4.3 alongside the simulation-based predictions of Deguchi and Tsurusaki[34] and Rybenkov et al[19] (Supplementary Section 4.5.7). Neither of these simulations include complex

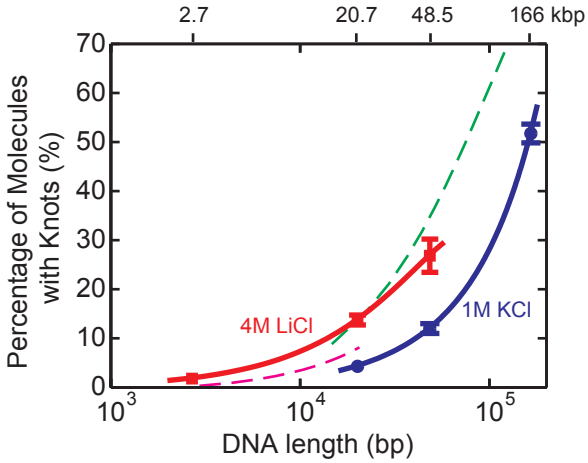


Figure 4.3: Percentage of events with knots as a function of DNA length for measurements carried out in 4M LiCl (red-squares) and 1M KCl (blue-circles). The 1M KCl data provide a lower limit of the knot occurrence due to its low resolution. The dashed lines represent knot occurrence predictions based on simulations by Deguchi and Tsurusaki[34] (green) and Rybenkov et al.[19] (magenta) (Supplementary Section 4.5.7). The solid lines are cubic spline interpolations.

knots and thus can only serve as a lower bound since the amount of more complex knots increases significantly as a function of the polymer length. For 2686, 20678, 48502, and 165648 bp DNA molecules these simulations predict knotting rates of 0.2%, 14%, 35%, and 80% respectively. Qualitatively we thus clearly see that the simulation-based predictions show the same overall trend as the experiments. Since our measurements are carried out in very high ionic strength conditions where the electrostatic screening is very strong, we expect the effective diameter to be very close to the physical diameter of DNA. This leads to higher knotting probabilities relative to those expected in the case of lower ionic strength regimes, such as those found at physiological conditions[19].

Since these measurements are carried out on linear molecules, we investigated the possibility of knots slipping out before being captured. Once a knot reaches the pore, translocation can only occur after the knotted strand has been bent to a size set by the pore diameter, i.e. on scales below the persistence length of dsDNA. If this process does not occur quickly enough, the low-friction DNA-DNA interactions could allow a knot to slip out[12, 30, 35] by remaining at the pore entrance while the non-knotted DNA strand translocates through. In circular DNA molecules, however, the closed curve topology prevents such slipping out and knots are intrinsically trapped. We investigated the position of the observed knots in 20.7 kbp relaxed circular molecules where any knots contained inside would not be able to escape, although within a translocation event they might slip until the end. Figure 4.4 shows the normalized center position of the knots (τ_p/τ_{DNA}) in a 20 nm pore at 4M LiCl for applied voltages of 100 mV and 200 mV. We observe that knots occur at random positions for 100 mV, whereas at 200 mV they indeed occur with a

strong preference for the end of the molecule that translocates through the pore last. We thus see a clear indication that the higher voltage causes knots to slide towards the end of the molecule. As a consequence, the numbers for the knot occurrence in linear molecules (Figure 4.3) provide a reasonable value at low applied voltage (100 mV), but should be treated as a lower bound to the equilibrium knotting levels at higher voltages, where a significant amount of knots may slip out. In order to quantify the amount of knots that slip out, we carried out experiments using lambda phage DNA (48.5 kbp). This molecule has a 12 bp complementary overhang allowing it to exist in both linear and circular forms depending on temperature and salt conditions[36]. We heated solutions of lambda DNA and quickly cooled them while in 2M LiCl, to form mixed populations of linear and circular DNA from which to compare the knotting occurrence. Previously Rybenkov et al.[19] observed no temperature dependence for the knotting probability, allowing us to attribute any differences between the knotting occurrence in the two populations to knots slipping out. We observed a 55% higher knotting occurrence in the circular molecules compared to the linear ones (Supplementary Section 4.5.6). These results indicate that knots are able to slip out of linear DNA molecules during their translocation through nanopores.

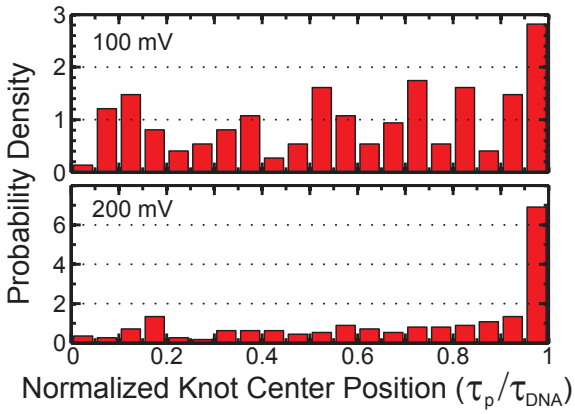


Figure 4.4: Normalized center position of knots observed in 20.7 kbp relaxed circular molecules at 100 mV (top) and 200 mV (bottom). τ_p/τ_{DNA} = 0 denotes the start of the translocation and τ_p/τ_{DNA} = 1 the end of the translocation event. As the voltage is increased, we see knots sliding towards the end of the molecule.

We estimated the size of knots from the time required for the knots to translocate through the nanopore. Figure 4.5 shows the measured time duration of knots on a log scale for linear 20.7 kbp molecules at 100 mV. The vertical dashed line indicates the filtering distortion point ($2T_r$) while the solid line is a fit to the model described below. We estimate the size of the knots to be in the tens of nm at both 100 mV and 200 mV (Supplementary Section 4.5.4) applied voltage, assuming mean translocation velocities (Supplementary Section 4.5.5) of 872 nm/ms and 1937 nm/ms respectively, as shown on the top x-axis of Figure 4.5. Note that such a small knot size indicates that the DNA knots in these long DNA polymers are re-

markably tight. The numbers may underestimate the size of the knots somewhat, especially for those which occur at the end of the translocation process where we know that the velocity is higher than average[37]. Measurements on circular versions of the same molecules in the same conditions reveal similar distributions of very tight knots (Supplementary Section 4.5.4). From all our data, we conclude that the majority of DNA knots have a size below 100 nm. The observation of these tight knots provides evidence for the occurrence of metastable tight knots which have been predicted to exist at equilibrium in long polymers[10]. Given the nature of the translocation process, it remains to be determined how much these knots are being tightened relative to their equilibrium sizes although the similar numbers for the data at 100 vs 200 mV and linear vs circular DNA suggest that the tight knots are intrinsic.

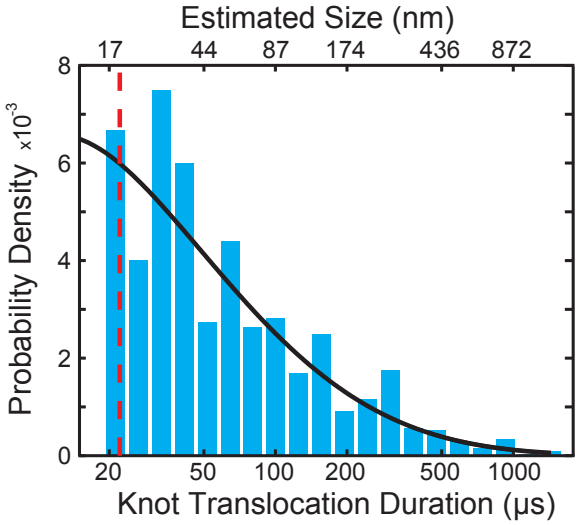


Figure 4.5: Translocation duration (τ) of knots observed in 20.7 kbp linear molecules at 100 mV, plotted on a logarithmic time scale. Solid line is a fit to the model described in the main text, while the vertical dashed line indicates the filtering distortion point ($2T_r$). The top x-axis provides an estimated knot size scale based on the mean translocation velocity of each population. The majority of the knots have very short durations and thus small sizes, indicating that we are observing primarily tight knots.

The knot translocation duration distribution $P(\tau)$ is consistent with a model based on the previously proposed[10] free energy penalty for knot formation, i.e., $P(\tau) = A \cdot exp(-c_1/\tau - c_2\tau^{1/3})$, where A is the normalization factor and the two terms in the exponential reflect the bending energy and confinement entropy contributions, respectively. Although this distribution was originally derived for the knot length and knot size, we adopt it here to characterize the knot translocation duration which is assumed to scale linearly with knot size. Fits were made to the data of 20678 bp DNA in 4M LiCl at 100 mV in 20 nm pores, see Figure 4.5. This yields c_2 = 0.61$\pm$0.09, while it is hard to obtain reliable estimates for c_1, as fitted values range from 0.3 to 18. Physically, coefficient c_1 controls the decay of prob-

ability for very small times τ, i.e., for very tight knots, and the large uncertainty in its magnitude is due to knots slipping out as well as the limited temporal resolution of the measurements[38]. The tail of the distribution is consistent with the functional form $exp(-c_2\tau^{1/3})$ which theoretically derives from DNA undulations inside the effective tube formed by the knot.

4.3. Conclusion

We have demonstrated that it is possible to detect knots in DNA molecules with solid-state nanopores and we have used this approach to estimate the knotting occurrence for a variety of DNA molecules with lengths far longer than previously possible. The single-molecule nature of the technique allows us to observe and analyze individual knots while still being able to generate population statistics. The measurements reveal that knots are capable of slipping out of linear molecules during translocation at high driving voltage, so care must be taken when comparing observed knotting rates to those predicted for molecules at equilibrium. Additionally, the knot translocation duration can be used to estimate the knot size, which is observed to be less than 100 nm for the majority of knots, although a further understanding of the translocation process is required to accurately relate this to the equilibrium knot size. From a nanopore applications perspective, efforts to sequence or detect DNA-bound proteins with nanopores will have to take into account the presence and effects of these knots. These results present a major step towards the ability to directly interrogate polymer knots and increasing our understanding of this ubiquitous phenomenon.

4.4. Methods

Solid-state nanopores were fabricated and used as described previously[39]. All buffers were pH 8 with 10 mM Tris and 1 mM EDTA. Data analysis was carried out using custom Matlab scripts described in detail elsewhere[40]. Lambda phage DNA and T4 GT7 DNA were purchased from Promega (Madison, WI) and Nippon Gene (Toyama, Japan) respectively. A 20.7 kbp plasmid was grown in XL10-Gold E.coli cells and midipreped. This plasmid was subsequently either linearized with BamHI or relaxed using the nt.BbvCI nickase. The resulting products where purified with phenol/chloroform and concentrated using ethanol precipitation.

Acknowledgement

The authors would like to thank Cristian Micheletti and Marco Di Stefano for discussions, Meng-Yue Wu for TEM drilling of nanopores, Jaco van der Torre for his help, and Rini Joseph and Stefan W. Kowalczyk for early experiments. This work was supported by the Netherlands Organisation for Scientific Research (NWO/OCW), as part of the Frontiers of Nanoscience program, and by the European Research Council under research grant NanoforBio (no. 247072), and by the Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) Academy Assistants Program. Y.R. and A.Y.G. would like to acknowledge support from the US-Israel Binational

Science foundation.

4.5. Supplementary Info

4.5.1. Knot Statistics

In Table 4.1 and 4.2 we provide the observed knot occurrence for various experimental conditions. Molecules with single knots have been sorted by the maximum amplitude (on top of the DNA blockade level I_1) observed within the knot duration. The knot occurrence in 48502 bp lambda-phage DNA in 1M KCl is estimated using the method described in Supplementary Section 4.5.2. In Table 4.3 we provide the observed folding rates (exclusive of knots).

Table 4.1: The knot occurrence data for linear dsDNA molecules in various conditions. The maximum bandwidth of the Axon Axopatch 200B amplifier used has been measured[41] to be around 52 kHz.

Salt	Pore size (nm)	Volt. (mV)	DNA length (bp)	% of events with a single knot of given amplitude			% of events with >1 knot	Total % of events with knots	Total # of events	BW (kHz)
				$2I_1$	$4I_1$	$>4I_1$				
2M LiCl	10	100	48502	13.7	1.2	1.2	1.2	17.2	262	15
		150		16.5	1.7	0.9	4.4	23.5	115	20
		200		25.5	2.6	0.0	1.3	29.3	157	20
		300		21.6	2.4	0.4	3.6	28.0	250	30
		400		20.7	1.3	1.7	4.3	28.0	232	40
		100		11.9	3.1	2.8	0.4	18.1	287	20
		200		17.8	3.2	1.8	3.7	26.5	219	30
		300		18.3	2.3	1.4	2.3	24.2	355	30
4M LiCl	20	100	2686	1.7	0.0	0.2	0.0	1.8	1084	30
		100	20678	11.4	1.8	0.9	0.4	14.5	228	60
		100		10.7	2.5	0.4	0.7	14.2	1022	30
		200		10.2	1.5	0.1	0.8	12.6	1078	60
		120	48502	22.6	3.8	0.9	1.9	29.2	114	25
		120		19.6	2.2	0.5	2.2	24.4	455	10
1M KCl	20	100	20678	3.6	0.2	0.2	0.3	4.3	1028	25
	20	100	48502	9.0	1.7	0.6	0.4	11.8	3993	30
		200		8.9	3.5	0.3	0.6	13.3	7209	70
		300		8.5	2.5	0.2	0.3	11.0	6256	100
	10	100	48502	8.5	2.5	1.0	0.6	12.6	1224	30
		200		9.0	2.7	2.2	0.0	13.9	992	100
		300		8.3	1.7	0.6	0.0	9.9	947	100
	20	100	165648	28.3	7.3	3.2	14.4	53.1	761	50
		200		29.5	7.2	1.6	12.3	50.7	1545	100

Table 4.2: The knot occurrence data for circular lambda dsDNA molecules cyclized in 2M LiCl.

Salt	Pore size (nm)	Volt. (mv)	DNA length (bp)	% of events with a single knot of given amplitude			% of events with > 1 knot	Total % of events with knots	Total # of events	BW (kHz)
				$2I_1$	$4I_1$	$> 4I_1$				
2M LiCl	10	100	48502	13.7	0.8	1.1	1.1	16.7	222	15
		150		27.0	1.4	0.0	4.1	32.4	74	20
		200		29.8	6.5	2.4	7.3	46.0	124	20
		300		27.6	3.8	1.9	12.4	45.7	210	25
		400		31.8	4.6	1.3	7.3	45.0	151	40

Table 4.3: The folding rates observed for different length molecules in various conditions. Data does not include knots. Folding rates are dependent on resolution. The better the resolution, the more short folds are observed at the start of the translocation process.

Buffer	Pore (nm)	Volt. (mV)	DNA length (bp)	Molecule type	Percentage of folded events (%)	BW (kHz)
2M LiCl	10	100	48502	Linear	87.8	15
		200			87.9	20
		300			88.8	30
4M LiCl	20	100	2686		53.8	30
		100	20678		82.3	30
		200			74.6	60
		120	48502		94.7	25
1M KCl	20	100	48502		63.5	30
		200			59.1	70
		300			53.8	100
	10	100			64.6	30
		200			58.9	100
		300			49.5	100
	20	100	165648		85.8	50
		200			82.3	100
2M LiCl	10	100	48502	Circular	3.2	15
		200			5.7	20
		300			3.8	30
4M LiCl	20	100	20678		4.3	60
		200			3.2	60

4.5.2. Lambda-DNA Measurements in 1M KCl

Measurements carried out in 1M KCl reveal the presence of knots, observed as short duration spikes within the DNA current blockade. These measurements suffer from two primary drawbacks. First, the resolution of these measurements is far lower than in 4M LiCl, as discussed in the main text, which means that many tight knots translocate without being observed. Secondly, analysis of these events is complicated by the presence of short folds at the start of the translocation event, as seen in the current trace in the lower-left part of Figure 4.6b. The latter effect can be observed in Figure 4.7, where the data has been split into spikes occurring during the first 10% of the translocation event (blue) and the last 90% (red). Due to this effect, we estimated the knot occurrence (Table 4.1) using only spikes which appeared the last 90% of the translocation event. When the spikes occurring in the first 10% of the event are included, we observe the values of the knotting occurrence to roughly double relative to the values provided in Table 4.1. On average we observe that spikes occurring within the first 10% of the event have lower amplitudes and longer dwell times compared to those in the latter 90%.

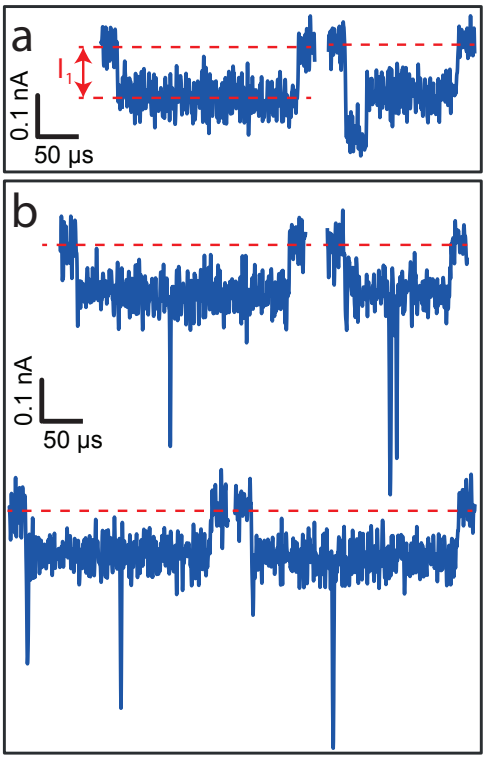


Figure 4.6: a) Typical current traces for unfolded and folded 48.5 kbp lambda DNA translocating through a 20 nm pore at 30 kHz bandwidth. b) Examples of translocation events exhibiting current spikes attributed to knotting.

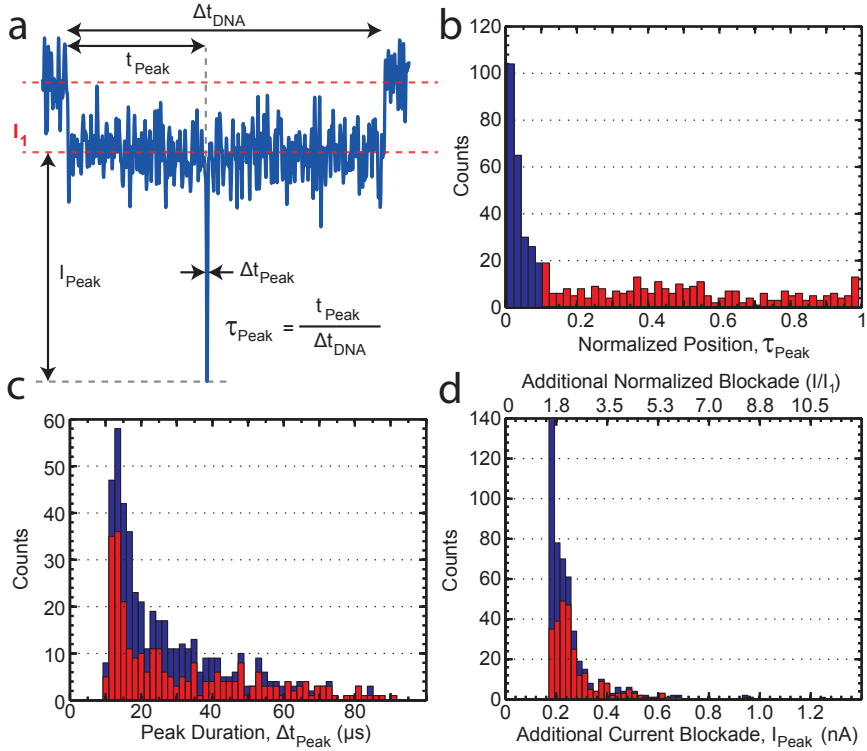


Figure 4.7: Statistics for current spikes with an amplitude (I_{peak}) larger than 1.5 I/I_1. Data is separated into current spikes occurring within 10% of the start of the translocation (blue) as well as spikes occurring in the latter 90% (red). a) A typical DNA translocation event showing the various measures used. b) The normalized position, showing a large population of spikes at the start of the translocation which is attributed to short folds, as well as a second population along the rest of the molecule. c) The FWHM dwell time of the current spikes with the distribution showing a peak around 15 μs, below the filtering distortion point at 22 μs. d) The additional current blockade for the observed current spikes.

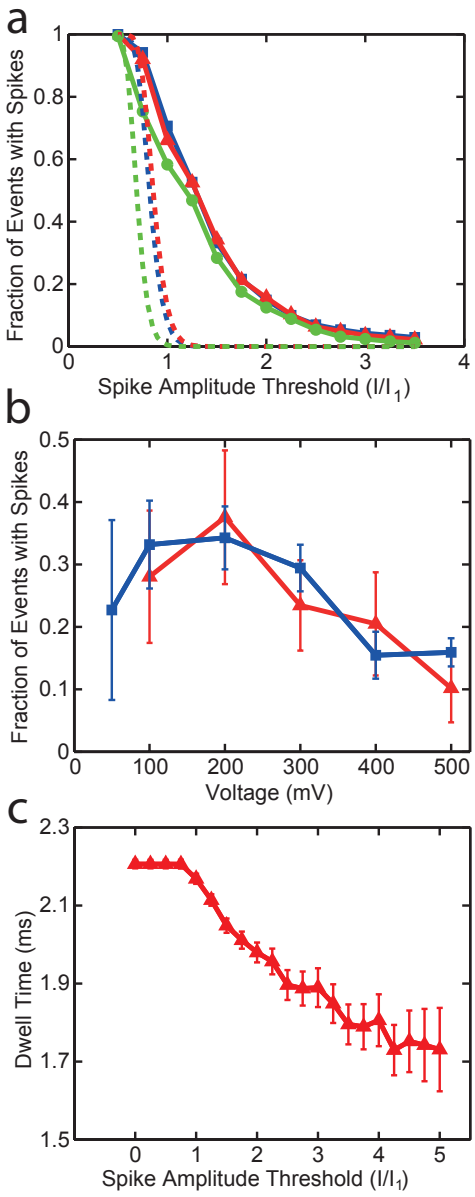


Figure 4.8: a) The fraction of events with spikes for lambda DNA in a 20 nm pore as a function of the spike amplitude threshold for three voltage levels 100 mV (blue-squares), 200 mV (red-triangles), and 300 mV (green-circles). The dashed lines indicate the amounts expected if only the Gaussian white noise were present. b) The fraction of events with spikes larger than 1.5 I/I_1 as a function of applied voltage for lambda DNA translocating through a 20 nm (blue-square) and 10 nm (red-triangle) pores. Error bars indicate the standard deviation. The decrease at higher voltages is attributed to the decreasing spatial resolution. c) The total translocation time for lambda DNA molecules at 100 mV in a 20 nm pore as a function of the spike amplitude threshold. Error bars indicate the standard error.

4.5.3. T4-DNA Measurements in 1M KCl

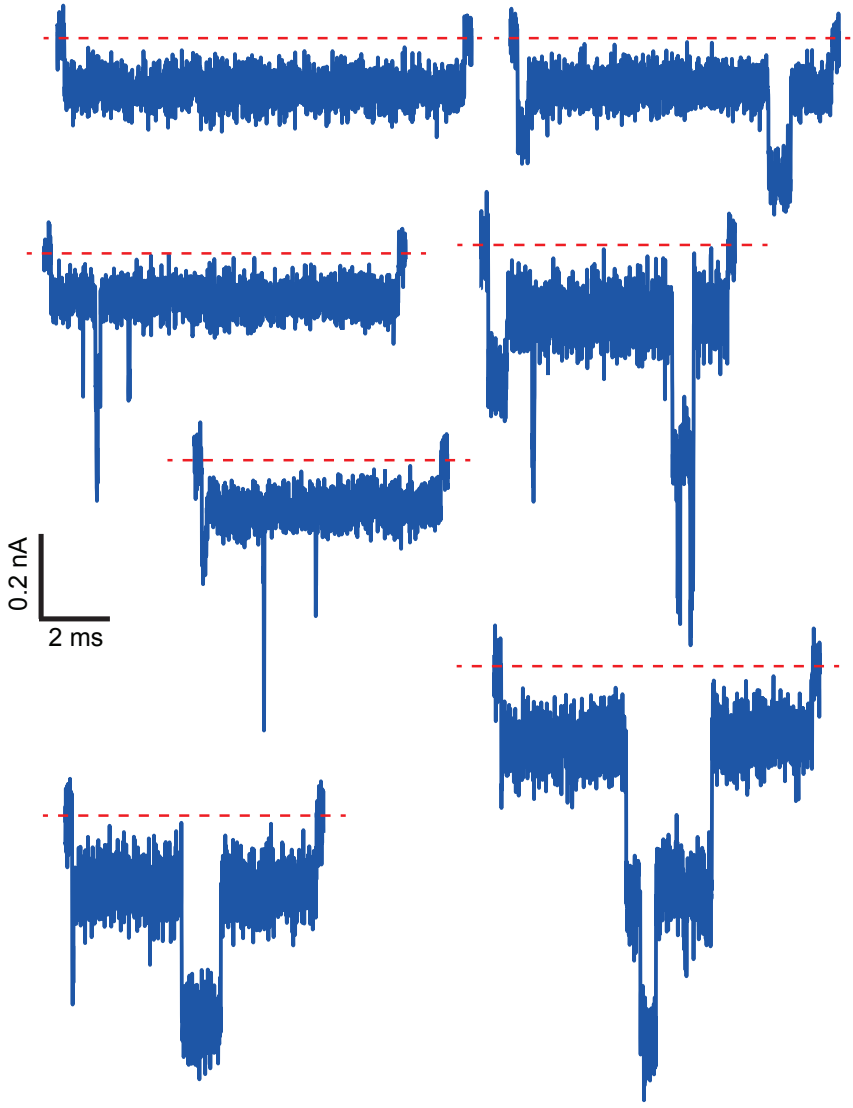


Figure 4.9: Example translocation events for 165.5 kbp T4 GT7 DNA translocating through a 20 nm pore at 50 kHz bandwidth in 1M KCl. The dashed red line indicates the current baseline level.

4.5.4. Knot Translocation Time for 20.7 kbp DNA in 4M LiCl

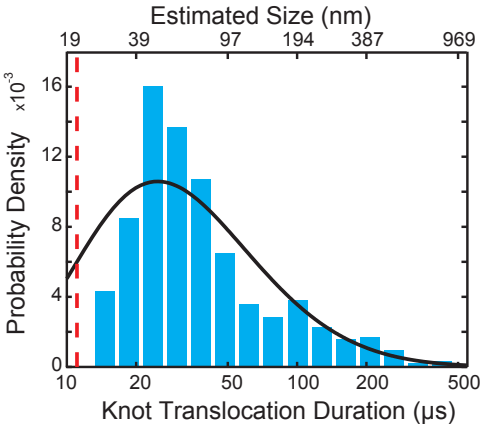


Figure 4.10: The duration (τ) of knots observed in 20.7 kbp linear molecules at 200 mV on a logarithmic time scale. This distribution may be distorted by knots slipping out during the translocation process.

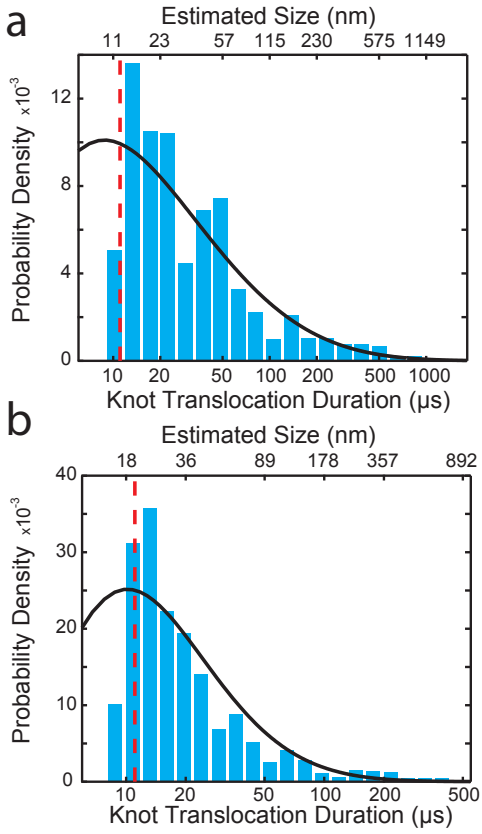


Figure 4.11: The duration (τ) of knots observed in 20.7 kbp relaxed circular molecules at a) 100 mV and b) 200 mV on a logarithmic time scale. Both distributions are cut-off by the limited temporal resolution of the measurements.

4.5.5. Mean Translocation Velocities

Table 4.4: The most probable translocation time and mean translocation velocity for different length molecules in various conditions.

Buffer	Pore (nm)	Volt. (mV)	DNA length (bp)	Molecule Type	Most probable translocation time (ms)	Mean translocation velocity (nm/ms)
2M LiCl	10	100	48502	Linear	10.90	1513
		200			4.76	3467
		300			2.95	5587
4M LiCl	20	100	2686		0.55	1673
		100	20678		8.06	872
		200			3.63	1937
		120	48502		16.50	999
1M KCl	20	100	20678		0.81	8680
		100	48502		2.10	7853
		200			1.04	15856
		300			0.71	23161
	10	100			1.49	11088
		200			0.75	21988
		300			0.49	33792
	20	100	165648		11.00	5120
		200			5.55	10146
2M LiCl	10	100	48502	Circular	5.15	1602
		200			2.35	3504
		300			1.59	5182
4M LiCl	20	100	20678		3.06	1149
		200			1.97	1784

4.5.6. Lambda-DNA Measurements in 2M LiCl

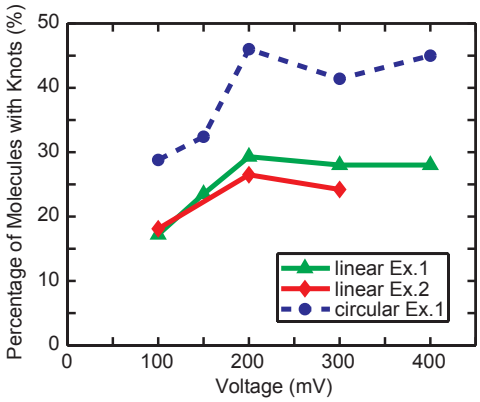


Figure 4.12: The percentage of knots observed for linear and circular lambda molecules translocating through a 10 nm pore in 2M LiCl as a function of voltage. We observe a 55% higher knotting occurrence in circular molecules compared to linear molecules, which is attributed to knots slipping out.

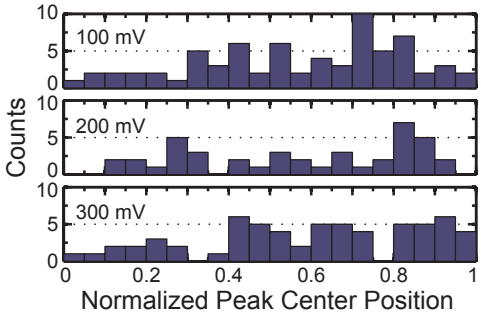


Figure 4.13: The normalized center position of knots observed in circular lambda-DNA molecules at 100 mV (top), 200 mV (middle), and 300 mV (bottom). The three distributions have mean values of 0.57, 0.58, and 0.60 for 100, 200, and 300 mV respectively.

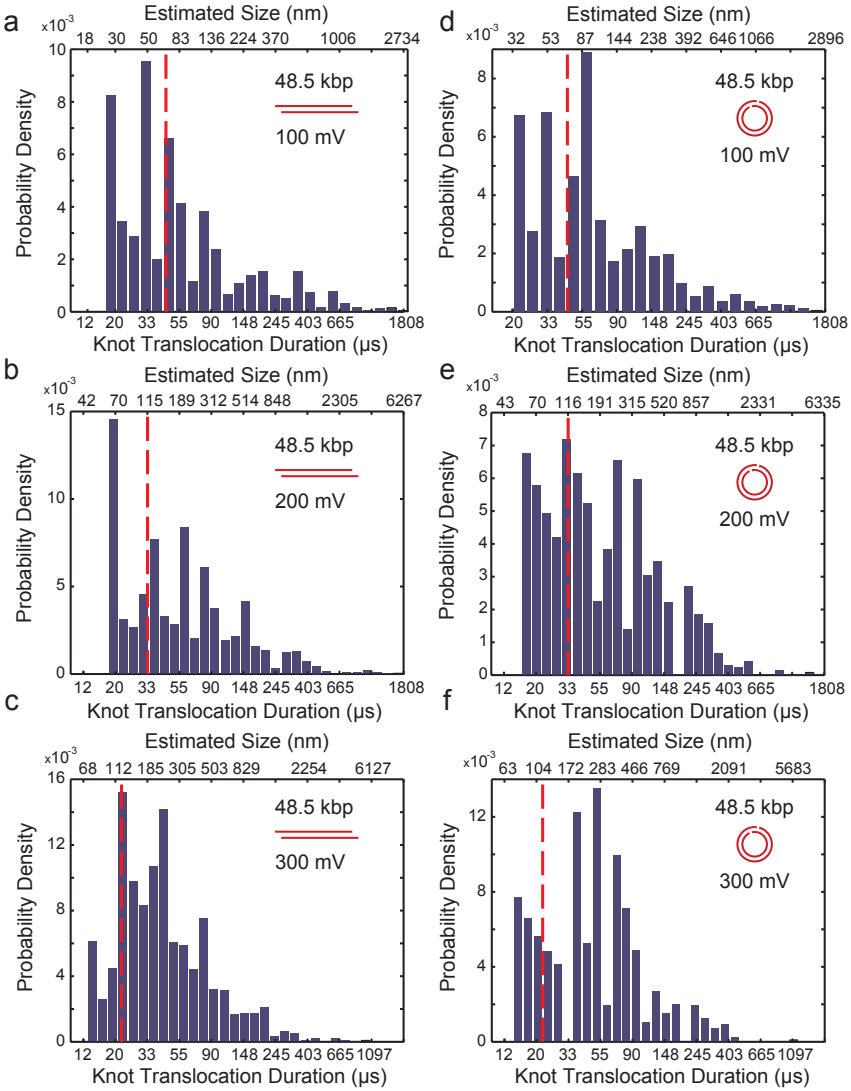


Figure 4.14: The translocation duration (τ) of knots observed in linear (a, b, c) and circular (d, e, f) molecules at 100 mV (a, d), 200 mV (b, e), and 300 mV (c, f) on a logarithmic time scale. The distributions are cut off by the limited resolution of these measurements. In these measurements we expect the best resolution (about 50 nm) to occur in the 100 mV measurements.

4.5.7. Simulation Based Predictions

The knotting probability as a function of DNA length shown in Figure 4.3 is co-plotted with two simulation-based predictions from previous publications. Here we provide the parameters used for these two models. The green dashed line in Figure 4.3 is based on Deguchi and Tsurusaki's[34] interpolation formula for the random knotting probability, which is given by

$$P_K(N) = C_K N^{\nu_K} exp(-N/N_K), \quad \text{for} \quad N >> 1 \tag{4.1}$$

where K is the knot type and N is the number of Kuhn segments in the polymer. The other three parameters C_K , ν_K, and N_K are fitting parameters dependent on the knot type. We refer readers to the original publication for a detailed description of this formula. In Figure 4.3, we plot the total knotting probability using their fitting parameters for the unknotted case ($K = 0$), which is given by

$$P_{total}(N) = 1 - P_0(N) = 1 - 1.03N^{0.007} exp(-N/335). \tag{4.2}$$

Deguchi and Tsurusaki's interpolation formula is valid for large values of N. In Figure 4.3, we plot their estimated knotting probability for values of $N > 50$ (15 kbp).

The second model (red dashed line in Figure 4.3) is based on fits for the Monte Carlo simulations of Rybenkov et al.[19] which are given by

$$P_K(N, d) = P_K(N, 0) exp(-r_t d/b), \tag{4.3}$$

where K is the knot type, N is the number of segments of Kuhn length b, d is the effective diameter of the polymer, $P_K(N, 0)$ is the knotting probability for zero effective diameter, and r_t is a proportionality constant they determined from simulations.

Table 4.5: Parameters used to determine the knotting probability.

Knot type (K)	Formula to estimate $P_K(N, 0)$	r_t
3_1	(1.818E-3)N - (1.091E-2)	22
4_1	(4.11E-4)N - (4.54E-3)	31
5_1	(1.44E-4)N - (2.17E-3)	42
5_2	(2.44E-4)N - (3.67E-3)	42

Using the values and formulas provided in Table 4.5 the knotting probability was estimated for a polymer with an effective diameter of 2.7 nm and Kuhn length of 100 nm using

$$P_{total}(N) = 0.552P_{3_1}(N, 0) + 0.433P_{4_1}(N, 0) + 0.322P_{5_1}(N, 0) + 0.322P_{5_2}(N, 0). \tag{4.4}$$

We plot this probability in Figure 4.3 until a maximum length of 20 kbp since Rybenkov et al carried out their Monte Carlo simulations for maximum chain lengths of $N = 60$ Kuhn segments.

References

[1] D. W. Sumners and S. Whittington, *Knots in self-avoiding walks,* Journal of Physics A: Mathematical and General **21**, 1689 (1988).

[2] A. Kawauchi, *Survey on knot theory* (Springer, 1996).

[3] D. Meluzzi, D. E. Smith, and G. Arya, *Biophysics of knotting,* Annual Review of Biophysics **39**, 349 (2010).

[4] P. Staczek and N. P. Higgins, *Gyrase and topo IV modulate chromosome domain size in vivo,* Molecular Microbiology **29**, 1435 (1998).

[5] A. Rodríguez-Campos, *DNA knotting abolishes in vitro chromatin assembly,* Journal of Biological Chemistry **271**, 14150 (1996).

[6] J. Portugal and A. Rodríguez-Campos, *T7 RNA polymerase cannot transcribe through a highly knotted DNA template,* Nucleic Acids Research **24**, 4890 (1996).

[7] A. Y. Grosberg, *A few notes about polymer knots,* Polymer Science Series A **51**, 70 (2009).

[8] R. Metzler, A. Hanke, P. G. Dommersnes, Y. Kantor, and M. Kardar, *Equilibrium shapes of flat knots,* Physical Review Letters **88**, 188101 (2002).

[9] E. Orlandini, A. L. Stella, and C. Vanderzande, *The size of knots in polymers,* Physical Biology **6**, 025012 (2009).

[10] A. Y. Grosberg and Y. Rabin, *Metastable tight knots in a wormlike polymer,* Physical Review Letters **99**, 217801 (2007).

[11] J. Tang, N. Du, and P. S. Doyle, *Compression and self-entanglement of single DNA molecules under uniform electric field,* Proceedings of the National Academy of Sciences **108**, 16153 (2011).

[12] X. R. Bao, H. J. Lee, and S. R. Quake, *Behavior of complex knots in single DNA molecules,* Physical Review Letters **91**, 265506 (2003).

[13] Y. Arai, R. Yasuda, K.-i. Akashi, Y. Harada, H. Miyata, K. Kinosita, and H. Itoh, *Tying a molecular knot with optical tweezers,* Nature **399**, 446 (1999).

[14] M. A. Krasnow, A. Stasiak, S. J. Spengler, F. Dean, T. Koller, and N. R. Cozzarelli, *Determination of the absolute handedness of knots and catenanes of DNA,* Nature **304**, 559 (1983).

[15] L. F. Liu, J. L. Davis, and R. Calendar, *Novel topologically knotted DNA from bacteriophage P4 capsids: studies with DNA topoisomerases,* Nucleic Acids Research **9**, 3979 (1981).

[16] S. Trigueros, J. Arsuaga, M. E. Vazquez, D. W. Sumners, and J. Roca, *Novel display of knotted DNA molecules by two-dimensional gel electrophoresis,* Nucleic Acids Research **29**, e67 (2001).

[17] S. A. Wasserman, J. M. Dungan, and N. R. Cozzarelli, *Discovery of a predicted DNA knot substantiates a model for site-specific recombination,* Science **229**, 171 (1985).

[18] K. Valencia and D. Buck, *Predicting knot and catenane type of products of site-specific recombination on twist knot substrates,* Journal of Molecular Biology **411**, 350 (2011).

[19] V. V. Rybenkov, N. R. Cozzarelli, and A. V. Vologodskii, *Probability of DNA knotting and the effective diameter of the DNA double helix,* Proceedings of the National Academy of Sciences **90**, 5307 (1993).

[20] S. Y. Shaw and J. C. Wang, *Knotting of a DNA chain during ring closure,* Science **260**, 533 (1993).

[21] E. Ercolini, F. Valle, J. Adamcik, G. Witz, R. Metzler, P. De Los Rios, J. Roca, and G. Dietler, *Fractal dimension and localization of DNA knots,* Physical Review Letters **98**, 058102 (2007).

[22] F. Haque, J. Li, H.-C. Wu, X.-J. Liang, and P. Guo, *Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA,* Nano Today **8**, 56 (2013).

[23] M. Wanunu, *Nanopores: A journey towards DNA sequencing,* Physics of Life Reviews **9**, 125 (2012).

[24] M. Muthukumar, *Mechanism of DNA transport through pores,* Annual Review of Biophysics and Biomolecular Structure **36**, 435 (2007).

[25] A. J. Storm, C. Storm, J. Chen, H. Zandbergen, J.-F. Joanny, and C. Dekker, *Fast DNA translocation through a solid-state nanopore,* Nano Letters **5**, 1193 (2005).

[26] C. Plesa, L. Cornelissen, M. W. Tuijtel, and C. Dekker, *Non-equilibrium folding of individual DNA molecules recaptured up to 1000 times in a solid state nanopore,* Nanotechnology **24**, 475101 (2013).

[27] M. Gershow and J. A. Golovchenko, *Recapturing and trapping single molecules with a solid-state nanopore,* Nat Nano **2**, 775 (2007).

[28] M. Mihovilovic, N. Hagerty, and D. Stein, *Statistics of DNA capture by a solid-state nanopore,* Physical Review Letters **110**, 028102 (2013).

[29] Y. Kantor and M. Kardar, *Anomalous dynamics of forced translocation,* Physical Review E **69**, 021806 (2004).
[30] A. Rosa, M. Di Ventra, and C. Micheletti, *Topological jamming of spontaneously knotted polyelectrolyte chains driven through a nanopore,* Physical Review Letters **109**, 118301 (2012).
[31] L. Huang and D. E. Makarov, *Translocation of a knotted polypeptide through a pore,* The Journal of Chemical Physics **129**, 121107 (2008).
[32] G. Ando, C. Hyun, J. Li, and T. Mitsui, *Directly observing the motion of DNA molecules near solid-state nanopores,* ACS Nano **6**, 10090 (2012).
[33] S. W. Kowalczyk, D. B. Wells, A. Aksimentiev, and C. Dekker, *Slowing down DNA translocation through a nanopore in lithium chloride,* Nano Letters **12**, 1038 (2012).
[34] T. Deguchi and K. Tsurusaki, *A statistical study of random knotting using the vassiliev invariants,* Journal of Knot Theory and Its Ramifications **03**, 321 (1994).
[35] A. Vologodskii, *Brownian dynamics simulation of knot diffusion along a stretched DNA molecule,* Biophysical Journal **90**, 1594 (2006).
[36] J. C. Wang and N. Davidson, *Thermodynamic and kinetic studies on the interconversion between the linear and circular forms of phage lambda DNA,* Journal of Molecular Biology **15**, 111 (1966).
[37] B. Lu, F. Albertorio, D. P. Hoogerheide, and J. A. Golovchenko, *Origins and consequences of velocity fluctuations during DNA passage through a nanopore,* Biophysical Journal **101**, 70 (2011).
[38] C. Plesa, S. W. Kowalczyk, R. Zinsmeester, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Fast translocation of proteins through solid state nanopores,* Nano Letters **13**, 658 (2013).
[39] X. J. A. Janssen, M. P. Jonsson, C. Plesa, G. V. Soni, C. Dekker, and N. H. Dekker, *Rapid manufacturing of low-noise membranes for nanopore sensors by trans -chip illumination lithography,* Nanotechnology **23**, 475302 (2012).
[40] C. Plesa and C. Dekker, *Data analysis methods for solid-state nanopores; submitted,* (2015).
[41] J. D. Uram, K. Ke, and M. Mayer, *Noise and bandwidth of current recordings from submicrometer pores and nanopores.* ACS Nano **2**, 857 (2008).

5

Ionic permeability and mechanical properties of DNA origami nanoplates on solid-state nanopores

While DNA origami is a popular and versatile platform, its structural properties are still poorly understood. In this study we use solid-state nanopores to investigate the ionic permeability and mechanical properties of DNA origami nanoplates. DNA origami nanoplates of various designs are docked onto solid-state nanopores where we subsequently measure their ionic conductance. The ionic permeability is found to be high for all origami nanoplates. We observe the conductance of docked nanoplates, relative to the bare nanopore conductance, to increase as a function of pore diameter, as well as to increase upon lowering the ionic strength. The honeycomb lattice nanoplate is found to have slightly better overall performance over other plate designs. After docking, we often observe spontaneous discrete jumps in the current, a process which can be attributed to mechanical buckling. All nanoplates show a non-linear current-voltage dependence with a lower conductance at higher applied voltages, which we attribute to a physical bending deformation of the nanoplates under the applied force. At sufficiently high voltage (force), the nanoplates are strongly deformed and can be pulled through the nanopore. These data show that DNA origami nanoplates are typically very permeable to ions and exhibit a number of unexpected mechanical properties which are interesting in their own right, but also need to be considered in the future design of DNA origami nanostructures.

This chapter has been published as: C. Plesa, A. N. Ananth, V. Linko, C. Gülcher, A. J. Katan, H. Dietz, and C. Dekker, *Ionic Permeability and Mechanical Properties of DNA Origami Nanoplates on Solid-State Nanopores*, ACS Nano **8**, 1, 35-43 (2013).

5.1. Introduction

Solid-state-nanopore research[1, 2] is an area which has been gaining much attention due to its potential applications in sequencing, biosensing, and as a tool for biophysics. In this technique an electric field is applied across a membrane containing a single pore. Charged molecules such as DNA experience an electrophoretic force which pulls them through the pore. Molecules are detected by the temporary reduction (or increase at low salt) in the ionic current which they cause as they pass through the pore. Despite recent advances[1, 2] in solid-state-nanopore research, biological pores still offer several advantages such as a higher signal-to-noise ratio[3], slower DNA translocation velocity[4], and the possibility for adding chemical modifications to the channel by creating mutant proteins[5]. In an attempt to combine the advantages of solid-state pores with those of biological pores, in 2010, our group demonstrated that a stable hybrid pore could be built by capturing an α-hemolysin protein pore with a DNA tail into a solid state nanopore[6]. This work subsequently opened up a new stream of research into combining biological structures with solid-state nanopores in order to open up new functionalities.

This approach can be extended to DNA origami, a technique introduced by Paul Rothemund in 2006 which allows complex shapes to be built from DNA by designing short oligonucleotide "staples" which upon hybridization bring two defined parts of a large single-stranded DNA template together[7]. Through the rational design of staples, the single-stranded DNA template can thus be folded into any desired shape. Since its introduction, DNA origami has grown into a robust technique capable of reliably producing 3D structures[8, 9] such as boxes, spheroids, and complex objects with curved surfaces[10, 11].

DNA origami can be used to create DNA origami nanopores. Recent literature has reported the creation of hybrid nanopores by capturing DNA origami nanoplates containing apertures[12] as well as funnel-like structures[13] onto solid state nanopores and glass nanocapillaries[14]. Synthetic lipid membrane channels[15, 16] have also been made using DNA origami nanostructures. The versatile approach, of docking DNA origami nanostructures onto solid-state nanopores, allows great control over both the geometry and the chemical functionality of the pore. Wei et al created a square honeycomb-lattice-based DNA-origami nanoplate with a central aperture. Their nanoplate contained a long ssDNA tail extending from the side of the aperture (close to the center of the nanoplate) to facilitate proper insertion. Initial work with this system showed translocation of DNA and protein through a docked nanoplate and stochastic sensing of target molecules using a bait-prey scheme. Bell et al focused on a hollow pyramid type structure with an aperture and a tail at the apex, reminiscent of the structure of α-hemolysin or MspA, with a top side larger than the diameter of the solid state pore. Their work showed the ability to capture and insert these structures into the pore and subsequently translocate dsDNA through docked nanostructures. These first experiments provided a good proof-of-concept and established that this technique can be used to add additional functionalities to solid-state nanopores. In addition, this approach can be used to study the inherent properties of DNA origami at the single-molecule

level and measure properties, such as ionic permeability, which cannot be determined with other methods.

Here, we carry out a fundamental investigation into the ion conduction and mechanical properties of DNA origami nanoplates, by docking them onto solid-state nanopores. The approach begins with electrophoretically capturing an origami plate (without an aperture) by use of a long centrally located DNA tail into a pore, as shown in Figure 5.1. The tail allows the guided insertion of the nanostructure with a well-defined orientation onto a solid-state nanopore. An example 3D cross section of a honeycomb lattice nanoplate docked onto a 20 nm nanopore is shown in Figure 5.1a. The docking of a nanoplate is observed in real time by a reduction in the current level as shown in Figure 5.1b. Once docked, a current-voltage sweep such as shown in Figure 5.1c, reveals a lower conductance when the nanoplate is docked (red curve) compared to the conductance for a bare nanopore (blue) taken before the nanoplate was added. Nanoplates can be undocked from the nanopore by reversing the polarity of the applied voltage or by pulling them through the pore by significantly increasing the voltage, as shown below. The docking time is found not to be an intrinsic property but dependent on the experiment. Most nanoplates would stay docked forever if kept at low voltages.

The ionic permeability of the nanoplate is an important parameter for origami nanopores to optimize because it sets the magnitude of the current blockade in sensing single biomolecules relative to the baseline current. The ideal nanoplate should have minimal leakage of ions through the nanoplate itself, with the majority of the ionic current given by ionic transport through the open aperture, leading to a high ratio between the excluded volume of the translocating molecule and the total volume available to carry ions. This would result in a high signal-to-noise ratio. Secondly, the nanoplate should have a high mechanical stability under any applied voltage. We investigate these issues by experimenting with different nanoplate designs, varying the diameter of the solid-state nanopores, probing a wide range of voltages, and altering buffer conditions. We find that the conductance of the nanoplates is high and increases as a function of pore diameter and ionic strength. Additionally we see interesting mechanical effects including deformation, buckling, and structural failure under increasing applied force.

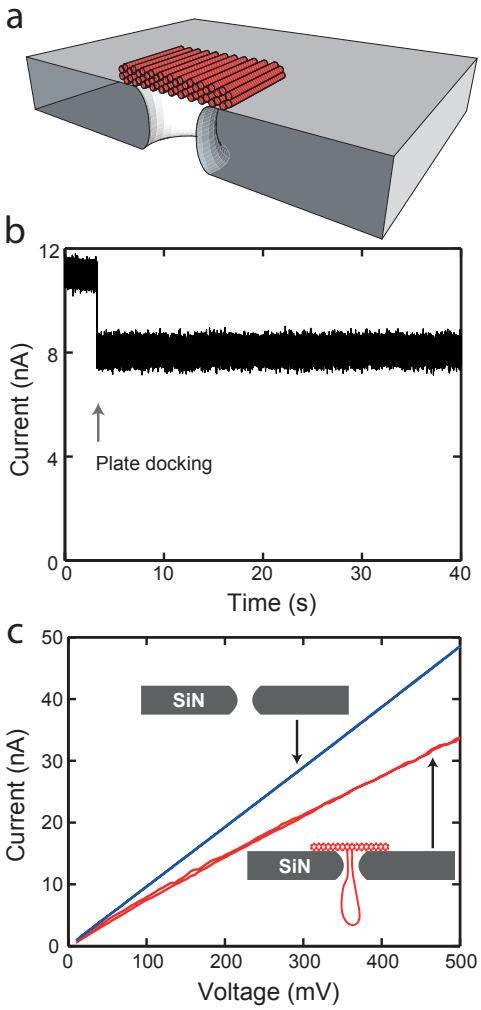


Figure 5.1: a) 3D representation showing a cross-sectional view of a Honeycomb lattice DNA nanoplate docked onto a SiN nanopore. The tail has been omitted for visual clarity. b) Current trace of a Honeycomb nanoplate captured onto a 14 nm pore at 100 mV. c) IV curve for a bare 14 nm SiN pore (blue) as well as for the same pore after a Honeycomb nanoplate was docked (red).

5.2. Results and Discussion

First, we investigated how the geometry of the DNA nanoplates influences their ionic permeability. Four different nanoplate designs were used, as shown in Figure 5.2a. None of the designs contained an aperture, to ensure that the measured current was only due to the ionic permeability of the nanoplate itself. The first design (HC) utilized a honeycomb-type DNA-packing architecture and has lateral dimensions of 45 nm x 52 nm with a thickness of 6.75 nm and an ideal spatial filling factor[12] of 0.605. The other designs are all based on the square-lattice rectangular DNA origami nanoplates introduced[7] by Paul Rothemund with varying thicknesses of one, two, or three DNA layers. The filling factor for these plates is expected to be below 0.785, as several studies[17, 18] have shown the DNA helices in the square-lattice are not close packed. The one-DNA-layer-thick Rothemund rectangle (RR) has lateral dimensions of 98 nm x 54 nm, the two-DNA-layer-thick rectangle (2LL) has 51 nm x 54 nm, and the three-DNA-layer-thick design (3LL) has 40 nm x 45 nm. Since each design is based on oligo-staple hybridization with a standard M13 single-stranded DNA template (RR and HC 7560 bases, 2LL and 3LL 7704 bases), the lateral dimensions of the designs are reduced as the thickness is increased. CanDo[19, 20] finite-element modeling was used to predict the flexibility of each plate (Supplementary Section 5.5.8). This revealed that the RR plate should be quite deformed, as evident in Figure 5.2a, due to a large twist along the plate, although thermal fluctuations and surface effects should be able to force it into a flat state. This twist is caused by the square lattice and was compensated for in the 2LL and 3LL designs by designing staples which skip hybridization to some bases of the scaffold[11]. The predicted plate stiffness, as established from the range of RMS fluctuations (provided in the brackets) over different parts of the structure can be ordered from highest to lowest as follows: HC (0.3 to 0.8 nm) ≈ 3LL (0.3 to 1.0 nm) > 2LL (0.6 to 1.7 nm) > RR (1.2 to 3.5 nm). In addition to the electrophysiological measurements which are the main focus of this work, the DNA nanoplates were characterized using TEM and high-speed liquid AFM. Negative-stain TEM was used to verify proper nanoplate assembly with averaged micrographs for each design shown in Figure 5.2c. The high-speed liquid AFM provides high-resolution images of the plates (Figure 5.2b) but also allowed us to confirm the good stability of each design in a variety of buffer conditions. Details of the TEM and AFM characterization have been provided in Supplementary Section 5.5.7.

We first report how the ionic conductance of the nanoplates varies among the different designs, how it depends on the nanopore diameter, and how the observed trends can be reproduced with a simple model. The diameter of the solid-state nanopores onto which nanoplates were docked was varied from 5 nm to 30 nm. Figure 5.3a shows the observed relative conductances (i.e. the conductance of the pore with the plate relative to that of the bare pore) for the Honeycomb (HC) nanoplate at 200 mV in 1 M KCl. The relative conductance decreases as the diameter on the solid-state nanopore is reduced. We see the relative conductance decrease from about 0.8 in large 30 nm pores down to below 0.6 in small 5 nm pores. Similar plots for the other nanoplate designs and at different voltages are shown in Supplementary Section 5.5.1.

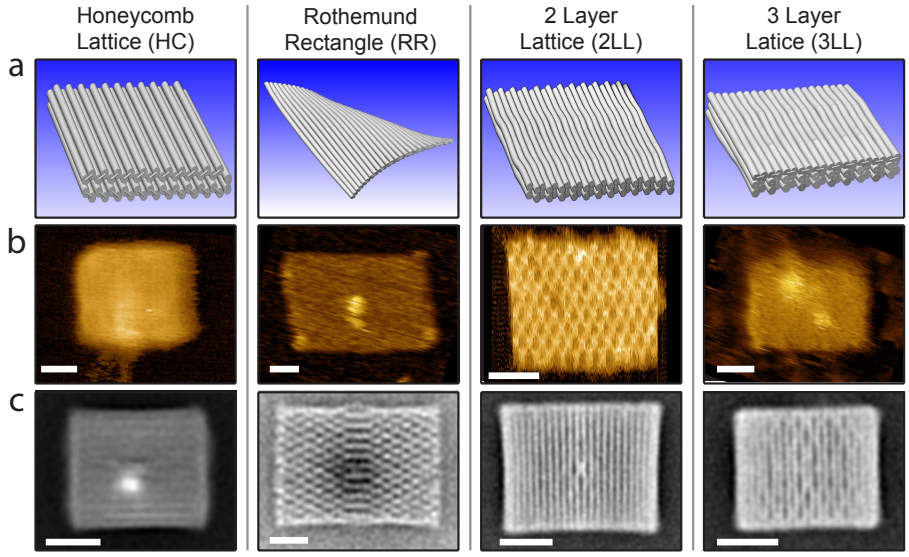


Figure 5.2: Properties of four different DNA origami nanoplates. a) 3D representations of each design. b) Liquid tapping-mode AFM scans for each nanoplate design. Variations in image resolution can be attributed to differences in AFM tip sharpness. c) TEM class averages of negative stain micrographs for each plate (RR adapted from Sobczak et al.[21]). All scale bars are 20 nm.

The solid line in Figure 5.3a represents a least-squares fit to the data using the model outlined below, with only one free parameter (α). Experimentally we measure G_{hybrid}, the conductance of the nanopore in series with the conductance of the nanoplate. As in previous work[12] this can be modeled as

$$G_{hybrid} = \left[\frac{1}{G_{pore}} + \frac{1}{G_{plate}}\right]^{-1} \tag{5.1}$$

where the conductance of the pore[22] is given by

$$G_{pore} = \kappa\left[\frac{4l_{pore}}{\pi d^2} + \frac{1}{d}\right]^{-1} \tag{5.2}$$

where κ is the conductivity of the buffer, l_{pore} is the effective thickness of the solid-state nanopore, and d is its diameter. Note that Eq. 5.2 also includes the access-resistance[22] contribution which was ignored in previous work[12]. We model the conductance of the nanoplate using

$$G_{plate} = \frac{\pi\alpha\kappa d^2}{4} \tag{5.3}$$

where α is a phenomenological parameter with units of inverse length given by

$$\alpha = \frac{f(V)}{l_{plate}} \tag{5.4}$$

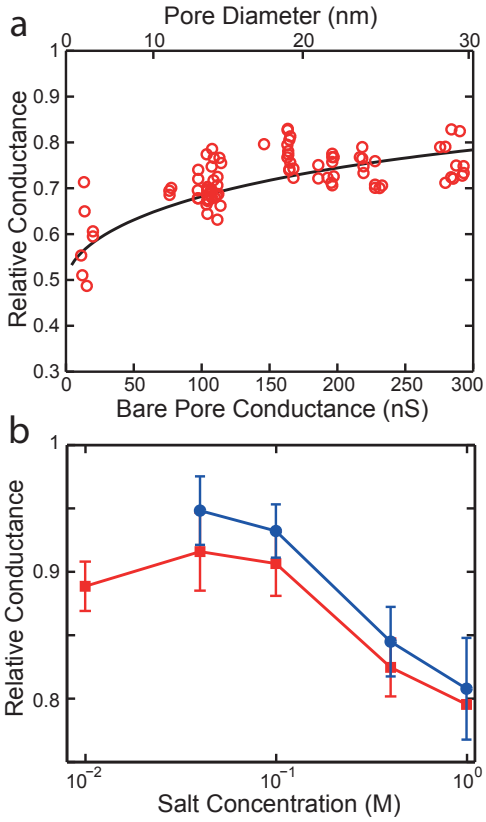


Figure 5.3: a) The relative conductance, G_{hybrid}/G_{pore} for the Honeycomb nanoplate design at 200 mV in 1 M KCl buffer plotted versus the pore conductance, which scales non-linearly with pore diameter (see scale at the top). Solid line denotes a fit of Eq. 5.6 giving α = 0.112 $\pm$ 0.06 nm^{-1}. b) Salt dependence of the relative conductance for the 2LL nanoplates docked onto 24 nm pores at 300 mV (red squares) and 200 mV (blue circles).

where l_{plate} is the nanoplate's thickness and $f(V)$ is a dimensionless function related to the applied voltage and the filling factor (F). If we ignore the (non-linear) voltage dependence discussed later on, $f(V)$ equals

$$f(V) = 1 - F \tag{5.5}$$

where F is the filling factor of the nanoplate. It is useful to examine the limiting behavior of α. When the nanoplate becomes very thick ($l_{plate} \to \infty$), α approaches zero. In the other limit of very thin plates, we encounter the minimum plate thickness which is set by the diameter of a single DNA helix (2.25nm). Here, the value of α approaches zero for a fully filled plate ($F = 1$), whereas it goes to 1/(2.25nm) as the filling factor goes to zero. In the context of optimizing the nanoplate design, the value of α should be as low as possible, since this represents the smallest leakage through the nanoplate, thus giving the highest signal-to-noise ratio in DNA translo-

cation experiments through a nanopore in the nanoplate. The nanoplate designs are experimentally best characterized by their relative conductance (RC), the ratio of G_{hybrid} to G_{pore}. Combining equations 5.1, 5.2, and 5.3, we thus obtain

$$RC = \frac{G_{hybrid}}{G_{pore}} = \frac{1}{1 + \frac{4}{\alpha(4l+\pi d)}} \tag{5.6}$$

This model nicely captures the observed trend of a decreasing relative conductance as the diameter on the solid-state nanopore is reduced, as seen in the solid line of Figure 5.3a. For the example in Figure 5.3a, $\alpha = 0.112 \pm 0.06$ nm^{-1}. Generally, we find values for α in the range from 0.10 to 0.18 nm^{-1}, see Figure 5.12, where we plot the value of the fit parameter α as a function of voltage. We find that the Honeycomb nanoplate has the lowest leakage (lowest α). Surprisingly, the one-layer-thick Rothemund Rectangle nanoplate has the next best characteristics, while the two and three-layer-thick nanoplates are worst. Although the differences are small, this observation is counterintuitive as we would expect the thicker nanoplates to have less leakage. Even though the higher leakage as the lateral dimensions of the origami nanoplate are decreased could indicate the presence of some leakage currents flowing in between the nanoplate and the solid-state nanopore substrate, we do not believe this to be a significant effect given the small size of the nanopores relative to the large size of the nanoplates involved as well as other factors discussed in detail in Supplementary Section 5.5.1. This phenomenon, of the thicker square-lattice plates being leakier, is therefore attributed to differences in the particular design of the nanoplates such as the arrangement of the oligo staples.

The dependence of the RC on the ionic strength of the buffer is shown in Figure 5.3b. The relative conductance is found to increase as the salt concentration is decreased. In other words, the nanoplates appear to become more leaky as the salt concentration drops. The relative conductance is observed to increase from around 0.8 at 1 M KCl up to 0.9 at salt concentration of 100 mM KCl or lower. Similar trends are observed for all other nanoplate designs and different pore sizes (see Supplementary Section 5.5.2 and 5.5.3). For the salt concentrations tested (1 M KCl, 400mM KCl, 100 mM KCl, 40mM KCl, and 10 mM KCl) the Debye screening length is 0.31 nm, 0.47 nm, 0.89 nm, 1.26 nm, and 1.82 nm, respectively. At 100 mM and lower, the Debye layer from neighboring DNA strands should begin to overlap in many areas of the DNA origami nanoplate. As the salt concentration is lowered from the (typically used) 1 M value, the reduced electrostatic screening could have two effects on a nanoplate's ionic permeability. First, increased electrostatic repulsion between neighboring strands could increase the size of the holes present in-between the crossover points[17], allowing more ions to pass though the structure, thus increasing the ionic permeability. This effect is a contributing factor, together with mechanically induced twist, in the diamond-like pattern clearly visible in the AFM and TEM characterization (Figure 5.2) of some plates, and has been observed in previous Cryo-EM[17] and AFM[23] studies of DNA origami structures. The structure swells up to the point where it is constrained by the oligo-staples. In addition to this, the higher effective negative charge and larger Debye screening lengths of the DNA leads to a higher cation selectivity, attracting more posi-

tive potassium (K^+) ions while repelling the negatively charged chloride (Cl^-) ions present in the solution[24]. This may result in higher conductivities than the bulk conductivity of the solution, as reported previously for small charged nanopores in low salt conditions[25]. The observed increase at low ionic strengths is likely due to a combination of these two effects. The measurements reveal that, while high salt concentrations should be preferred, the docking of nanoplates into nanopores can be detected in salt concentrations as low as 10 mM, opening up the possibility of conducting measurements on docked nanoplates at physiological conditions.

We found that the presence of magnesium in the buffer was unimportant for the behavior of the nanoplates. Although it has been recently shown that Magnesium divalent cations are not required for proper folding of the DNA origami structures[26], the need for a divalent cation such as magnesium in the buffer can be rationalized as a means to allow the negatively charged nanoplate to stick to the negatively charged surface (an effect well known from AFM literature) of the SiN surface at pH 8, where all measurements are carried out. All nanoplates were folded in a buffer containing 20 mM Mg. Before measuring, these stock solutions are diluted into the proper salt concentration with new buffers containing 11 mM Mg. As a control, we also diluted some samples into magnesium-free buffer resulting in a final Mg concentration of around 0.5 mM. No differences were observed in the relative conductance measurements for these samples compared to those carried out in buffer with 11 mM Mg (data not shown). Furthermore, no issues were encountered with the docking of the nanoplates, suggesting that the electrophoretic force and van der Waals adhesion are stronger than the electrostatic forces. Over typical experimental timescales of several hours, no deterioration in the stability of the nanoplates was observed. These results show that the presence of high concentrations of magnesium in the buffer is not a stringent requirement for either maintaining nanoplate stability or facilitating the docking of nanoplates into solid-state nanopores.

Next, we report an interesting observation that was not anticipated, namely the effects of mechanical deformation of the nanoplates. We find that docked nanoplates may undergo physical deformation as the voltage is increased and can even be pulled through the pore if sufficient force is applied. Once a nanoplate is docked into a pore, we can subject the nanoplate to a voltage ramp in order to probe its current-voltage (IV) characteristics. Voltage sweeps begin at 10 mV and go to 800 mV in steps of 5 mV (each 103 ms long), and subsequently we ramp back down to 10 mV. Examples of typical IV curves can be seen in Figure 5.1c for a HC nanoplate on a 14 nm pore and in Figure 5.4a for a 2LL nanoplate on a 9 nm pore. All nanoplates exhibit non-linear IV behavior with the conductance of the nanoplate being reduced as the voltage is increased. This trend can also be seen in the decrease of the fit parameter alpha as a function of voltage (Figure 5.12). We attribute this effect to the nanoplate undergoing increased physical deformation as the voltage is increased, which brings more DNA material into the pore (cf. sketch in the bottom right inset to Figure 5.4a).

This idea is supported by the fact that beyond a critical voltage value, the current level is observed to suddenly return to the bare pore level as shown in Figure

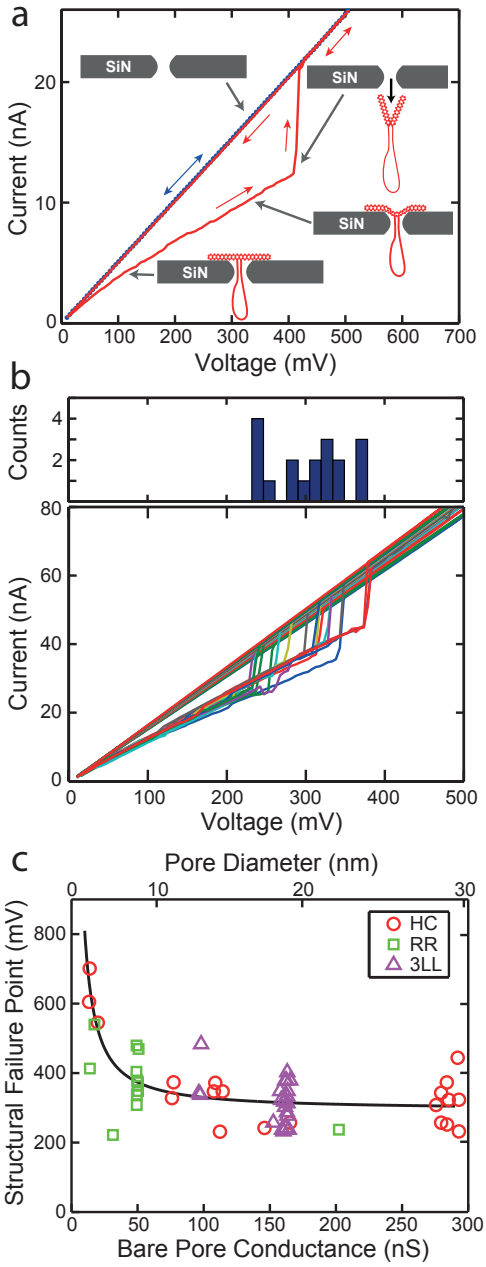


Figure 5.4: a) IV curve (red) showing a 2LL nanoplate undergoing structural failure and being pulled through a 9 nm pore. Once pulled through, the conductance of the pore returns to the level seen for the bare pore (blue dots). b) IV curves of 18 different 3LL nanoplates being pulled through a 19 nm pore. Top panel displays a histogram of the structural failure points. c) The structural failure points for three nanoplate designs docked onto different-diameter pores. It can be seen that smaller pores require a greater force to pull the nanoplate through. The solid line has been added to guide the eyes.

5.4a, after which the IV curve is linear, reproducible, and back to the conductance value for the bare nanopore, indicating that the nanoplate has been pulled through the pore, as also proven using recapture experiments described below. We term the voltage level at which this occurs the structural failure point (SFP). This effect is quite reproducible and we are able to repeatedly capture nanoplates and pull them through, as shown in Figure 5.4b, where 18 different 3LL nanoplates are pulled through a 19 nm pore. The histogram at the top shows that the mean SFP is at a voltage of 307 $\pm$ 51 mV. If we plot the observed SFP for different nanoplate designs in different pores, as shown in Figure 5.4c, we observe that, not surprisingly, a higher voltage, and thus higher force, is required to pull the nanoplates through smaller pores compared to larger pores. The dependence on pore diameter is in fact even stronger than apparent for Figure 5.4c, since our detection method misses many SFPs with a large value (exceeding our maximum voltage of 800 mV) in small pores and additionally, it misses SFPs with a small value in larger pores. In small pores (5 nm) it is often very difficult to pull a nanoplate through the pore (as evidenced by the low number of SFP events occurring). Instead, almost all nanoplates docked onto small pores remain stable up to the maximum applied voltage of 800 mV and are not pulled through. In large-diameter pores, some nanoplates are instantly pulled through already at the typical docking voltage (200 mV), thus creating short translocation events that are observed as spikes in the recorded current trace as they pass through the pore. The short spikes are observed to occur most frequently for the most flexible nanoplate, the single-layer Rothemund rectangle. This is expected because more flexible plates will undergo more deformation at a given level of applied force compared to stiffer plates. Figure 5.20 shows the spikes observed from RR nanoplates being pulled through a 24 nm pore at 200 mV and higher. This effect underlies the lack of data collected for the RR nanoplate in large-diameter pores, as evident in Figure 5.4c, Figure 5.6, and elsewhere. In addition to flexible plates, plates containing structural defects, such as those caused by missfolding are likely to be pulled through the pore very quickly.

In order to conclusively prove that the nanoplates were actually being pulled through the pore (rather than undocking back into the cis insertion chamber), we carried out recapture experiments where the electric field was reversed shortly after a spike was observed. If a nanoplate is being pulled through the pore, it should be recaptured if the electric field is reversed quickly enough and thus produce another spike. As shown in Figure 5.21, we indeed were able to recapture many nanoplates. The ability to recapture nanoplates when switching even 3 seconds after the translocation event at an applied voltage of 200 mV shows that the nanoplates were pulled through and undergo a slow drift-diffusion away from the pore. All our observations thus reveal that the nanoplates undergo mechanical deformation as the force applied on them is increased, and furthermore that the nanoplates can be completely pulled through the pore if sufficient force is applied. Flexible nanoplates can easily be pulled through larger pores, something which should be taken into account in future designs of origami nanopores.

Another surprising observation is that the current through a docked nanoplate can suddenly and randomly jump between discrete levels. Observations of many

such spontaneous jumps over time indicate that these jumps occur to and from quantized levels as shown in Figure 5.5a for a RR nanoplate docked in a 20 nm pore. These levels are clearly visible as discrete peaks in the histogram of the current trace, shown on the right. The magnitude of the jumps is up to about 10% of their relative conductance. Examples for other plate designs are provided in Supplementary Section 5.5.5. These jumps are also often seen during IV sweeps, as shown in Figure 5.5b. Here an HC nanoplate was docked in a 10 nm pore and subjected to a voltage ramp during which it transitioned through three distinct levels before being pulled through at the SFP. Subsequently another HC plate was docked in the same pore and showed similar behavior when subjected to the same voltage ramp. Interestingly, the conductance switched among the same discrete values for these two independent plates (cf. dotted green lines). We often observe these level jumps occurring in a docked nanoplate at constant voltage just before a nanoplate is pulled through, as shown in Figure 5.23 and Figure 5.24.

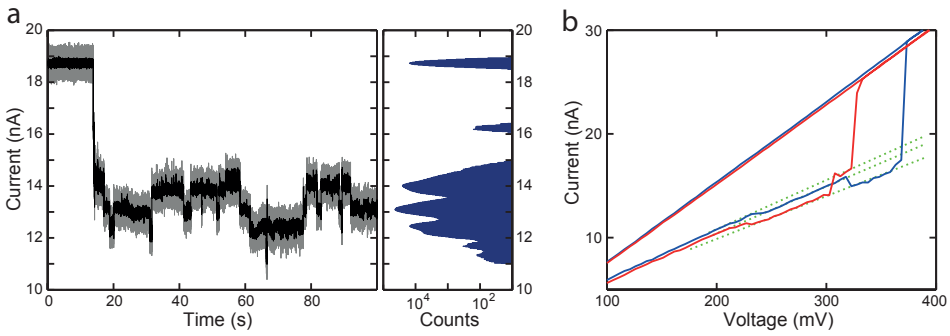


Figure 5.5: a) Current is seen to jump among a number of discrete levels for a Rothemund rectangle nanoplate docked onto a 20 nm pore at 100mV. The black trace shows data filtered at 20 kHz while the gray data is unfiltered. The histogram on the right shows the frequency of all of the current values in the trace on a log scale, with each peak representing a discrete level. b) IV curve of two different Honeycomb nanoplates docked onto the same 10 nm pore. As the voltage is ramped up, the two nanoplates each jump through three distinct levels before being pulled through. Three dotted green lines have been added for visualization purposes.

All these observations indicate that the source of these jumps is a physical change in the nanoplate structure. Although not every nanoplate tested exhibits these jumps, statistics at 100 mV on the occurrence reveal the majority (69% of HC, 60% of RR, 44% of 2LL, and 75% of 3LL nanoplates) exhibit these jumps. At 200 mV applied voltage these percentages all increase, to 82%, 80%, 68%, and 76% respectively, as shown in Figure 5.25. The 2LL nanoplate thus appears to exhibit a smaller effect compared to the other plates. We have previously seen that the 2LL nanoplate, unlike the other nanoplates, shows a very weak voltage dependence for its conductance (Figure 5.12). Ideally for origami nanopore experiments, these jumps should be absent or as small as possible in order to distinguish them from the signals produced by translocations.

Several different mechanisms could explain the physical origin of these jumps. Jumps occurring right after the nanoplate has docked could be attributed to the

nanoplate orienting itself into the most energetically favorable position. We have indeed observed some events, which represent only a small fraction of the total events, where some jumps occurred within a short time after docking, after which the nanoplate remains at a stable current level for a long period of time (minutes) or indefinitely. The majority of jumps, however, continue occurring long after the nanoplate has docked and can be attributed to mechanical buckling of the nanoplate. Such buckling could be associated with various distinct mechanical modes of the plates. Temporary melting of the strands hybridized to the M13 template seems less likely since we see no evidence that ssDNA staples are ripped off of the nanoplates as jumps occur in both directions and will often return to the original conductance level after a number of jumps. It should be noted that Langeker et al[15] attributed part of the current gating they observed in their DNA origami nanochannel to a similar effect. Although melted strands could in principle rehybridize, they are unlikely to do so in the presence of the high electric field of the nanopore where the highly charged melted strand is pulled away from its hybridization counterpart. For unzipping of the strands, the orientation of the staples is also very important, as it is known that a lower force is required to unzip a DNA helix if it is applied to the 5′ and 3′ ends[27] of complementary strands compared to the 5′ and 5′ ends[28]. In summary, we see discrete jumps in the current levels in many docked nanoplates, an effect which may be attributed mechanical readjustments such as nanoplate re-orientation and mechanical buckling.

How does the tail of the nanoplates affect the relative conductance measurements? The DNA origami nanostructures tested all contained a tail, 648 bp to 798 bp in length, protruding from the center of the nanoplate in order to facilitate their proper insertion into the solid-state nanopore. Since this single-stranded tail will form a blob of secondary structures, as confirmed by AFM and in previous nanopore experiments[29], we attempted to investigate if this blob impacts the observed conductance by comparing it to an identical nanoplate design where the single-stranded loop had been hybridized with short oligos into a fully double-stranded loop which due to its large persistence length should stick out far from the pore. Comparison of these two tail designs in a Honeycomb nanoplate showed no significant differences (Supplementary Section 5.5.6). Based on the observations, we conclude that the tail design does not affect the conductance of the nanoplate.

5.3. Conclusion

The experimental results in this work present a detailed study of ionic permeability through DNA origami objects. The conductance data reveal that the Honeycomb structure is the best design for minimizing the leakage through the nanoplate, while the 2LL nanoplate appears to have the best mechanical stability. A number of methods could be explored in the future to reduce the ionic permeability of the nanoplates. For example, groove binders and possibly intercalators could provide an interesting route. We have seen that, counterintuitively, creating a thicker nanoplate (up to 3 DNA helices thick) does not necessarily reduce the resulting ionic leakage through the nanoplate. The availability of ssDNA templates longer than M13 could allow much thicker structures to be created. Such structures could

reveal whether there is an inherent minimum ionic permeability of DNA nanostructures which is independent of thickness, as our data suggests. Investigation into the behavior of the nanoplates in different ionic regimes revealed an increased leakage through the nanoplates as the monovalent salt concentration was reduced, whereas the reduction of the magnesium concentration was shown to have no detectable effect.

DNA origami is beginning to move beyond the early static structures towards active machines with advanced functionalities. The design of such machines requires insight into the dynamics of these structures, as explored in this study. We have seen that the force applied on the nanoplates can deform the structures, cause them to buckle, and even pull them through the pore if sufficient voltage is applied. Alternative versions of the designs tested here, with different oligo-stapling or alternative scaffold routing may reveal how much the observed effects are dependent on minute design details connected to the stapling. Mechanical defects can be designed into plates to observe their effect on the plates mechanical properties such as the SFP. The combination of solid-state nanopores with DNA origami structures promises to open up a number of new possibilities that cannot be realized with other methods, including the ability to simultaneously integrate multiple receptors, binding sites, enzymes, protein, and simple mechanical machines directly within or adjacent to the nanopore.

5.4. Methods

Nanopore fabrication and measurements. SiN pores were fabricated as described previously[30]. Nanopore membranes were mounted in a PEEK flowcell separating two aqueous chambers into which Ag/AgCl electrodes were inserted. Solutions of nanoplates at approximately 200 nM concentration were added to the CIS reservoir and a voltage was subsequently applied. The measurement buffer (unless otherwise stated) consisted of 1 M KCl, 10 mM Tris, 1 mM EDTA, 11 mM Mg at pH 8. Ionic currents were detected using an Axopatch 200B amplifier at 100 kHz bandwidth and digitized with a DAQ card at 500 kHz. Current traces were analyzed using Matlab and Clampfit.

AFM and TEM. AFM measurements were carried out under solution on a RIBM High-Speed AFM 1.0. Imaging buffers contained 10 mM Tris pH 7.6, 10 mM $MgCl_2$ and KCl concentrations varying between 10 and 1000 mM. No significant differences were observed between different KCl concentrations (Figure 5.30). Samples were prepared by applying a drop of origami nanostructures in their folding buffer to plates of freshly cleaved muscovite mica. After an incubation time of 10 minutes, unadhered origamis were gently rinsed off with imaging buffer. Without drying, the samples were then transferred to the liquid cell of the AFM and imaged in tapping mode using Nanoworld USC-f1.5-k0.6 cantilevers. The TEM protocol and image processing followed the method described in Wei et al.[12]

Nanoplate design and assembly. The structures were designed using caDNAno v 0.2. DNA staple oligonucleotide strands were prepared by solid-phase chemical synthesis (Eurofins MWG) with Eurofins MWG high purity salt-free purification grade. The folding buffer was 5 mM Tris-base, 1 mM EDTA, 20 mM $MgCl_2$ (except

for RR 12.5 mM), and 5 mM NaCl at pH8. For the scaffold strand either 7560 (RR & HC) or 7704 (2LL & 3LL) base long M13mp18-phage-derived genomic DNA was used (20 nM). Staple strands were added with 10x excess (200 nM). The RR plate was folded in a thermocycler by ramping the temperature from 95°C to 25°C at 66 s/°C. For HC, 2LL, 3LL plates the temperature was ramped from 65°C to 60°C at 15 min/°C and then 59°C to 40°C with 3 h steps. All plates were subsequently stored at 4°C. The quality of folding was verified by 2% agarose gel electrophoresis (running buffer 0.5xTBE+11 mM $MgCl_2$) showing that all structures fold with an acceptable yield. All the samples were 4x filter-purified after folding by using the folding buffers indicated above. Briefly, 50 µL of folded sample and 450 µL of buffer were added to Amicon Ultra 0.5 mL filter having molecular weight cutoff of 100 kDa. The sample was centrifuged with 14,000 rcf for 3 minutes. This step was repeated 3 times by adding 450 µL of buffer each round (final round 5 min centrifugation). Then the structures were collected to a fresh tube by placing the filter upside down and centrifuging with 1,000 rcf.

Acknowledgement

We would like to thank Meng-Yue Wu for TEM drilling and imaging of nanopores. H.D. acknowledges support from the European Research Council (starting grant # GA 256270) and the Deutsche Forschungsgemeinschaft via the Cluster Nanosystems Initiative Munich. V.L. acknowledges support through The Emil Aaltonen Foundation. This work was funded in part by the European Research Council research grant NanoforBio (no. 247072).

5.5. Supplementary Info

5.5.1. Nanoplate Conductance in 1M KCl

Here we present the relative conductances for all the nanoplate designs docked onto nanopores of various diameters. First, the data for individual nanoplates is shown at 200 mV (Figures 5.6-5.8). Subsequently data for all designs is shown at 100 mV, 200 mV, and 300 mV (Figures 5.9-5.11). Relative conductances were determined using the mean value of the current after nanoplate docking divided by the mean value of the bare pore current, as described in Eq. 5.6. Solid lines represent fits using the model described in the main manuscript. The bottom x-axis is the measured bare pore conductance for that respective data point, while the top x-axis is the estimated diameter of the nanopore based on the formula

$$d = \frac{1}{2\pi\kappa}\left(\pi G_{pore} + \sqrt{\left(\pi G_{pore}\right)^2 + 16\pi\kappa l G_{pore}}\right) \tag{5.7}$$

where the measured conductivity of the buffer κ is 13.5 S/m and the effective thickness is taken to be 8.6 nm as previously determined.[22] We observe the relative conductance increase as the nanopore diameter is increased from 5 nm to 30 nm, as predicted by Eq. 5.6. The voltage dependence of the fit parameter α is shown in Figure 5.12. The value of α is observed to decrease, indicating that the

plates become less permeable, as the applied voltage is increased. A number of IV curves for each type of plate, docked onto various size nanopores are shown in Figures 5.13-5.16.

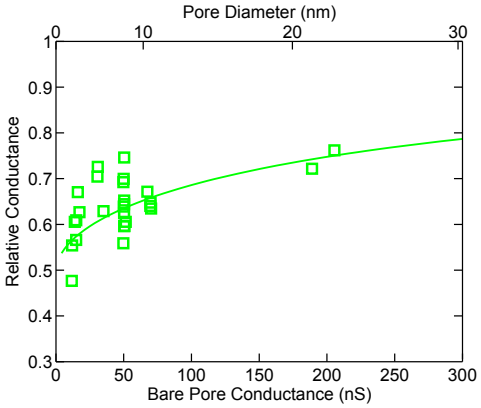


Figure 5.6: Relative conductance for Rothemund Rectangle (RR) nanoplates at 200 mV in 1M KCl.

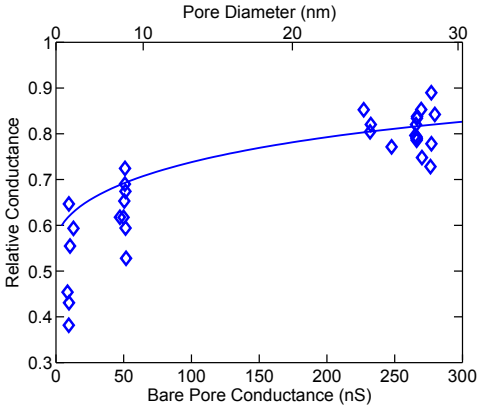


Figure 5.7: Relative conductance for 2 Layer Lattice (2LL) nanoplates at 200 mV in 1M KCl.

In Figure 5.12 we see the counter-intuitive trend that thicker plates seem to have a higher leakage. Several observations and considerations indicate that leakage currents passing underneath the plate do not contribute significantly to the nanopore current: (1) Since the tails that protrude from the center of the DNA nanoplate are used to thread the plate into the pore, it is very unlikely that the nanoplate will be off center by more than the radius of the pore, relative to the axis of the pore. This ensures full coverage of the pore, leaving no path for a leakage current between the SiN surface and the nanoplate in all but the very largest nanopores. Moreover, if an effect of non-complete coverage were present, we would expect our model to underestimate the relative conductance in large pores, since the presence of

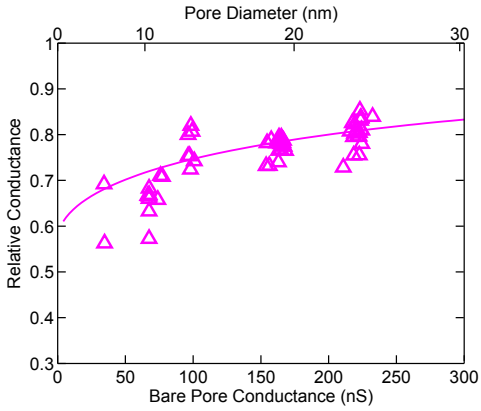


Figure 5.8: Relative conductance for 3 Layer Lattice (3LL) nanoplates at 200 mV in 1M KCl.

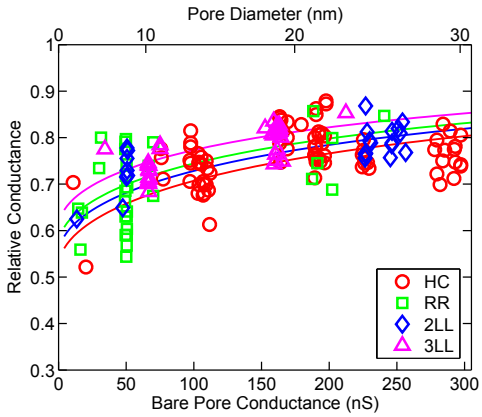


Figure 5.9: Relative conductances for the four different nanoplate designs at 100 mV in 1M KCl.

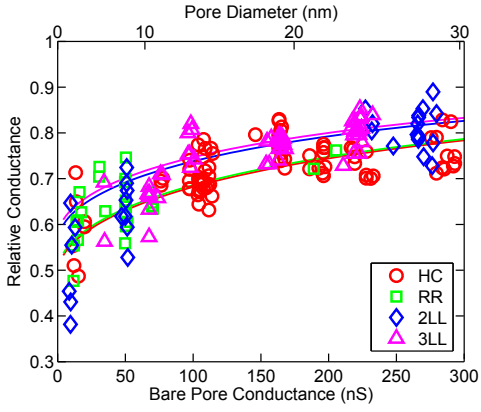


Figure 5.10: Relative conductances for the four different nanoplate designs at 200 mV in 1M KCl.

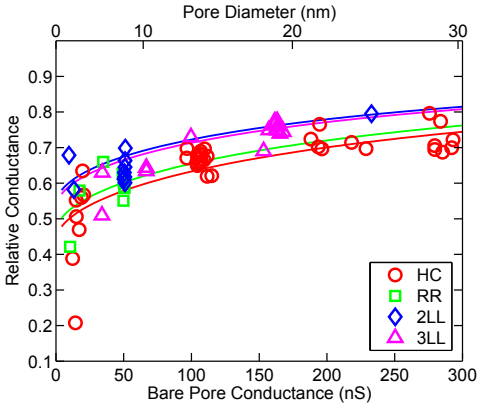


Figure 5.11: Relative conductances for the four different nanoplate designs at 300 mV.

leakage currents would increase the value of the relative conductance observed. No such effect is observed in our data. (2) We can take a different approach that might lead to leaking currents and assume that the nanoplate is properly centered on the nanopore but that there is a small gap remaining between the SiN surface and the bottom surface of the plate which allows ions to flow through. The leakage current will then involve a surface-current contribution. The ratio between the bulk conduction through the ion permeable nanoplate and the surface leakage component increases as the pore becomes smaller which implies that the leakage currents would have a much larger effect in small pores. As a rough quantitative estimate of

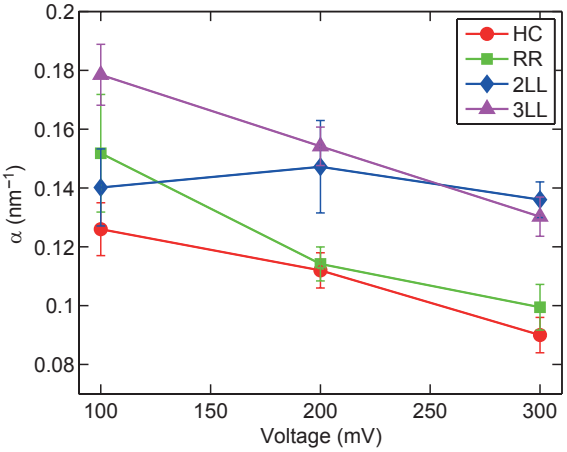


Figure 5.12: Dependence of the fit parameter α on the applied voltage. Smaller values of α represent less leakage through the nanoplate. We observe, with the exception of the 2LL nanoplate, that the leakage decreases as the voltage is increased. Error bars represent 95% confidence intervals, i.e. 2 standard deviations from the least squares fit.

this type of effect, let's assume that a potential leakage current between the plate and the SiN surface is due to the contribution of the counterions shielding the surface change. Using equation 1 from Smeets et al.[31] we calculate the contribution of the counterion current relative to the bulk ionic current. At 1M KCl for a 20 nm pore, 5.8% of the total current (14 nS of 251 nS total) is due to the counterion current. If the pore diameter is reduced to 5 nm, the surface contribution increases to 19.6% (3.6 nS of 18.3 nS total). So we would expect the contribution of a potential leakage current to increase 3.4x between 20 nm and 5 nm. This is not at all what we observe. (3) The presence of leakage currents would have the effect of increasing the relative conductance, and we thus expect that our model would give smaller relative conductances compared to experimental data. Accordingly, the gap between model and experimental values would be largest at the smallest pore values. In fact we see just the opposite: Looking at Fig 5.7, 5.8, 5.9, and 5.10, it seems that the model over-estimates the relative conductance. Hence, also this observation indicates that a significant leakage current is not present.

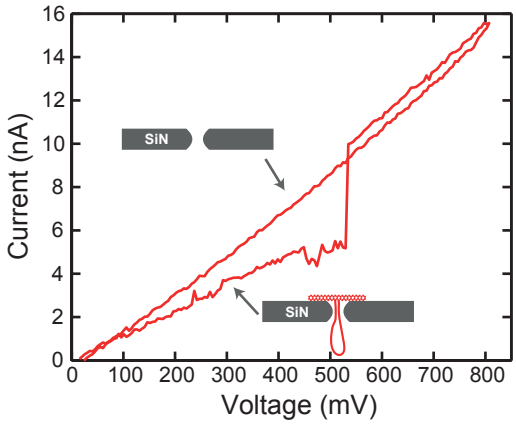


Figure 5.13: IV curve for an RR nanoplate docked onto a 4.5 nm pore. The nanoplate is pulled through the nanopore at 540 mV, after which the current returns to the bare nanopore values.

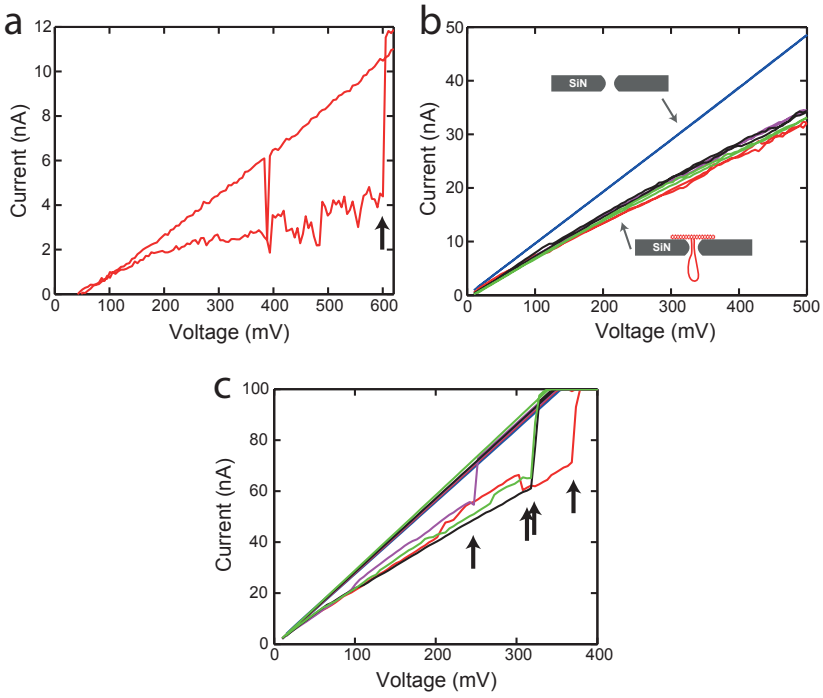


Figure 5.14: IV curves for HC plates docked onto three different size pores. Black arrows indicate the voltages at which the nanoplates were pulled through the nanopore. The uncertainty in the current data is 0.25 nA (STD). a) A plate on a 4 nm pore. b) Four plates on a 14 nm pore. c) Four plates on a 28 nm pore.

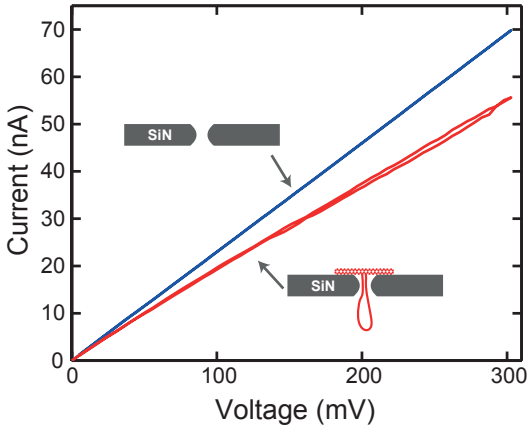


Figure 5.15: IV curve of a 2LL plate docked onto a 26 nm pore (red). The blue curve shows the IV characteristics for the bare nanopore.

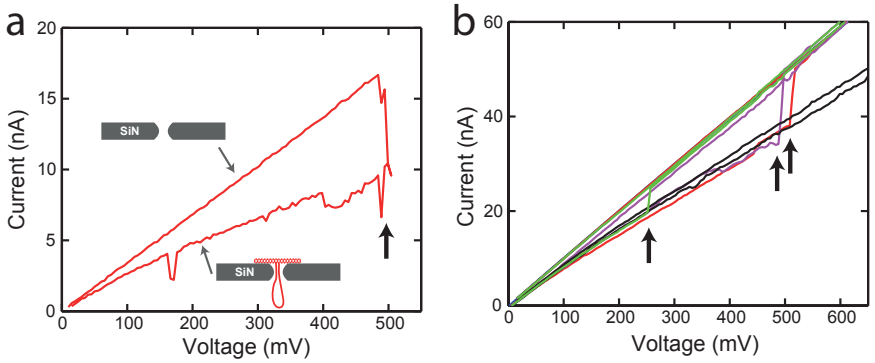


Figure 5.16: IV curves for 3LL nanoplates docked onto two different size pores. The black arrows indicate the voltage at which a nanoplate was pulled through the pore. a) 6.6 nm pore b) 13 nm pore.

5.5.2. Nanoplate Conductance in 100 mM KCl

The relative conductances for all the nanoplate designs docked into nanopores of various diameters in 100 mM KCl, 10 mM Tris, 1 mM EDTA, and 11 mM Mg^{2+}. The data for all designs is plotted at 200 mV and 300 mV. We observe a similar trend as seen with the data taken in 1M KCl, i.e., a smaller RC for smaller pore diameters. The conductance model used to fit the data at 1M KCl, gives a poor fit at 100 mM KCl. The model should, at low salt, be extended by integrating the contributions of surface charge, similar to previous work by Smeets et al.[31]

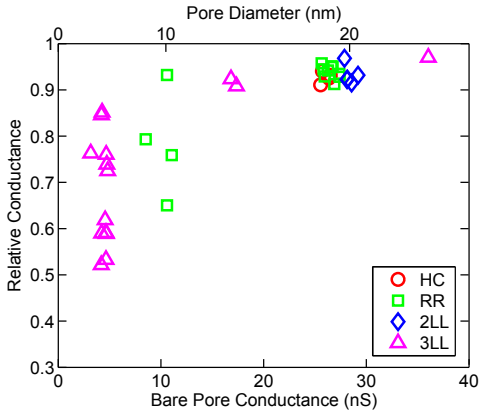


Figure 5.17: Relative conductances for the different nanoplate designs at 200 mV in 100 mM KCl.

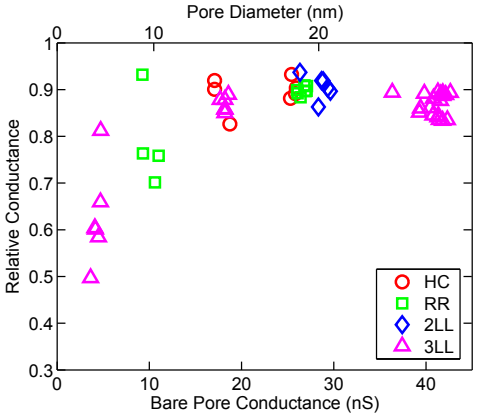


Figure 5.18: Relative conductances for the different nanoplate designs at 300 mV in 100 mM KCl.

5.5.3. Salt Dependence of a Honeycomb Nanoplate

The relative conductance of all nanoplates is observed to increase as the ionic strength is decreased from 1M to 100 mM. This trend is shown for HC nanoplate data in Figure 5.19 for three different applied voltage levels. Similar trends are observed for 3LL and RR nanoplates (data not shown).

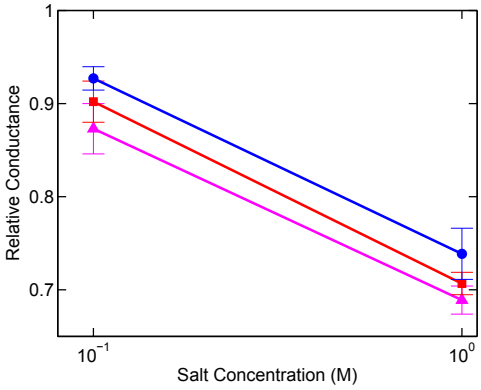


Figure 5.19: The relative conductance observed at low (100 mM) and high (1M) KCl concentrations for Honeycomb (HC) nanoplates docked onto 20 nm pores at 200 mV (top-blue), 300 mV (middle-red), and 400 mV (bottom-magenta).

5.5.4. Spike Events and Nanoplate Recaptures

If a nanoplate is very flexible and the nanopore's diameter is sufficiently large, the nanoplate is instantly pulled through the nanopore instead of being docked. These events show up in current traces as fast, high amplitude spikes as shown in Figure 5.20 for RR nanoplates passing through a 24 nm pore at several voltages. We successfully carried out recapture experiments, shown in Figure 5.21, in order to

confirm that the nanoplates are being pulled through the nanopores.

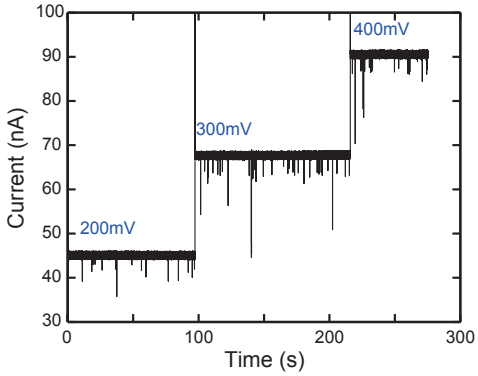


Figure 5.20: Short spike events observed for RR nanoplates in a 24 nm pore at voltages of 200 mV or higher.

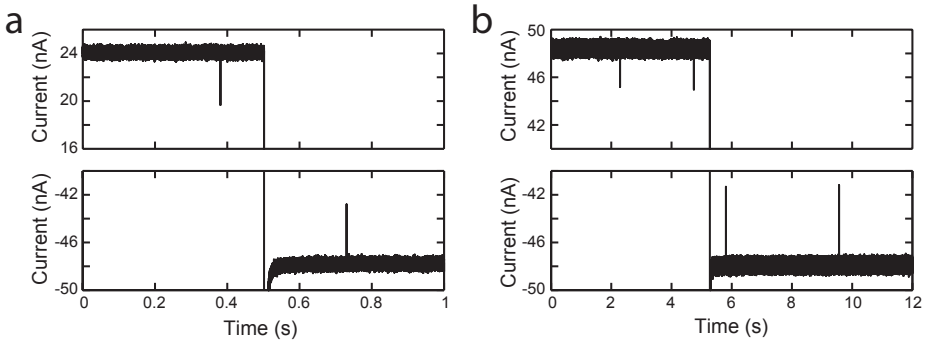


Figure 5.21: The recapture of RR nanoplates that are pulled through a 24 nm pore. (a) An event is observed at a voltage of 100 mV. The voltage is then switched to -200 mV, 120 ms after the first event. The nanoplate is recaptured 228 ms after switching. (b) Two short events are observed at 200 mV applied voltage. The polarity is reversed 2.98 s after the first event and 530 ms after the second. After reversal two recapture events are observed at 530 ms and 4.29 s after switching.

5.5.5. Multi-Level Conductance in Docked Nanoplates

Current traces from docked nanoplates often exhibit spontaneous sudden jumps in the current level. Examples of this multi-level conductance effect are shown in Figure 5.22 for HC nanoplates in several different pores. The phenomenon of multiple conductance levels was observed with all nanoplate designs, in all pore diameters, at all salt concentrations tested. The source of this effect is attributed to mechanical buckling and re-orientation of the nanoplate. This is supported by the observation that current level jumps are often observed before a docked nanoplate is pulled through the nanopore, as shown in Figure 5.23 and Figure 5.24.

Statistics on docked nanoplates reveal that, except for the 2LL nanoplate, 60% to 75% of docked nanoplates have multiple levels at 100 mV applied voltage, as shown in Figure 5.25. The effect occurs more frequently as the voltage is increased. The magnitude of these jumps, given as the change in relative conductance normalized by the average relative conductance, is shown in Figure 5.26. It varies from 0.05 to 0.07 at 100 mV, and increases to 0.08 to 0.1 at 200 mV.

In considering possible sources for the observed current jumps we also considered the free staple oligos present in the DNA nanoplate solution after purification. The possibility of free staples causing the observed current jumps can be ruled out by several experimental observations. (1) We observe that not all nanoplates show the current jumps, even within the same experiment with the same buffer containing excess staples. For example, in Figure 5.25 at 100 mV, at least 25% of events show no jumping behavior. If excess staples were the cause we would expect all plates to show some jumping behavior. (2) We typically see current jumps occur in both directions, i.e. towards higher and towards lower current values. If current jumps were due to staples approaching the plate, the observance of both downward and upward current jumps would mean that staples which are brought to the nanoplate either pass through the plate, which is extremely unlikely, or return back into solution which is also unlikely due to the high electric fields. (3) Let us nevertheless assume that free staples temporarily get stuck to the surface of the DNA nanoplate. What magnitude current drop would we expect? The magnitude of the jumps we observe ranges from 5% to 10% of the baseline value (Figure 5.26). If we assume optimistic values for the hydration volume of the free staples (3 nm radius of gyration) and use the standard volume exclusion formula to calculate the expected current blockade produced by a free staple inside the pore we find values that vary from 1.5% for 20 nm pores, to 36% for 5 nm pores. This does not match our observations which show no significant pore size dependence for the current jumps. Furthermore, excluded volume analysis states that there should be no voltage dependency for the normalized blockade, while our data shows that the normalized blockade increases with voltage for most plates (Figure 5.26). (4) Using pessimistic retention values of 30% for the low-MW oligos and 95% for the high-MW DNA nanoplates, if we start with 20 nM scaffold DNA and 200 nM staples, we estimate final concentrations of 14 nM nanoplates and 14 nM staples after 4x purification (assuming a retention volume of 58uL). Such a free staple concentration of 14 nM should produce an event rate of about 4Hz. The observation that the frequency of buckling is observed to vary substantially among different plates in the same experiment, suggests that the free staples are not the source of buckling.

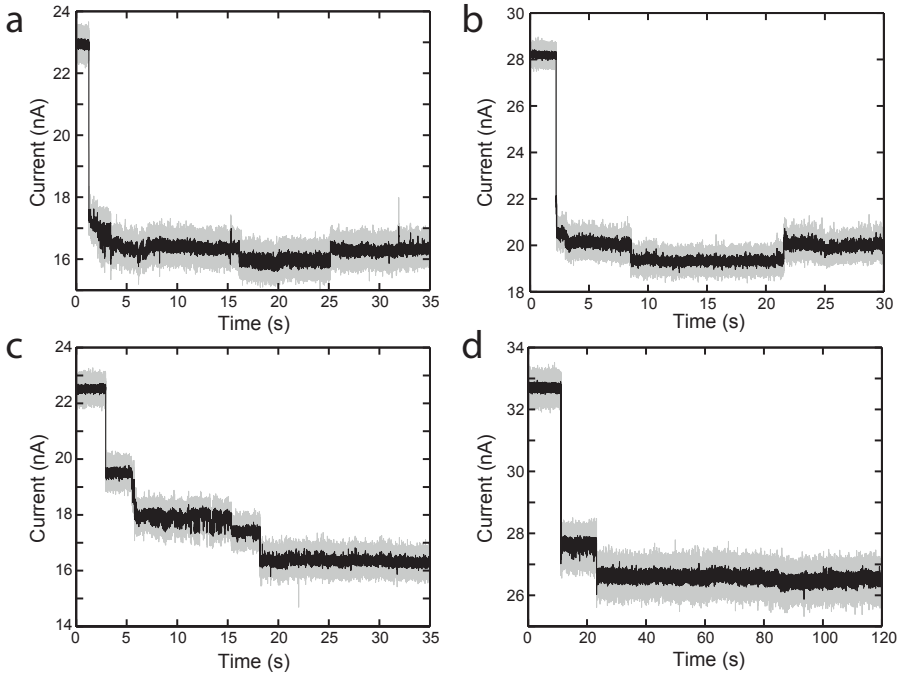


Figure 5.22: Examples of Honeycomb nanoplate docking events in several different pores exhibiting multiple conductance levels.

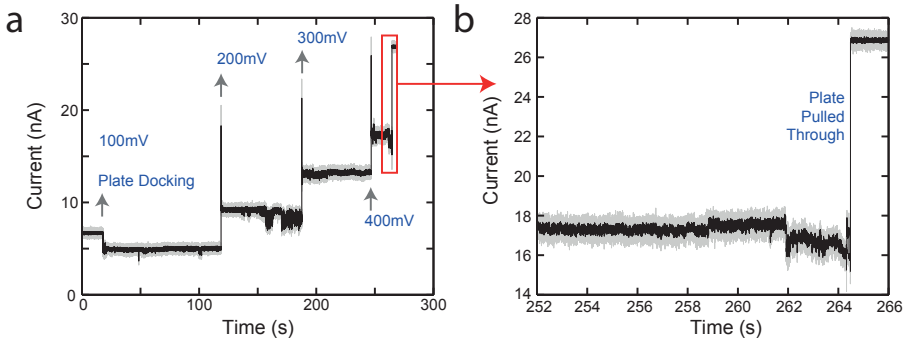


Figure 5.23: a) A 3 Layer Lattice nanoplate is captured in a 10 nm pore at 100 mV. The applied voltage is then increased to 200 mV, 300 mV, and 400 mV. The nanoplate is finally pulled through the pore after the voltage was set to 400 mV. b) Close-up of the trace in the seconds before the nanoplate is pulled through. The presence of multiple levels in the current trace right before the nanoplate is pulled through (Figure 5.24) supports the hypothesis that these levels are related to mechanical buckling of the nanoplates.

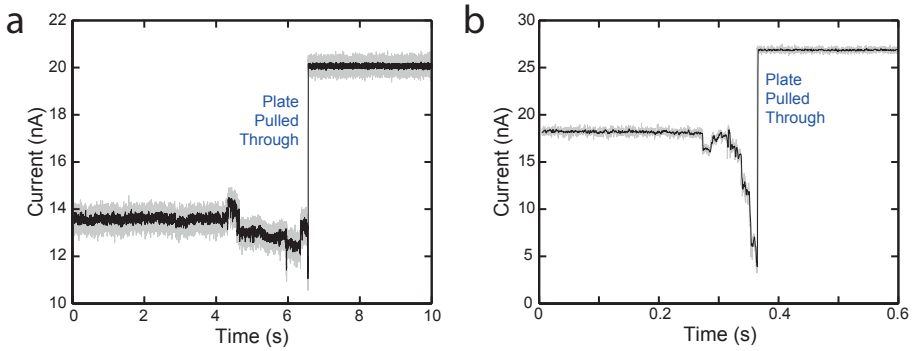


Figure 5.24: Examples of two other nanoplates exhibiting, similar to the data in Figure 5.23, multiple conductance levels right before being pulled through the pore.

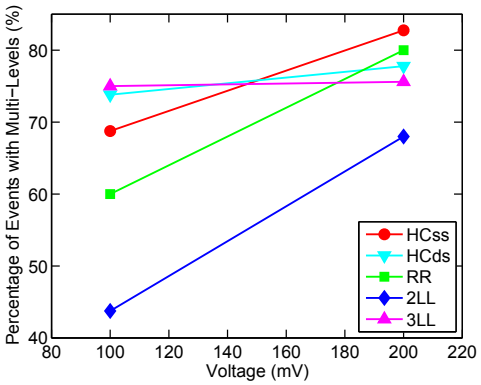


Figure 5.25: Percentage of events which displayed conductance levels jumps, while docked into a nanopore, as a function of voltage. In all cases we see the percentage of events with jumps to increase as the voltage is increased.

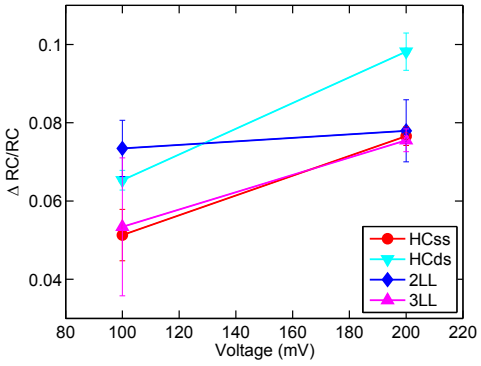


Figure 5.26: The normalized magnitude of the conductance levels shifts (ΔRC/RC) as a function of voltage. The magnitude of the jumps increases as the voltage is increased, except for the 2LL nanoplate.

5.5.6. Comparison of Single-Stranded and Double-Stranded DNA Tails

Two HC nanoplates, one with a ssDNA tail and another with a fully hybridized ds-DNA tail, were compared, as shown in Figure 5.27, in order to determine if the nanoplate's tail influenced the observed conductance values. The slightly lower relative conductance seen for the dsDNA tail can be explained by its larger excluded volume, since the tail is threaded through the pore while the nanoplate is docked.

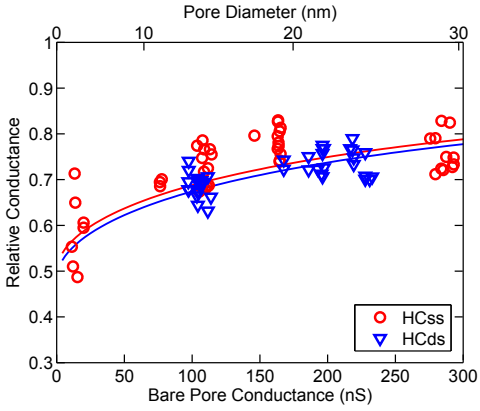


Figure 5.27: Relative conductance for Honeycomb nanoplates with single-stranded and double-stranded DNA tails docked at 200 mV in 1M KCl. The slightly lower conductance of the nanoplate with a dsDNA tail is attributed to the tail's higher excluded volume.

5.5.7. TEM and AFM Characterization

The long single stranded tail forms a large blob above the nanoplate as visible in Figures 5.28 and 5.29. The nanoplates were also scanned with AFM under a variety of different ionic conditions and found to be stable (Figure 5.30). Montages of the TEM micrographs for all plates are shown in Figures 5.31 to 5.34.

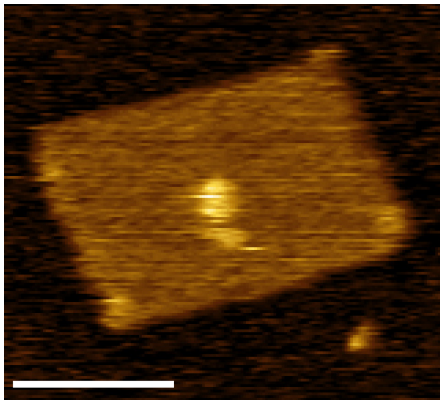


Figure 5.28: An AFM image of a RR nanoplate with the ssDNA tail visible. The scale bar is 50 nm.

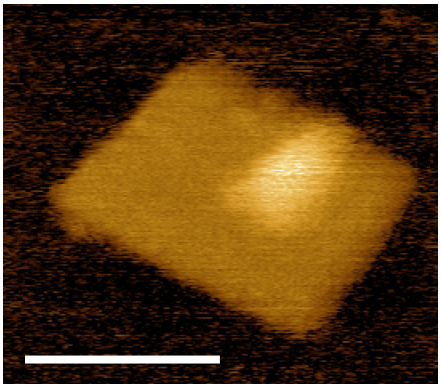


Figure 5.29: An AFM image of a 2LL nanoplate with the ssDNA tail visible. The scale bar is 50 nm.

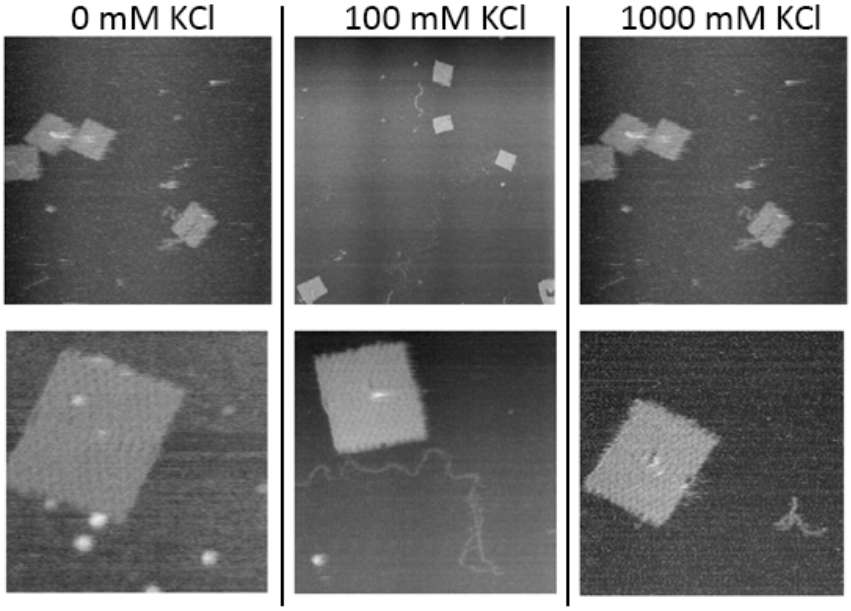


Figure 5.30: The RR nanoplate imaged at three different ionic strengths. No significant differences are noticeable between the different ionic conditions. Left: 0 mM KCl, Center: 100mM KCl, Right: 1000mM KCl.

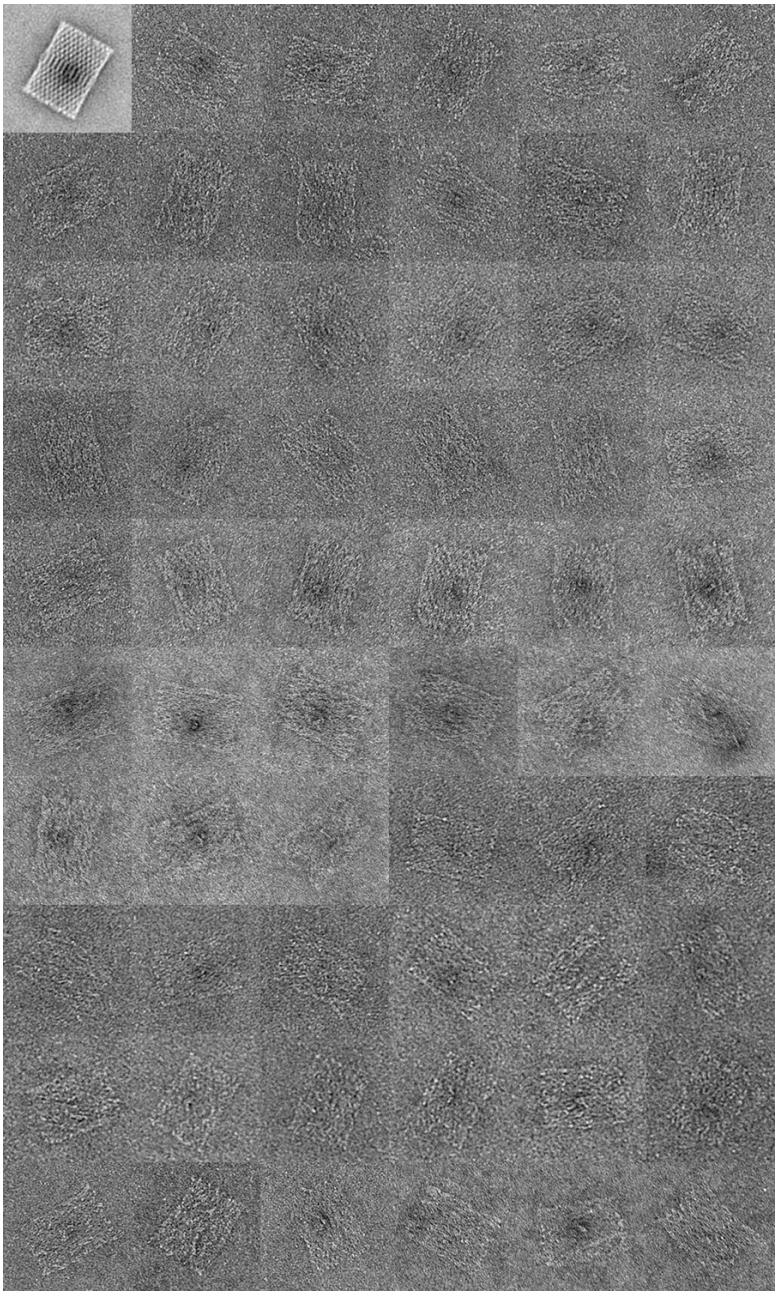


Figure 5.31: The negative stain micrographs of the RR nanoplate, with the average shown in the top left. Each micrograph is 137nm x 137nm. Adapted from Sobczak et al.[21]

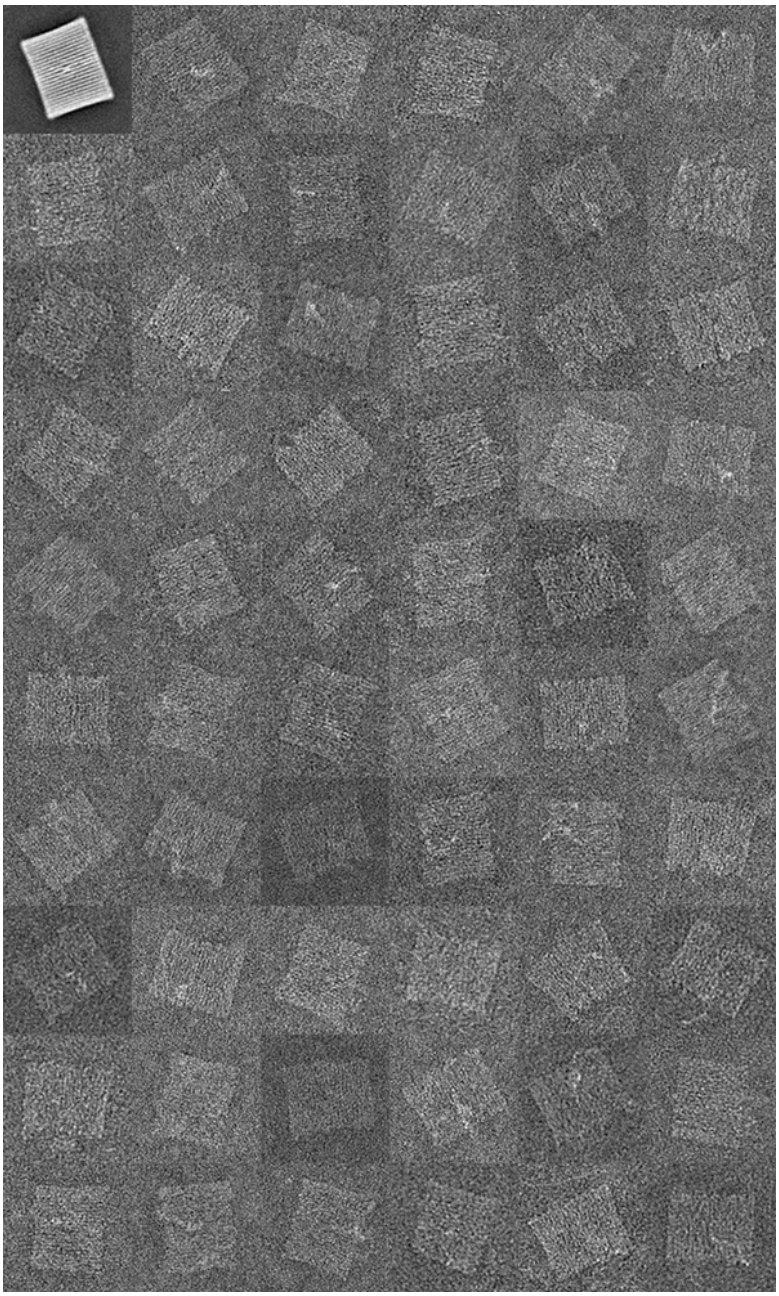


Figure 5.32: The negative stain micrographs of the 2LL nanoplate, with the average shown in the top left. Each micrograph is 91nm x 91nm.

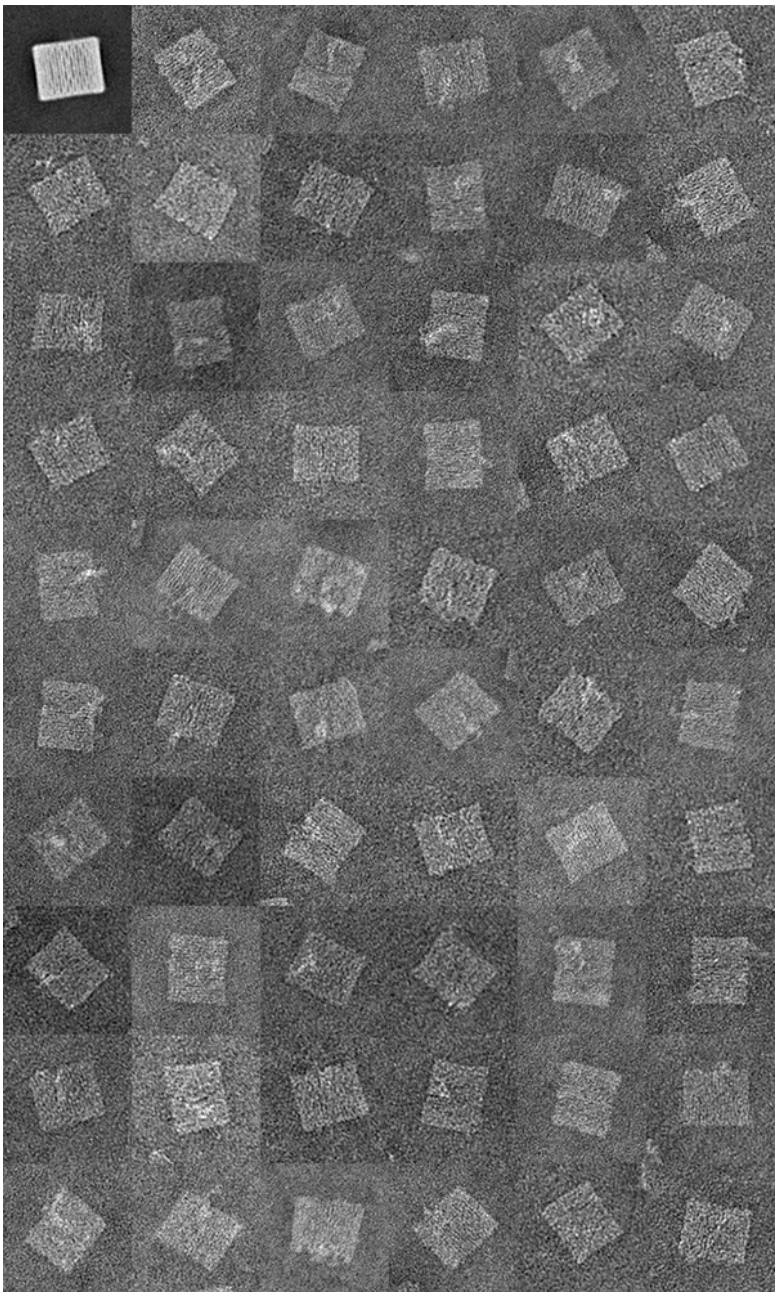


Figure 5.33: The negative stain micrographs of the 3LL nanoplate, with the average shown in the top left. Each micrograph is 91nm x 91nm.

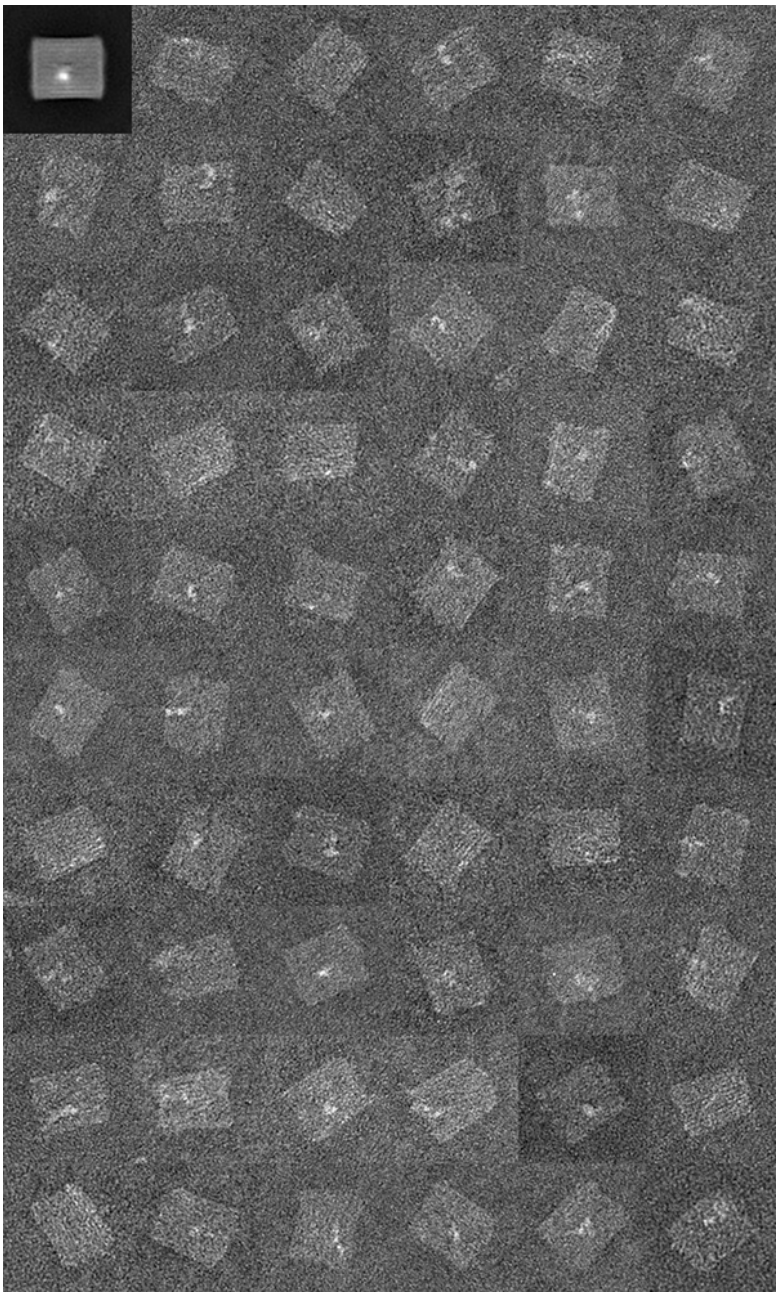


Figure 5.34: The negative stain micrographs of the HC nanoplate, with the average shown in the top left. Each micrograph is 91nm x 91nm.

5.5.8. Nanoplate Design Details

A detailed description of each nanoplate design is provided. CanDo modeling results are shown for each plate (Figure 5.35 to 5.38) as heat maps of the RMS fluctuations in different regions of each object. Briefly, these fluctuations are determined by applying external forces to an origami object which has been modeled as a series of elastic rods where cross-overs are rigid constraints and observing the structural relaxation using finite element analysis.[19] as well as a gel-electrophoresis characterization of assembled plates (Figure 5.39). Finally, the scaffold routing diagrams for each plate design as generated by caDNAno as well as the sequences of all of the oligo-staples used to generate each nanoplate are available in the online supplementary information for the accompanying publication[32].

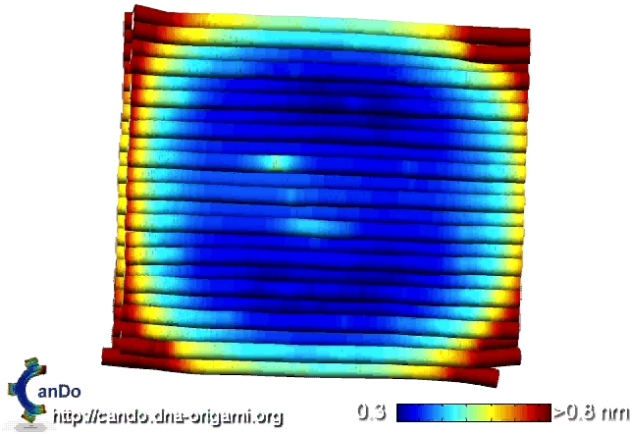


Figure 5.35: CanDo simulation of the HC plate. This plate is predicted to exhibit the lowest flexibility of all of the designs, as evidenced by the small RMS fluctuations present over much of the plate.

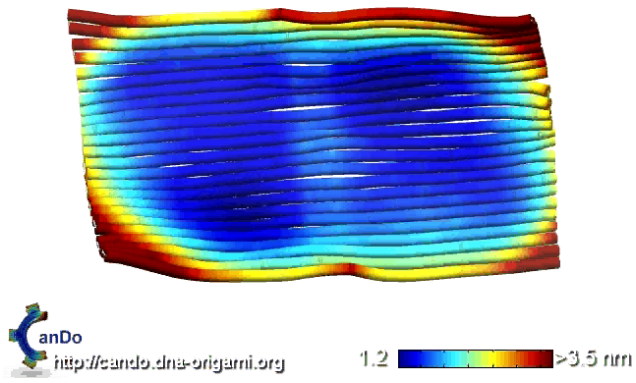


Figure 5.36: CanDo simulation of the RR plate. This plate is predicted to exhibit the highest flexibility of all of the designs.

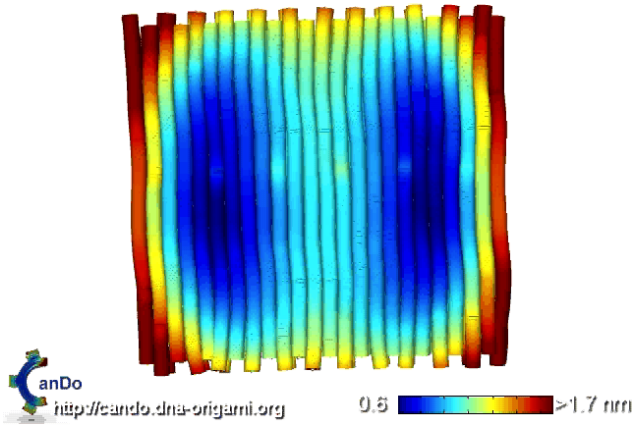


Figure 5.37: CanDo simulation of a 2LL plate. This plate is predicted to exhibit a lower flexibility than the RR plate, but higher than 3LL or HC.

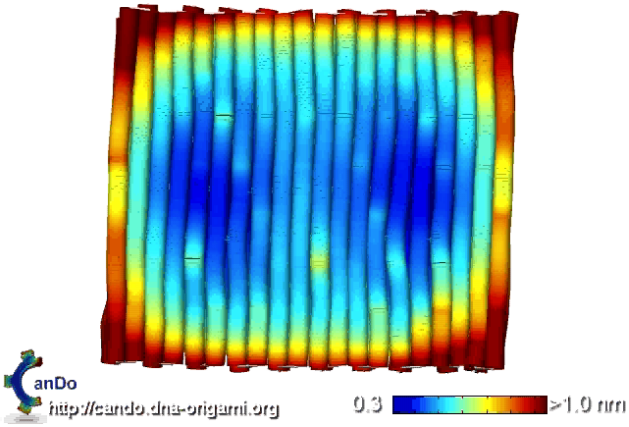


Figure 5.38: CanDo simulation of a 3LL plate. This plate is predicted to exhibit a low flexibility, slightly larger than the HC.

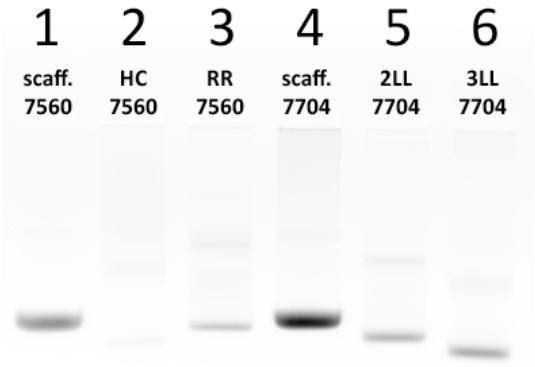


Figure 5.39: The spin-filtered nanoplates run on a 2% agarose gel containing EtBr. From left to right: scaffold 7560, HC (7560), RR (7560), scaffold 7704, 2L (7704), 3L (7704). The scaffold is a reference only for the running speed of structures, the concentrations are not comparable. All the excess strands have been filtered out as described in the methods.

References

[1] M. Wanunu, *Nanopores: A journey towards DNA sequencing,* Physics of Life Reviews **9**, 125 (2012).

[2] F. Haque, J. Li, H.-C. Wu, X.-J. Liang, and P. Guo, *Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA,* Nano Today **8**, 56 (2013).

[3] V. Tabard-Cossa, D. Trivedi, M. Wiggin, N. Jetha, and A. Marziali, *Noise analysis and reduction in solid-state nanopores,* Nanotechnology **18**, 305505 (2007).

[4] B. M. Venkatesan and R. Bashir, *Nanopore sensors for nucleic acid analysis,* Nat Nano **6**, 615 (2011).

[5] D. Rotem, L. Jayasinghe, M. Salichou, and H. Bayley, *Protein detection by nanopores equipped with aptamers,* Journal of the American Chemical Society **134**, 2781 (2012).

[6] A. R. Hall, A. Scott, D. Rotem, K. K. Mehta, H. Bayley, and C. Dekker, *Hybrid pore formation by directed insertion of [alpha]-hemolysin into solid-state nanopores,* Nat Nano **5**, 874 (2010).

[7] P. W. K. Rothemund, *Folding DNA to create nanoscale shapes and patterns,* Nature **440**, 297 (2006).

[8] E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf, and J. Kjems, *Self-assembly of a nanoscale DNA box with a controllable lid,* Nature **459**, 73 (2009).

[9] S. M. Douglas, H. Dietz, T. Liedl, B. Hogberg, F. Graf, and W. M. Shih, *Self-assembly of DNA into nanoscale three-dimensional shapes,* Nature **459**, 414 (2009).

[10] D. Han, S. Pal, J. Nangreave, Z. Deng, Y. Liu, and H. Yan, *DNA origami with complex curvatures in three-dimensional space,* Science **332**, 342 (2011).

[11] H. Dietz, S. M. Douglas, and W. M. Shih, *Folding DNA into twisted and curved nanoscale shapes,* Science **325**, 725 (2009).

[12] R. Wei, T. G. Martin, U. Rant, and H. Dietz, *DNA origami gatekeepers for solid-state nanopores,* Angewandte Chemie, International Edition **51**, 4864 (2012).

[13] N. A. W. Bell, C. R. Engst, M. Ablay, G. Divitini, C. Ducati, T. Liedl, and U. F. Keyser, *DNA origami nanopores,* Nano Letters **12**, 512 (2012).

[14] S. Hernández-Ainsa, N. A. W. Bell, V. V. Thacker, K. Göpfrich, K. Misiunas, M. E. Fuentes-Perez, F. Moreno-Herrero, and U. F. Keyser, *DNA origami nanopores for controlling DNA translocation,* ACS Nano **7**, 6024 (2013).

[15] M. Langecker, V. Arnaut, T. G. Martin, J. List, S. Renner, M. Mayer, H. Dietz, and F. C. Simmel, *Synthetic lipid membrane channels formed by designed DNA nanostructures,* Science **338**, 932 (2012).

[16] J. R. Burns, E. Stulz, and S. Howorka, *Self-assembled DNA nanopores that span lipid bilayers,* Nano Letters **13**, 2351 (2013).

[17] X.-c. Bai, T. G. Martin, S. H. W. Scheres, and H. Dietz, *Cryo-EM structure of a 3d DNA-origami object,* Proceedings of the National Academy of Sciences of the United States of America **109**, 20012 (2012).

[18] Y. Ke, S. M. Douglas, M. Liu, J. Sharma, A. Cheng, A. Leung, Y. Liu, W. M. Shih, and H. Yan, *Multilayer DNA origami packed on a square lattice,* Journal of the American Chemical Society **131**, 15903 (2009).

[19] C. E. Castro, F. Kilchherr, D.-N. Kim, E. L. Shiao, T. Wauer, P. Wortmann, M. Bathe, and H. Dietz, *A primer to scaffolded DNA origami,* Nat Meth **8**, 221 (2011).

[20] D.-N. Kim, F. Kilchherr, H. Dietz, and M. Bathe, *Quantitative prediction of 3D solution shape and flexibility of nucleic acid nanostructures,* Nucleic Acids Research **40**, 2862 (2012).

[21] J.-P. J. Sobczak, T. G. Martin, T. Gerling, and H. Dietz, *Rapid folding of DNA into nanoscale shapes at constant temperature,* Science **338**, 1458 (2012).

[22] S. W. Kowalczyk, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Modeling the conductance and DNA blockade of solid-state nanopores,* Nanotechnology **22**, 315101 (2011).

[23] J. M. Arbona, J.-P. Aimé, and J. Elezgaray, *Modeling the mechanical properties of DNA nanostructures,* Physical Review E: Statistical, Nonlinear, and Soft Matter Physics **86**, 051912 (2012).

[24] R. B. Schoch, J. Han, and P. Renaud, *Transport phenomena in nanofluidics,* Reviews of Modern Physics **80**, 839 (2008).

[25] C. Ho, R. Qiao, J. B. Heng, A. Chatterjee, R. J. Timp, N. R. Aluru, and G. Timp, *Electrolytic transport through a synthetic nanometer-diameter pore,* Proceedings of the National Academy of Sciences

of the United States of America **102**, 10445 (2005).
[26] T. G. Martin and H. Dietz, *Magnesium-free self-assembly of multi-layer DNA objects,* Nat Commun **3**, 1103 (2012).
[27] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, *Mechanical separation of the complementary strands of DNA,* Proceedings of the National Academy of Sciences of the United States of America **94**, 11935 (1997).
[28] T. Strunz, K. Oroszlan, R. Schäfer, and H.-J. Güntherodt, *Dynamic force spectroscopy of single DNA molecules,* Proceedings of the National Academy of Sciences of the United States of America **96**, 11277 (1999).
[29] S. W. Kowalczyk, M. W. Tuijtel, S. P. Donkers, and C. Dekker, *Unraveling single-stranded DNA in a solid-state nanopore,* Nano Letters **10**, 1414 (2010).
[30] X. J. A. Janssen, M. P. Jonsson, C. Plesa, G. V. Soni, C. Dekker, and N. H. Dekker, *Rapid manufacturing of low-noise membranes for nanopore sensors by trans -chip illumination lithography,* Nanotechnology **23**, 475302 (2012).
[31] R. M. M. Smeets, U. F. Keyser, D. Krapf, M.-Y. Wu, N. H. Dekker, and C. Dekker, *Salt dependence of ion transport and DNA translocation through solid-state nanopores,* Nano Letters **6**, 89 (2006).
[32] C. Plesa, A. N. Ananth, V. Linko, C. Gülcher, A. J. Katan, H. Dietz, and C. Dekker, *Ionic permeability and mechanical properties of DNA origami nanoplates on solid-state nanopores,* ACS Nano **8**, 35 (2013).

Part II

DNA-Protein

6

Fast translocation of proteins through solid-state nanopores

Measurements on protein translocation through solid-state nanopores reveal anomalous (non-Smoluchowski) transport behavior, as evidenced by extremely low detected event rates, i.e., the capture rates are orders of magnitude smaller than what is theoretically expected. Systematic experimental measurements of the event rate dependence on the diffusion constant are performed by translocating proteins ranging in size from 6 kDa to 660 kDa. The discrepancy is observed to be significantly larger for smaller proteins, which move faster and have a lower signal-to-noise ratio. This is further confirmed by measuring the event rate dependence on the pore size and concentration for a large 540 kDa protein and a small 37 kDa protein, where only the large protein follows the expected behavior. We dismiss various possible causes for this phenomenon, and conclude that it is due to a combination of the limited temporal resolution and low signal-to-noise ratio. A one-dimensional first-passage time-distribution model supports this and suggests that the bulk of the proteins translocate on timescales faster than can be detected. We discuss the implications for protein characterization using solid-state nanopores and highlight several possible routes to address this problem.

This chapter has been published as: C. Plesa, S. W. Kowalczyk, R. Zinsmeester, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Fast Translocation of Proteins through Solid State Nanopores*, Nano Letters **13**, 2, 658-663 (2013).

6.1. Introduction

Over the past decade, nanopores have become a very popular method to study analytes at the single-molecule level. While the focus has been on nucleic acids[1, 2], proteins and protein DNA-complexes increasingly are becoming a prime target of investigation[3–18].

There are two notable differences which make proteins more difficult to characterize than DNA by use of nanopore measurements. The well-defined structure of proteins, the source of their biological function, is normally globular, with lengths in all dimensions that are comparable to or smaller than the thickness of the pore. DNA and other polynucleotides, however, typically have one dimension much larger than the length of the pore, leading to much longer dwell times inside the nanopore compared to proteins. Secondly, proteins have a variety of different charges, that are distributed unevenly throughout their structure. Again, this contrasts the case for DNA which is highly negatively charged with a uniform charge distribution along its length. These differences make nanopore experiments on proteins more challenging than on DNA. This study focuses at the issues caused by the protein's short translocation times and the implications this has for characterization of proteins with solid-state nanopores.

Figure 6.1 shows the setup and data for a typical nanopore protein measurement. A voltage is applied across the pore, which electrokinetically drives a protein through it, temporarily blocking the ionic current. The dynamics of the translocation process can be split into two stages, the capture step and the actual translocation through the nanopore. The former involves the biomolecule diffusing to the pore, which can be characterized by the rate at which molecules are captured. This process is governed by the diffusion constant of the analyte, assisted by the applied electric field. The rate J at which the analytes arrive at the pore entrance due to free diffusion is given by the Smoluchowski rate equation [19]

$$J = 2\pi c D r_p \tag{6.1}$$

where c is the bulk analyte concentration, D the diffusion constant, and r_p the radius of the pore. With these three values at hand, we can calculate the expected event rate based on this very well established equation.

6.2. Results and Discussion

We collected data for the event rate of a large variety of proteins [3–5, 7, 9–18, 20, 21] both from in-house experiments as well as from surveying published literature. This yielded a data set with 37 different protein-translocation experiments. It includes experiments on 12 different protein species conducted using SiN pores ranging in diameter from 10 nm to 57 nm at voltages from 50 mV to 200 mV. Values for the diffusion constant were determined separately for each of the protein species involved, as described in Supplementary Section 6.4.2. Figure 6.2 shows the ratio of the observed event rate to that predicted by Smoluchowski for this data set, as a function of the concentration multiplied by the pore radius (left plot) and of the molecular weight (right plot).

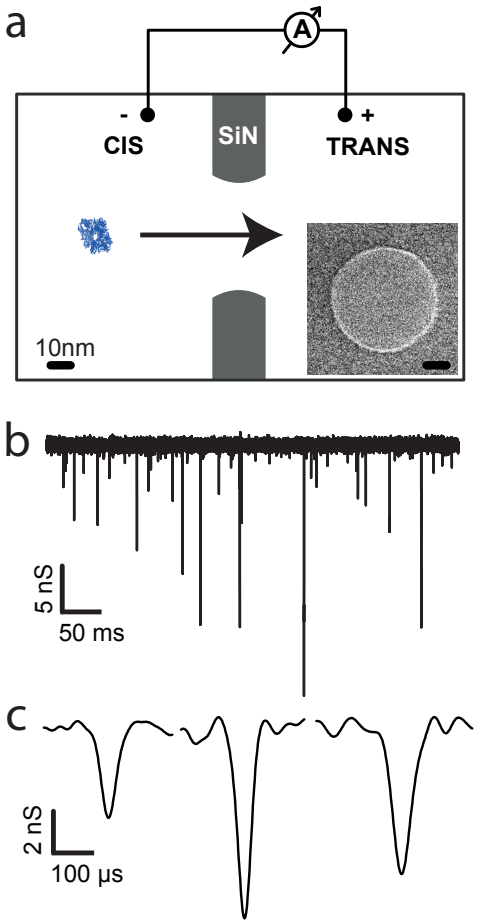


Figure 6.1: a) Schematic illustration of the nanopore setup, with the 20 nm thick SiN membrane and a β-Galactosidase protein shown to scale. Insert: TEM image of one of the 40 nm nanopores used in this study. b) Current trace of β-Galactosidase translocating through a 40 nm pore. c) Representative translocation events, shown at better temporal resolution.

If free diffusion would govern the capture rate, we would expect a ratio of 1, irrespective of the values of c, D, or r_p. For field-assisted captures, we even expect this ratio to be greater than 1 in all experiments due to the electrophoretic force driving the protein into the pore and thus increasing the capture rate. Instead we find that while the ratio is indeed larger than 1 for dsDNA translocation, it is up to five orders of magnitude lower than expected for proteins, with larger and larger discrepancies for larger cr_p. When the same data is shown plotted against the molecular weight (Figure 6.2 right) of the particular species being measured, it becomes clear that the ratio becomes significantly lower as the size of the protein becomes smaller. Note that while the smaller proteins translocate faster and have a

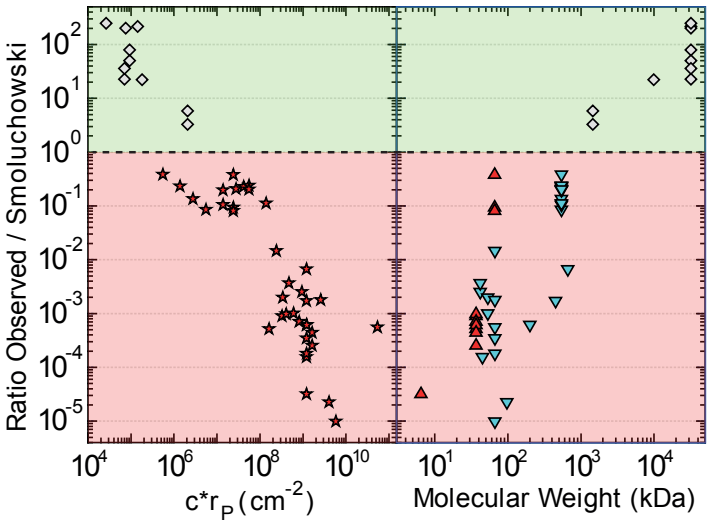


Figure 6.2: (Left) The ratio of the observed event rate to the event rate predicted by the Smoluchowski rate equation for dsDNA (white diamonds) and protein (red stars). We expect all points to be in the green area with a ratio larger than one. Instead, we see that all protein measurements yield a ratio less than 1, with some values even up to five orders of magnitude smaller. This study focuses on explaining this discrepancy. (Right) Same data set versus the molecular weight. Proteins have been separated into positively charged (red triangles) and negatively charged (blue triangles) proteins. The ratio becomes worse as the size of the protein becomes smaller, which can be attributed to their faster diffusion as well as smaller excluded volumes.

lower amplitude blockade due to their smaller excluded volume, it does not explain the spectacular failure of the Smoluchowski equation to describe the rate of their capture by nanopores.

How to explain this large discrepancy? Several possible explanations come to mind: (1) Adsorption [5, 14, 17, 18] could explain lower capture rates, if the protein of interest would get trapped on the sidewalls of the flowcell or on the SiN membrane. In order to rule this out, two PMMA flowcells containing SiN membranes were incubated for one hour with 2.5 μM BSA or 2 μM RecA, respectively. Subsequently, the protein solutions were recovered from the flowcells and the protein concentration was determined using a spectrophotometer. No detectable change was observed in the bulk protein concentrations of either solution. Additionally, we modified our analysis procedure to track possible changes in the event rate over the course of the translocation experiment, expecting the observed event rate to decrease if more proteins were absorbed and the bulk concentration was lowered. Most proteins were found to exhibit a constant event rate (e.g. Figure 6.5). Some drop was observed for Avidin, which had its event rate steadily lowered, until becoming less than half its initial value after 20 min (Supplementary Section 6.4.3). While this shows that adsorption may sometimes be a contributing factor in devices in which the ratio of surface to volume is not negligible, its effect is too small to explain the low event rates observed. (2) Aggregation of the proteins might be an-

other contributing factor, but it is unlikely to be a dominant effect, given the large number of different protein species and the broad range of concentrations tested. Moreover, dynamic-light-scattering measurements of the hydrodynamic radius for several of the proteins showed no evidence of aggregation. (3) Electroosmosis is known to play an important role in protein translocation of nanopores, and can even be responsible for the translocation of proteins occurring at a polarity opposite to what is expected from simple electrostatics[3]. For typical pH values used in these experiments, from 7 to 8, the SiN surface of the pore is known to be negatively charged[3], with a magnitude that depends on a number of factors, particularly the SiN membrane treatment. For negatively charged protein, the electroosmotic force opposes the translocation of the proteins. The fact, however, that very low event rates are also observed for positively charged proteins, which are aided by the electrosmotic force, rules this out as the primary source of the event rate discrepancy. (4) One might question our assumption that the capture process is diffusion controlled. Indeed, barrier-limited capture was previously reported in protein capture into small pores[9, 11]. However, we focus our study on large (> 10 nm) nanopores with diameters much larger than the protein size, and we observe no evidence (such as exponential voltage dependence) of barrier-limited transport.

Finally, (5) a possible explanation is the finite temporal resolution in nanopore experiments. Indeed, we conclude that limited temporal resolution is the most likely explanation for the anomalously low event rates observed in virtually all nanopore probing experiments on proteins. The idea is that proteins move through the nanopore so swiftly that they escape detection. The limited temporal resolution can be attributed to two sources. First, the bandwidth of the amplifier. The commonly used Axon Axopatch 200B amplifier has a bandwidth of about 52 kHz.[22] Secondly, the signal-to-noise ratio (SNR) of the translocation events is limited, and recorded signals are filtered in order to distinguish events from the noise inherently present in the system. This distorts the shape of the events with a duration shorter than twice the rise time of the filter used ($T_r = 0.332/BW$ for a Gaussian low pass filter (LPF)), which lowers their amplitude and can significantly decrease the apparent dwell time[18]. This is illustrated in Figure 6.3a, where rectangular pulses with durations from 200 μs to 10 μs are filtered with a 10 kHz Gaussian LPF. If an event has a duration sufficiently below $2T_r$, equal to 66 μs in this case, its amplitude drops below the detection level, the analysis software treats it as noise, the protein translocation is not noticed, and the event rate is underestimated. We investigated this effect for Thyroglobulin (660 kDa), one of the proteins used in the experimental section of this work (Supplementary Section 6.4.6). The expected conductance blockade was determined using its excluded volume, as others have done in the past[4, 5, 12, 15]. Pulses with this conductance blockade and ranging in duration from 1 μs to 100 μs were then filtered in order to determine the minimum detectable pulse duration for the characteristic noise level in our pores. The minimum detectable pulse duration (MDPD) was estimated to be 25 μs, which is larger than the 20 μs temporal resolution of our amplifier. Since the MDPD becomes larger as proteins become smaller, we conclude that the SNR is nearly always the limiting factor setting the temporal resolution. In other words, the smaller the protein is,

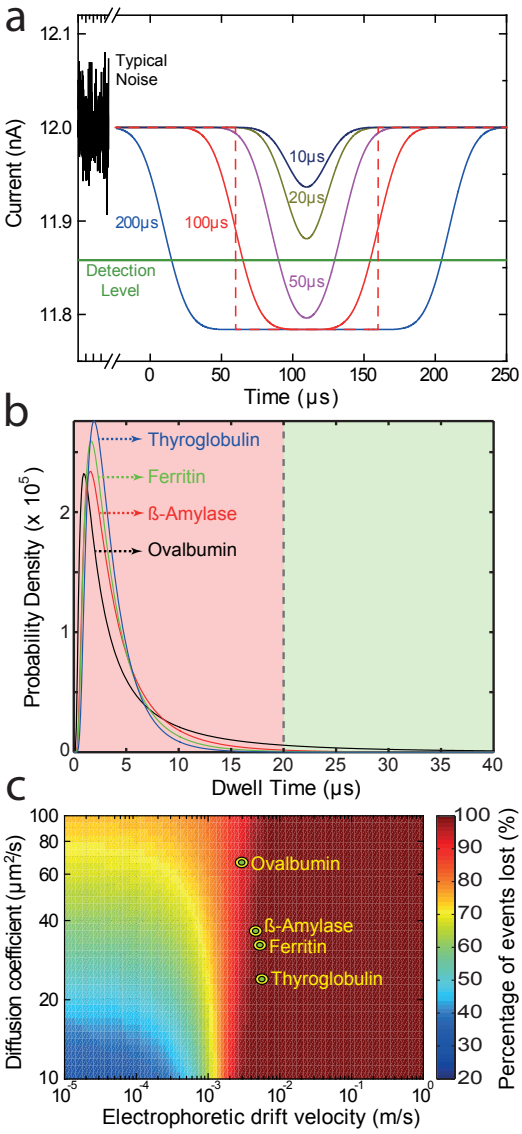


Figure 6.3: a) The distortion of rectangular pulses of 10 to 200 μs duration by a 10 kHz Gaussian filter. For a 100 μs pulse duration the dashed red line shows an ideal pulse (delayed 40 μs for visual clarity), before filtering, while the solid red line shows the same 100 μs pulse after filtering. All pulses have the same amplitude before filtering. Due to the limited rise time of the filter, any pulse with a duration less than twice the filter rise time ($2T_r$ = 66 μs) is distorted. b) A simulated dwell-time histogram showing visible (green) and lost (red) events given a resolution limit of 20 μs. At this resolution we expect the smallest protein, Ovalbumin, to show the most events. If the temporal resolution would be improved to below 6 μs, however, the largest protein, Thyroglobulin, would produce the most observable events. c) The percentage of events below the temporal resolution (color scale) for various values of the diffusion coefficient and electrophoretic drift velocity. The estimated positions of four proteins are shown. We expect the majority of events for these proteins to be lost.

the more events we will miss as they are shorter than the MDPD. This implies that we should observe a decrease in the event rate with decreasing protein size, as indeed is observed in Figure 6.2 (right panel).

We now quantitatively estimate how many events are lost due to the finite temporal resolution. We do so using the first-passage time distribution (FPTD) for one-dimensional diffusion as derived for nanopores by Talaga et al[11]. This model provides the probability that a protein with an electrophoretic drift velocity v, set by the voltage drop across the pore, will translocate with a given dwell time t, as detailed in Supplementary Section 6.4.5. This was used to generate dwell time distributions for various values of the diffusion constant and electrophoretic drift velocity. Four of these distributions, corresponding to four of the proteins used in the experimental part of our study, are shown in Figure 6.3b. For each of these proteins diffusion constants are determined using the method of He and Niemeyer[23]. This involves using the molecular weight and radius of gyration as correlation parameters to determine the diffusion constant. The electrophoretic drift velocities of the proteins were estimated based on their Stokes radii and their net charge, as determined from zeta-potential measurements[24] that we carried out on these proteins (Supplementary Section 6.4.4). As Figure 6.3b displays, the analysis reveals that smaller proteins (e.g. Ovalbumin) have a most probable dwell time which is smaller compared to the larger proteins (e.g. Thyroglobulin), but the tail of their distribution is larger. Counterintuitively, therefore, smaller proteins can therefore yield more detectable events. This effect is expected to reverse as the temporal resolution of the system becomes smaller than 6 μs which however is smaller than the temporal resolution of commercial amplifiers. The amount of events lost for a particular protein is set by its size, net charge, and the minimum time resolvable by the system. The fraction of events below the temporal resolution (set as 20 μs) within the typical range of values for the diffusion constant and the electrophoretic drift velocity is shown in Figure 6.3c. As the positions of the proteins on this diagram show, the vast majority of protein translocations will not be observed under normal conditions. This suggests that any events that are observed either belong to the far tail of the dwell time distribution, or are held long enough inside the pore due to protein-pore interactions to be resolved.

In order to experimentally verify that the limited temporal resolution is the cause of the discrepancy observed, the event rate dependence was investigated separately for each of the parameters in the Smoluchowski rate equation. The dependence on the diffusion constant was studied by translocating five proteins (Aprotinin 6.5 kDa, Ovalbumin 45 kDa, β-Amylase 200 kDa, Ferritin 450 kDa, Thyroglobulin 660 kDa) of varying size, each at a concentration of 1 μM, through a 40 nm SiN pore, in 1 M NaCl with pH 7.5, at an applied voltage of 50 mV. Based on Smoluchowski's equation as well as our FPTD simulation, one would expect the event rate to increase with increasing D, since proteins should be able to diffuse to the pore entrance faster. Experimentally, however, we see just the opposite, as shown in Figure 6.4a: the event rate decreases as the size of the protein becomes smaller and D thus grows. The observed decrease must be attributed to the decrease in the SNR as the proteins get smaller (Figure 6.9b). Aprotinin, the smallest protein

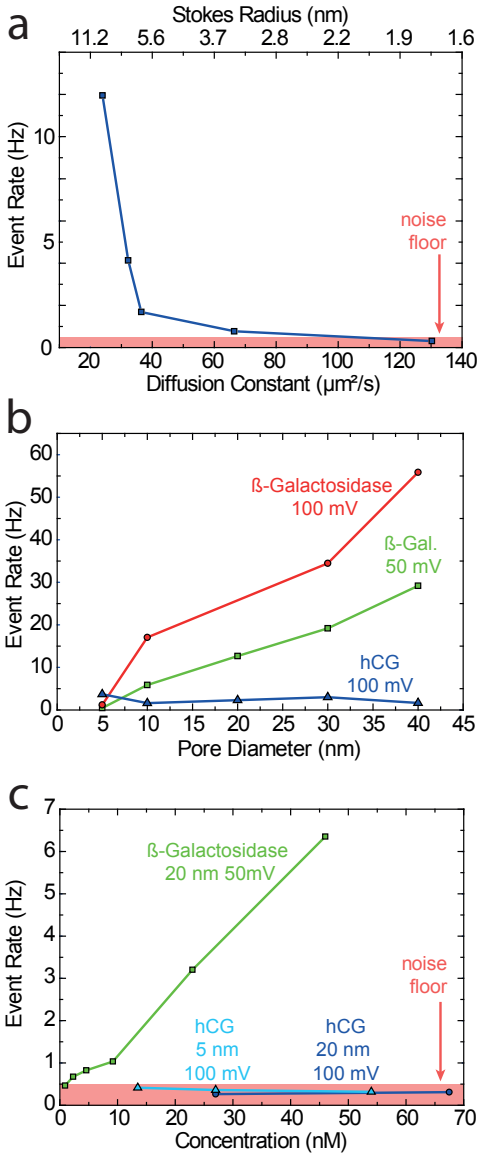


Figure 6.4: a) The event rate of protein translocation through a 40 nm nanopore for Aprotinin (6.5 kDa), Ovalbumin (45 kDa), β-Amylase (200 kDa), Ferritin (450 kDa), Thyroglobulin (660 kDa) at 1 μM concentration. The event rate is observed to decrease as the protein becomes smaller, i.e., when the diffusion constant becomes larger. This is counter to the predictions by both Smoluchowski and the FPTD model and is attributed to the reduction in the SNR as the proteins become smaller. b) The pore-size dependence of the event rate for 46 nM 540 kDa β-Galactosidase and 1.35 μM 37 kDa hCG. The event rate of β-Galactosidase increases linearly with pore size, as expected, while no change is observed for hCG. c) The concentration dependence of the event rate for these proteins. The event rate of β-Galactosidase increases linearly with concentration, as predicted by Smoluchowski. The event rate of hCG remains within the noise floor for all measurements.

tested, had an event rate within the noise floor of the system. This noise floor of false positives is determined by the rate of noise peaks with sufficient amplitude to be detected as apparent events, and is estimated to be between 0.3 and 0.5 Hz for our system (Supplementary Section 6.4.7). The noise floor is shown in the panels of Fig. 6.4 as a light red area at the bottom, except in panel b where it has been omitted for visual clarity. In order to test the dependence of the rate on the pore-diameter, a large 540 kDa protein at 46 nM concentration, β-Galactosidase, and a small 37 kDa protein at 1.35 μM concentration, human chorionic gonadotropin (hCG) were translocated through pores varying in diameter from 5 nm to 40 nm, as shown in Figure 6.4b. The event rate increased linearly with pore diameter for the large β-Galactosidase, as predicted by Smoluchowski. But the event rate for hCG remained low and relatively constant, indicating non-Smoluchowski behavior. Finally, the concentration dependence for these two proteins was investigated, by applying increasingly concentrated solutions to a flow cell containing a 20 nm pore, as shown in Figure 6.4c. Again, β-Galactosidase followed the expected Smoluchowski behavior, whereas the hCG's event rate remained within the noise floor of the system. In an attempt to investigate the dependence of an increased SNR for the hCG, we also checked J for two concentrations with a 5 nm pore (light blue data points in Figure 6.4c), but we did not observe any difference in behavior. By further increasing the hCG concentration, we finally did observe the event rate rise above the noise floor at very high concentrations above 1 μM.

6.3. Conclusion

What are the implications and possible solutions to this issue of a very significantly reduced event rate for protein translocation through nanopores? The findings presented in this study suggest that most of the protein translocation events that are observed are either due to protein-pore interactions or represent the far tail of the residence-time distribution. In patch-clamp techniques, analysis software such as Qub[25] routinely takes into account events that are missed due to limited temporal resolution in the estimation of kinetic parameters but this has, to our knowledge, so far not been incorporated for nanopores. Some previous studies[26, 27] have attributed low event rates in small antibiotic molecules translocating through membrane pores to limited temporal resolution. Additionally this subject has been touched upon briefly in the past for protein translocation through solid state nanopores[3, 9, 11, 15]. If nanopores are to be used in biosensors, one requirement would be the ability to determine the concentration of analyte in the sample being probed. Given the high percentage of lost events, which also varies widely with protein mass and charge, it seems unlikely that this can be achieved without significant technological improvements.

Solutions to address this temporal issue can be divided into two categories: (i) slowing down the speed of the biomolecules, or (ii) improving the bandwidth of the amplifier utilized. Traditionally the former approach has involved altering buffer properties such as the viscosity or the temperature, both of which have the drawback of simultaneously reducing the amplitude of the signal[28]. For dsDNA, it has been shown that an 18°K drop in temperature can increase the dwell time by

1.7x while decreasing the conductance blockade by 1.5x[28]. Similarly, increasing the viscosity by 4.4x can increase the dwell time by 3.7x while reducing the signal by 2.2x[28]. In the case of proteins, the pH of the buffer can be modified to be close to the isoelectric point of the protein being studied. This will slow down the translocation velocity by reducing the charge of the protein[15]. Alternatively, the protein of interest can be immobilized within the detection volume of the pore for long periods of time using either surface modification[29–31] of the pore or designed DNA-origami structures inserted into the pore[32, 33]. Both these approaches implement a bait and prey type system. More recently, Yusko et al[12] have demonstrated an elegant technique which involves coating the nanopore sidewalls with a fluidic lipid bi-layer and anchoring the target analyte to the bi-layer. Both Yusko et al and others[34] have shown that the diffusion of these anchored analytes is dominated by the in-plane diffusion constant of the lipids, which is two orders of magnitude smaller then the typical diffusion constant for proteins in free solution. As this approach does not modify the inner measurement channel, there is no loss in signal. The second possible solution involves improving the electronics of the system, in order to directly measure the translocation signal at higher bandwidth. Several low-noise high-bandwidth amplifier designs have been proposed in the past few years[35–38]. Some of these have been designed specifically for use with nanopores and have been demonstrated to allow measurements on short DNA oligos that have previously been invisible. Assuming we apply one of these amplifiers, say with a bandwidth of 1 MHz, to our protein measurements, our FPTD model predicts that for a temporal resolution of 1 μs, we would expect to observe most proteins and only 6% of events to be lost. Unfortunately, however, the increased bandwidth comes with an increase in noise at the higher frequencies, so realistically this will be difficult to achieve. As previously noted, the temporal resolution is limited by the SNR, meaning that the best resolution will occur in experiments with large proteins, small pores, and low-noise membranes.

While all of these methods may help improve the number of translocation events visible, none of these solutions seem to completely address the problem at hand. It is thus clear that the issues discussed in this paper will remain a problem for protein translocation experiments with nanopores well into the foreseeable future.

Acknowledgement

We would like to thank Christophe Danelon and Aleksei Aksimentiev for productive discussions, as well as Khin Sam Ly for contributions to nanopore measurements. This work was funded in part by the European Research Council research grant NanoforBio (no. 247072).

6.4. Supplementary Info

6.4.1. Materials and Methods

Solid-state nanopores were fabricated as previously described[39, 40]. Nanopores were treated with oxygen plasma for 1 min on both sides, before each measurement. Nanopore containing membranes were mounted into PMMA or PEEK flowcells

containing separate aqueous reservoirs on each side of the chip into which Ag/AgCl electrodes were inserted. The flowcell and the amplifier headstage were placed into a Faraday cage. The ionic current was recorded with an Axopatch 200B amplifier with the low-pass filter set to 100 kHz, the resulting signal was digitized at 500 kHz. Unless otherwise noted the traces were digitally filtered with a 10 kHz Gaussian low pass filter. All analysis was done with a set of custom Matlab scripts with GUI front ends, to be described elsewhere. Event rates were calculated by doing an exponential fit to the distribution of times between the leading edges of events. All buffers used 10 mM Tris-HCl with 1mM EDTA at the pH values given in Supplementary Section 6.4.2.

Aprotinin (Bovine lung), Ovalbumin (Hen egg), Ferritin (Horse spleen), and Thyroglobulin (Bovine thyroid) were all purchased from GE Healthcare. β-Amylase (Sweet potato) was acquired from Sigma-Aldrich. β-Galactosidase was purchased from Roche Applied Science. hCG (human pregnancy urine) was from Abcam. Streptavidin and Avidin were from Invitrogen. BSA, RecA, and λ-DNA were from New England Biolabs.

6.4.2. Experimental translocation data for proteins and dsDNA

Tables 6.1, 6.2, and 6.3 provide the experimental data used in Figure 6.2. Experiments without a reference indicate work done as part of this study.

With the exception of Avidin, all diffusion constants were determined using the equation of He and Niemeyer[23]

$$D = \frac{6.85 \cdot 10^{-8} T}{\eta \sqrt{M^{1/3} R_g}} \tag{6.2}$$

where M is the molecular weight of the protein.

Using the worm-like chain model in the limit $l_p << l_c$, the radius of gyration is given by

$$\langle R_g \rangle = \sqrt{\frac{l_p l_c}{3}} \tag{6.3}$$

From the Zimm model for linear polymers we have the ratio between the radius of gyration and the hydrodynamic radius as

$$\frac{R_g}{R_H} = \frac{8}{3}\sqrt{\pi} \cong 1.51 \tag{6.4}$$

Table 6.1: Experimental parameters for protein translocation experiments.

Protein	c	Buffer	pH	r_p	V	J_{obs}	J_{Smo}	J_{obs}/J_{Smo}	Ref.
	nM			nm	mV	Hz	Hz		
MBP	1560	1MKCl	7.5	10	50	11	4400	2.5E-03	[9]
MBP	780	1MKCl	7.5	10	50	8	2200	3.6E-03	[9]
Aprotinin	1000	1MNaCl	7.5	20	50	0.310	9850	3.1E-05	
Ovalbumin	1000	1MNaCl	7.5	20	50	0.77	5030	1.5E-04	
β-Amylase	1000	1MNaCl	7.5	20	50	1.69	2760	6.1E-04	
Ferritin	1000	1MNaCl	7.5	20	50	4.13	2440	1.7E-03	
Thyroglob.	1000	1MNaCl	7.5	20	50	11.95	1810	6.6E-03	
β-Gal.	0.92	1MKCl	8	10	50	0.465	1.23	3.8E-01	
β-Gal.	2.3	1MKCl	8	10	50	0.704	3.06	2.3E-01	
β-Gal.	4.6	1MKCl	8	10	50	0.825	6.13	1.3E-01	
β-Gal.	9.2	1MKCl	8	10	50	1.037	12.3	8.5E-02	
β-Gal.	23	1MKCl	8	10	50	3.2	30.6	1.0E-01	
β-Gal.	46	1MKCl	8	5	50	5.96	30.6	1.9E-01	
β-Gal.	46	1MKCl	8	10	50	12.66	61.3	2.1E-01	
β-Gal.	46	1MKCl	8	15	50	20.76	91.9	2.3E-01	
β-Gal.	46	1MKCl	8	20	50	29.23	123	2.4E-01	
β-Gal.	92	1MKCl	8	10	100	24.8	123	2.0E-01	
β-Gal.	230	1MKCl	8	10	50	34	306	1.1E-01	
hCG	270	1MKCl	8	10	100	0.34	666	5.1E-04	
hCG	540	1MKCl	8	10	100	1.18	1330	8.9E-04	
hCG	1350	1MKCl	8	5	100	1.6	1670	9.6E-04	
hCG	1350	1MKCl	8	10	100	2.32	3330	7.0E-04	
hCG	1350	1MKCl	8	15	100	3	5000	6.0E-04	
hCG	1350	1MKCl	8	20	100	1.67	6660	2.5E-04	
hCG	2700	1MKCl	8	10	100	2.92	6660	4.4E-04	
Streptav.	566	1MKCl	8	10	100	2.82	1440	2.0E-03	
Streptav.	1000	1MKCl	6.6	10	150	2.54	2540	1.0E-03	
Avidin	40	1MKCl	8	10	100	9.84	106	9.3E-02	
Avidin	40	1MKCl	8	10	50	8.48	106	8.0E-02	
Avidin	40	0.05MKCl	8	10	150	40	106	3.8E-01	[3]
BSA	1500	1MKCl	8	29	100	17	9590	1.8E-03	[5]
BSA	100k	0.4MKCl	7	9	200	111	202k	5.5E-04	[4]
BSA	4200	0.15MKCl	8	23	50	0.21	21.7k	9.7E-06	[7]
Importin-β	2900	0.15MKCl	8	23	50	0.27	12.2k	2.2E-05	[7]
BSA	400	1MKCl	8	10	100	12.96	897	1.4E-02	
BSA	2000	1MKCl	8	10	50	0.8	4480	1.8E-04	
BSA	2000	1MKCl	8	10	100	1.532	4480	3.4E-04	

Table 6.2: Protein information. † R_g of Avidin was determined using HydroPro[41].

Protein	M.W.	Stokes Radius (R_s)	Radius of Gyration (R_g)	PDB ID	Diffusion Constant (D)
	kDa	nm	nm		m^2/s
MBP	42	3.0 [42]	2.2 [43]	1URD	7.45E-11
Aprotinin	6.5	1.64	1.34 [44]	3LDM	1.30E-10
Ovalbumin	45	3.05 [45]	2.7 [46]	1OVA	6.65E-11
β-Amylase	200	4.8 [47]	5.4 [48]	1BTC	3.65E-11
Ferritin	450	6.1 [45]	5.33 [49]	1FHA	3.22E-11
Thyroglobulin	660	8.5 [45]	8.5 [48]		2.39E-11
β-Gal.	540	6.86 [50]	4.2 [51]	1BGL	3.52E-11
hCG	36.7	3.3 [52]	3.0 [53]	1HRP	6.52E-11
Streptavidin	53	3.69 [54]	2.5 [55]	2GH7	6.72E-11
Avidin	66	3.37	2.14†	1AVD	7.00E-11
BSA	66.46	3.48 [56]	2.98 [57]	3V03	5.93E-11
Importin-β	97		3.97 [58]	2P8Q	4.82E-11

Combining these two with the Stokes-Einstein equation

$$D = \frac{k_B T}{6\pi\eta R_H} = \frac{8\sqrt{\pi}k_B T}{18\pi\eta R_g} = \frac{8\sqrt{\pi}k_B T}{18\pi\eta\sqrt{\frac{l_p l_c}{3}}} \tag{6.5}$$

and using values of 50 nm for the persistence length l_c of dsDNA and $\eta = 8.9 \cdot 10^{-4} Pa \cdot s$ for 1M KCl we obtain

Table 6.3: dsDNA translocation experimental data. ‡ This study used Al_2O_3 membranes.

DNA length	c	Buffer	pH	r_p	V	J_{obs}	J_{Smo}	J_{obs}/J_{Smo}	Ref.
bp	nM			nm	mV	Hz	Hz		
2200	3.44	1M KCl	8	10	100	4.4	1.35	3.27	
2200	7.2	1M KCl	8	4.75	100	7.83	1.34	5.85	
48514	0.155	1M KCl	8	10	50	0.65	0.0129	50.34	
48514	0.155	1M KCl	8	10	100	1.03	0.0129	79.77	
48514	0.312	1M KCl	8	7.65	120	4.3	0.0199	215.99	
48514	0.125	1M KCl	8	10	120	2.1	0.0104	201.74	
48514	0.062	1M KCl	8	7	120	0.9	0.00364	247.02	
15000	1	1M KCl	8.5	3	100	1	0.0449	22.25	[21]
48514	0.156	1M KCl	8	7.5‡	200	0.22	0.00975	22.572	[20]
48514	0.156	1M KCl	8	7.5‡	300	0.35	0.00975	35.91	[20]

$$D = \frac{1.155 \cdot 10^{-18}}{R_g} = \frac{8.947 \cdot 10^{-15}}{\sqrt{l_c}} \tag{6.6}$$

Dynamic light scattering measurements were performed on several proteins in the same high salt (1 M) buffer used in the nanopore experiments. The measured radii, shown in Table 6.4, were consistent with values reported in literature and no evidence of aggregation was observed. Additionally the diffusion constant was calculated from the measured hydrodynamic radii using the Stokes-Einstein equation.

Table 6.4: The hydrodynamic radii of several proteins in 1 M salt, as measured by DLS.

Protein	Hydrodynamic Radius (R_H)	Diffusion Constant (D)
	nm	m^2/s
Ovalbumin	3.25	6.90E-11
β-Amylase	5.05	4.40E-11
Ferritin	6.75	3.30E-11
Thyroglobulin	9.1	2.50E-11
hCG	3.5	7.00E-11

6.4.3. Tracking Adsorption by Monitoring Changes in the Event Rate over Time

As a way to asses the amount of adsorption on to the flowcell walls and the SiN membrane, we tracked the event rate as a function of experimental time. Current traces were split up into 3 second segments and the event rate was determined within each of these segments. The positively charged protein Avidin, showed the largest drop over time, with the event rate reduced from 8.5 Hz at the start of the experiment to 3.4 Hz after 1233 seconds, as can be seen in Figure 6.5b. This was exceptional as there were no significant changes in event rates of other proteins, as shown for example in the Ferritin trace in Figure 6.5a.

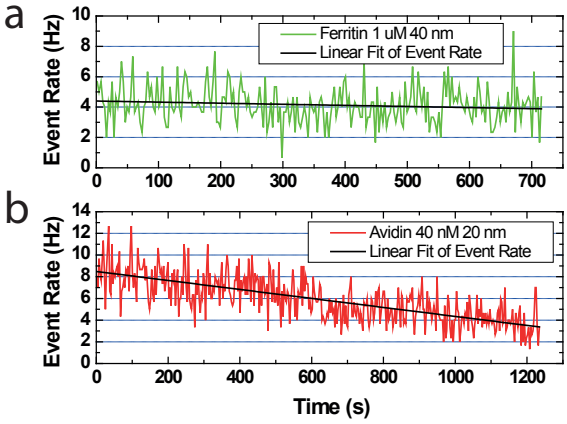


Figure 6.5: The event rate as a function of time for (a) Ferritin and (b) Avidin translocation experiments.

6.4.4. Protein Charge Measurements

In order to determine the charge of the proteins and calculate their electrophoretic drift velocities, the protein's zeta-potentials were measured using a Malvern Instruments Zetasizer Nano ZS, see Table 6.5.

Table 6.5: Zeta-potential, net valence, and electrophoretic drift velocity.

Protein	Zeta Potential (ζ_p)	Protein Valence	Electrophoretic Drift Velocity (ν)
	mV	|z|	m/s
Ovalbumin	-11.7	0.83	2.96E-03
β-Amylase	-12.3	2.05	4.63E-03
Ferritin	-11.4	2.99	5.32E-03
Thyroglobulin	-8.9	4.44	5.67E-03

The zeta potential (ζ_p) is related to the net valence of the protein (z) through the equation[24]

$$z = \frac{\zeta_p \varepsilon R_H \left[1 + \kappa \left(R_H + R_b\right)\right]}{e \left(1 + \kappa R_b\right)} \tag{6.7}$$

where R_b is the average radius of the salt ions (taken to be 0.25 nm), ε is the dielectric constant of the medium, and κ is the Debye–Hückel inverse screening length given by $\kappa = 3.289\sqrt{I}$ in units of nm^{-1}, where I is the ionic strength in M. The values of the net valence were not rounded to nearest integer, due to the possibility that multiple populations with different charged states may exist simultaneously (although β-Amylase and Ferritin valence values clearly correspond to -2e

and -3e respectively). This has been observed previously for Avidin[15] and was attributed to heterogeneity caused by various post-translational protein modification processes. The electrophoretic drift velocity is determined using

$$v = \mu_p E = \mu_p \frac{V_p}{h_t} = \frac{e|z|V_p}{6\pi\eta R_H h_t} \tag{6.8}$$

where V_p is the portion of the applied voltage that drops over the pore (taking into account the access resistance) and h_t is the effective thickness of the pore.

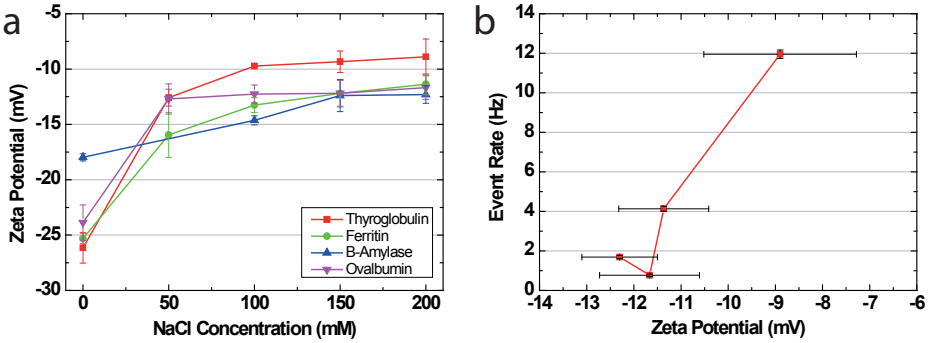


Figure 6.6: a) Zeta-potential measurements of the proteins used in this experiment as a function of increasing NaCl concentrations with pH 7.5. The largest ionic strength which could be measured was 200mM NaCl. No data could be taken for Aprotinin due to its small size. b) The event rates observed as a function of the protein's zeta-potential.

6.4.5. First Passage Time Model

The distributions shown in Figure 6.3b of the main paper were generated by using the diffusion constants and electrophoretic drift velocities calculated earlier and the cumulative distribution function of the 1D first passage time model as derived by Talaga et al.[11]

$$CDF_{FPTD}(t) = \frac{1}{2} erfc\left(\frac{h_t - vt}{2\sqrt{Dt}}\right) \tag{6.9}$$

with an effective membrane thickness (h_t) of 20 nm. Next the PDF(t) was recovered by inverse transform sampling this cumulative distribution a total of 1,000,000 times. A histogram of this data was generated and normalized. The surface plot shown in Figure 6.3c was calculated by computing the value of $CDF_{FPTD}(20\mu s)$ over the parameter space shown.

After the publication of this chapter, a correction to the 1D first passage time model derived by Talaga et al.[11] was published by Ling and Ling[59]. This correction results in only a small change in the distributions. None of the conclusions are affected. Figure 6.3b and 6.3c have been updated in this thesis using the correct form of the cumulative distribution function[60]

$$CDF_{FPTD}(t) = \frac{1}{2}\left(e^{h_t v/D} erfc\left(\frac{h_t + vt}{2\sqrt{Dt}}\right) + erfc\left(\frac{h_t - vt}{2\sqrt{Dt}}\right)\right) \tag{6.10}$$

and the probability density function[59]

$$PDF_{FPTD}(t) = \frac{h_t}{\sqrt{4\pi D t^3}} e^{-(h_t - vt)^2/4Dt} \tag{6.11}$$

6.4.6. Filtering Effects

The effects of filtering on the resulting capture rates were investigated. The dependence of the event rate on the filtering frequency for the set of experiments in Figure 6.4a of the main paper, is shown in Figure 6.7. While the amplitude of the event rate for each protein varies with the filter frequency, all of the proteins show the same behavior. This means that the trend seen in the event rate the dependence on protein size (Figure 6.4a) remains the same for all filter settings and with only the absolute values of the event rates changing. The experimentally measured detection level for the 40 nm pores used in these measurements is shown in Figure 6.8 as a function of the filtering frequency. This shows the minimum amplitude necessary to detect an event. Any events with an amplitude below the threshold (red line) will not be detected.

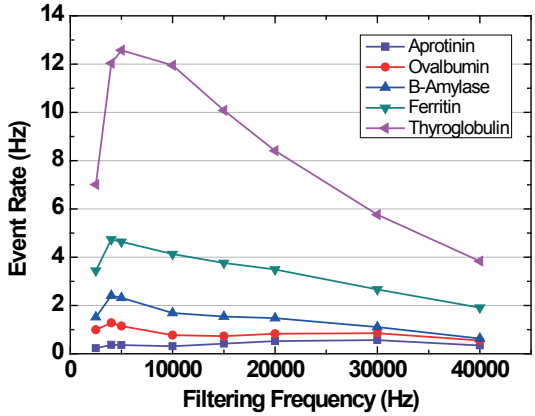


Figure 6.7: The event rate as a function of the filtering frequency used in the analysis.

Furthermore, Figure 6.9a shows the effect of filtering pulses having an initial amplitude corresponding to the expected blockade for Thyroglobulin (based on its excluded volume). In other words, we generate 4.32 nS pulses ranging in duration from 1 μs to 100 μs. Each of these pulses is then filtered with a 10 kHz Gaussian filter and the resulting filtered pulse amplitude is determined. The intersection of this curve with the threshold line provides the minimum detectable pulse duration (MDPD). This value is determined to be 25 μs for Thyroglobulin. Figure 6.9b compares all aspects of the experiment from a theoretical perspective. First, the black

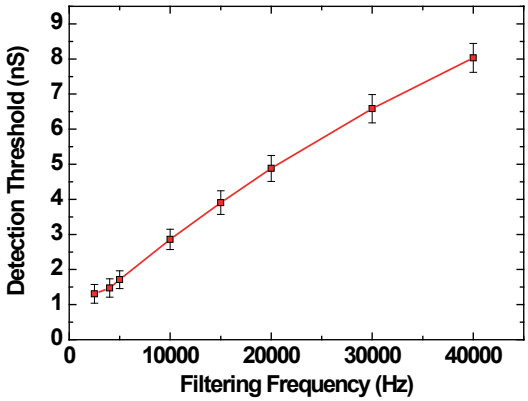


Figure 6.8: The level of the detection threshold, set as 4.5σ, as a function of the filter frequency as determined experimentally for 40 nm SiN nanopores. This level can be lowered by painting the membrane with a layer of PDMS[61] to reduce the noise.

squares show the expected dependence on the diffusion constant, for the five proteins measured in Figure 6.4a, as predicted by the Smoluchowski rate equation. We see the event rate increase with the diffusion constant. Secondly, the red circles represent the prediction of the one-dimensional first-passage time-distribution obtained by multiplying the event rates given by the Smoluchowski rate equation with the probability that an event will translocate within a time larger than 20 μs, the amplifier's temporal limit:

$$J_{FPTD(20\mu s)} = J_{Smo}\left[1 - CDF_{FPTD}(20\mu s)\right] \tag{6.12}$$

This also predicts an increase in the event rate as the diffusion constant increases. Finally, the blue triangles show the same dependence when the SNR ratio is taken into account. The MDPD was determined for Thyroglobulin and Ferritin and used to define the limit in the first-passage time-distribution:

$$J_{FPTD(MDPD)} = J_{Smo}\left[1 - CDF_{FPTD}(MDPD)\right] \tag{6.13}$$

We now see a decrease in the event rate as the diffusion constant increases, which matches the experimentally observed trend (Figure 6.4a).

For Thyroglobulin the expected amplitude was determined to be 4.32 nS using a shape factor[62] of 1.55 and a partial specific volume[63] of 0.723. For Ferritin the expected amplitude was determined to be 2.86 nS using a shape factor of 1.5 and a partial specific volume[64] of 0.738. In both cases a conductivity of 9.24 S/m was used for 1M NaCl and the expected current drop for each protein was calculated using the same method as Yusko et al.[65]

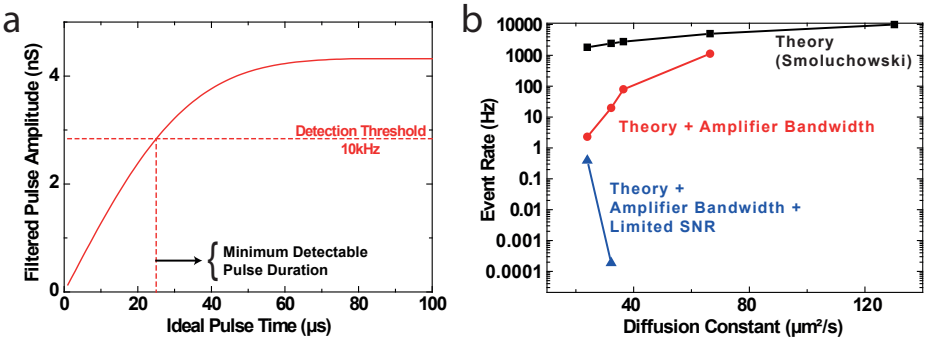


Figure 6.9: a) Pulses calculated for Thyroglobulin, ΔG = 4.32 nS are Gaussian low-pass filtered to 10 kHz and 5 kHz. The amplitude of the filtered pulse is shown versus pulse duration and compared to the detection threshold. Any event with a pulse duration smaller than the point at which the curve drops below the threshold level will not be detected. b) A comparison of three theoretical models: the Smoluchowski rate equation (black squares), the first-passage time-distribution with amplifier limit of 20 μs (red circles), the first-passage time-distribution with the temporal limit given by the limited SNR of each protein (blue triangles).

Table 6.6: The data used in Figure 6.9b.

Protein	D	J_{Smo}	At 20μs			At MPDT	
			1-CDF	J_{FPTD}	MDPD	1-CDF	J_{FPTD}
	m^2/s	Hz		Hz	µs		Hz
Aprotinin	1.30E-10	4.4					
Ovalbumin	6.65E-11	7.83	0.223546	1123.9			
β-Amylase	3.65E-11	0.65	0.028665	79.1			
Ferritin	3.22E-11	1.03	0.008055	19.6	70	7.76E-08	1.89E-04
Thyroglob.	2.39E-11	4.3	0.001272	2.3	25	2.17E-04	3.92E-01

6.4.7. Noise Floor

The noise floor of our system is defined by the rate of false events due to the noise inherently present in the system. For Gaussian distributed noise, the rate of false events is given by[66]

$$J_{False} = k f_c e^{\left(-\frac{\Phi^2}{2\sigma^2}\right)} \tag{6.14}$$

where k is a constant (0.849 flat noise; 1.25 f^2 noise), f_c is the bandwidth (10 kHz), Φ is the detection threshold, and σ is the rms current at this bandwidth. In our experiments the false event rate is expected to lie between 0.34 and 0.5 Hz. A detection threshold of 4.5σ was used in our analysis in order to capture as many low SNR events as possible.

References

[1] C. Dekker, *Solid-state nanopores,* Nat Nano **2**, 209 (2007).

[2] B. M. Venkatesan and R. Bashir, *Nanopore sensors for nucleic acid analysis,* Nat Nano **6**, 615 (2011).

[3] M. Firnkes, D. Pedone, J. Knezevic, M. Döblinger, and U. Rant, *Electrically facilitated translocations of proteins through silicon nitride nanopores: conjoint and competitive action of diffusion, electrophoresis, and electroosmosis.* Nano Letters **10**, 2162 (2010).

[4] D. Fologea, B. Ledden, D. S. McNabb, and J. Li, *Electrical characterization of protein molecules by a solid-state nanopore.* Applied Physics Letters **91**, 539011 (2007).

[5] A. Han, G. Schürmann, G. Mondin, R. A. Bitterli, N. G. Hegelbach, N. F. de Rooij, and U. Staufer, *Sensing protein molecules using nanofabricated pores,* Applied Physics Letters **88**, 093901 (2006).

[6] S. W. Kowalczyk, A. R. Hall, and C. Dekker, *Detection of local protein structures along DNA using solid-state nanopores,* Nano Letters **10**, 324 (2009).

[7] S. W. Kowalczyk, L. Kapinos, T. R. Blosser, T. Magalhaes, P. van Nies, Y. H. LimRoderick, and C. Dekker, *Single-molecule transport across an individual biomimetic nuclear pore complex,* Nat Nano **6**, 433 (2011).

[8] B. Ledden, D. Fologea, D. S. Talaga, and J. Li, *Sensing single protein molecules with solid-state nanopores,* Proteins , 129 (2011).

[9] A. Oukhaled, B. Cressiot, L. Bacri, M. Pastoriza-Gallego, J.-M. Betton, E. Bourhis, R. Jede, J. Gierak, L. Auvray, and J. Pelta, *Dynamics of completely unfolded and native proteins through solid-state nanopores as a function of electric driving force.* ACS Nano **5**, 3628 (2011).

[10] L. T. Sexton, L. P. Horne, S. a. Sherrill, G. W. Bishop, L. a. Baker, and C. R. Martin, *Resistive-pulse studies of proteins and protein/antibody complexes using a conical nanotube sensor.* Journal of the American Chemical Society **129**, 13144 (2007).

[11] D. S. Talaga and J. Li, *Single-molecule protein unfolding in solid state nanopores.* Journal of the American Chemical Society **131**, 9287 (2009).

[12] E. C. Yusko, J. M. Johnson, S. Majd, P. Prangkio, R. C. Rollings, J. Li, J. Yang, and M. Mayer, *Controlling protein translocation through nanopores with bio-inspired fluid walls.* Nat Nano **6**, 253 (2011).

[13] A. Spiering, S. Getfert, A. Sischka, P. Reimann, and D. Anselmetti, *Nanopore translocation dynamics of a single DNA-bound protein,* Nano Letters **11**, 2978 (2011).

[14] K. J. Freedman, M. Jürgens, A. Prabhu, C. W. Ahn, P. Jemth, J. B. Edel, and M. J. Kim, *Chemical, thermal, and electric field induced unfolding of single protein molecules studied using nanopores,* Analytical Chemistry **83**, 5137 (2011).

[15] A. Han, M. Creus, G. Schürmann, V. Linder, T. R. Ward, N. F. de Rooij, and U. Staufer, *Label-free detection of single protein molecules and protein-protein interactions using synthetic nanopores.* Analytical Chemistry **80**, 4651 (2008).

[16] R. Stefureac, D. Trivedi, A. Marziali, and J. S. Lee, *Evidence that small proteins translocate through silicon nitride pores in a folded conformation,* Journal of Physics: Condensed Matter **22**, 454133 (2010).

[17] D. J. Niedzwiecki, J. Grazul, and L. Movileanu, *Single-molecule observation of protein adsorption onto an inorganic surface,* Journal of the American Chemical Society **132**, 10816 (2010).

[18] D. Pedone, M. Firnkes, and U. Rant, *Data analysis of translocation events in nanopore experiments.* Analytical Chemistry **81**, 9689 (2009).

[19] M. v. Smoluchowski, Z. Phys. Chem. **92** (1917).

[20] P. Chen, J. Gu, E. Brandin, Y.-R. Kim, Q. Wang, and D. Branton, *Probing single DNA molecule transport using fabricated nanopores,* Nano Letters **4**, 2293 (2004).

[21] C. A. Merchant, K. Healy, M. Wanunu, V. Ray, N. Peterman, J. Bartel, M. D. Fischbein, K. Venta, Z. Luo, A. T. C. Johnson, and M. Drndic, *DNA translocation through graphene nanopores,* Nano Letters **10**, 2915 (2010).

[22] J. D. Uram, K. Ke, and M. Mayer, *Noise and bandwidth of current recordings from submicrometer pores and nanopores.* ACS Nano **2**, 857 (2008).

[23] L. He and B. Niemeyer, *A novel correlation for protein diffusion coefficients based on molecular weight and radius of gyration,* Biotechnology Progress **19**, 544 (2003).

[24] D. J. Winzor, *Determination of the net charge (valence) of a protein: a fundamental but elusive*

parameter, Analytical Biochemistry **325**, 1 (2004).
[25] F. Qin, A. Auerbach, and F. Sachs, *Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events,* Biophysical Journal **70**, 264 (1996).
[26] K. Mahendran, C. Chimerel, T. Mach, and M. Winterhalter, *Antibiotic translocation through membrane channels: temperature-dependent ion current fluctuation for catching the fast events,* European Biophysics Journal **38**, 1141 (2009).
[27] K. Mahendran, P. R. Singh, J. Arning, S. Stolte, U. Kleinekathöfer, and M. Winterhalter, *Permeation through nanochannels: revealing fast kinetics,* Journal of Physics: Condensed Matter **22**, 454131 (2010).
[28] D. Fologea, J. Uplinger, B. Thomas, D. S. McNabb, and J. Li, *Slowing DNA translocation in a solid-state nanopore.* Nano Letters **5**, 1734 (2005).
[29] M. Ali, P. Ramirez, M. N. Tahir, S. Mafe, Z. Siwy, R. Neumann, W. Tremel, and W. Ensinger, *Biomolecular conjugation inside synthetic polymer nanopores via glycoprotein-lectin interactions,* Nanoscale **3**, 1894 (2011).
[30] S. M. Iqbal, D. Akin, and R. Bashir, *Solid-state nanopore channels with DNA selectivity,* Nat Nano **2**, 243 (2007).
[31] R. Wei, V. Gatterdam, R. Wieneke, R. Tampe, and U. Rant, *Stochastic sensing of proteins with receptor-modified solid-state nanopores,* Nat Nano **7**, 257 (2012).
[32] N. A. W. Bell, C. R. Engst, M. Ablay, G. Divitini, C. Ducati, T. Liedl, and U. F. Keyser, *DNA origami nanopores,* Nano Letters **12**, 512 (2012).
[33] R. Wei, T. G. Martin, U. Rant, and H. Dietz, *DNA origami gatekeepers for solid-state nanopores,* Angewandte Chemie, International Edition **51**, 4864 (2012).
[34] L. Zhang, K. Dammann, S. C. Bae, and S. Granick, *Ligand-receptor binding on nanoparticle-stabilized liposome surfaces,* Soft Matter **3**, 551 (2007).
[35] G. Ferrari, M. Farina, F. Guagliardo, M. Carminati, and M. Sampietro, *Ultra-low-noise CMOS current preamplifier from DC to 1MHz,* Electronics Letters **45**, 1278 (2009).
[36] G. Ferrari, F. Gozzini, A. Molari, and M. Sampietro, *Transimpedance amplifier for high sensitivity current measurements on nanodevices,* Solid-State Circuits, IEEE Journal of **44**, 1609 (2009).
[37] J. K. Rosenstein, M. Wanunu, C. A. Merchant, M. Drndic, and K. L. Shepard, *Integrated nanopore sensing platform with sub-microsecond temporal resolution,* Nat Meth **9**, 487 (2012).
[38] A. Uddin, S. Yemenicioglu, C.-H. Chen, E. Corgliano, K. Milaninia, F. Xia, K. Plaxco, and L. Theogarajan, *Biosensing with integrated CMOS nanopores,* Proc. SPIE 8460, Biosensing and Nanomedicine V **8460**, 846010 (2012).
[39] D. Krapf, M.-Y. Wu, R. M. M. Smeets, H. W. Zandbergen, C. Dekker, and S. G. Lemay, *Fabrication and characterization of nanopore-based electrodes with radii down to 2 nm,* Nano Letters **6**, 105 (2005).
[40] X. J. A. Janssen, M. P. Jonsson, C. Plesa, G. V. Soni, C. Dekker, and N. H. Dekker, *Rapid manufacturing of low-noise membranes for nanopore sensors by trans -chip illumination lithography,* Nanotechnology **23**, 475302 (2012).
[41] A. Ortega, D. Amorós, and J. García de la Torre, *Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models,* Biophysical journal **101**, 892 (2011).
[42] P. A. Wearsch and C. V. Nicchitta, *Endoplasmic reticulum chaperone grp94 subunit assembly is regulated through a defined oligomerization domain ,* Biochemistry **35**, 16760 (1996).
[43] B. H. Shilton, M. M. Flocco, M. Nilsson, and S. L. Mowbray, *Conformational changes of three periplasmic receptors for bacterial chemotaxis and transport: The maltose-, glucose/galactose- and ribose-binding proteins,* Journal of Molecular Biology **264**, 350 (1996).
[44] M. S. Appavou, G. Gibrat, and M. C. Bellissent-Funel, *Influence of pressure on structure and dynamics of bovine pancreatic trypsin inhibitor (BPTI): Small angle and quasi-elastic neutron scattering studies,* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1764**, 414 (2006).
[45] G. D. Fasman, *Practical handbook of biochemistry and molecular biology / edited by Gerald D. Fasman* (CRC Press, Boca Raton, Fla.:, 1989).
[46] T. Matsumoto and H. Inoue, *Association state, overall structure, and surface roughness of native ovalbumin molecules in aqueous solutions at various ionic concentrations,* Journal of Colloid and Interface Science **160**, 105 (1993).
[47] L. K. Miller, F. B. Tuazon, E.-M. Niu, and M. R. Sherman, *Human breast tumor estrogen receptor:*

Effects of molybdate and electrophoretic analyses, Endocrinology **108**, 1369 (1981).
[48] J. Espeut, A. Gaussen, P. Bieling, V. Morin, S. Prieto, D. Fesquet, T. Surrey, and A. Abrieu, *Phosphorylation relieves autoinhibition of the kinetochore motor cenp-e,* Molecular Cell **29**, 637 (2008).
[49] P. Zipper, M. Kriechbaum, and H. Durchschlag, *Small-angle x-ray scattering studies on the polydispersity of iron micelles in ferritin,* J. Phys. IV France **03**, C8 (1993).
[50] M. Potschka, *Universal calibration of gel permeation chromatography and determination of molecular shape in solution,* Analytical Biochemistry **162**, 47 (1987).
[51] N. Nurdin, L. Canaple, A. Bartkowiak, B. Desvergne, and D. Hunkeler, *Capsule permeability via polymer and protein ingress/egress,* Journal of Applied Polymer Science **75**, 1165 (2000).
[52] C. S. Bambra, S. S. Lynch, G. R. Foxcroft, G. Robinson, and E. C. Amoroso, *Purification and characterization of guinea-pig chorionic gonadotrophin,* J Reprod Fertil **71**, 227 (1984).
[53] I. Pilz, E. Schwarz, W. Durchschein, A. Light, and M. Sela, *Effect of cleaving interchain disulfide bridges on the radius of gyration and maximum length of anti-poly(d-alanyl) antibodies before and after reaction with tetraalanine hapten,* Proceedings of the National Academy of Sciences of the United States of America **77**, 117 (1980).
[54] S.-C. Wu and S.-L. Wong, *Engineering soluble monomeric streptavidin with reversible biotin binding capability,* Journal of Biological Chemistry **280**, 23225 (2005).
[55] C. Desruisseaux, D. Long, G. Drouin, and G. W. Slater, *Electrophoresis of composite molecular objects. 1. relation between friction, charge, and ionic strength in free solution,* Macromolecules **34**, 44 (2000).
[56] S. Ikeda and K. Nishinari, *Intermolecular forces in bovine serum albumin solutions exhibiting solid-like mechanical behaviors,* Biomacromolecules **1**, 757 (2000).
[57] J. W. Anderegg, W. W. Beeman, S. Shulman, and P. Kaesberg, *An investigation of the size, shape and hydration of serum albumin by small-angle x-ray scattering1,* Journal of the American Chemical Society **77**, 2927 (1955).
[58] N. Roman, B. Kirkby, M. Marfori, B. Kobe, and J. K. Forwood, *Crystallization of the flexible nuclear import receptor importin-[beta] in the unliganded state,* Acta Crystallographica Section F **65**, 625 (2009).
[59] D. Y. Ling and X. S. Ling, *On the distribution of DNA translocation times in solid-state nanopores: an analysis using Schrodinger's first-passage-time theory,* Journal of Physics: Condensed Matter **25**, 375102 (2013).
[60] D. S. Talaga and J. Li, *Correction to "single-molecule protein unfolding in solid state nanopores",* Journal of the American Chemical Society **135**, 13220 (2013).
[61] V. Tabard-Cossa, D. Trivedi, M. Wiggin, N. Jetha, and A. Marziali, *Noise analysis and reduction in solid-state nanopores,* Nanotechnology **18**, 305505 (2007).
[62] H. Brumberger, J. L. Lipton, G. Dorfman, and E. Nakano, *Small-angle x-ray scattering of pig thyroglobulin in solution,* Biochemical Journal **143**, 495 (1974).
[63] Y. Derrien, R. Michel, K. O. Pedersen, and J. Roche, *Recherches sur la preparation et sur les proprietes de la thyroglobuline pure. ii,* Biochimica et Biophysica Acta **3**, 436 (1949).
[64] C. de Haën, *Molecular weight standards for calibration of gel filtration and sodium dodecyl sulfate-polyacrylamide gel electrophoresis: Ferritin and apoferritin,* Analytical Biochemistry **166**, 235 (1987).
[65] E. C. Yusko, P. Prangkio, D. Sept, R. C. Rollings, J. Li, and M. Mayer, *Single-particle characterization of aβ oligomers in solution,* ACS Nano **6**, 5909 (2012).
[66] B. Sakmann and E. Neher, *Single-Channel Recording* (Springer Science+Business Media, 2009).

7

Detection of single DNA-bound proteins using solid-state nanopores

Detection of DNA-bound proteins is a challenge that is well suited to solid-state nanopores as they provide a linear readout of the DNA and DNA-protein volume in the pore constriction along the entire length of a molecule. Here we demonstrate that we can realize the detection of even individual DNA-bound proteins at the single-DNA-molecule level using solid state nanopores. We introduce and use a new model system of anti-DNA antibodies bound to lambda phage DNA. This system provides several advantages since the antibodies bind individually, tolerate high salt concentrations, and due to their positive charge, will not translocate through the pore unless bound to the DNA. Translocation of DNA-antibody samples reveals the presence of short 15 μs current spikes within the DNA traces, with amplitudes that are about 4.5 times larger than that of dsDNA. Given the resolution limits of the method, we conclude that transient interactions between the pore and the antibodies are the primary mechanism by which bound antibodies are observed. This work provides a proof-of-concept for how nanopores could be used for future sensing applications.

C. Plesa, J. W. Ruitenberg, M. J. Witteveen, and C. Dekker, *Detection of single DNA-bound proteins using solid state nanopores*, in preparation.

7.1. Introduction

Solid-state nanopores have facilitated the development of a number of novel techniques to study biological samples. Although the majority of the work has focused on DNA and proteins, there has been some research into DNA-protein complexes as well. Some of these studies have investigated the detection of DNA-bound nucleosomes[1], biotinylated DNA-oligomer-bound monovalent streptavidin[2], or RNAP-DNA transcription complexes[3] with short (<1kbp) strands where no positional information could be determined. Other studies have used small diameter solid-state pores[4] or biological pores[5, 6], where the DNA-protein complex is too large to translocate though the pore, to determine the presence of the complex through its effects on the characteristics of the translocation events. Previous studies in our lab with longer DNA strands focused on RecA[7, 8]. This DNA repair protein binds cooperatively to DNA, forming long patches which makes it difficult to observe an isolated single-protein complex. Here we present our experimental efforts to map out *single* bound-protein positions on long DNA strands using solid-state nanopores in standard measurement conditions. We show that transient protein-pore interactions allow us to detect the presence of individual DNA-bound proteins.

In our solid-state nanopore measurements, a 20 nm thick silicon nitride membrane containing a 20 nm diameter nanopore is placed in-between two reservoirs containing a buffered 1M KCl salt solution and Ag/AgCl electrodes. Upon the application of an electric field, negatively charged biomolecules are electrophoretically driven from the one chamber (Cis) towards the positive electrode in the other chamber (Trans) as shown in Figure 7.1a, while positively charged biomolecules experience the opposite effect. The passage of a biomolecule through the nanopore results in a temporary reduction in the ionic current (Figure 7.1d). The magnitude of this reduction is, at these high-salt concentrations, determined by the volume taken up by the part of the molecule that resides in the pore. Consequently, in the case of a protein bound along a long piece of DNA, the passage of DNA produces a distinct blockade level (I_1) while any bound protein is expected to appear as an additional temporary increase in the blockade level, on top of the DNA blockade level.

We introduce a new model system to study this concept in detail. It is based on anti-DNA antibodies to serve as a proof-of-principle for this technique. Anti-DNA antibodies were first discovered in patients with systemic lupus erythematosus in the late 1950s[9, 10]. These antibodies bind randomly and independently to DNA with a measured K_D of 90 nM for 160 kDa mouse monoclonal IgG2a antibodies against dsDNA in physiological conditions[11]. We carried out nanopore measurements (Supplementary Section 7.5.1) using only free antibodies and determined that they are positively charged at pH 8 in 1M KCl. This property ensures that any free antibodies present in the solution on the Cis side will translocate away from the nanopore unless they are bound to a DNA molecule, eliminating the possibility that any of the observed events are due to a free unbound antibody co-translocating with a DNA molecule. We verified the binding of the antibodies to DNA molecules with AFM (Figure 7.1b) and gel-shift assays (Figure 7.1c). Our nanopore based

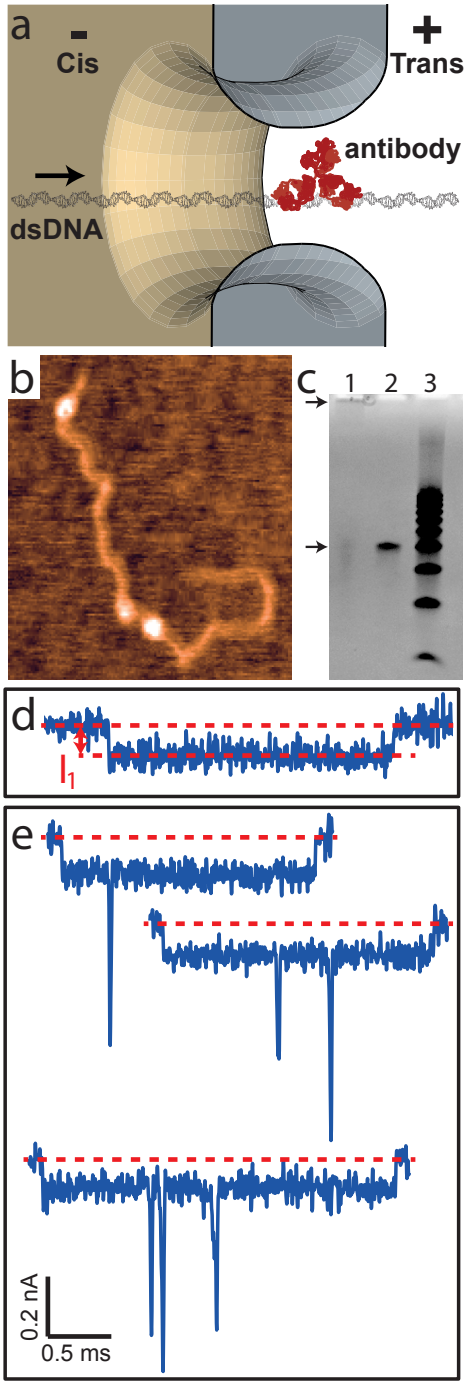


Figure 7.1: a) Schematic representation of a dsDNA molecule with a bound antibody translocating through a 20 nm pore. b) AFM image of a 2.2 kbp DNA molecule with three antibodies attached. c) Gel-shift-assay showing that antibodies binding to a 40 bp DNA fragment do not traverse into the gel in lane 1, free DNA in lane 2, and a 10 bp ladder in lane 3. d) Typical unfolded lambda DNA translocation event. e) Examples of translocation events containing spikes associated to anti-DNA antibodies. The applied voltage is 100 mV and the nanopore diameter is 20 nm.

measurements also showed that the antibody-DNA complexes do not dissociate quickly in high salt conditions, as discussed later. This is an important point, since most proteins do not bind DNA in the 1M salt concentrations used, since the electrostatic screening is so high. These commercially available antibodies provide a simple system which can be used to explore the capabilities and limits of mapping local structures along DNA molecules using solid-state nanopores.

7.2. Results and Discussion

Nanopore measurements on DNA-antibody mixtures reveal the presence of current spikes on top of the DNA blockade signals, which can be directly correlated to the addition of the antibodies, see Fig. 7.1e. Solutions containing both anti-DNA antibodies and lambda phage 48.5 kbp DNA were incubated at 37°C in low salt conditions, before being diluted to 1M KCl and added to the Cis chamber of the nanopore flowcell. Analysis of the resulting translocation events revealed events with very short-duration large-amplitude current spikes present within the DNA blockades as shown in Figure 7.1e. Analysis was carried out by first selecting unfolded events[12]. Spikes were subsequently detected if they crossed a minimum amplitude threshold defined from the local open pore current before the start of the DNA translocation event; basically an analysis that comes down to detecting events within events. This helps differentiate the high amplitude spikes produced by the antibodies from smaller amplitude folds and knots always observed in high-bandwidth measurements on long DNA in large pores[13]. Statistics on the percentage of events with spikes reveal a clear correlation between the addition of antibodies and the appearance of the spikes.

Figure 7.2a shows the fraction of DNA events containing spikes as the minimum amplitude threshold is increased. Two independent antibody-DNA experiments (red and blue) show the presence of many large amplitude spikes, compared with the DNA-only control (green). We chose to use a spike detection threshold of $3.5I_1$ as it differentiates quite well between spikes caused by antibodies and spikes observed in DNA-only experiments, which are due to DNA knots and folds[14]. Figure 7.2b shows the fraction of events with spikes of amplitude larger than $3.5I_1$ at voltages ranging from 100 mV to 400 mV for experiments with both DNA and antibodies as well as control experiments containing only DNA. At 100 mV approximately 82% of events have at least one spike present, while only 6% of the events in the DNA-only controls have spikes, a clear indication that the observed current spikes can be attributed to bound antibodies. This fraction is observed to decrease as a function of applied voltage. We attribute this effect to both the increased force reducing the lifetime of transient antibody-pore interactions and the decreased spatial resolution since DNA and DNA-protein translocate faster but the 52 kHz amplifier bandwidth[15] cannot be increased further. These experimental results demonstrate that it is possible to detect single DNA-bound proteins with solid-state nanopores.

We investigated the time dependence of the percentage of events with spikes and found no signs of significant DNA-antibody complex dissociation. This stability is remarkable, particularly since these nanopore experiments are done at high salt

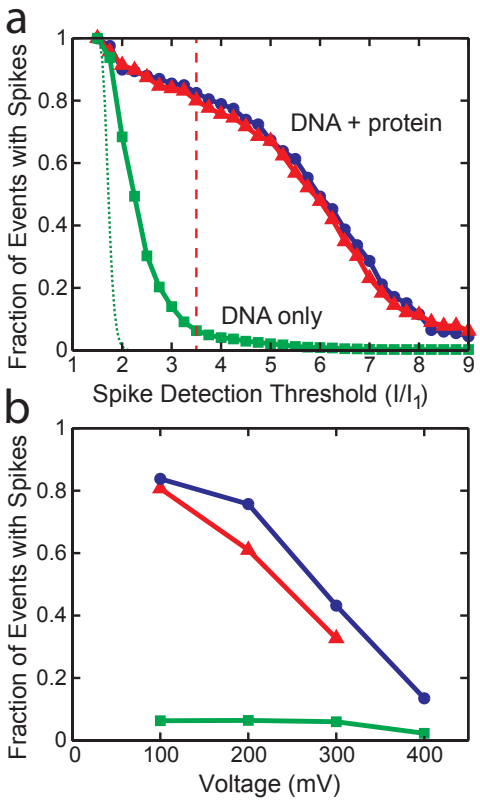


Figure 7.2: a) The fraction of events with spikes as a function of the spike detection threshold for two experiments with antibodies and lambda DNA (red and blue) as well as a control experiment with only lambda DNA (green). The dotted green line represents the "ghost" spikes that would be detected due to Gaussian noise. The dashed red line is the $3.5I_1$ spike detection threshold used in this study to differentiate antibody-induced spikes from spikes also present in the DNA-only control. b) The fraction of events with spikes as a function of applied voltage for two experiments with antibodies and lambda DNA (red and blue) as well as a control experiment with only lambda DNA (green).

concentrations (1M KCl). DNA-protein interactions tend to be electrostatic in nature. Due to this, exposure to high ionic concentrations such as those used here (1M KCl), typically lead to significant reductions in the DNA-protein binding affinity constants. However, Supplementary Figure 7.8 shows the percentage of events with spikes as a function of time for the first 30 minutes of two independent experiments. No significant reduction is observed, indicating that the antibody-DNA complex has a high salt tolerance allowing measurements in 1M KCl.

Analysis of the spike position is consistent with a random binding process, while the spike duration and amplitude show that the antibody passage occurs very quickly and is at the limit of what is resolvable. Figure 7.3a shows the position of the current spikes, normalized using the translocation time of the full DNA event. The spikes are found to be distributed randomly over the entire duration of the

events, as is expected since the antibodies bind at random positions along the molecule. Their typical amplitude is 4.5 times higher than the current drop produced by a single dsDNA molecule (Figure 7.3d) at 100 mV applied voltage. This amplitude is around 2 times lower than the blockade expected from the excluded volume of the antibody (Supplementary Section 7.5.4), indicating that the spikes are being distorted by the filtering due to their very short duration[14]. Indeed, the FWHM dwell time Δt_{Peak} of the current spikes was found to be approximately 15 μs (Figure 7.3c), which is below the amplitude distortion threshold of the Gaussian low-pass-filter used and right on the edge of the resolution of our system (17 μs). We observe that the amplitude is higher for longer events. Indeed, analyzing a DNA-antibody dataset measured at our 40 kHz bandwidth, we observe an average amplitude of 0.45 ±0.01 nA for spikes that have a FWHM time below 17 μs (the filter's distortion point), but 0.61 ±0.02 nA for those with FWHM times longer than 17 μs. These observations suggest that we are likely only observing a fraction of the total amount of bound antibodies.

What current signatures should we expect to see for bound antibodies in these conditions? The evidence suggests that the current spikes are due to transient interactions between the DNA-antibody complex with the nanopore, effectively holding the complex within the nanopore long enough for it to be resolved. As a simple straightforward estimate we can use the most-probable-translocation-time of the DNA events (1.5 ms) and the known DNA length (16.5 μm) to estimate a mean translocation velocity of around 11 nm/μs. This implies that, on average, a duration of 15 μs to corresponds to roughly 165 nm, which is far larger than the size of the antibody. In other words, the duration of the observed current spikes is significantly longer than what we would expect from a freely translocating antibody-DNA complex. Furthermore, assuming that there are no strong high-salt-induced dissociation effects, we would expect to have hundreds of antibodies bound to each lambda DNA molecule based on the results of the AFM characterization and the antibody-DNA concentrations used, whereas we only observe an average of about two spikes in each event in the nanopore translocation experiments (insert Figure 7.3d). These observations strongly indicate that the antibodies are interacting with the SiN pore surface and holding the complex bound inside the pore for a sufficiently long time to detect it. Indeed, this idea is further supported by the occasional presence of events with very long blockades with amplitudes we would expect from the antibody-DNA complex (Supplementary Section 7.5.5) sticking inside the pore for a longer period of time.

The hypothesis that the current spikes are due to transient interactions between bound antibodies and the pore is consistent with an observed increase in the total duration of the DNA translocation time as a function of the number of spikes that are observed within the event. Figure 7.4 shows the average dwell time of DNA events without spikes to be 1.48 ±0.08 ms. This increases to an average of 1.60 ±0.07, 1.78 ±0.07, 2.02 ±0.10, 2.2 ±0.14 ms for events with one to four spikes, respectively. The linear fit to this data, shown in Figure 7.4, has a slope of 189 μs/spike. This surprisingly high value prompts the question: If the average FWHM dwell time of a spike is 15 μs, then why is the average time of the DNA translocation

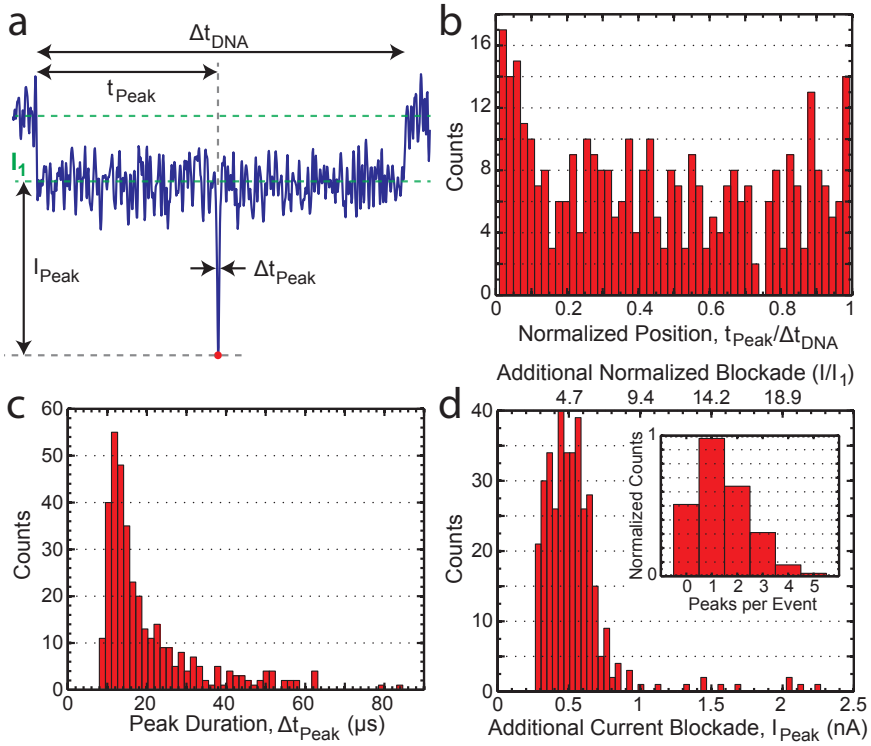


Figure 7.3: Statistics for spikes with an amplitude above 3.5 I/I_1 observed on lambda DNA translocating through a 20 nm pore at 100 mV. a) Schematic of the data extracted for each spike, which includes the amplitude (I_{Peak}) as measured from the first DNA level (I_1), the FWHM duration (Δt_{Peak}), the absolute position (t_{Peak}), and the normalized position (τ_{Peak}). b) Normalized position ($t_{Peak}/\Delta t_{DNA}$) distribution shows uniform binding all over the length of the molecule. Part of the population at the start is attributed to brief large amplitude folds which sometimes occur at the start of the translocation process. c) FWHM dwell time (Δt_{Peak}) histogram for the current spikes showing a peak around 15 μs. d) Amplitude (I_{Peak}) of the current spikes showing a peak around 0.5 nA (4.5x larger than $I_1 = 0.11$ nA). Insert: distribution of the number peaks-per-event.

increased by a much higher value of 189 μs per spike? This observation suggests that the spikes have a higher probability of being observed within DNA events with inherently longer translocation times[16, 17]. In other words, DNA events that take longer to translocate and thus move at a slower speed, increase the probability of a transient protein-pore interaction occurring. This clear increase in the total dwell time is consistent with a temporary interaction between the DNA-antibody complex and the pore.

Antibody-pore interactions may be mediated by the F_c region of the antibody. We created F_{ab} fragments from the anti-DNA antibodies using a standard Papain/Protein A technique. These smaller (50 kDa versus the 150 kDa of IgG2a) F_{ab} fragments were incubated with lambda DNA and the DNA-F_{ab} constructs were translocated through smaller 10 nm pores. No statistically significant differences were seen

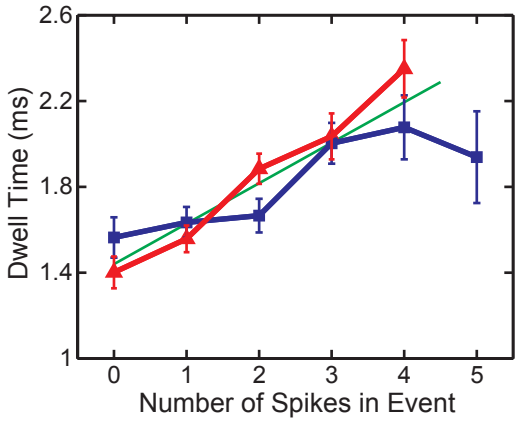


Figure 7.4: The average dwell time and standard error of DNA translocation events as a function of the number of spikes present within the event. The solid line is a linear fit to the average of the two datasets and has a slope of 0.189 ms/spike.

between these samples and DNA-only controls. This may indicate that the transient interactions observed in the IgG2a antibodies is mediated by the F_c region, which is no longer present in the F_{ab} fragments.

Could the observed current spikes be the result of mechanisms other than transient protein-pore interactions? We considered and ruled out a number of other possible alternatives to explain the observed current spikes. **1)** Transient DNA-pore interactions are a known issue in solid-state nanopore experiments. In this mechanism the DNA and the pore temporarily interact. If such an interaction occurs in the close vicinity of a bound protein that protein may become visible if the duration of the interaction is greater than about 15 μs. Although this is consistent with the observed increase in dwell time as a function of the number of observed spikes, the interaction would have to occur right at the site of a bound antibody. It would also have to occur multiple times for the events with multiple spikes and it does not explain why no current spikes are visible in the case of the F_{ab} fragments. Furthermore, we would expect the observed amplitude to have a much wider distribution since the protein could sit at varying distances away from the pore constriction[18]. **2)** The local velocity during the course of a DNA translocation event is known to fluctuate[16, 17, 19]. Could the DNA temporarily slow down to the extent that the antibodies are visible? We estimate that the local velocity would have to slow down by a factor of around 16x relative to the mean translocation velocity, something unlikely to occur so consistently in all events. **3)** Each antibody has two binding sites, bringing up the possibility that they could be forming loops in the DNA molecule reminiscent of those formed by Lac repressor[20]. This, however, also seems unlikely since as soon as the first site has bound, the probability of the second site binding to a nearby piece of DNA is far higher than for any DNA segment far away. Furthermore DNA loops would have amplitudes of $2I_1$, well below the observed spike amplitudes. **4)** Could two antibodies bind very closely together to form a

complex larger than the diameter of the pore? Such a large complex could get stuck at the pore constriction long enough to produce a spike signal. To check this hypothesis, we translocated the same antibody-DNA mixture through a pore with a 35 nm diameter. This large diameter would be sufficient to accommodate two or more antibodies and the DNA molecule. We were, however, still able to observe the current spikes in this situation, suggesting that the formation of a bulky complex is not the mechanism. All these considerations and observations together render the proposed mechanism of transient protein-pore interactions as the most probable.

7.3. Conclusion

In a series of proof-of-principle experiments, we have demonstrated that individual DNA-bound protein can be detected with solid-state nanopores. Although this work focused on anti-DNA antibodies which bind randomly, it should be possible to extend this technique using cross-linked site-specific bound proteins and primary antibodies, to visualize specific complexes. Several obstacles must be overcome for this technique to be generally applicable to any protein-DNA system. Since measurements at physiological salt conditions do not provide a high enough signal-to-noise ratio, the technique is currently limited to proteins which are resistant to high-salt (1M) conditions with a slow timescale of dissociation, or proteins which have been covalently cross-linked to the DNA. Alternatively, a non-ionic current based readout method may allow measurements at physiological conditions in the future. A further complication is the fast translocation velocity relative to the maximum measurement bandwidth achievable, which makes the detection of a single protein very challenging. The recent development of high-bandwidth low-noise amplifiers may help improve this limitation[21, 22]. Although it has been shown that optical tweezers can be used to slow down the translocation process[23, 24], a simpler technique is necessary for practical applications. Finally, in order to determine the positions of bound protein, the current signals which are recorded in the temporal domain must be converted to spatial information using a currently unknown mapping function which is dependent on the local velocity profile along the length of the DNA. Although both simulations[16] as well as our recent study on recaptured DNA molecules[17] suggest that the non-linear local velocity profile is dependent on the conformation of the DNA molecule at the start of the translocation process, the local velocity profile has yet to be measured directly[19]. None of these obstacles present a fundamental roadblock to the continued development of this technique. With further improvements, it should be possible to develop a nanopore technique for identifying DNA-binding protein which is complementary to approaches such as chromatin immune-precipitation (ChIP) and DNA adenine methyltransferase identification (DamID).

7.4. Methods

Nanopores. SiN membranes were fabricated and 20 nm diameter nanopores were drilled with a TEM as described previously[25]. After TEM drilling, membranes we manually painted with a layer of PDMS in order to reduce the capacitance and im-

prove the signal-to-noise ratio. Chips were mounted in a PMMA flowcell, after which the two reservoirs were filled with 1M KCl 10mM Tris 1mM EDTA solution at pH8. The current was recorded with a standard electrophysiology setup consisting of an Axopatch 200B amplifier, digitized with a Digidata 1322A DAQ, and subsequently analyzed in Matlab[12]. The nanopore measurements presented have a bandwidth of at least 30 kHz at 100 mV and are limited by the Axopatch's bandwidth of 52 kHz at higher applied voltages[15].

DNA with bound antibodies. Anti-DNA antibodies (HYB331-01) purchased from Abcam (Cambridge, UK) at a concentration of 625 nM were incubated with 0.25 nM lambda DNA (Promega) (at a 932:1 ratio) in 18.75 mM NaCl, 2 mM Tris, pH8 for 10 min at 37°C. Right before starting a nanopore measurement, 20 μL of this DNA-antibody mixture was added to 10 μL of 3M KCl, 30 mM Tris, 3 mM EDTA at pH8 for final concentrations of 166 pM DNA and 156 nM antibodies.

Acknowledgement

The authors would like to thank Meng-Yue Wu for TEM drilling of nanopores and Christophe Danelon for helpful discussions. This work was supported by the Netherlands Organisation for Scientific Research (NWO/OCW), as part of the Frontiers of Nanoscience program, and a European Research Council Advanced grant NanoforBio (no. 247072), and by the Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) Academy Assistants Program.

7.5. Supplementary Info

7.5.1. Nanopore characterization of anti-DNA antibodies

Anti-DNA Antibodies (HYB331-01) were diluted into 1M KCl TE pH8 solution to a final concentration of 312 nM. This solution was measured with a 20 nm diameter SiN nanopore. Figure 7.5 shows traces taken at +100 mV and -100mV applied voltage, demonstrating that the antibodies are positively charged at pH8. Figure 7.6 shows a scatter density plot and current histogram from a dataset of 8815 antibody translocation events at 100 mV and 30 kHz bandwidth. The events produced by the antibodies have a most probable translocation time of 13 μs and an most probable current blockade of 210 pA. These short translocation times are on the edge of the temporal resolution for this technique, as discussed in detail in a previous publication[14].

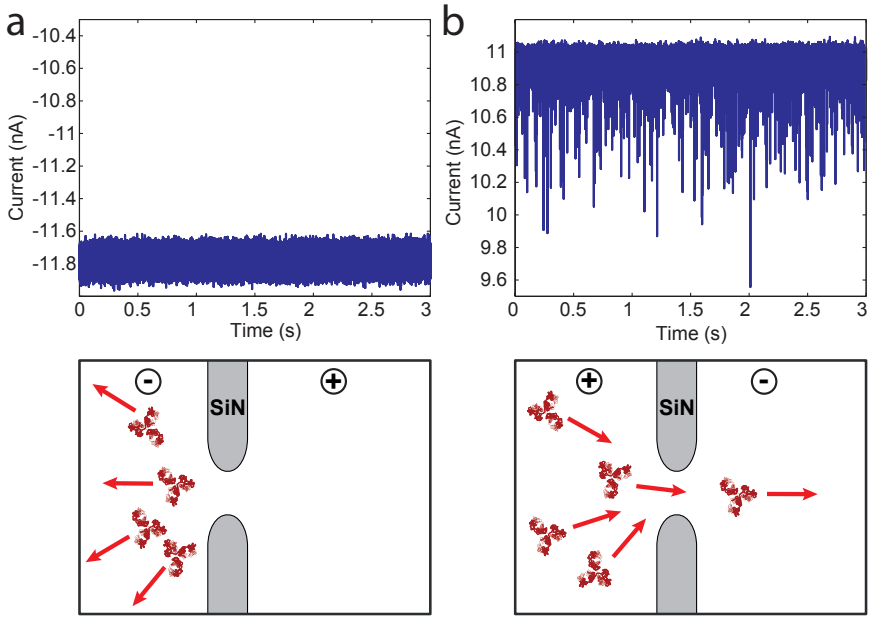


Figure 7.5: Current traces at a) -100 mV and b) +100mV applied voltage for 312 nM IgG2a mouse antibodies translocating through a 20 nm pore. The large difference in event rates between the two polarities indicates that these antibodies are positively charged. Below each trace is a schematic representation of each situation.

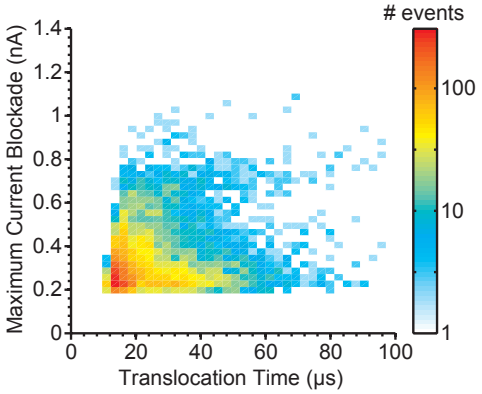


Figure 7.6: Translocation data from an 8815-event dataset of antibodies translocating through a 20 nm pore at 100 mV and 40 kHz bandwidth. Scatter density plot showing the maximum current blockades and translocation times. The distributions peak at around 210 pA and 13 μs.

7.5.2. AFM characterization of anti-DNA antibodies bound to 2.2kbp DNA

We imaged antibody-DNA complexes using AFM to verify binding and observe the number of bound antibodies. A solution of 2.97 nM anti-DNA antibodies, 0.7 nM 2.2kbp DNA in 10 mM NaCl, 1 mM Tris pH 8, and 10 mM MgCl was incubated on freshly cleaved Mica for 5 minutes and imaged using tapping mode AFM with an Olympus AC160TS tip. We observed, on average, two antibodies bound on each 2.2 kbp molecule.

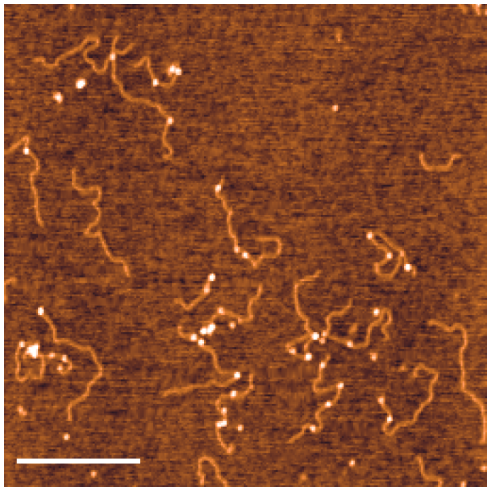


Figure 7.7: Tapping mode AFM scan of 2.2kb DNA molecules with bound antibodies on a mica surface. The scale bar represents 500 nm.

7.5.3. Time dependence of detection of events with spikes

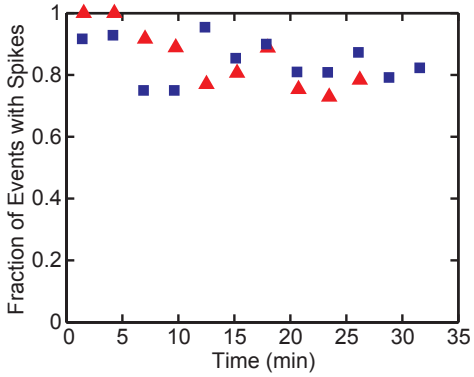


Figure 7.8: Time dependence of the fraction of events with spikes detected at 100 mV for the first 30 min after the antibody-DNA solution was diluted into high (1M KCl) salt. No significant dissociation is observed on this timescale.

7.5.4. Estimated antibody current blockade

The current blockade produced by the antibody translocating through the pore can be estimated using the excluded volume technique as in a number of previous studies[1, 26, 27]. The expression used is

$$\Delta I = \frac{\gamma \sigma V_a V_{excluded}}{\left(h_p + \frac{\pi d_p}{4}\right)^2} f(d_p, d_m) \tag{7.1}$$

where γ is the shape factor taken as 1.5 (even though these antibodies are not a perfect sphere[27]), σ is the bulk buffer conductivity (10.5 S/m for 1M KCl), V_a is the applied voltage set as 100 mV, $V_{excluded}$ is the volume of the antibody (taken as 347 nm^3)[27], h_p is the effective thickness of the pore set to 8.6 nm based on previous work[28], d_p is the diameter of the pore set to 20 nm, and f is a correction factor taken to be 1. Based on these values we would expect a current blockade of around 0.92 nA, well above what we observe in both the free antibody translocation (0.21 nA) and in the DNA-bound antibody case (0.5 nA). These observations can be explained by the fact that the majority of translocation times in both of these cases is below the filtering distortion frequency (22 μs for 30 kHz and 16.6 μs for 40 kHz) as discussed in detail previously[14].

7.5.5. Long blockade Antibody-DNA events

Among the events observed, occasionally we observe some events with a very long duration blockade event, which can be attributed to long-lasting antibody-pore interaction events (sticking). Figure 7.9 provides some examples.

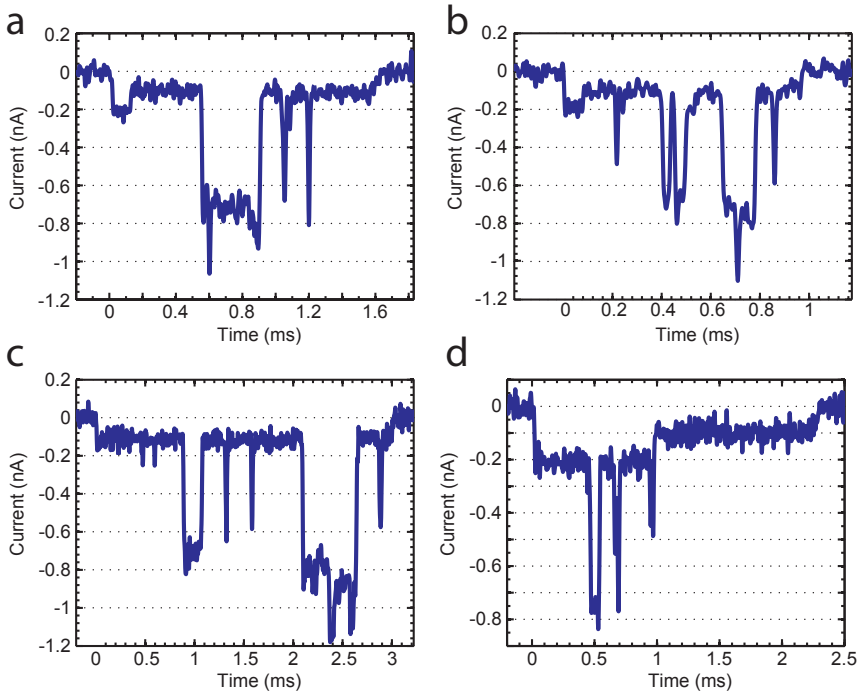


Figure 7.9: Four example events with long duration blockades present within the DNA event. These long blockades are attributed to sticking between the antibody-DNA complex and the pore.

References

[1] G. V. Soni and C. Dekker, *Detection of nucleosomal substructures using solid-state nanopores,* Nano Letters **12**, 3180 (2012).

[2] A. T. Carlsen, O. K. Zahid, J. A. Ruzicka, E. W. Taylor, and A. R. Hall, *Selective detection and quantification of modified DNA with solid-state nanopores,* Nano Letters **14**, 5488 (2014).

[3] C. Raillon, P. Cousin, F. Traversi, E. Garcia-Cordero, N. Hernandez, and A. Radenovic, *Nanopore detection of single molecule RNAP–DNA transcription complex,* Nano Letters **12**, 1157 (2012).

[4] A. Ivankin, S. Carson, S. R. M. Kinney, and M. Wanunu, *Fast, label-free force spectroscopy of histone–DNA interactions in individual nucleosomes using nanopores,* Journal of the American Chemical Society **135**, 15350 (2013).

[5] B. Hornblower, A. Coombs, R. D. Whitaker, A. Kolomeisky, S. J. Picone, A. Meller, and M. Akeson, *Single-molecule analysis of DNA-protein complexes using nanopores,* Nat Meth **4**, 315 (2007).

[6] S. Benner, R. J. A. Chen, N. A. Wilson, R. Abu-Shumays, N. Hurt, K. R. Lieberman, D. W. Deamer, W. B. Dunbar, and M. Akeson, *Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore,* Nat Nano **2**, 718 (2007).

[7] S. W. Kowalczyk, A. R. Hall, and C. Dekker, *Detection of local protein structures along DNA using solid-state nanopores,* Nano Letters **10**, 324 (2009).

[8] R. M. M. Smeets, S. W. Kowalczyk, A. R. Hall, N. H. Dekker, and C. Dekker, *Translocation of RecA-coated double-stranded DNA through solid-state nanopores,* Nano Letters **9**, 3089 (2008).

[9] R. Ceppellini, E. Polli, and F. Celada, *A DNA-reacting factor in serum of a patient with lupus erythematosus diffusus,* Experimental Biology and Medicine **96**, 572 (1957).

[10] G. L. Asherson, *Antibodies against nuclear and cytoplasmic cell constituents in systemic lupus erythematosus and other diseases,* British journal of experimental pathology **40**, 209 (1959).

[11] N. H. H. Heegaard, D. T. Olsen, and K.-L. P. Larsen, *Immuno-capillary electrophoresis for the characterization of a monoclonal antibody against DNA,* Journal of Chromatography A **744**, 285 (1996).

[12] C. Plesa and C. Dekker, *Data analysis methods for solid-state nanopores; submitted,* (2015).

[13] C. Plesa, D. Verschueren, J. W. Ruitenberg, M. J. Witteveen, M. P. Jonsson, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Direct observation of DNA knots using solid state nanopores; submitted,* (2015).

[14] C. Plesa, S. W. Kowalczyk, R. Zinsmeester, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Fast translocation of proteins through solid state nanopores,* Nano Letters **13**, 658 (2013).

[15] J. D. Uram, K. Ke, and M. Mayer, *Noise and bandwidth of current recordings from submicrometer pores and nanopores.* ACS Nano **2**, 857 (2008).

[16] B. Lu, F. Albertorio, D. P. Hoogerheide, and J. A. Golovchenko, *Origins and consequences of velocity fluctuations during DNA passage through a nanopore,* Biophysical Journal **101**, 70 (2011).

[17] C. Plesa, L. Cornelissen, M. W. Tuijtel, and C. Dekker, *Non-equilibrium folding of individual DNA molecules recaptured up to 1000 times in a solid state nanopore,* Nanotechnology **24**, 475101 (2013).

[18] R. Wei, V. Gatterdam, R. Wieneke, R. Tampe, and U. Rant, *Stochastic sensing of proteins with receptor-modified solid-state nanopores,* Nat Nano **7**, 257 (2012).

[19] C. Plesa, N. van Loo, P. Ketterer, H. Dietz, and C. Dekker, *Velocity of DNA during translocation through a solid state nanopore; in press,* Nano Letters (2014).

[20] S. Oehler, E. R. Eismann, H. Krämer, and B. Müller-Hill, *The three operators of the lac operon cooperate in repression,* The EMBO journal **9**, 973 (1990).

[21] J. K. Rosenstein, M. Wanunu, C. A. Merchant, M. Drndic, and K. L. Shepard, *Integrated nanopore sensing platform with sub-microsecond temporal resolution,* Nat Meth **9**, 487 (2012).

[22] A. Uddin, S. Yemenicioglu, C.-H. Chen, E. Corigliano, K. Milaninia, and L. Theogarajan, *Integration of solid-state nanopores in a 0.5 um CMOS foundry process,* Nanotechnology **24**, 155501 (2013).

[23] U. F. Keyser, B. N. Koeleman, S. van Dorp, D. Krapf, R. M. M. Smeets, S. G. Lemay, N. H. Dekker, and C. Dekker, *Direct force measurements on DNA in a solid-state nanopore,* Nat Phys **2**, 473 (2006).

[24] A. Spiering, S. Getfert, A. Sischka, P. Reimann, and D. Anselmetti, *Nanopore translocation dynamics of a single DNA-bound protein,* Nano Letters **11**, 2978 (2011).

[25] X. J. A. Janssen, M. P. Jonsson, C. Plesa, G. V. Soni, C. Dekker, and N. H. Dekker, *Rapid manufacturing of low-noise membranes for nanopore sensors by trans -chip illumination lithography,*

Nanotechnology **23**, 475302 (2012).
[26] D. S. Talaga and J. Li, *Single-molecule protein unfolding in solid state nanopores.* Journal of the American Chemical Society **131**, 9287 (2009).
[27] E. C. Yusko, J. M. Johnson, S. Majd, P. Prangkio, R. C. Rollings, J. Li, J. Yang, and M. Mayer, *Controlling protein translocation through nanopores with bio-inspired fluid walls.* Nat Nano **6**, 253 (2011).
[28] S. W. Kowalczyk, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Modeling the conductance and DNA blockade of solid-state nanopores,* Nanotechnology **22**, 315101 (2011).

Part III

Methods

8

DNA nanopore translocation in glutamate solutions

Nanopore experiments have traditionally been carried out with chloride-based solutions. Here we introduce silver/silver-glutamate-based electrochemistry as an alternative. We show that it has a linear response at typical voltages and can be used to detect DNA translocations through a nanopore. The glutamate anion also acts as a redox-capable thickening agent, with high-viscosity solutions capable of slowing down the DNA translocation process by up to 11 times, with a corresponding 7 times reduction in signal. These results demonstrate that glutamate can replace chloride as the primary anion in nanopore resistive pulse sensing.

C. Plesa, N. van Loo, and C. Dekker, *DNA nanopore translocation in glutamate solutions*, in preparation.

8.1. Introduction

Solid-state nanopores constitute an emerging biophysical technique which has been used to study DNA[1], protein[2], and DNA-protein complexes[3]. Although this area has been primarily pushed by the promise of its use for low-cost sequencing of long nucleic acid polymers, many other applications have opened up. Numerous studies have been published on methods to improve various aspects of this technique with different pore types[4, 5], membrane materials[6–8], new salts[9, 10], and passivation techniques[11]. Although central to their operation, alternative electrochemical reactions for detecting the ionic current have not been pursued. In this study we introduce a novel electrochemical reaction based on glutamate (Glu) anions and demonstrate its potential for studying DNA translocations through solid-state nanopores. This approach provides a way to avoid the use of chloride and slow down the translocation velocity using high-viscosity solutions.

We first provide a brief overview of the solid-state nanopore method. Nanopores are typically fabricated in a 20 nm thick free-standing SiN membrane by focusing an electron beam on the membrane using a transmission electron microscope (TEM). This allows simultaneous in-situ imaging of the pore as it is being created. Membranes with pores are subsequently placed in a flow cell such that it separates two aqueous reservoirs filled with salt solution, as shown in Figure 8.1a. Chlorinated silver electrodes are placed into the solution on each side and an electric potential is applied. The resulting ionic current is measured with a low-noise amplifier. Since DNA is negatively charged, it is electrophoretically driven through the nanopore. When a DNA molecule passes through the pore it causes a transient drop in the ionic current, since the molecule reduces the pore volume normally used by ions for transport. The duration of the blockade can be related to the length of the polymer, while the amplitude of the blockade is proportional to the volume of the molecule currently in the pore constriction.

We now address the electrochemistry used for the ionic current detection. In silver/silver-chloride based electrochemistry, a chlorinated silver wire is placed in a chloride-based solution and an electric potential is applied. At the anode, chloride ions are oxidized into solid-phase silver chloride through the reaction $Ag(s) + Cl^{-}(aq) \rightarrow AgCl(s) + e^{-}$ while at the cathode silver chloride is reduced and chloride ions are released into the solution via the reaction $AgCl(s) + e^{-} \rightarrow Ag(s) + Cl^{-}(aq)$. The measured ionic current is linear with voltage at potentials below 1V. The ease of electrode preparation, made by placing silver wire in a bleach solution, makes this a simple and effective experimental approach. With one exception, all nanopore publications that we are aware of use silver-chloride based electrochemistry together with either alkali metal halide solutions (LiCl[9], NaCl[3], KCl[11], RbCl[12], CsCl[13]) or organic-salts (BMIM-Cl)[10]. The only non-chloride electrochemistry nanopore study looked at ionic liquids using platinum electrodes[14].

L-Glutamic acid (Glu) plays a fundamental role in biology as one of the amino acids used in protein synthesis. In many organisms glutamate, not chloride, is the major intracellular anion[15]. Although many of its properties have been studied in detail, we are not aware of any literature on non-enzymatic glutamate-based electrochemistry. In 1974, an article by Tabei et al[16] described a method for

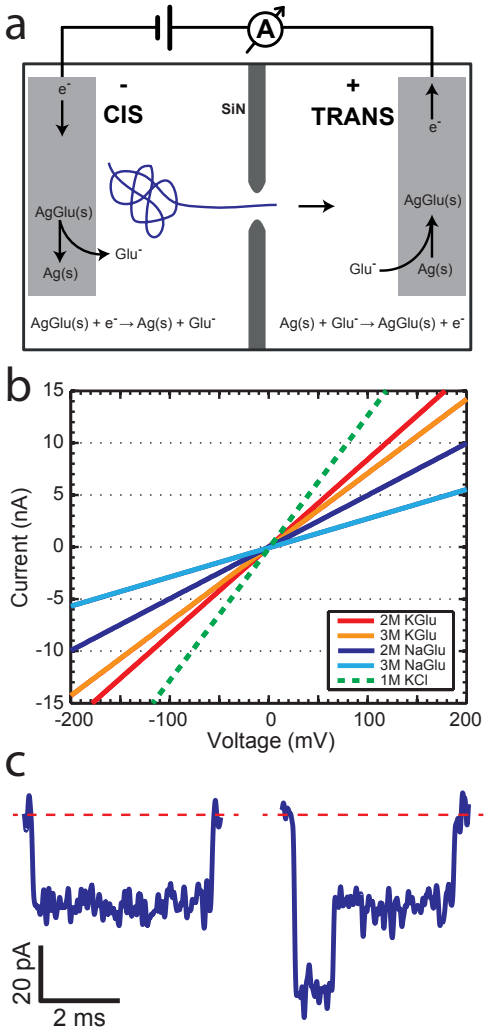


Figure 8.1: a) Schematic illustration of the nanopore setup with the redox reactions occurring at the electrodes. b) The current voltage response for 20 nm pores is observed to be linear over the range -800 mV to 800 mV for NaGlu and KGlu at all concentrations tested. The pore conductance is higher for KGlu solutions compared to NaGlu. For both salts, the pore conductance is higher at 2M compared to 3M concentration. c) Typical translocation events for a 20.7 kbp linear DNA molecule observed in a 16 nm pore and 3M KGlu at 100 mV.

making printed circuits through the photolysis of silver glutamate. From this work, we know that silver glutamate is a stable, albeit photosensitive, solid-phase compound, much like silver chloride. We hypothesized that glutamate could be used to replace chloride in the electrochemical reactions used in nanopore studies. Analogously to the silver chloride case, we expect similar electrochemical reactions to take place at the anode

$$Ag(s) + Glu^{-}(aq) \rightarrow AgGlu(s) + e^{-} \tag{8.1}$$

and the cathode

$$AgGlu(s) + e^{-} \rightarrow Ag(s) + Glu^{-}(aq) \quad . \tag{8.2}$$

In such a way, glutamate could be used as a redox-capable thickening agent to increase the viscosity of a solution while still maintaining the required electrochemistry.

8.2. Results

We now show that glutamate solutions produce a linear current-voltage response and can be used for detecting DNA nanopore translocations. We prepared silver glutamate electrodes by electrolysis of a 99.99% silver wire (anode) and platinum wire (cathode) in a concentrated 3M potassium glutamate solution. These electrodes were then used, following standard procedures, in solid-state nanopore experiments. Figure 8.1b shows the typical current-voltage curves observed for potassium glutamate (KGlu) and sodium glutamate (NaGlu) solutions at 2M and 3M concentration in 20 nm pores. We have also included a typical response observed for 1M KCl (dashed green line) for comparison. The IV responses were observed to be linear in the tested range of -800 mV to 800 mV. Additionally, the open pore conductance was observed to be higher at 2M compared to 3M, in both KGlu and NaGlu, as explained below in discussing the conductivity measurements. Upon addition of a DNA solution to the flowcell, typical DNA translocation events could be observed in the current trace, as shown in Figure 8.1c. These results demonstrate the feasibility of using glutamate-based salts for nanopore experiments.

To understand the pore conductance, we measured the conductivity of concentrated solutions of glutamate salts. Figure 8.2a shows the conductivity at 23°C as a function of concentration for KGlu, NaGlu, LiGlu, alongside literature values[17] for KCl, NaCl, and LiCl at 25°C. The maximum concentrations tested were limited by the solubility of the salts which were found to be around 1M for LiGlu, 3.5M for KGlu, and 3.3M for NaGlu. The conductivity of KGlu was observed to be significantly higher than NaGlu, with values for KGlu of 6.57 S/m and 5.33 S/m at 2M and 3M compared to 3.76 S/m and 2.69 S/m, respectively, for NaGlu. The conductivities of both solutions were observed to have maxima, which occurred around 2.0 for KGlu and 1.7 for NaGlu. This is a common phenomenon at very high electrolyte concentrations and is attributed to the small inter-ionic separations which lead to strong ion-ion interactions that begin to reduce the conductivity of solutions above a certain concentration. Since we would like to maximize the ionic current signal, larger

conductivities are desired and these measurements reveal that KGlu is preferable over NaGlu. As expected, the conductivities observed for the glutamate solutions were smaller than the corresponding chloride solutions at the same concentrations.

We also investigated the viscosity of glutamate salt solutions, relevant for the translocation speed of traversing molecules. A Brookfield DV2TLVCP Wells-Brookfield cone/plate viscometer was used to measure the viscosities of KGlu, NaGlu, and LiGlu at 23°C. Viscosity is an important parameter in nanopore measurements since increasing viscosity results in higher drag forces and longer translocation times. In Figure 8.2b the viscosity of LiGlu, KGlu, and NaGlu is shown as a function of solution concentration alongside literature values for KCl[18], NaCl[18], and LiCl[19] at 25°C. As expected, the glutamate solutions have a significantly higher viscosity than the chloride solutions at the same concentration. We measure a viscosity of 2.8 cP and 6.6 cP for KGlu at 2M and 3M, and even higher viscosities of 4.1 cP and 11.3 cP, respectively, for NaGlu. The larger viscosities of the glutamate solutions compared to the chloride solutions can be explained by the much larger size of the glutamate (MW 147.13 g/mol) anion compared to chloride (MW 35.45 g/mol) and highlight its potential as a thickening agent.

Finally, we studied the characteristics of DNA translocation through nanopores in high concentration glutamate solutions to see how much the translocation velocity and conductance blockade were reduced. Linear 20.7 kbp DNA molecules were translocated through nanopores in 2M and 3M solutions of KGlu and NaGlu at 100 mV. Figure 8.3a shows a typical heat plot showing the event's conductance blockade and translocation time in 2M KGlu and a 20 nm pore. As expected from the viscosity measurements, events were slower in NaGlu compared to KGlu, as shown in Figure 8.3b. In 2M and 3M KGlu, events had a 3.3X and 5.8X longer translocation time relative to 1M KCl, while the translocation in 3M NaGlu took 11.3X longer. The price paid for this slower translocation is a reduction in the blockade signal, as already observed in the conductivity measurements. In Figure 8.3c we show the DNA conductance blockade in 2M and 3M solutions of NaGlu and KGlu. We observe blockade amplitudes of 0.35 nS and 0.31 nS in 2M and 3M KGlu, which are around 3.6X and 4X smaller than the DNA blockade in 1M KCl (1.25 nS). This becomes worse for the NaGlu, with blockades of 0.22 nS and 0.18 nS, respectively. One common approach to increase the conductance blockade is to decrease the diameter of the nanopores used. Figure 8.3d shows the DNA blockade for 3M KGlu in 10, 16, and 20 nm pores. When the pore diameter is reduced from 20 nm to 10 nm we see a 1.5X increase in the DNA blockade, to 0.47 nS. The observed trend matches the prediction from a simple geometric model (dashed line) using the measured solution conductivity, pore diameter, and membrane thickness[20]. At very small pore sizes we eventually expect to see a decrease in the pore conductance, as predicted by hindered transport theory[21] for large ions. These initial experiments demonstrate the viability of using glutamate-based solvents for nanopore measurements.

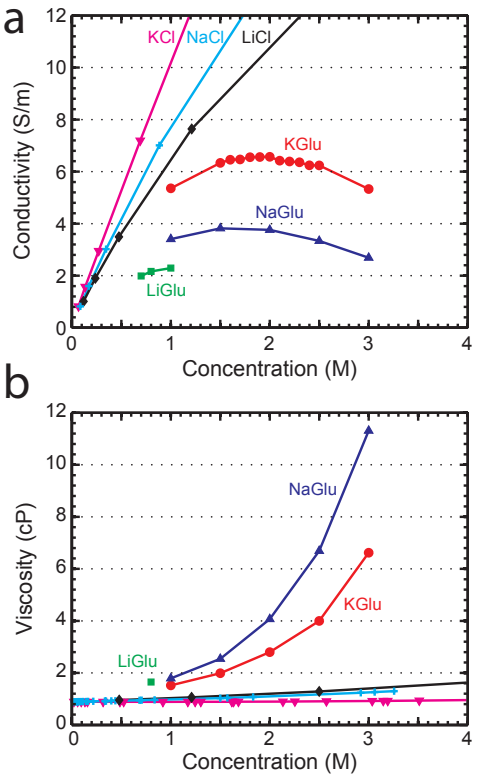


Figure 8.2: a) The conductivity of the glutamate solutions is lower than the equivalent chloride solutions and reaches a maximum between 1.5M and 2M. LiGlu could not be dissolved at concentrations above 1M. b) The viscosity of glutamate solutions can be significantly higher than corresponding chloride solutions demonstrating their potential advantage as thickening agents.

8.3. Conclusion

In this preliminary study we have demonstrated that glutamate-based electrochemistry can replace traditional chloride-based reactions for use in nanopore measurements. Potassium glutamate provides a higher conductivity, compared to sodium glutamate, with a maximum of 6.57 S/m at around 2M concentration. The use of lithium glutamate is limited by its maximum solubility of around 1M. DNA translocation through glutamate solutions can be slowed down by 3.3X to 11.3X relative to 1M KCl, while the conductance blockade signal is reduced by 3.6X and 7X respectively. These numbers are similar to those from a study carried out by adding glycerol into 1.5M KCl solutions[22]. There, a 5.3X increase in viscosity resulted in a 4.4X increase in the translocation time and a 3.1X decrease in the blockade amplitude, relative to the values for 1.5M KCl. In addition to slowing down DNA translocation, the use of glutamate instead of chloride may provide a better environment for gold electrodes in three terminal devices, an application which is now under investigation.

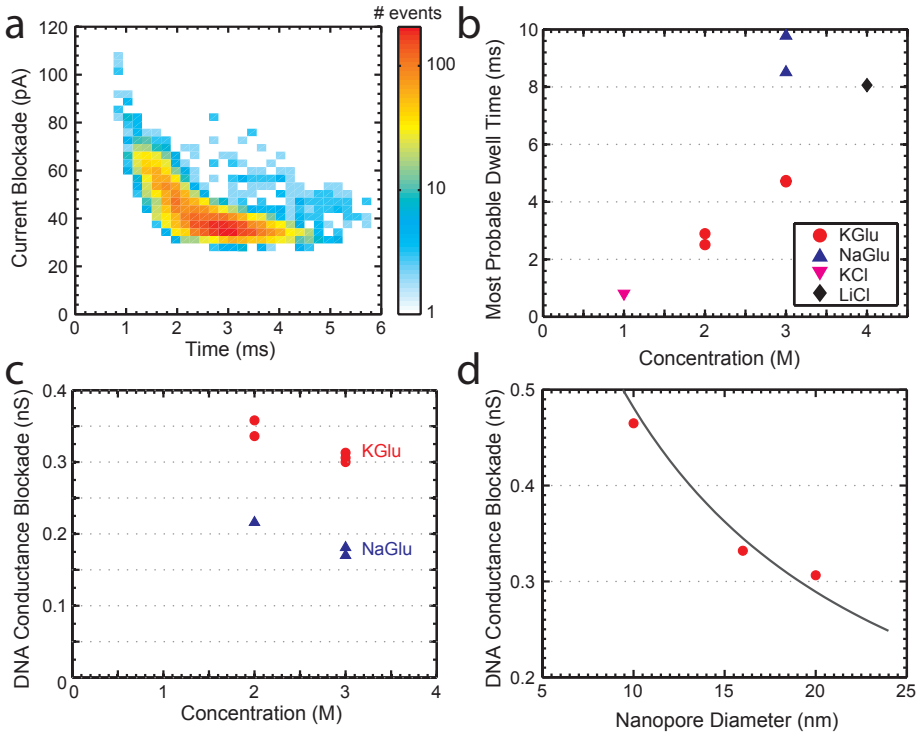


Figure 8.3: a) Heat plot showing the conductance blockade and translocation time of 20.7 kbp DNA translocating through a 20 nm pore in 2M KGlu. b) The most probable dwell times observed for 20.7 kbp DNA at 100 mV in a variety of buffers. c) DNA conductance blockade of 2M and 3M solutions of KGlu and NaGlu at 100 mV in 20 nm pores. d) DNA conductance blockade in 3M KGlu as a function of the pore diameter. The solid line is the dependence predicted by the geometric model of Kowalczyk et al.[20]

8.4. Methods

Free standing SiN membranes were fabricated using trans-chip illumination lithography[23]. Pores were drilled with a Philips CM200-FEG TEM. Ionic currents were measured with an Axopatch 200B and digitized at 500kHz. Signals were subsequently analysed with the Transalyzer Matlab package[24]. Potassium glutamate, sodium glutamate, and L-glutamic acid was purchased from Sigma-Aldrich (St. Louis, MO, USA). Lithium glutamate was made by Tocris Bioscience (Bristol, UK) following a procedure similar to Wiesbrock et al[25]. Briefly high purity L-α-glutamic acid was dissolved in HPLC grade water, after which an equivalent amount of lithium hydroxide was added. The resulting clear solution was then freeze dried to yield the solid salt. The product was characterized with micro analysis, infrared spectroscopy, proton NMR, and mass spectrometry. High purity silver and platinum wire was purchased from Advent Research Materials (Oxford, UK).

References

[1] M. Wanunu, *Nanopores: A journey towards DNA sequencing,* Physics of Life Reviews **9**, 125 (2012).

[2] C. Plesa, S. W. Kowalczyk, R. Zinsmeester, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Fast translocation of proteins through solid state nanopores,* Nano Letters **13**, 658 (2013).

[3] A. T. Carlsen, O. K. Zahid, J. A. Ruzicka, E. W. Taylor, and A. R. Hall, *Selective detection and quantification of modified DNA with solid-state nanopores,* Nano Letters **14**, 5488 (2014).

[4] L. J. Steinbock, O. Otto, C. Chimerel, J. Gornall, and U. F. Keyser, *Detecting DNA folding with nanocapillaries,* Nano Letters **10**, 2493 (2010).

[5] E. A. Heins, Z. S. Siwy, L. A. Baker, and C. R. Martin, *Detecting single porphyrin molecules in a conically shaped synthetic nanopore,* Nano Letters **5**, 1824 (2005).

[6] J. Larkin, R. Henley, D. C. Bell, T. Cohen-Karni, J. K. Rosenstein, and M. Wanunu, *Slow DNA transport through nanopores in hafnium oxide membranes,* ACS Nano **7**, 10121 (2013).

[7] B. M. Venkatesan, A. B. Shah, J.-M. Zuo, and R. Bashir, *DNA sensing using nanocrystalline surface-enhanced Al2O3 nanopore sensors,* Advanced Functional Materials **20**, 1266 (2010).

[8] G. F. Schneider, S. W. Kowalczyk, V. E. Calado, G. Pandraud, H. W. Zandbergen, L. M. K. Vandersypen, and C. Dekker, *DNA translocation through graphene nanopores,* Nano Letters **10**, 3163 (2010).

[9] S. W. Kowalczyk, D. B. Wells, A. Aksimentiev, and C. Dekker, *Slowing down DNA translocation through a nanopore in lithium chloride,* Nano Letters **12**, 1038 (2012).

[10] R. S. S. de Zoysa, D. A. Jayawardhana, Q. Zhao, D. Wang, D. W. Armstrong, and X. Guan, *Slowing DNA translocation through nanopores using a solution containing organic salts,* The Journal of Physical Chemistry B **113**, 13332 (2009).

[11] E. C. Yusko, J. M. Johnson, S. Majd, P. Prangkio, R. C. Rollings, J. Li, J. Yang, and M. Mayer, *Controlling protein translocation through nanopores with bio-inspired fluid walls.* Nat Nano **6**, 253 (2011).

[12] M. F. Breton, F. Discala, L. Bacri, D. Foster, J. Pelta, and A. Oukhaled, *Exploration of neutral versus polyelectrolyte behavior of poly(ethylene glycol)s in alkali ion solutions using single-nanopore recording,* The Journal of Physical Chemistry Letters **4**, 2202 (2013).

[13] G. V. Soni, A. Singer, Z. Yu, Y. Sun, B. McNally, and A. Meller, *Synchronous optical and electrical detection of biomolecules traversing through solid-state nanopores,* Review of Scientific Instruments **81**, (2010).

[14] M. Davenport, A. Rodriguez, K. J. Shea, and Z. S. Siwy, *Squeezing ionic liquids through nanopores,* Nano Letters **9**, 2125 (2009).

[15] S. Cayley, B. A. Lewis, H. J. Guttman, and M. T. Record Jr, *Characterization of the cytoplasm of escherichia coli K-12 as a function of external osmolarity: Implications for protein-DNA interactions in vivo,* Journal of Molecular Biology **222**, 281 (1991).

[16] H. Tabei, S. Nara, and K. Matsuyama, *Method for forming a printed circuit by photolysis of silver salt of organic acid,* Journal of The Electrochemical Society **121**, 67 (1974).

[17] G. D. Fasman, *Practical handbook of biochemistry and molecular biology / edited by Gerald D. Fasman* (CRC Press, Boca Raton, Fla.:, 1989).

[18] Z. Hai-Lang and H. Shi-Jun, *Viscosity and density of water + sodium chloride + potassium chloride solutions at 298.15 k,* Journal of Chemical and Engineering Data **41**, 516 (1996).

[19] I. M. Abdulagatov, A. B. Zeinalova, and N. D. Azizov, *Experimental viscosity b-coefficients of aqueous LiCl solutions,* Journal of Molecular Liquids **126**, 75 (2006).

[20] S. W. Kowalczyk, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Modeling the conductance and DNA blockade of solid-state nanopores,* Nanotechnology **22**, 315101 (2011).

[21] W. M. Deen, *Hindered transport of large molecules in liquid-filled pores,* AIChE Journal **33**, 1409 (1987).

[22] D. Fologea, J. Uplinger, B. Thomas, D. S. McNabb, and J. Li, *Slowing DNA translocation in a solid-state nanopore.* Nano Letters **5**, 1734 (2005).

[23] X. J. A. Janssen, M. P. Jonsson, C. Plesa, G. V. Soni, C. Dekker, and N. H. Dekker, *Rapid manufacturing of low-noise membranes for nanopore sensors by trans -chip illumination lithography,* Nanotechnology **23**, 475302 (2012).

[24] C. Plesa and C. Dekker, *Data analysis methods for solid-state nanopores; submitted,* (2015).

[25] F. Wiesbrock and H. Schmidbaur, *Lithium l-hydrogen-[alpha]-glutamate: A layer structure with*

asymmetrical tunnels formed by nets with two different macrocycles, CrystEngComm **5**, 262 (2003).

9

Data analysis methods for solid-state nanopores

We describe a number of techniques for the analysis of solid-state nanopore ionic current traces and introduce a new package of Matlab analysis scripts with GUI frontends. We discuss methods for the detection of the local baseline and propose a new detection algorithm which bypasses some of the classical weaknesses of moving-average detection. Our new approach removes detected events and re-creates an ideal event-free baseline which is subsequently used to recalculate the local baseline. Iterative operation of this algorithm causes both the moving average of the baseline current and its standard deviation to converge to their correct values. We explain different approaches to selecting events and building event populations, and we show the value of keeping track of the changes in parameters such as the event rate and the pore resistance throughout the course of the experiment. Finally, we introduce a new technique for separating unfolded events and detecting current spikes present within translocation events. This open source software package is available online at: `http://ceesdekkerlab.tudelft.nl/downloads/`

C. Plesa and C. Dekker, *Data analysis methods for solid-state nanopores*, submitted.

9.1. Introduction

Over the past decade, there has been tremendous growth and progress in research on solid-state nanopores[1, 2]. In this technique, a membrane containing a nanometer-scale pore is placed in-between two chambers containing an electrolyte solution, as shown in Figure 9.1a. An electric field is applied across the membrane and charged molecules, such as DNA, present in the solution experience an electrophoretic force which pulls them towards the pore and causes them to translocate through. As a molecule translocates through the nanopore, it temporarily blocks the current and this causes a temporary resistive pulse, as shown in Figure 9.1b. Typically, the duration of the pulse contains information about the length of the molecule while its amplitude is dependent on the molecule's cross-sectional volume.

With the steep development of this field has come the need for signal-processing tools specifically suited to this niche. While many different techniques exist for analysis of nanopore current traces, the majority of data analysis is done on custom software which differs from lab to lab, although some approaches have recently been published[3–5]. Arjmandi et al[3] have discussed the advantages of wavelets over low-pass filtering, particularly in the accurate recovery of the dwell time and amplitude of translocation events. Raillon et al[5] have proposed a new level-fitting algorithm based on the cumulative-sums algorithm. Pedone et al[4] focused on the accurate analysis of short pulses, which is a common issue in experiments aimed at detecting proteins and short DNA.

In this paper, we describe the many aspects of nanopore data analysis as combined in one single comprehensive new Matlab GUI-based package named Transalyzer. We also introduce novel approaches for detecting the local baseline, extracting current peaks present within events, and we describe various analysis strategies for specific scenarios. Our analysis procedure is split into three successive stages, with each stage utilizing parameters determined in the previous stage, as shown in Figure 9.1c. The first stage (GUI_detect) determines the local baseline, rms noise level (σ), it detects each translocation event, and determines its basic properties such as duration, current blockade level, and integrated area [event charge deficit (ECD)]. In the second stage (GUI_events), mixed event populations are sorted and population level statistics are generated, such as the most probable dwell time, blockade level, and event rate. The third and final stage (GUI_localstructures) reanalyzes each event in a given population for the presence of local structures such as bound protein or knots. This analysis pipeline allows us to address the large variability encountered in different types of experiments.

9.2. Event detection and characterization

The analysis procedure begins with the detection of translocation events within a noisy baseline. As most other labs, we use a thresholding algorithm to extract events. In this approach, events are identified if they cross a threshold (typically 5σ) away from the local baseline level. The threshold is defined by multiplying a peak detection factor and the rms noise level (σ), as shown in Figure 9.2. The

peak detection factor is chosen large enough to minimize the number of noise spikes captured, while simultaneously low enough to capture as many translocation events as possible. Successful detection of translocation events requires proper identification of the local value of the baseline and the noise level (σ). A variety of factors complicate the determination of these two values, including: (1) inherently unstable baselines, (2) very large event rates, (3) pore clogging, and (4) successive closely-spaced events. Here, we describe how we have addressed some of these issues in our analysis software, which has been used to analyze a large variety of experimental data.

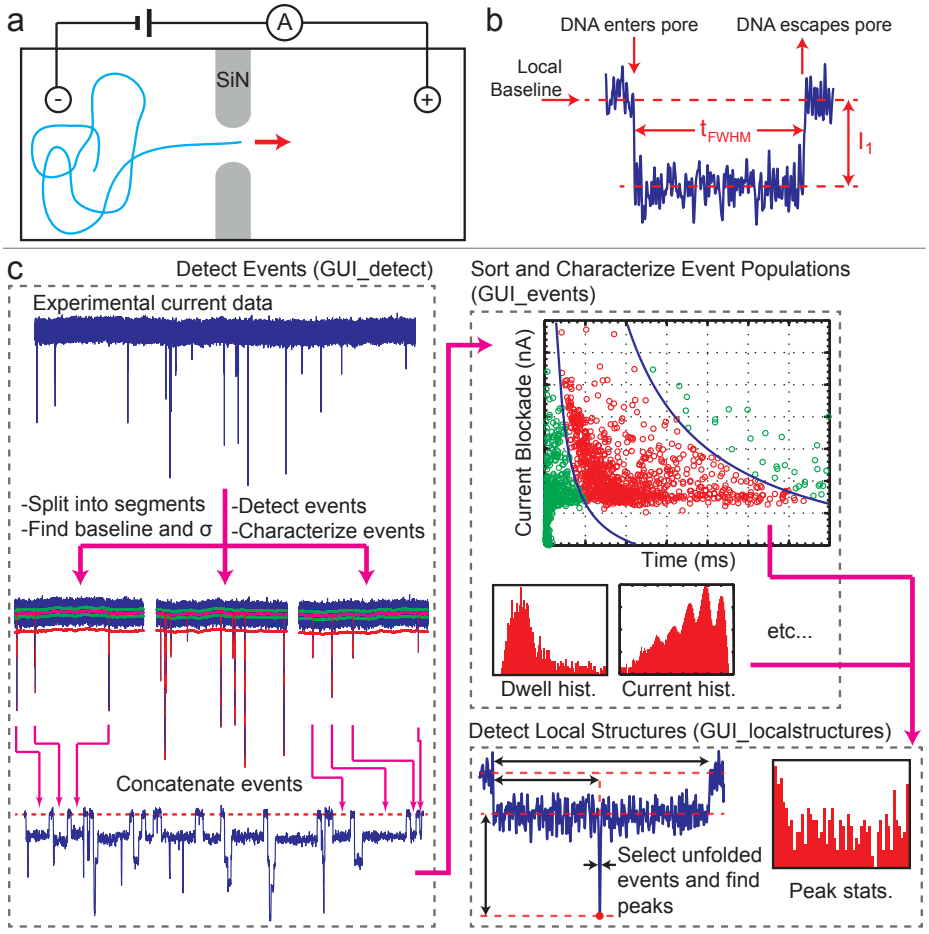


Figure 9.1: a) Illustration of a typical nanopore setup. b) Current signal produced by a translocating DNA molecule. c) Schematic of the typical analysis procedure of a nanopore current trace, which is divided into three parts. The first part splits a current file, detects the events in each segment, characterizes each event, and concatenates all found events. The second part sorts and characterizes the event populations. The final part can be used to sort and re-analyze events for the presence of local structures and to generate relevant statistics.

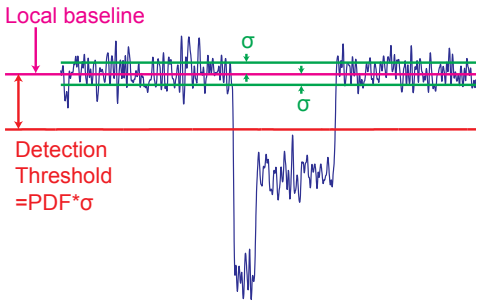


Figure 9.2: A typical threshold detection scheme involves finding the local baseline and rms noise level (σ). A detection threshold is set as a constant (PDF; peak detection factor) multiplied by the rms noise level, away from the local baseline. Events are detected by finding points where the current trace crosses the detection level. The trace can be analyzed backward from the first crossing point and forward from the next crossing point to find the points where the current crosses the local baseline, which define the start and stop of the event.

9.2.1. Baseline detection

Traditionally, baseline detection is performed by calculating a moving average, with the window size optimized to the maximum period of time over which the baseline value is allowed to fluctuate by some chosen amount. Issues affecting proper baseline detection can be distinguished based on the baseline's stability. In the case of a stable baseline, the size of the window can be kept quite large (say, 30k points at 500k samples/s) and this can provide an accurate value in most cases. In cases where the baseline is unstable, however, the window size must be kept small (<6k points) in order to track the baseline fluctuations. In both of these cases, particularly the latter, the moving average can become an inaccurate representation of the local baseline due to the fact that previous events influence the value of the local baseline. This effect is especially noticeable if the event rate is very high, leading to many closely spaced events, or if the event durations are a significant fraction of the window size. We introduce a simple algorithm to deal with all these issues in Section 9.2.3.

9.2.2. Noise level determination

A number of techniques exist for determining the noise level. In the case of a stable baseline, with short-duration events at a low event rate, the trace file can be split into small segments and the global standard deviation can be used. A more accurate method, which works well with high-event-rate data sets, is to first determine the standard deviation in a small moving window (typically 1000 points in size). The values of the standard deviation for all of the windows in a trace segment can be put into a histogram where the bin width is defined by the precision required. The center of the main peak in the resulting histogram typically provides an accurate value of the standard deviation within the trace.

9.2.3. Iterative detection algorithm

In order to overcome the limitations of the thresholding approach, we designed and implemented a new algorithm shown in Figure 9.3a. This approach involves iterating through the thresholding algorithm multiples times in order to decouple the moving average calculation of the local baseline from the influence of previous events. At the end of each iteration a new current trace is generated where the duration of each detected event is replaced by the value of the local baseline at the start of the event. An assumption is thus made that the baseline value does not change significantly ($>\sigma$) over an average event's translocation time. This new trace is subsequently used to recalculate the moving average (using the same approach described in section 9.2.1), the rms noise level, and detect the events again. In Figure 9.3b we show a comparison of the value of the local baseline for a 20k moving average, a 5k moving average, and a 5k moving average with 2 iterations of the algorithm. In this simulation, three events are placed close together. Using the small 5k-point moving average results in a very inaccurate value of the local baseline for the second and third events. Increasing the window size to 20k points improves the accuracy but fails to completely eliminate this effect. The proposed algorithm quickly converges to the correct value with each iteration, while allowing small window sizes to be used, as shown by the 5k-point 2-iteration trace. Similarly for the calculation of the standard deviation, a second trace is created after each iteration were the events are removed and this is subsequently used to determine the new value of the standard deviation. The iterative algorithm is capable of handling event rates where the average time between events is twice as small as the size of the moving window used. So a 5k point window, corresponding to 10 ms at 500k samples/s, can be used on data with event rates of 200 Hz, as long as the average translocation time of the events is several times smaller than the size of the moving window. Much higher event rates must be addressed on a case-by-case basis, although these situations are typically avoided because they can lead to multiple molecules within the pore simultaneously, which can significantly complicate analysis. In the future several alternative implementations of the iterative algorithm could be used to handle more unstable baselines, at the cost of increased computational time. This could include using both the forward and backward moving averages to determine starting and ending points for the event, and interpolating the change in the baseline that occurred over the course of the event.

How does the iterative algorithm perform when analyzing experimental data? We can quantify the improvement in the value determined for the local baseline by introducing a new measure $<I_{\Delta B}>$. This is calculated by first finding the mean value of the fifty points preceding the start of this event and subsequently determining the difference between this mean and the value of the local baseline (from the moving average) for each particular event. We take the absolute value of this difference and determine the mean ($<I_{\Delta B}>$) and standard deviation (STD) of the resulting distribution. If the value of the baseline improves, we expect the value of $<I_{\Delta B}>$ to reduce and the spread of its distribution to become more narrow. We applied this approach to several DNA and protein experimental datasets and reanalyzed each

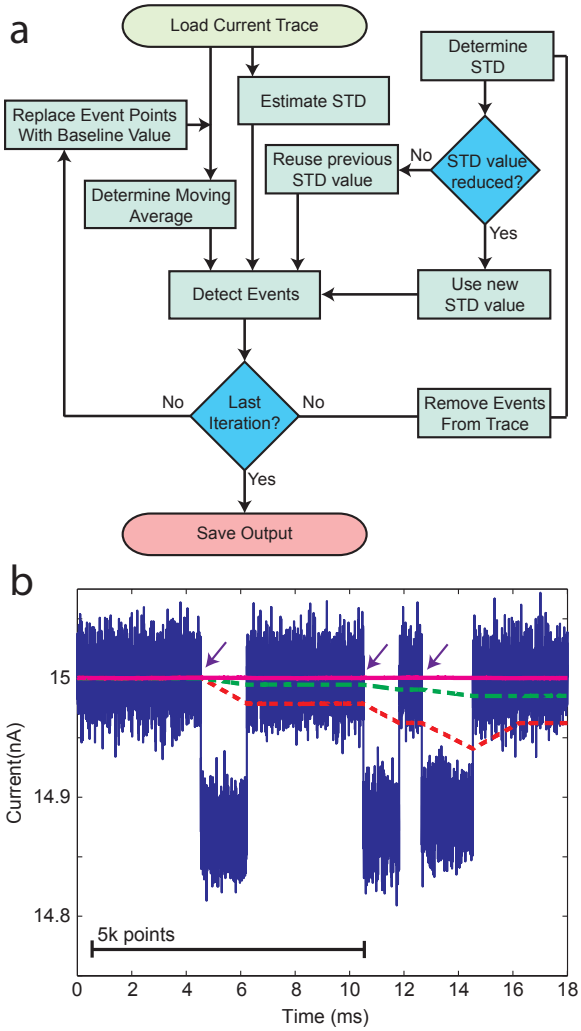


Figure 9.3: a) Flowchart of the iterative detection algorithm. After each iteration, information about the events found feeds back into the next iteration to improve the value of the local baseline and rms noise level. b) A simulation of three closely spaced translocation events, with the local baseline determined using three different techniques. Moving averages of 20k and 5k points, represented by the dashed green and red lines, fail to properly determine the local baseline of the 2nd and 3rd events because the moving average is influenced by the previous events. The solid magenta line shows the same 5k-point moving average after 2 iterations of the detection algorithm, demonstrating that it is able to accurately determine the value of the local baseline despite using a small window size.

dataset using 0, 1, or 2 iterations of the algorithm, with the results shown in Table 9.1. We observed reductions in $<I_{\Delta B}>$ and STD after one iteration in all cases, with further iterations bringing minimal improvements. The larger improvements observed in DNA experiments, can be attributed to the longer duration of these events compared to proteins, which leads to larger changes in the moving average. Although the changes may appear small, these values are averaged over thousands of events. This simple algorithm can thus improve the analysis results and overcome the issues associated with thresholding detection.

Table 9.1: In order to quantify the improvement in the calculation of the baseline we determine $I_{\Delta B}$, which is the absolute value of the difference between the mean of the fifty points preceding the start of the event and its local baseline value. The mean ($<I_{\Delta B}>$) and standard deviation of $I_{\Delta B}$ values in each dataset is shown. All experiments were carried out in 1M KCl, filtered at 10 kHz, and analyzed with a 5k point moving average. Improvements in the value determined for the baseline results in lower values of $<I_{\Delta B}>$ and its STD. In all cases we see an improvement after one iteration, with further iterations bringing only minimal improvements.

Dataset		Num. of Events	Event rate (Hz)	Num. of Iter.	$<I_{\Delta B}>$ (pA)	STD (pA)
A	λ DNA, 20 nm pore, 100 mV	1975	4.8	0	6.3	6.2
				1	5.5	4.7
				2	5.4	4.4
				3	5.4	4.4
B	λ DNA, 10 nm pore, 500 mV	1477	10.0	0	23.4	23.7
				1	20.9	20.1
				2	20.9	20.1
C	T4 DNA, 20 nm pore, 100 mV	1287	0.3	0	4.7	4.1
				1	4.6	3.7
				2	4.6	3.7
D	IgG antibody, 20 nm pore, 100 mV	10221	162.2	0	6.6	6.7
				1	6.0	6.0
				2	5.9	6.0
E	99kDa protein, 16 nm pore, 100 mV	6009	35.8	0	7.0	6.2
				1	6.9	6.1
				2	6.9	6.1

9.2.4. Noise level determination

Proper determination of each event's characteristics (duration, blockade, and ECD) can be complicated by many types of physical phenomena and data-handling effects, depending on the type of experiment, including: short events prone to filtering distortions[4], low SNR, long tail events, folding[6, 7], events were the current increases rather than decreases during translocation[8], hybrid events were the current both decreases and increases[9], events where the molecule docks

onto the pore before translocating[10–13], knotting[14], mixed populations, pore growth over time[15], biomolecule-pore interactions[16], protein-DNA interactions, and the presence of short DNA fragments. For the event duration, using the full-width-half-maximum value (in conjunction with a Gaussian low-pass filter), provides the most accurate translocation time value, even in the light of various distortions introduced by filtering[3, 4]. The blockade level for very short duration events (short DNA or proteins) is best represented by the maximum blockade value. For longer events, dividing the ECD by the FWHM time provides the best representation of the blockade level for many different types of events. For blockades with well-defined levels such as large folds[6], level-fitting software such as OpenNanopore[5] can be used. We have added an export function into Transalyzer capable of exporting event databases into OpenNanopore, effectively acting as an event pre-processor. Our software allows the user to select between multiple analysis techniques to determine the translocation time and blockade level of each population, since multiple populations can coexist within the same experiment.

Due to the many different types of event blockades possible, we allow the user to select between three different types: current increase, current decrease, and hybrid (decrease and increase). This feature can also be used in situations where there is a very low SNR by exploiting the fact that noise is symmetric around the baseline while translocation events (typically) are not. In our approach, the same dataset is analyzed twice, once assuming current increase and again using current decrease. Differences in the properties of the resulting populations, such as the event rate, provide strong evidence that translocation events are present, even when it is difficult to differentiate individual events from noise.

9.3. Population sorting and characterization

Events can be sorted using a number of different criteria into different populations. Our software allows the user to set a minimum and maximum translocation time, current blockade, local baseline level, event number, and event charge deficit in order to select out an event population. In most situations, the event charge deficit (ECD) has a Gaussian distribution for a population of molecules with homogeneous length. A non-Gaussian or distorted ECD distribution can be caused by significant molecule fragmentation, strong biomolecule-pore interactions, low SNR, the presence of docking levels, or overlapping populations. Importantly, selecting a population using the ECD allows folded events to be included in the selection. If folding is not possible, due to the nature of the analyte (nanoparticle, globular protein, etc..) or because the size of the pore is too small, selection using the translocation time can also be useful. Unfolded events (i.e. events with no extra peaks present) can be selected by looking at the maximum amplitude distribution, where similar to current histograms, events contribute to Gaussian peaks depending on the folding, with the first peak corresponding to unfolded events. Selection on event number can be used in time-dependent processes where conditions change during the experiment. Finally, selecting using each event's local baseline allows the quick removal of clogs as well as, if preferred, translocation events which occurred while the pore was partially blocked. Once an event population is selected,

it can be characterized using well-established properties such as the most probable translocation time, the most probable blockade amplitude, the most probable ECD, and the event rate.

Keeping track of how properties change over time during an experiment can be quite useful in many instances. Fluctuations in the event rate as a function of time can indicate the presence of a number of processes: Sudden changes in the event rate can indicate the presence of a clog or partial pore blockage. A slow decrease in the event rate over time suggests possible adsorption of the analyte to the pore membrane or flowcell walls, as can, for example, occur with DNA sticking to SiN in the presence of divalent cations. A gradual increase in the event rate at the start of the experiment which subsequently reaches a plateau level is indicative of poor mixing conditions in the flowcell, an effect noticeable with high-viscosity buffers. Tracking of the absolute value of the baseline as a function of time can be used to quantify effects such as pore growth. Indeed, for long-duration experiments where the baseline value is observed to change significantly over the course of the experiment, the amplitude of the events should be normalized by the value of the local baseline in order to be comparable to each other. This issue is particularly relevant in measurements on small diameter pores. These issues highlight the benefit of tracking how global properties change over the course of an experiment.

9.4. Local structures detection

Finally, we briefly describe how the presence of small current spikes within a translocation event can be detected. Such analysis can, for example, be useful for experiments involving DNA-bound proteins or DNA knots. We begin by separating events containing large folds from unfolded events that contain local spikes. This is accomplished by looking at the area occupied by the current trace in between the first two DNA blockade levels (I_1 and I_2). The first blockade level (I_1) is the most probable blockade level with only a single (double-stranded) DNA molecule inside the pore, while the second blockade level (I_2) is the most probable blockade level when two DNA molecules are in the nanopore simultaneously. These two levels can be determined from their respective peaks in a current histogram. Figure 9.4ab provides two example events, one unfolded and one with a large fold at the start. The area occupied by the current trace (between I_1 and I_2) is shown in red in Figure 9.4cd, while the product ($I_1 t_{FWHM}$) of the DNA blockade level (I_1) and the FWHM translocation time of the event (t_{FWHM}) is represented by a green rectangle. The area occupied by the current (red) is normalized using this value (green) to produce the normalized charge deficit between I_1 and I_2 (NCD_{1-2}). Events with large folds have a large value of NCD_{1-2} while unfolded events with spikes have smaller values. For example, the event in Figure 9.4a has NCD_{1-2} = 0.125 while the folded example of Figure 9.4b has NCD_{1-2} = 0.350. Circular molecules produce NCD_{1-2} values close to 1. Figure 9.4e shows a typical distribution of NCD_{1-2} values for an experiment of DNA with bound proteins that translocate through a 20 nm pore. In order to determine a cutoff between folded and unfolded events, we look at known folding rates from DNA-only experiments. For example, in 1M KCl at 30 kHz bandwidth in a 20 nm pore (i.e. the same conditions of the experiment of Fig. 9.4e), lambda-

phage DNA is observed to have approximately 36% of events unfolded. Figure 9.4f shows the normalized cumulative sum of the NCD_{1-2} distribution. A horizontal blue line has been added at a value of 0.36; a vertical blue line defined by the point of intersection (in this case at NCD_{1-2} = 0.22) between the curve and 0.36 provides the cutoff value used to define events as unfolded or folded. Once a dataset is generated with only unfolded events, we then detect peaks present within the DNA event. Essentially our analysis comes down to detecting events within events. For each peak detected we record the temporal position, the position normalized with the total event duration, the peak FWHM, and the peak amplitude. This simple approach allows for the quick separation of folded and unfolded events and the subsequent detection of any local structures present.

9.5. Discussion and conclusions

We have described a number of analysis techniques implemented in our analysis software and provided a number of examples for specific scenarios. Unlike previous works, we have addressed various effects which occur throughout the analysis procedure. The iterative detection algorithm that we have described provides a simple way to overcome issues typically encountered when using the thresholding detection approach. Furthermore, we have outlined a new method for separating folded events from unfolded events containing current spikes, which is particularly useful in the detection of local structures.

Our Transalyzer analysis package has been licensed under the New BSD Licence, which encourages further development and modification by other labs by imposing minimal restrictions on its modification and redistribution. It is freely available for download from our lab website (`http://ceesdekkerlab.tudelft.nl/downloads/`). A Subversion repository has also been created on Google Code (`http://code.google.com/p/transalyzer/`) to encourage future improvements, additions, and code modifications by other labs.

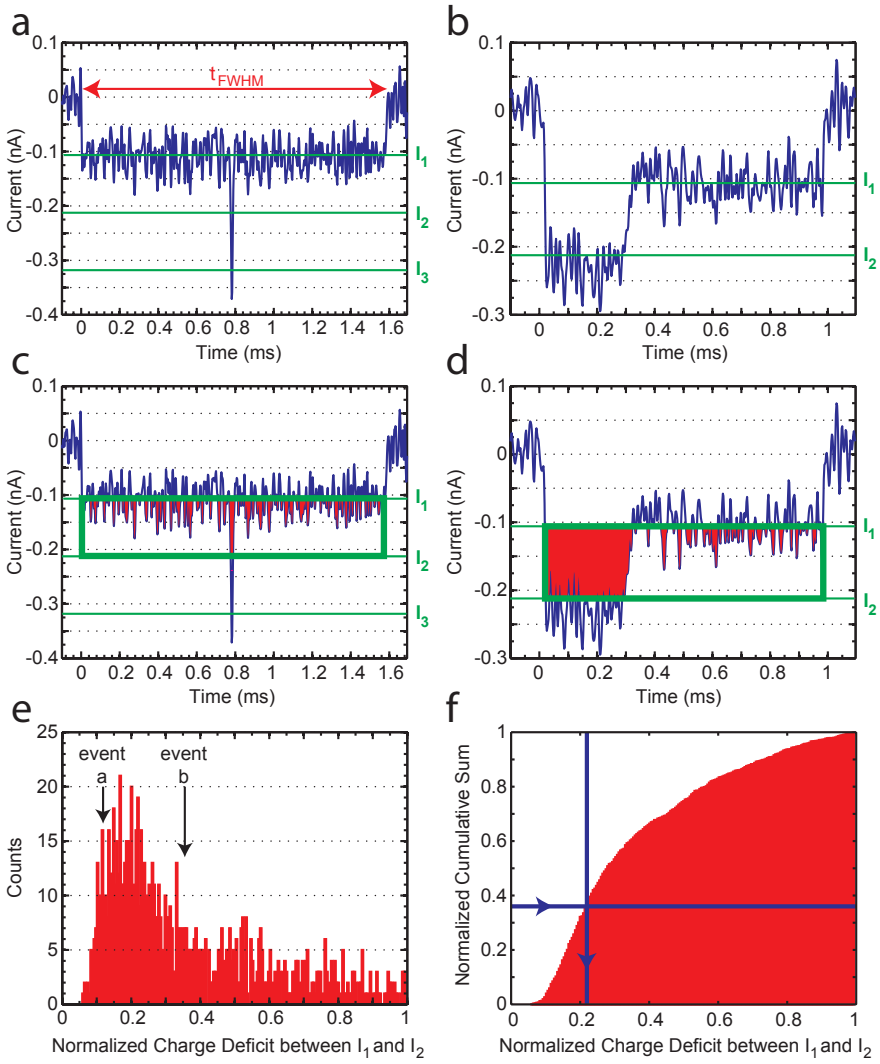


Figure 9.4: Detection of local peaks within events. a) Current trace of an unfolded event with a single spike. b) Current trace with a folded event. The horizontal green lines represent multiples of the single dsDNA blockade level (I_1) as determined using a current histogram. c-d) The same events as in a-b with the area in between the first two blockade levels highlighted. The integral (charge deficit) of the current trace between these two levels is shown shaded in red. This charge deficit is normalized by the total area given by the product ($I_1 t_{FWHM}$) of the DNA blockade level I_1 and the FWHM translocation time t_{FWHM} of the event, shown as a green rectangle. The resulting value is termed the normalized charge deficit between I_1 and I_2 (NCD_{1-2}). Events with folds have higher NCD_{1-2} values. e) Typical distribution of NCD_{1-2} values for a protein-DNA experiment where DNA events contain short spikes, along with the positions of the two example events of panel a and b. f) Normalized cumulative histogram for the distribution shown in e. The vertical line shows the proportion of events which are typically unfolded in these conditions as determined using DNA-only control experiments. The vertical line is the intercept of the normalized cumulative sum with this, and is used to determine the maximum NCD_{1-2} value allowed for an event to be considered unfolded.

Acknowledgments

We thank Auke Booij and Mathijs Rozemuller for code contributions. We also thank Sanne de Jongh, Ruben Zinsmeester, Bruno van den Toorn, Stephanie Heerema, Magnus Jonsson, Adithya Ananth, Daniel Verschueren, Francesca Nicoli, and Claire Hurkmans for testing, reporting bugs, and suggesting new features. This work was supported in part by the European Research Council Advanced Grant NanoforBio (no. 247072) as well as the Netherlands Organisation for Scientific Research (NWO/OCW), as part of the Frontiers of Nanoscience program.

References

[1] C. Dekker, *Solid-state nanopores,* Nat Nano **2**, 209 (2007).

[2] M. Wanunu, *Nanopores: A journey towards DNA sequencing,* Physics of Life Reviews **9**, 125 (2012).

[3] N. Arjmandi, W. V. Roy, L. Lagae, and G. Borghs, *Improved algorithms for nanopore signal processing,* arXiv (2012).

[4] D. Pedone, M. Firnkes, and U. Rant, *Data analysis of translocation events in nanopore experiments.* Analytical Chemistry **81**, 9689 (2009).

[5] C. Raillon, P. Granjon, M. Graf, L. J. Steinbock, and A. Radenovic, *Fast and automatic processing of multi-level events in nanopore translocation experiments,* Nanoscale **4**, 4916 (2012).

[6] C. Plesa, L. Cornelissen, M. W. Tuijtel, and C. Dekker, *Non-equilibrium folding of individual DNA molecules recaptured up to 1000 times in a solid state nanopore,* Nanotechnology **24**, 475101 (2013).

[7] A. J. Storm, J. H. Chen, H. W. Zandbergen, and C. Dekker, *Translocation of double-strand DNA through a silicon oxide nanopore,* Physical Review E: Statistical, Nonlinear, and Soft Matter Physics **71**, 051903 (2005).

[8] R. M. M. Smeets, U. F. Keyser, D. Krapf, M.-Y. Wu, N. H. Dekker, and C. Dekker, *Salt dependence of ion transport and DNA translocation through solid-state nanopores,* Nano Letters **6**, 89 (2006).

[9] S. W. Kowalczyk and C. Dekker, *Measurement of the docking time of a DNA molecule onto a solid-state nanopore,* Nano Letters **12**, 4159 (2012).

[10] A. T. Carlsen, O. K. Zahid, J. Ruzicka, E. W. Taylor, and A. R. Hall, *Interpreting the conductance blockades of DNA translocations through solid-state nanopores,* ACS Nano **8**, 4754 (2014).

[11] C. Plesa, A. N. Ananth, V. Linko, C. Gülcher, A. J. Katan, H. Dietz, and C. Dekker, *Ionic permeability and mechanical properties of DNA origami nanoplates on solid-state nanopores,* ACS Nano **8**, 35 (2013).

[12] R. Wei, T. G. Martin, U. Rant, and H. Dietz, *DNA origami gatekeepers for solid-state nanopores,* Angewandte Chemie, International Edition **51**, 4864 (2012).

[13] J. K. Rosenstein, M. Wanunu, C. A. Merchant, M. Drndic, and K. L. Shepard, *Integrated nanopore sensing platform with sub-microsecond temporal resolution,* Nat Meth **9**, 487 (2012).

[14] A. Rosa, M. Di Ventra, and C. Micheletti, *Topological jamming of spontaneously knotted polyelectrolyte chains driven through a nanopore,* Physical Review Letters **109**, 118301 (2012).

[15] M. v. d. Hout, A. R. Hall, M. Y. Wu, H. W. Zandbergen, C. Dekker, and N. H. Dekker, *Controlling nanopore size, shape and stability,* Nanotechnology **21**, 115304 (2010).

[16] C. Plesa, S. W. Kowalczyk, R. Zinsmeester, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Fast translocation of proteins through solid state nanopores,* Nano Letters **13**, 658 (2013).

Summary

Solid-state nanopores are small nanometer-scale holes in thin membranes. When used to separate two chambers containing salt solution, any biomolecule passing from one chamber to the other is forced to pass through the pore constriction. An electric field applied across the membrane is used to create an ionic current and electrophoretically drive charged molecules through the pore. As a molecule translocates through the pore, it causes a temporary reduction in the ionic current which is measured with a low-noise amplifier. The current blockade is a unique signature containing information about the volume and length of the molecule. Many translocation events are recorded in each experiment and the resulting distributions are analysed statistically. In most cases, high-concentration salt solutions like 1M KCl are used as the measurement buffer, since they provide the highest signal-to-noise ratios. Higher concentrations and different cations such as lithium can be used to slow down the translocation velocity at the cost of a reduced event rate. The work in this thesis uses nanopores with diameters in the range from 5 nm to 40 nm. At these diameters DNA molecules can also translocate through in a folded configuration, where the capture point is not at an end but somewhere along the length of the polymer. The simple but versatile nanopore technique can be applied and modified in many ways, revealing a wide variety of phenomena.

In the first part of this thesis, we focus on translocation of DNA through solid-state nanopores. Instead of probing thousands of individual molecules using a nanopore, as typically done, one can measure one individual molecule thousands of times by switching the electric field very quickly after a molecule has passed through the pore. This is the approach taken in **Chapter 2**, where we recapture four different types of molecules hundreds or even thousands of times and analyse their behavior. The recaptured translocation events are characterized by very strong folding of the DNA molecule. This reflects an inherent trade-off in this approach. If the molecule is recaptured too quickly, it will still be in a high-density non-equilibrium state close to the pore, which results in a lot of folding, while if too much time is allowed to pass, the molecule will diffuse away and escape into the solution. This dense non-equilibrium state is present on a timescale defined by the time required for the polymer to relax back to equilibrium. We observe that the amount of folding indeed is dependent on how quickly the molecule is recaptured relative to this relaxation time. Furthermore, we also note that the total translocation time correlates with the molecule's initial conformation at the start of the

translocation, which can be controlled by changing the recapture time relative to the relaxation time. Molecules with more extended conformations are found to have longer translocation times. These observations highlight some of the characteristics of this method and pave the way for a number of other possible applications.

One of the major open questions on the mechanism of nanopore translocation is how the local velocity changes as a function of the length of the molecule that is translocating through. The shape of this velocity profile is vital to be able to convert the measured signals which are recorded in the temporal domain into positional information. One can imagine that a large DNA blob, such as the one at the start of the translocation process, induces a much larger drag force compared to the situation close to the end of the translocation process when only a few DNA segments remain to be pulled through. In **Chapter 3** we address this problem by assembling linear DNA origami constructs with markers at known positions in order to measure the velocities of different segments along the molecule as it translocates through. We observe significant intramolecular velocity fluctuations which are attributed to changes in the drag force as the DNA blob unfolds. Furthermore, we also see significant intermolecular fluctuations between different molecules of the same length as well as between equal-length molecules translocating through different pores of the same diameter. A speed up is observed at the end of the translocation process which is consistent with the proposed model of the DNA blob unfolding. The measured velocity profile allows us to estimate the uncertainty present when determining the position of a feature along the DNA and highlights the importance of carrying out ensemble measurements.

In the next chapter, we shift attention to knotting. Have you ever placed some earphones into your pocket, only to find them completely tangled when you pull them out again? A similar effect occurs in long polymers such as DNA, which can form knots in equilibrium conditions. The amount of molecules with a knot is dependent on the persistence length (flexibility) of the polymer, the ionic conditions, and the polymer's length. Previous methods to study DNA knotting were limited to short circular molecules which could be separated on a gel. In **Chapter 4** we demonstrate that solid-state nanopores can be used to detect DNA knots in both linear and circular molecules up to 166 kbp in length. Knots are found to be very tight, with sizes below 100 nm, which requires the use of high-resolution measurements. The translocation process can cause knots to slide out of linear molecules or get pushed towards the end in the case of circular molecules at high applied forces. The knotting probability is observed to increase quickly with length and is in agreement with simulation-based predictions. These measurements highlight the usefulness of this approach in probing DNA knotting phenomena.

One approach to controlling the geometry and chemical functionalities of a nanopore is to dock a DNA origami nanoplate on top of a nanopore. In order to achieve the highest signal-to-noise ratio for these types of measurements, the ionic permeability of the nanoplates must be minimized, such that most of the ions pass through the aperture rather than the DNA origami nanoplate itself. In **Chapter 5** we characterize four different DNA origami nanoplate designs by docking them on top of nanopores of different diameter and measuring their conductance

relative to the open pore conductance. We find that all designs show high ionic permeability with only small differences observed between different designs. The ionic permeability is found to increase with pore diameter and increase at lower ionic strengths. A number of interesting mechanical effects are observed. The DNA nanoplates can physically deform under high applied force and even be pulled through the nanopore if enough force is applied. Some DNA nanoplates are also observed to spontaneously mechanically buckle, a phenomenon which manifests as discrete jumps in the current level of docked nanoplates. This characterization should facilitate the development of new applications using DNA origami nanoplates docked onto nanopores.

In part 2 of this thesis we introduce proteins. Proteins have several key differences compared to DNA which makes them more challenging to work with. Most proteins are globular with a much smaller size which is able to fit within the volume of the nanopore constriction. Unlike DNA, they can have various charges which can be distributed in a non-uniform manner across their structure. These differences, in particular their small size, can make them difficult to detect. In **Chapter 6** we examine the event rates for a number of different proteins translocating through nanopores and find that they are significantly smaller than expected. This discrepancy, which can be orders of magnitude in size, is larger for smaller proteins which both diffuse faster and produce smaller current blockades. We investigate this phenomenon experimentally by looking at how the event rate changes with pore size and protein concentration for both large and small proteins. We conclude that the majority of proteins pass through the pore undetected due to a combination of limited temporal resolution and low signal-to-noise ratios. We consider the potential implications of this on nanopore-based protein detection and discuss several possible solutions.

Within the cell, DNA is never bare but rather covered with many different proteins having different functions. Detection of DNA-bound protein is an interesting application with many potential uses. In **Chapter 7** we demonstrate the detection of single DNA-bound protein using anti-DNA antibodies. These antibodies bind randomly onto DNA, can withstand high-salt concentrations, and only translocate through the pore when bound to DNA, due to their positive charge. The bound antibodies produce current spikes within the DNA current blockades which have very short durations and high amplitudes. Given the limited resolution of the measurement, we conclude that antibodies are primarily observed due to transient protein-pore interactions. Further developments could allow this type of detection approach to be generalized to other DNA-binding protein.

In the last part of the thesis we look at new methods for nanopore sensing. Nearly all nanopore experiments carried out in literature have used silver/silver-chloride electrochemistry. In **Chapter 8** we introduce an alternative electrochemisty based on silver/silver-glutamate. We observe that this set of reactions produces a linear response and can be used to detect DNA translocations. Due to its much larger size, the glutamate anion can act as a thickening agent in high-concentration solutions. We use this approach to slow down DNA translocation by up to 11 times, with a reciprocal reduction in signal of up to 7 times. These results demonstrate

that this is a viable alternative approach to existing methods.

Data analysis is an important part of the nanopore sensing process. In **Chapter 9** we discuss a number of techniques for analysing nanopore data. A new detection algorithm is introduced which aims to bypass some of the typical issues with moving-average-based detection. This is achieved by removing detected events from the current trace and recalculating the moving average. This approach is shown to quickly converge to correct values after several iterations. We also detail how different event populations can be sorted and how different analysis approaches can be used to infer certain experimental properties. A new technique for separating folded and unfolded events is also introduced which should be useful in the detection of local structures.

The wide variety of results presented in this thesis demonstrate the versatility of the solid-state nanopore approach and highlight some of its potential applications. We expect many follow-up studies to continue developing some of the topics explored in this thesis, particularly the use of nanopores to determine polymer properties, the combination of DNA-origami nanostructures and nanopores, the use of nanopores to study knotting phenomenon in long polymers, and detection of DNA-bound protein. This young field of research will continue to experience tremendous growth and development well into the foreseeable future.

Samenvatting

'Solid-state nanopores' (vaste-stof nanogaatjes) zijn gaatjes van enkele nanometers in een dun membraan van bijvoorbeeld silicium nitride. Zo'n membraan met een nanogaatjes kan worden gebruikt om twee compartimenten met een zoutoplossing te scheiden. Wanneer er biomoleculen in de oplossing zitten, kunnen deze zich via het gaatje van het ene compartiment naar het andere bewegen. Als men een elektrisch veld over het membraan aanlegt, wordt er een ionenstroom gegenereerd en worden geladen moleculen via elektroforese door het nanogaatje gedreven. Een molecuul dat door het gaatje beweegt resulteert dan in een tijdelijke reductie van de ionenstroom, die wordt gemeten met een lage-ruis versterker. Deze tijdelijke stroomafname bevat unieke informatie over het volume en de lengte van het molecuul. Door veel van dit type molecuultranslocaties te meten kan de statistische distributie worden bepaald en bestudeerd. Vaak wordt er gebruik gemaakt van een hoge zoutconcentratie in de oplossing, zeg 1 M KCl, omdat dit de hoogste signaal-ruis verhouding oplevert. Nog hogere zoutconcentraties en andere kationen, zoals lithium, kunnen worden gebruikt om de translocatiesnelheid van een molecuul te verlagen, maar heeft als nadeel dat de frequentie van de translocaties afneemt. Voor het werk dat in deze thesis wordt beschreven zijn nanogaatjes met diameters van 5 tot 40 nm gebruikt. DNA moleculen kunnen gaatjes van deze diameters ook in gevouwen toestand passeren. In dat geval penetreert niet het uiteinde van het molecuul als eerste het gaatje, maar een punt ergens in het midden van het DNA polymeermolecuul. De simpele en veelzijdige 'nanopore' techniek kan op veel verschillende manieren worden toegepast en gemodificeerd, waarmee een breed scala aan fenomenen kan worden onthuld.

In het eerste deel van deze thesis bespreken we DNA translocaties door vaste-stof nanogaatjes. In plaats van duizenden verschillende moleculen te meten door een enkel nanogaatje, zoals meestal gedaan wordt, kan men ook een en hetzelfde molecuul duizend maal opnieuw meten. Dit kan door het elektrisch veld snel om te keren wanneer het molecuul net het gaatje gepasseerd is. Deze techniek wordt besproken in **Hoofdstuk 2**, waar we vier verschillende typen DNA moleculen honderd en soms zelfs duizend keer terugvangen om hun gedrag te kunnen analyseren. De DNA moleculen uit deze metingen, waarbij we de moleculen na passage direct weer opnieuw transloceren, worden gekarakteriseerd door een hoge mate van vouwing in het DNA, iets waar rekening mee moet worden gehouden wanneer men deze techniek gebruikt. Wanneer het molecuul te snel wordt terug gevangen, zal het zich in hoge dichtheid, in een niet-evenwichtstoestand bevinden vlakbij de pore, wat resulteert in een veelvuldig opgevouwen structuur. Wanneer men echter te lang wacht met het omschakelen van het elektrisch veld, dan diffunderen de moleculen weg en ontsnappen deze weer in de oplossing. De niet-evenwichtstoestand van het molecuul dichtbij het gaatje duurt voort gedurende een tijd die het poly-

meer nodig heeft om te relaxeren naar evenwicht. We zien inderdaad dat de mate van vouwing direct afhankelijk is van de relaxatietijd. Ook zien we dat de totale translocatietijd correleert met de initiële formatie van het molecuul bij aanvang van de translocatie. Dit wordt gecontroleerd door de tijd waarin het molecuul opnieuw wordt gevangen te variëren ten opzichte van de relaxatietijd. Langere moleculen hebben een langere translocatietijd. Deze observaties laten zien wat de karakteristieke eigenschappen zijn van dit type 'ping-pong' metingen en vormen een basis voor een aantal toepassingen.

Een van de belangrijkste open vragen voor onderzoekers die werken aan de ontwikkeling van nanogaatjes is hoe de lokale snelheid verandert tijdend het proces van translocatie. De vorm van dit snelheidsprofiel is essentieel voor de omzetting van de gemeten signalen die opgenomen worden in het tijdsdomein, terwijl we in het algemeen juist geïnteresseerd zijn in de positie-informatie. Men kan zich voorstellen dat een grote DNA kluwen, zoals aanwezig bij de start van de translocatie, een veel grotere wrijvingsweerstand veroorzaakt in vergelijking met de situatie vlak voor het eind van de translocatie, wanneer er nog slechts een aantal DNA segmenten resteren die door het gaatje getrokken moeten worden. In **Hoofdstuk 3** komt dit probleem aan de orde. Hier wordt de snelheid van de verschillende DNA segmenten gemeten met behulp van een lineair DNA origami construct met markers op gegeven posities. We observeren significante intramoleculaire fluctuaties, die waarschijnlijk veroorzaakt worden door de verandering in weerstand wanneer het DNA zich ontvouwt. Tevens zien we ook significante intermoleculaire fluctuaties tussen moleculen van dezelfde lengte, alsmede aanzienlijke fluctuaties tussen moleculen van dezelfde lengte die door verschillende gaatjes van dezelfde diameter getrokken worden. Aan het einde van de translocatieproces voor een molecuul wordt een verhoogde snelheid geobserveerd, wat consistent is met het voorgestelde model van het uitvouwen van de DNA kluwen. Het gemeten snelheidsprofiel maakt het mogelijk om een schatting te maken van de onzekerheid van de plaatsbepaling van een bepaald object langs het DNA. Ook benadrukken de snelheidsfluctuaties in deze experimenten het belang van het uitvoeren van experimenten aan een grote hoeveelheid moleculen.

In het volgende hoofdstuk, verschuiven we de aandacht naar knopen. Heeft u dat ook wel eens: u stopt uw oordopjes in uw zak en wanneer u ze er vervolgens weer uit haalt, zit het snoer volledig in de war. Een soortgelijk effect treedt op in lange polymeren zoals DNA, die knopen kunnen vormen in evenwichtstoestand. Het aantal moleculen met een knoop hangt af van de persistentie lengte (buigzaamheid) van het polymeer, de zoutconcentratie van de oplossing en de lengte van het polymeer. Eerdere methoden om knopen in DNA te onderzoeken waren gelimiteerd tot het gebruik van korte circulaire moleculen die op een gel gescheiden konden worden. In **Hoofdstuk 4** demonstreren wij dat vaste-stof nanogaatjes gebruikt kunnen worden voor de detectie van DNA knopen in zowel lineair als circulaire moleculen van zeer grote lengtes (maximaal 166 kbp). De waargenomen knopen blijken heel strak te zijn, met een grootte die kleiner is dan 100nm. Observatie ervan vereist daarom hoge resolutie metingen. Het translocatieproces kan er toe leiden dat knopen uit het molecuul schuiven in het geval van lineaire molecu-

len, terwijl knopen naar het eind van het molecuul worden geschoven voor circulaire moleculen bij hoge kracht. Wij observeren een waarschijnlijkheid tot knoopvorming die snel toe neemt naarmate de lengte van het molecuul toe neemt, en die overeenkomt met op simulatie gebaseerde voorspellingen. De resultaten benadrukken de waarde van deze aanpak in het onderzoek naar het fenomeen van DNA knopen.

Een nieuwe manier om de geometrie en chemische functionaliteit van een nanogaatje te regelen is het plaatsen van een DNA origami nanoplaat op een nanogaatje. Ten einde een hoge signaal-ruis verhouding te verkrijgen voor dit type meting is het noodzakelijk dat de doorlaatbaarheid van de nanoplaat voor ionen minimaal is, zodat de ionen door de opening in de nanoplaat passeren, in plaats van door de DNA origami nanoplaat zelf. In **Hoofdstuk 5** karakteriseren we vier verschillende types DNA origami nanoplaten door ze op nanogaatjes van verschillende diameter te verankeren en dan de geleiding te bepalen ten opzichte van de geleiding van het nanogaatje zonder nanoplaat. Onze resultaten laten een hoge ionische doorlaatbaarheid zien met slechts kleine verschillen tussen de verschillende ontwerpen. We vonden dat de doorlaatbaarheid voor ionen toeneemt met de diameter van het gaatje en toeneemt bij lagere zoutconcentraties. Een aantal interessante mechanische effecten werden waargenomen. De DNA nanoplaat kan fysiek vervormen onder hoge kracht en kan zelfs door het nanogaatje heengetrokken worden als er genoeg kracht word aangebracht. Enkele DNA nanoplaten vertoonden spontane mechanische knikken, een fenomeen dat zich manifesteert als een discrete stapjes in de elektrische stroom. Onze eerste karakterisering van deze DNA origami structuren opent de weg naar de ontwikkeling van nieuwe toepassingen van DNA origami nanoplaten op nanogaatjes.

In deel 2 van dit proefschrift introduceren we eiwitten. Eiwitten kenmerken zich door een aantal belangrijke verschillen in vergelijking met DNA die er toe leiden dat het een grotere uitdaging is om deze te detecteren met vaste-stof nanogaatjes. De meeste eiwitten zijn bolvormig en veel kleiner dan DNA, waardoor ze geheel binnen het volume van de nanogaatje passen. In tegenstelling tot DNA kunnen eiwitten zowel positieve als negatieve ladingen hebben, die niet uniform verdeeld zijn over hun oppervlak. Deze verschillen, voornamelijk het feit dat eiwitten zo klein zijn, kunnen ervoor zorgen dat eiwitten lastig te detecteren zijn. In **Hoofdstuk 6** onderzoeken we de frequentie waarmee translocaties door een nanogaatje plaatsvinden voor een aantal verschillende eiwitten, en we vinden dat deze significant lager is dan verwacht. Dit verschil, dat ordes van grootte kan zijn, is groter voor kleinere eiwitten die zowel sneller diffunderen als kleinere stroomblokkades teweegbrengen. We onderzoeken dit fenomeen experimenteel, door te kijken naar de verandering van de frequentie als functie van de grootte van het nanogaatje en als functie van de eiwitconcentratie voor zowel grote als kleine eiwitten. We concluderen dat de meerderheid van de eiwitten door het nanogaatje gaan zonder een meetbaar signaal af te geven, hetgeen veroorzaakt wordt door een combinatie van een eindige tijdsresolutie en een lage signaal-ruis verhouding. We bekijken de potentiële implicaties hiervan op eiwit detectie met nanogaatjes en we bespreken verschillende mogelijke oplossingen.

In de cel is DNA nooit helemaal kaal, maar juist bedekt met veel verschillende

eiwitten die diverse functies hebben. Detectie van aan DNA gebonden eiwitten is interessant vanwege vele potentiële toepassingen. In **Hoofdstuk 7** tonen we detectie van individuele DNA-gebonden eiwitten aan met behulp van anti-DNA antilichamen. Deze antilichamen binden DNA op willekeurige plekken, kunnen tegen hoge zoutconcentraties en transloceren vanwege hun positieve lading alleen door een nanogaatje wanneer ze aan DNA gebonden zijn. De gebonden antilichamen geven zeer korte stroompieken met een hoge amplitude binnen de door DNA veroorzaakte stroomblokkades. In het licht van de beperkingen in de resolutie van onze metingen concluderen we dat de detectie van antilichamen voornamelijk mogelijk wordt gemaakt door kortdurende interacties tussen het eiwit en het nanogaatje. Verdere ontwikkelingen kunnen de detectiemethode geschikt maken voor metingen aan andere DNA-bindende eiwitten.

In het laatste gedeelte van dit proefschrift kijken we naar nieuwe detectiemethodes voor nanogaatjes. Bijna alle experimenten met nanogaatjes in de literatuur gebruiken zilver/zilver-chloride elektrochemie. In **Hoofdstuk 8** introduceren we een elektrochemisch alternatief gebaseerd op zilver/zilver-glutamaat. Deze set van reacties produceert een lineaire respons en kan worden gebruikt voor de detectie van DNA translocaties. Omdat het glutamaat anion groter is, werkt het als een verdikker in hoge-concentratie oplossingen. We gebruiken deze benadering om DNA tot wel 11 keer langzamer te laten transloceren, waarbij het signaal met een factor 7 reduceert. Deze resultaten tonen aan dat deze aanpak een alternatief is voor bestaande methodes.

Data analyse is een belangrijk onderdeel van detectie met behulp van nanogaatjes. In **Hoofdstuk 9** bespreken we een aantal technieken voor de analyse van nanogaatjesdata. We introduceren een nieuw detectie algoritme dat enkele typische problemen omzeilt die geassocieerd zijn met detectie gebaseerd op het voortschrijdende gemiddelde. We verhelpen dit probleem door gedetecteerde events uit de curve te verwijderen en vervolgens het voortschrijdende gemiddelde te herberekenen. Na enkele iteraties convergeert deze methode snel naar de correcte waardes. We beschrijven ook hoe verschillende populaties in de data gesorteerd kunnen worden en hoe verschillende analysemethodes gebruikt kunnen worden om een aantal verschillende experimentele eigenschappen te bepalen. Daarnaast introduceren we een nieuwe techniek voor het scheiden van gevouwen en niet-gevouwen events. Deze techniek zou waardevol kunnen zijn voor de detectie van lokale structuren.

De grote verscheidenheid aan resultaten gepresenteerd in dit proefschrift demonstreert de veelzijdigheid van vaste stof nanogaatjes en onderstreept een aantal potentiële toepassingen. Ik verwacht dat er veel studies uit dit onderzoek zullen voortvloeien die doorgaan op de ontwikkelingen die in dit proefschrift verkend zijn, in het bijzonder wat betreft het gebruik van nanogaatjes voor het bepalen van polymeereigenschappen, de combinatie van DNA origami nanostructuren en nanogaatjes, het gebruik van nanogaatjes voor het bestuderen van knopen in lange polymeren en de detectie van DNA gebonden eiwitten. Ik verwacht dat dit jonge onderzoeksveld een verdere groei en ontwikkeling zal doormaken in de nabije toekomst.

Acknowledgements

A handful of centuries ago it was still possible to board a ship venturing into the unknown and sail the uncharted oceans until a speck appeared on the horizon, *Land, ho!*, a moment of discovery. In the last four years I've had my own amazing adventure in a strange invisible world all around us which has been filled with such thrilling moments. Along the way I've met many amazing people, had a ton of fun, and learned a lot about both science and myself. Of course no such voyage could be complete without the help of many others who I would like to recognize here.

As any long voyage, science is fraught with difficulties, challenges, and frustrations so a constant supply of enthusiasm is needed. Few people I've ever met have as much infectious enthusiasm as Cees. I'd like to thank Cees for his support throughout my time here, all of his great advice, and his constant friendly attitude. You gave me all the freedom I needed to push my boundaries which allowed me to work on many different ideas and try things I would have not otherwise done (Who says nanopore people can't work with cells?). I will always remember our fast-paced heated debates next to the whiteboard and admire how quickly you could cut through the noise and get to the heart of the issue. Finally, thanks for establishing such a great place to work. Where else could I have said that I've been canyoning in the Pyrenees, flying a helicopter above the Veluwe, wading through the zee, or riding a speedboat across a remote lake towards the face of a glacier?

BN is a remarkable community where I've had great times with a lot of people both within the lab and outside. My first year here was a roller coaster, and two weeks after I arrived, I got a crash course in lab traditions when Marijn came into my office saying, *"Here put this suit on, we're filming a movie. Oh, you can't drive a manual car? Don't worry, just try not to crash it.*" Outside the lab, Monday nights were spent on the football pitch. Never a dull game with Real RKC. Any other day, it was foosball, with the usual suspects, you know who you are. Lots of deals were sealed, trash talk with no limits, and every now and then the floor was cleaned. Fridays came TeePee and for days ending in 'y' there was always 't Klooster. Between all of this, my thesis work was pushed along by a constant stream of advice from postdocs, PhDs, and friends who I'd all like to acknowledge. Iwijn, thanks for the constant supply of bad humor. I've never seen anyone so good at convincing people to patent stuff, let's find something else to patent and good luck at Cornell. Rutger thanks for the pointers on data analysis, good luck in Utrecht. Xander, thanks for all that fabrication work. Snippert, thanks for bringing in the biologist's perspective. Gautam, I can always count on your help regardless of the problem, best of luck at RRI. Allard, thanks for all the advice, for both AFM and other. Tim and Meghan, always a great dinner and a delicious turkey, thanks. Burnham, let's face it, things sound better with a British accent, thanks. Siddarth, thanks for the interesting discussions. Yaron, thanks for making the world a better

place. Greg, who knew chemists could be so much fun? I'm still waiting for that drink with a message floating on top of the water and *poiu, poiu* is still stuck in my head, thanks for that. Every lab needs someone like Magnus to keep the place classy, be it fashion (with bow ties both big and very small), craft brews, or a steady supply of Slagsmålsklubben. Thanks for all the great times and I wish you the best in Linköping, I'm sure you'll make a great professor. Speaking of classy, Dulin, thanks for always restoring the balance with that uncomparable humor, don't ever change, best of luck in Oxford. When all else fails, Lipfert can fix it with his endless knowledge of all science topics. Thanks for all the advice and convincing Tessa to come to BN. Everyone else in the nanopore group: Adi, Stephanie, Daniel, Laura thanks for all of your input and contributions. Keep up the great work. Adi, if we keep eating pizza and going to TeePee, we'll eventually find a way to deal with those pesky elephants. Stephanie, graphene is a tricky thing, but I'm sure you can get it to work. Daniel, not everyone can look good with a wig and a dress, but you pulled it off, thanks for the entertainment, its been fun working with you. Nanopores can be problematic. Stefan, thanks for all of the tips on getting things working, it made life much easier. To the EMM gang, Fabai, Andrew, Hugo and Jess. Fabai, after sharing an office for so many years a lot of memories add up, parties, dumplings, music, and more. Thank you for your friendship throughout the years. I am grateful to have you as a *paranimf*. Andrew, I really enjoyed our talks where every topic is open and there are no limits on new ideas. Hugo and Jess, thanks for all the parties, good luck with everything ahead. Thanks to all my other fellow PhD students. Felix, the one and only. Thanks for your constant supply of entertainment, long discussions at 't Klooster, leading our glorious football team to victory, and the thousand other things you do. I'm honored to have you as a *paranimf*. Marijn, thanks for all the advice, hope you're having a great time in Vancouver, good luck with Steven. Rifka, thanks for all of your help, particularity in the last few months. Bojk, keep those great editorials coming. Kuba, thanks for reminding me that art and science are two sides of the same coin and demonstrating that the Bermuda Triangle is not just in the Atlantic. Jorine, good luck taming those sneaky yeast cells. Thanks also to Pauline, Maarten, Charl, Sriram, Regis, Mahipal, Jetty, Natalia, Roy, Zohreh, Orkide, Eugen, Ruben, Alicia, Dominik and all the other students at BN. Any day is a good day for a party, just ask Lajko, Tom, and Benedetta. Good luck with your PhDs. I'd also like to thank all of my students: Sanne, Ruben, Erik, Maarten, Ludo, Menno, Just, Maryse, and Nick for all their experimental work and their large contributions to this thesis. I hope you learned as much from me as I did from you. Special thanks to Tessa, Stephanie, Jorine, and Felix for helping me translate parts of this thesis into Dutch. Within BN I've received help and advice from many other people as well. Thanks Christophe for your help, particularly when I got started with the lipid work. Thanks also to Nynke, Anne, Elio, Bertus, Martin, Marie-Eve, Chirlmin, and Timon for advice and great discussions during my time here. Thanks to Susanne, Serge, Bronwen, and Jaco for making countless constructs and bearing with me while I tried to learn some of the protocols myself. Thank you Dimitri and Jelle, who needs a 3D printer with them around? Meng-Yue thanks for drilling all those pores. Thanks Jaap, Sacha, and Roland for help with all computers, pumps, and electronics issues

throughout the years. Thanks Jacob for your advice and for providing a constant supply of inspirational artwork. Thanks also to Amanda, Jolijn, Dijana, Emmylou, and all the other staff. In the past several years I've spent a significant amount of time with the local iGEMers and it's been an amazing experience. Where else can you spend months helping students brainstorm and develop all sorts of crazy ideas limited only by your imagination? I would love science even more if this was always the normal approach. Thanks Anne, Esengül, Daniel, and Emrah for the great team spirit and all of the memories along the way.

Scientifically, I've had collaborators without whom many projects would not have been possible. When things don't quite make sense Alexander Grosberg and Yitzhak Rabin were always there to suggest possible explanations. Shura, our conversations in Warsaw were very illuminating and I enjoy your humor. Aleksei, thanks for pointing me in the right direction. Hendrik, thanks for bringing your expertise and your input along the way. Without it, the fishing line would probably still be just an idea on paper. The nanopore community is a great and intellectually vibrant place, where I've had stimulating discussions at conferences with more people than I can name. Finally, I'd like to thank all of the members of my defence committee: Nynke Dekker, Marileen Dogterom, Henny Zandbergen, Hendrik Dietz, Ulrich Keyser, Kees Storm, and Giovanni Maglia for taking the time review this thesis.

I owe a debt of gratitude to my family, who guided me on this path since I was very young. I'd like to thank my parents for all of their support throughout the years. You gave me my sense of curiosity without which I would have never become a scientist. Thanks for giving me such a unique childhood and all of the small things which brought me to this point, from visits to Antipa, *General Physics* on the bookshelf, playing around in the hydraulics lab, to climbing weather stations in the Karakoram. Dad, thanks for showing me the mysteries of the world around us and teaching me that the greatest discoveries lie at the edge of our knowledge. Mom, thanks for always being there to help and reminding me that there is more to life than science.

Tessa, where to begin? If I had been asked four years ago what I thought my greatest achievement would be in four years time, I would have probably said completing this thesis. Now as I write these last few lines, I can say it was bringing you into my life. Thanks for always being there by my side and for providing a constant distraction. I can't wait for our next adventure together. I love you.

Calin Plesa
Delft, December 2014

Curriculum Vitæ

Calin Plesa

18-03-1984	Born in Bucharest, Romania.
2002–2008	Bachelor of Applied Science in Engineering Physics (Honours) Simon Fraser University, Burnaby, British Columbia, Canada
2008–2010	Erasmus Mundus Master of Nanoscience and Nanotechnology Technische Universiteit Delft, Delft, The Netherlands (1^{st} year) Chalmers University of Technology, Göteborg, Sweden (2^{nd} year) Thesis: Assembly and characterization of two-dimensional DNA nanostructures using synthetic three-way oligonucleotides Promotor: Prof. dr. B. Nordén
2010–2015	Ph.D. Bionanoscience Technische Universiteit Delft, Delft, The Netherlands Thesis: Solid-state nanopores for probing DNA and protein Promotor: Prof. dr. C. Dekker

List of Publications

11. **C. Plesa**, J. W. Ruitenberg, M. J. Witteveen, and C. Dekker, *Detection of single DNA-bound proteins using solid state nanopores*, manuscript in preparation.

10. **C. Plesa**, D. Verschueren, J. W. Ruitenberg, M. J. Witteveen, M. P. Jonsson, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Direct observation of DNA knots using solid state nanopores*, submitted.

9. **C. Plesa** and C. Dekker, *Data analysis methods for solid-state nanopores*, submitted.

8. **C. Plesa**, N. van Loo, P. Ketterer, H. Dietz, and C. Dekker, *Velocity of DNA during translocation through a solid state nanopore*, Nano Letters (2014), in press.

7. **C. Plesa**, A. N. Ananth, V. Linko, C. Gülcher, A. J. Katan, H. Dietz, and C. Dekker, *Ionic Permeability and Mechanical Properties of DNA Origami Nanoplates on Solid-State Nanopores*, ACS Nano **8**, 1, 35-43 (2013).

6. **C. Plesa**, L. Cornelissen, M. W. Tuijtel, and C. Dekker, *Non-equilibrium folding of individual DNA molecules recaptured up to 1000 times in a solid state nanopore*, Nanotechnology **24**, 47, 475101 (2013).

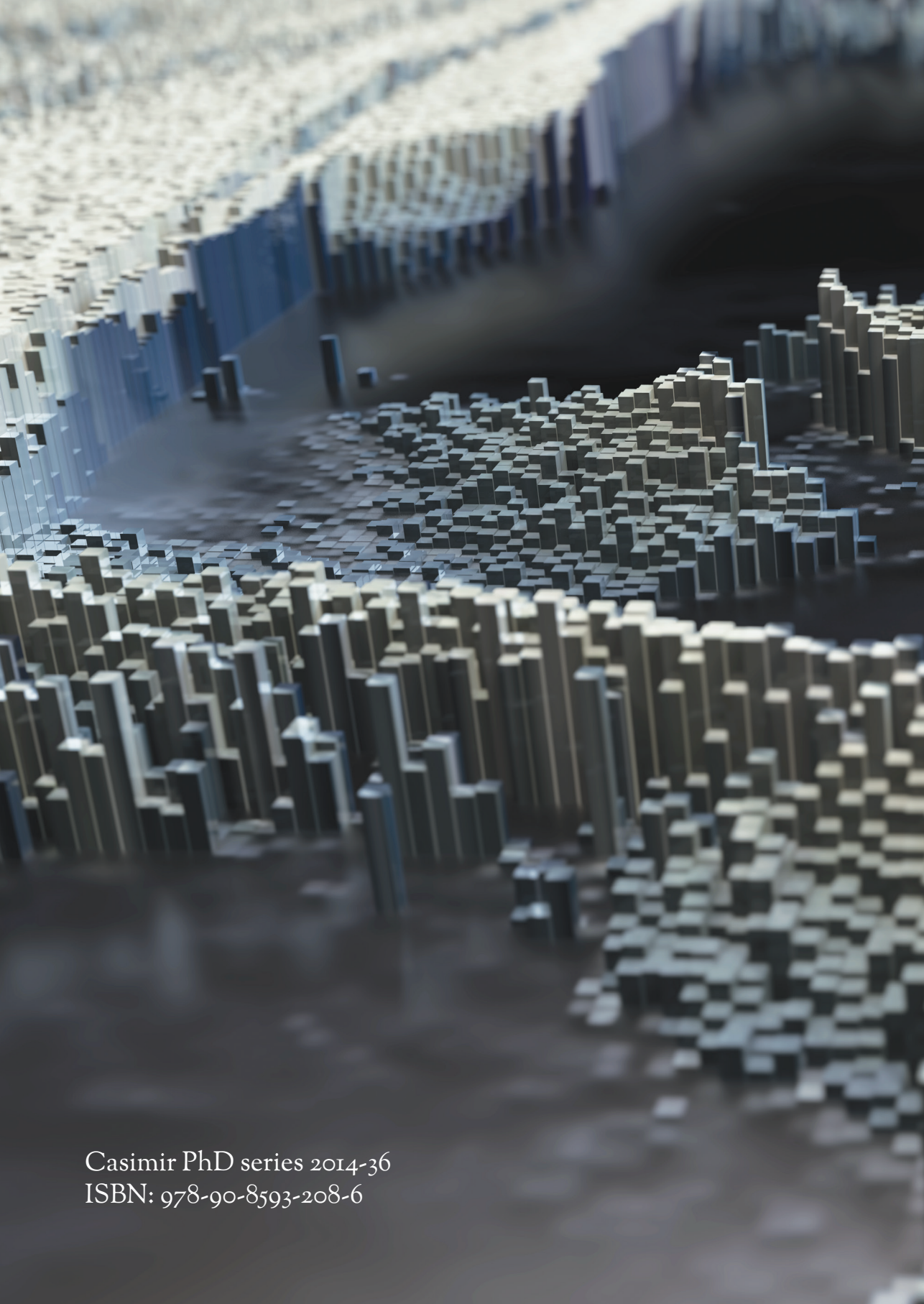
5. **C. Plesa**, S. W. Kowalczyk, R. Zinsmeester, A. Y. Grosberg, Y. Rabin, and C. Dekker, *Fast Translocation of Proteins through Solid State Nanopores*, Nano Letters **13**, 2, 658-663 (2013).

4. X. J.A. Janssen, M. P. Jonsson, **C. Plesa**, G. V. Soni, C. Dekker and N. H. Dekker, *Rapid manufacturing of low-noise membranes for nanopore sensors by trans-chip illumination lithography*, Nanotechnology **23**, 475302 (2012).

3. E. P. Lundberg, **C. Plesa**, L. M. Wilhelmsson, P. Lincoln, T. Brown, and B. Nordén, *Nanofabrication Yields. Hybridization and Click-Fixation of Polycyclic DNA Nanoassemblies*, ACS Nano **5**, 9, 7565-7575 (2011).

2. J. M. Dykes, **C. Plesa**, G. H. Chapman, *Enhancing direct-write laser control techniques for bimetallic grayscale photomasks*, Proc. of SPIE **6883**, 688312:1-12 (2008).

1. J. M. Dykes, **C. Plesa**, C. Choo, G. H. Chapman, *Bimetallic thermal resists potential for double-exposure immersion lithography and grayscale photomasks*, Proc. of SPIE **6730**, 673040:1-10 (2007).



Casimir PhD series 2014-36
ISBN: 978-90-8593-208-6