NANOSCALE CONTROL IN BIOLOGICAL AND SYNTHETIC SYSTEMS

BY

DAVID BAIER WELLS

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Physics in the Graduate College of the University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Robert M Clegg, Chair
Associate Professor Oleksii Aksimentiev, Advisor
Associate Professor Emadeddin Tajkhorshid
Professor John D Stack
Abstract

The ability to control a system is at the heart of experimental science. Modern experimental methods have pushed the length scale at which control is possible down to the nanometer level. Indeed, methods for manipulating single molecules have in some ways outstripped the ability to observe the systems under study. The “computational microscope” of the molecular dynamics (MD) simulation method provides insight into the behavior of systems at the atomic level, enabling the visualization of systems far beyond the limits of any experimental method. Moreover, MD facilitates experimentation with a virtually unlimited level of control. In this dissertation, I describe my efforts to enhance the ability to control MD simulations, and to use MD simulations to illuminate experimental systems. I have implemented and subsequently enhanced a flexible and powerful method for applying force in MD simulations, and have used this method to investigate microtubules and DNA translocation through the biological nanopore α-hemolysin. I have also employed MD simulations to explore methods for enhancing experimental control over the translocation of DNA through synthetic nanopores.
The uninterrupted and pervasive interaction of scientific discovery and industrial application has fructified both science and industry . . . the gist of scientific knowledge is control of natural energies. — John Dewey
Acknowledgments

First, I would like to thank my Doctoral Committee: Bob Clegg, Emad Tajkhorshid, John Stack, and Alek Aksimentiev. I would like to thank my advisor, Alek Aksimentiev, in particular. Your drive, ambition, and passion for your work are palpable, and I thank you for passing on some of this passion and work ethic.

The work in this dissertation is the result not just of my work, but of everyone I have worked with. I thank the developer team of Prof. Klaus Schulten’s group. I am especially grateful to Jim Philips and John Stone: You were both very helpful throughout my work on NAMD, and I benefitted greatly from your deep experiences, not only as they pertained to NAMD, but to more general issues of computation. My work on G-SMD would not have been possible without you. I thank Stefan Kowalczyk and Cees Dekker for an interesting and fruitful collaboration on the work presented in Chapter 4. I would also like to thank Dr. Kenneth H. Downing for providing the cryo-EM map of an MT used in Chapter 2.

I thank all current and former members of the Aksimentiev group. In particular, many thanks go to my “brothers in arms” [1]: Jeff Comer, Rogan Carr, and Chris Maffeo. We went through a lot of ups and downs together, and suffered through many an [LTX]RAC proposal. I will forever have fond memories of Jaded Babies, Inc. and TBGL Makefile ponies.

I must thank current and former members of my bowling team, the Urbana Achievers/Los Derechos: Nick Bronn, Paco Jain, Erik Bushey, Jen Hsin, Alice Quisno, Gene Basden, and Rogan Carr. You kept me sane over the years and got me through the strikes and gutters of grad school.

I thank my family for their support: my mother, Pam; my father, Neil; my brothers, Mark, Edward, and Alan; my sister, Dawn; and my girlfriend, Luana. I love you all very much! Without your constant support, I would not have been able to achieve what I have. I am particularly grateful to Luana, who has probably suffered the most during the past months as I’ve finished this dissertation.

The work in this dissertation was supported by grants from the National Institutes of Health (PHS 5 P41 RR05969 and R01-HG003713), the National Science Foundation (PHY0822613 and DMR-0955959), and the Department of Physics at the University of Illinois. I gladly acknowledge supercomputer time provided at the Pittsburgh Supercomputer Center, the National Center for Supercomputing Applications, and the
Texas Advanced Computing Center through TeraGrid/XSEDE Allocation grant MCA05S028, as well as time provided on the Turing and Taub Clusters (UIUC).
# Table of Contents

List of Abbreviations ................................................................. vii

Chapter 1 Introduction ........................................................... 1
  1.1 Control in simulations ...................................................... 2
  1.2 Grid-steered molecular dynamics ........................................ 3
  1.3 Control in experiment ...................................................... 5
  1.4 The future of simulation .................................................... 6

Chapter 2 Grid-steered Molecular Dynamics and microtubule mechanics .......... 7
  2.1 Introduction ................................................................. 7
  2.2 Results and Discussion .................................................... 9
  2.3 Conclusion ...................................................................... 23

Chapter 3 Grid-steered Molecular Dynamics and accelerating DNA translocation . 24
  3.1 Introduction ................................................................. 24
  3.2 Results .......................................................................... 26
  3.3 Discussion and Conclusions ............................................... 36

Chapter 4 Slowing down DNA translocation ........................................ 38
  4.1 Introduction ................................................................. 38
  4.2 Summary of Experimental Results ........................................ 39
  4.3 MD Results ..................................................................... 39
  4.4 Discussion ...................................................................... 45

Chapter 5 DNA in graphene nanopores .............................................. 47
  5.1 Introduction ................................................................. 47
  5.2 Results .......................................................................... 49
  5.3 Discussion ...................................................................... 59

Chapter 6 Other applications of Grid-steered Molecular Dynamics .................. 60
  6.1 Modeling biomolecules ...................................................... 60
  6.2 Thermal gradients ............................................................ 61

Appendix A Implementation of Grid-steered Molecular Dynamics .................... 63

Appendix B Methodology ................................................................ 65
  B.1 Grid-steered Molecular Dynamics and microtubule mechanics .................. 65
  B.2 Grid-steered Molecular Dynamics and accelerating DNA translocation .......... 70
  B.3 Slowing down DNA translocation ............................................ 73
  B.4 DNA in graphene nanopores ............................................... 74

Appendix C Publications .................................................................. 77

References ..................................................................................... 78
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHARMM</td>
<td>Chemistry at Harvard Macromolecular Mechanics, a force field used in molecular dynamics simulation.</td>
</tr>
<tr>
<td>cryo-EM</td>
<td>Cryo-electron microscopy, an experimental technique which can provide low- to medium-resolution (&gt; 5 Å) structures of proteins and other biomolecules.</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid, a biological polymer which is used to encode the genomes of all life on Earth.</td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double-stranded deoxyribonucleic acid.</td>
</tr>
<tr>
<td>G-SMD</td>
<td>Grid-Steered Molecular Dynamics, a simulation method in which a force calculated from a potential grid is applied to atoms.</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics, a simulation method which individually models atoms interacting according to a force field.</td>
</tr>
<tr>
<td>MT</td>
<td>Microtubule, a large protein filament found in cells.</td>
</tr>
<tr>
<td>N</td>
<td>Normal; used to refer to a system of two protofilaments in which lateral contacts are between like tubulin monomers, i.e. α-tubulin–α-tubulin and β-tubulin–β-tubulin. Compare S below.</td>
</tr>
<tr>
<td>NAMD</td>
<td>Nanoscale Molecular Dynamics, a computer program for molecular dynamics simulation.</td>
</tr>
<tr>
<td>PF</td>
<td>Protofilament, a number of which combine to form a microtubule.</td>
</tr>
<tr>
<td>S</td>
<td>Seam; used to refer to a system of two protofilaments in which lateral contacts are between unlike tubulin monomers, i.e. α-tubulin–β-tubulin. Compare N above.</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single-stranded deoxyribonucleic acid.</td>
</tr>
<tr>
<td>VMD</td>
<td>Visual Molecular Dynamics, a computer program for the visualization and analysis of molecular dynamics simulation data.</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The history of humankind is the story of the pursuit of control over nature. Four hundred thousand years ago, homo erectus first controlled fire [2], separating man from beast and beginning our long struggle to bend the universe to our will.

Science is a manifestation of this pursuit of control. First, many scientific pursuits have been motivated by a specific problem, be it disease, hunger, or war. Moreover, knowledge itself may be considered the ultimate form of control [3]. When one knows the laws that govern a phenomenon, one may predict and manipulate it, and therefore have power over it.

Control likewise plays a fundamental role in the mechanics of science. The scientific method may be expressed as: Observe, hypothesize, test, repeat. While our ability to hypothesize may arguably be considered constant, our capabilities vis-à-vis the other substantive steps, i.e. observe and test, have improved radically since the Enlightenment. The ability to introduce ever more precise perturbations to a system, and correspondingly to record ever more precise measurements of the system’s response, have driven the profound scientific and technological revolution still underway.

The biological sciences have arguably benefitted the most from these advances. In contrast to the (purposely) simple systems traditionally studied in the physical sciences, a biological system can span a huge range of time and length scales. Historically, without the ability to probe the nanometer length scales at which the basic mechanisms of life operate, biology has largely been a science of description and categorization. In the last decades, however, biology has undergone a quantitative revolution as experimental techniques have improved dramatically. With the advent of techniques such as atomic force microscopy (AFM), optical tweezers, fluorescence techniques such as Förster resonance energy transfer (FRET), electron microscopy, X-ray crystallography, nuclear magnetic resonance (NMR), and DNA sequencing efforts such as the Human Genome Project [4], the quantity of biological data now available is staggering, and is being produced at an ever-accelerating rate.

The information revolution of the last 30 years has seen the rise of a third prong to accompany those of observation and testing: computation. Exponential growth in computing capacity combined with the
development of sophisticated simulation methods have made simulation an important tool whose influence is rapidly growing. Simulation, in a sense, offers the ultimate level of control: within the limits of the underlying model, the perturbations applied are under essentially perfect control, and the resulting signal data is likewise essentially perfect.

One such simulation method is molecular dynamics (MD). In the MD method, atoms are modeled as charged point masses with van der Waals radii, and covalent bonds are essentially springs, with additional energy terms for the angles made between bonds and the twist of bonds (dihedral term.) The parameters defining the aforementioned interactions constitute the force field. A given force field, such as the CHARMM force field [5], is parameterized using experimental and/or quantum simulation data to best reproduce a set of biologically relevant observables, such as solvation energy [6]. Atomic positions are updated every timestep (usually 1 or 2 fs) according to the calculated forces and Newton’s laws of motion. Thus, the physics involved in MD is purely classical and, with appropriate approximations for non-bonded terms, the time to perform a simulation scales almost linearly in the number of atoms, allowing simulations of systems of biologically relevant size while still retaining atomic-level detail. The MD method is capable of simulating up to 100 million atoms [7] and up to a millisecond in duration [8, 9]. And because the position of every atom is available, MD can produce a dazzlingly detailed view of otherwise inscrutable systems, prompting MD to be referred to as a computational microscope [10].

1.1 Control in simulations

Using only the forces defined by the force field (and perhaps also a thermostat to control the temperature) a simulation moves toward thermodynamic equilibrium. As is the case for their experimental counterparts, however, simulation systems are often more interesting and useful when we poke at them. In MD simulations, this is accomplished via external forces.

A basic example of an external force is an external electric field. In an MD simulation, an external electric field is simply an extra force which is scaled by the charge of each atom:

\[ f_i = q_i E_{\text{external}} \]  \hspace{1cm} (1.1)

where \( f_i \) is the force on atom \( i \), \( q_i \) is the charge of atom \( i \), and \( E_{\text{external}} \) is the external electric field. Membrane simulations, in particular, make extensive use of external electric fields to mimic the transmembrane biases often used in experiment [11].

Another method for applying perturbations in molecular dynamics simulations is called Steered Molecular
Dynamics (SMD) [12, 13]. Like external electric fields, the method was designed to mimic experiment, in this case an AFM tip [12]. In the SMD method, a force (usually harmonic) is applied to either a single atom or a group of atoms. To steer the target atoms, the potential is moved, usually at constant velocity:

\[ f(t) = -k [x(t) - (x(0) + v_{SMD}t)] \] (1.2)

where \( f(t) \) is the applied force, \( x(t) \) is the target position at time \( t \), and \( v_{SMD} \) is the velocity of the potential. The force is akin to attaching one end of a spring to the target atoms and pulling the other. SMD was originally used to study the binding strength of biotin to streptavidin [12]. It has since been applied to a variety of problems including other binding pairs [14, 15], transmembrane permeation [16–18], unfolding of proteins [19–21], and others.

### 1.2 Grid-steered molecular dynamics

While external electric fields and SMD are extremely useful, the forms of the forces one can apply with them are very specific. Grid-steered molecular dynamics (G-SMD) is a much more general method of applying forces, which I have implemented and subsequently improved in the molecular dynamics program NAMD [22, 23].

In the G-SMD method, an arbitrary, three-dimensional potential is applied to the desired target atoms. As the name suggests, the potential is defined on a grid. The force on a particular atom \( i \) is determined by the gradient of the potential \( V \) and the coupling (charge) \( q_i \) of the atom:

\[ f_i = -q_i S \cdot \nabla V(x_i) \] (1.3)

where \( S \) is a diagonal scaling matrix.

In the original implementation, only one grid could be used in a given simulation, and the grid defining the potential \( V \) was required to be regular, i.e. the grid point spacing along each grid dimension was constant. Subsequent improvements have added support for multiple grids, multi-resolution grids, reloading of grids mid-simulation, and a faster, light-weight version called G-SMD Lite. Multiple grids allow the simultaneous forcing of different subsets of atoms by different potentials. Multi-resolution grids allows more efficient modeling of highly non-uniform potentials, particularly potentials with gradients spanning a wide range of values, e.g. a van der Waals potential. G-SMD Lite trades memory and accuracy for speed, and has smoothed the path for a graphics processing unit (GPU) implementation of G-SMD.
The flexibility of G-SMD has resulted in a wide range of applications. The G-SMD feature of NAMD has been used for transmembrane transport [22], molecular dynamics flexible fitting (MDFF) [24–27], DNA-nanopore interactions [28], fine-graining of coarse-grained simulation data [29], non-uniform electric fields [30, 31], exotic boundary conditions [32], and pressure-driven flow [33].

In Chapter 2, I describe the application of G-SMD to the largest cytoskeletal filament in the eukaryotic cell, the microtubule (MT). In that chapter, the fine control provided by G-SMD is used to construct an all-atom model of a complete MT by combining data from electron crystallography [34, 35] and cryo-electron microscopy (cryo-EM) [36]. The crystallographic data is high resolution, providing coordinates for all structured atoms in the tubulin dimer, and is suitable for MD simulation. However, due to the crystallization process, the data provide no information about the interaction of multiple dimers in the context of a complete MT. The cryo-EM data, in contrast, is lower resolution, lacking all-atom detail but showing the large-scale structure of an MT. By creating an energy potential grid derived from the cryo-EM data and applying that grid to all-atom tubulin dimers from the crystallography data, G-SMD allows us to get the best of both worlds: the all-atom detail of the crystal data, with the large-scale structure of the cryo-EM data. The resulting structure allowed me to conduct the first all-atom MD simulations of a microtubule, which I used to study the MT’s elastic properties. The flexible control afforded by G-SMD allowed the construction of a new model by combining previous models spanning multiple length scales, a capability which could be of tremendous importance in biology.

In Chapter 3, I describe the application of G-SMD to the transmembrane transport of DNA through the biological nanopore α-hemolysin, again highlighting the flexibility and power of the method. The unique features of α-hemolysin [37] have resulted in a tremendous amount of interest [38–63]. Of particular interest are the high stability of the pore, the fact that it self-assembles into a lipid bilayer membrane, and that its dimensions are such that single-stranded DNA (ssDNA) can fit through the pore, but double-stranded DNA (dsDNA) cannot. It is therefore very well-suited to the exciting and novel idea of nanopore sequencing [1]. In fact, α-hemolysin is the basis for the first commercial nanopore DNA sequencing platform [64].

MD simulation can provide an extremely detailed picture of the interaction between α-hemolysin, ssDNA, and ions [65]. However, the timescale for ssDNA translocation under experimental conditions (\( \sim 1 \mu s/base \)) is too long to collect statistically significant translocation data. Standard methods for accelerating the translocation process, such as increasing the electric field or using SMD, are shown to be inadequate. Using G-SMD, however, we are able to essentially enhance the driving electric field for the DNA only, which we show to radically accelerate the translocation process while maintaining a realistic conformation. Through the fine and flexible control of G-SMD, whole new classes of problems become amenable to MD.
1.3 Control in experiment

Control in simulation is tremendously useful. Ultimately, however, the information obtained in simulations must be fed back into experiment. Experiment is one step closer to practical application, and MD can provide extremely valuable insight into experimental systems.

In Chapter 4, I continue my work on DNA nanopore systems. Nanopore DNA sequencing promises a revolution in sequencing technology and medicine [1]. The principle of nanopore sequencing is straightforward: an external electric field is used to drive negatively charged DNA through a nanopore. As DNA transits the pore, its nucleotide sequence is determined by measuring changes in the nanopore ionic current. By reading the sequence directly from genomic DNA, nanopore methods promise to offer single-molecule, label-free DNA sequencing with virtually unlimited read lengths [66,67], overcoming many of the drawbacks of so-called next-generation sequencing platforms [68,69].

A persistent problem with the idea of DNA nanopore sequencing, however, is the speed of DNA translocation. Specifically, the DNA translocates too quickly to make base-specific measurements of the ionic current [70]. Translocation may be slowed by lowering the transmembrane potential, but this also reduces the ionic current signal, and so nothing is gained. Our experimental collaborators showed, however, that a simple change of ionic solution from the standard KCl to LiCl results in a dramatic reduction in DNA translocation speed without sacrificing the readout signal [70]. Using MD simulation, I investigate the microscopic mechanism behind this surprising discovery. The ability to reduce the DNA translocation speed so simply gives yet more control to the experimentalist, and may help lead to a profound change in our health and knowledge of ourselves.

In Chapter 5, I address another aspect of experimental control in DNA nanopore systems. Solid-state nanopores [71] offer a number of practical advantages over their biological counterparts. However, conventional synthetic membranes are generally over 10 nm thick, meaning the pore is occupied by many DNA bases at the same time. Thus, making single-base-sensitive measurements with such pores is extremely difficult [67]. Graphene [72] has recently been used as an alternative membrane material for DNA nanopore systems [73–76]. Graphene nanopores share many advantages of conventional solid-state nanopore systems while offering atomically precise control over the nanopore dimensions, in particular the membrane thickness, which can be made smaller than the distance between neighboring DNA nucleotides in a DNA strand. Using MD, I show that, with the proper choice of membrane thickness and pore geometry, the translocation dynamics of ssDNA are highly controlled, and that translocation occurs predominantly in single-base steps. Thus, we see that MD is a powerful tool to guide the experimentalist in the eternal quest for greater control.

Finally, in Chapter 6, I discuss ongoing work surrounding G-SMD. I briefly describe application of the
method to the modeling of biomolecules, and to the facilitation of thermal gradients in MD simulations.

1.4 The future of simulation

In this dissertation, I show a number of powerful aspects of MD and their contributions to greater nanoscale control of biological and synthetic systems. With the trajectory of computational power dictated by Moore’s Law, MD and other simulation techniques promise to be of ever greater influence. Thus, the major limitation will be the accuracy of the underlying mathematical model. While the detail provided by MD is great, the classical “balls and springs” model is a radical simplification of reality. Major progress has already been made on multiple fronts, notably the modeling of polarizability [77] and the inclusion of quantum effects. As simulation models become more and more realistic, it seems safe to say that the spectacular control afforded by them is bound to displace more traditional methods of inquiry.
Chapter 2

Grid-steered Molecular Dynamics and microtubule mechanics


2.1 Introduction

Microtubules (MTs) are ubiquitous biological filaments found in all eukaryotic cells. The largest type of cytoskeletal filament, MTs are critical to many cellular functions, including positioning of the centriole, providing a track for biological motors to enable organelle motility, forming the mitotic spindle during M-phase of eukaryotic cell division, and providing the mechanical core of eukaryotic flagella, cilia, and the axons of neuronal cells [78]. Because of their essential role in cellular function, especially during cell division, MTs are the target of many cancer treatments. By disrupting the MT structure, drugs such as taxol and vinblastine dramatically alter the dynamics of MT assembly and disassembly, either stabilizing (in the case of taxol) or destabilizing (vinblastine) the structure and leading to cell death [79]. Several applications of MTs in bionanotechnology have been suggested as well [80,81].

The basic building block of an MT is the αβ-tubulin dimer, shown in Figure 2.1 a. The dimers associate end-to-end, forming so-called protofilaments (PFs). Because of the heterogeneous nature of the αβ-tubulin dimer, PFs have a polarity. An MT, then, consists of a number of parallel PFs, and is itself polar, with a plus and minus end, as shown in Figure 2.1 b. In vivo, MTs are most often composed of 13 PFs, in which case the PFs are parallel to the MT axis. In vitro MTs, on the other hand, may comprise anywhere from 11 to 17 PFs [82].

The all-atom structure of αβ-tubulin was solved some time ago [34], and later refined [35], showing α- and β-tubulin to be structurally very similar to one another, and revealing the binding site of taxol, which was used to stabilize the structure during crystallization. The tubulin in these studies was crystallized in antiparallel sheets, rather than parallel tubes as in an MT. Thus, while the longitudinal bonds within the
Figure 2.1: Microscopic model of a tubulin dimer and a microtubule (MT). (a) The αβ-tubulin dimer. α-tubulin (bottom) and β-tubulin (top) are shown in cartoon representation, and GTP and GDP are shown in licorice representation. (b) All-atom model of an MT. The protofilaments (PFs) form α-β lateral contacts at the seam of the MT (indicated by the arrow) and α-α and β-β lateral contacts in the rest of the MT. These two types of the PF contacts are studied using two systems: N (normal) and S (seam), shown in the insets.

PFs presumably represent those found in MTs, the same cannot be said for the lateral bonds between PFs. These lateral bonds are thought to play a vital role in MT structure and dynamics, as highlighted by the position of taxol very near the lateral interface [35]. More recently, the structure of a complete MT was solved at 8 Å resolution using cryo-electron microscopy (cryo-EM) techniques [36]. While not of atomic resolution, this map is detailed enough to show the arrangement of the secondary structural elements of tubulin in MT form.

Due to their structural importance, the mechanical properties of MTs have been extensively studied [83–88]. However, because of the small size of MTs and their high rigidity, the experimental studies have mainly relied either on measurement of flexural rigidity using MT bending or buckling [83–85], or on some variation of radial indentation of the MT wall [86–88]. In both cases, the Young’s modulus is only arrived at indirectly, and depends on the choice of elastic model used during analysis. Indeed, estimates of the Young’s modulus have varied wildly, spanning almost four orders of magnitude [86,89]. More sophisticated, anisotropic elastic models have helped reconcile discrepancies in the reported moduli [89], but precise investigation of MT elasticity remains a daunting experimental challenge.

The molecular dynamics (MD) method allows the elastic properties of a biopolymer to be determined computationally [19,20]. Successful examples include systems such as titin [19], ankyrin [20], cadherin [20], actin [90], spectrin [91], collagen [21], and DNA [92]. Although elastic [89,93], molecular-mechanics [94] and finite-element [89] models can be used to interpret the results of experiments, relating the atomic structure of an MT to its mechanical properties requires an all-atom approach. There have already been a number of
all-atom studies that investigated the energetics of inter-PF interactions [95,96], the effect of GTP hydrolysis on intrinsic dimer bending [97], the effect of the drug taxol on tubulin flexibility [98], and the elastic moduli of individual tubulins [99]. Nevertheless, modeling the entire MT has not been (until now) accomplished because of the sheer size of the system and the lack of an atomic resolution structure.

In this chapter, using a variation of the molecular dynamics flexible fitting technique [24], we built an all-atom model of a complete MT by fitting the atomic model of αβ-tubulin into the cryo-EM map of a complete MT. We used the obtained structure to study the mechanical properties of a complete MT subject to longitudinal extension and compression, radial indentation and twist deformations. The calculated elastic moduli were found to be in agreement with available experimental data. This study is one of the first to utilize cryo-EM-fitted structures for all-atom molecular dynamics, demonstrating the usefulness of the technique and expanding the range of systems amenable to simulation using MD.

2.2 Results and Discussion

To build the atomic resolution model of a complete MT, we constructed two 2-PF systems, referred to as N (for normal) and S (for seam), that represented the two types of inter-PF interactions found in a complete MT, Figure 2.1b. Specifically, the N system features lateral bonds between like monomers, i.e. α-α and β-β bonds, while the S system features lateral bonds between unlike monomers, i.e. α-β and β-α bonds.

Construction and equilibration of the N and S systems

The procedure for construction of the N and S systems is outlined in Figure 2.2a. The crystal structure for αβ-tubulin was first rigidly fit to the cryo-EM density map to produce the initial structures. Flexible fitting was then performed to conform the atomic coordinates of the PFs to the cryo-EM density map of the complete MT. These steps are detailed in Section B.1.

Following flexible fitting, the N and S systems were equilibrated at constant pressure, with axial periodic dimension \( L_z \) fixed at the value suggested by the crystal structure and no external forces applied, for 15 ns (equilibration-I). We found the \( P_{zz} \) component of the stress tensor in both the N and S systems to be \( \sim 100 \) bar, much higher than other diagonal components \( P_{xx} \approx P_{yy} \approx 1 \) bar, indicating that the filaments were under compression. Therefore, we performed 2-ns axial stress adjustment simulations of the N and S systems, during which \( P_{zz}^{\text{target}} \), the Langevin piston target for the \( P_{zz} \) component, was linearly decreased from 100 to 1 bar in 1-bar increments. This was followed by 35-ns (N) and 20-ns (S) equilibration in the NpT ensemble (equilibration-II). The pressure and extension of the systems during the pressure adjustment...
simulation and the subsequent equilibration are shown in Figure 2.3. During the equilibration, \( L_z \) was observed to attain an equilibrium value of \( \sim 84 \, \text{Å} \) in both systems. Equilibration-III, a continuation of equilibration-I, was performed for 25 and 20 ns for the N and S systems, respectively. The structures obtained at the end of equilibration-III were used to build the complete MT system.

![Diagram](image)

Figure 2.2: Construction of the 2-PF systems. (a) Outline of the procedure. (b, c) N (b) and S (c) all-atom systems. Through periodic boundary conditions, the PFs are effectively infinite. Periodic images of the protein are shown translucently. \( \alpha \)-tubulin and \( \beta \)-tubulin are shown in cartoon representation, GTP and GDP in licorice representation, and ions as spheres. Boxes indicate the periodic cell. (d, e) Backbone RMSD of N (d) and S (e) system from the initial structures. Traces are shown for \( \alpha_1, \beta_1, \alpha_2, \) and \( \beta_2 \) as labeled. Vertical dashed lines indicate different simulation phases, from left to right: flexible fitting, equilibration with \( L_z \) fixed to the crystal value (equilibration-I), \( P_{zz} \) adjustment, and equilibration with \( L_z \) free to change (equilibration-II). (f, g) Backbone RMSD of N (f) and S (g) system during extension of equilibration with \( L_z \) fixed (equilibration-III.) The color scheme is the same as in (d) and (e).
Structural changes and stability of the N and S systems

Average root mean square deviation (RMSD) of backbone coordinates from the starting structure (crystal structure plus reconstructed loop) of the N and S systems during flexible fitting, equilibration-I, pressure adjustment, and equilibration-II are shown in Figure 2.2, d–e, while data for equilibration-III is shown in Figure 2.2, f–g. α-tubulin is seen to have a higher overall RMSD than β-tubulin, which is quite stable with a backbone RMSD of ∼2 Å. The difference between α- and β-tubulin can be attributed to the rebuilt residues 35–60 in α-tubulin (see Section B.1), as the RMSD plots become virtually identical when these residues are excluded from the calculation. Backbone RMSD values from equilibration-III are similar to those observed during the pressure adjustment and equilibration-II simulations.

The RMSD of individual residues at various stages of the simulation are shown in Figure 2.4. After flexible fitting, the N system showed only small changes other than in residues 35–60 of α-tubulin, while the S system displayed changes in the M-loop (residues 272–288) at the PFs’ interface. After flexible fitting, the M-loop displayed significant change (RMSD ≳ 5 Å), maintained in all subsequent simulations. Other regions of significant change were the H6-H7 loop (residues 216–224) at the longitudinal interface, the S9-S10 loop (residues 357–372) in the N system, and helix H4 in α-tubulin.

To determine if the fitting procedure created a stable structure, we computed the root mean square fluctuation (RMSF) of the interface residues, shown in Figure 2.5, a–b. The interfaces of both systems began with RMSF values of ≳ 1.0 Å, and attained lower values (≲ 0.9 Å) by the end of equilibration-II and equilibration-III. The low RMSF values imply a reasonably equilibrated, stable structure of the PF interface.
Figure 2.4: Per residue RMSD of the N and S protofilaments’ backbone atoms from the respective initial structures. The data for the N and S systems are shown in the left and right columns, respectively. The RMSD after each stage of the simulation is shown. The data is shown for residues 1–205 of \( \alpha_1 \) (solid black line) and \( \beta_1 \) (solid red line), and 206–439 of \( \alpha_2 \) (dashed black line) and 206–437 of \( \beta_2 \) (dashed red line). This set of residues is shown in red in (i). To compute the RMSD, the structure of each monomer was first aligned with the reference structure. For each residue, the RMSD was then calculated for backbone atoms averaged over the last ns of the trajectory.
The number of inter-PF contacts (shown in Figure 2.5, c–j), in general, was seen to increase through the course of the simulations. The contacts present within the first 1 ns of equilibration-I were approximately maintained, while new contacts formed in both the N and S systems.

Figure 2.5: Properties of the protofilament interface in the N and S systems. (a, b) Interface RMSF, calculated using Cα atoms of the interface residues for the N (black) and S (red) systems. To compute RMSF, a reference structure was determined for each 2.5-ns window as the time average after structural alignment. The RMSF within the window is then the RMSD of the Cα coordinates from that window’s reference structure averaged over the window. For these calculations, the interface residues were defined as those whose heavy atoms resided within 5 Å of the adjacent monomer. (c–j) Number of unique residue-residue contact pairs between α1-α2 (c, d) and β1-β2 (e, f) in the N system, and α1-β2 (g, h) and β1-α2 (i, j) in the S system. Two residues with heavy atoms within 5 Å of each other were considered to be in contact. Each contact plot shows the number of persistent (black), new (red), and total (green) number of contacts. The persistent contacts are defined as those present in the first 1 ns of equilibration-I. Vertical dotted lines indicate the beginning and end of pressure adjustment.

**Axial extension and compression of the N and S systems**

To determine the mechanical response of the N and S systems to axial extension and compression, we carried out a set of simulations in which $P_{zz}^{\text{target}}$ was changed gradually during the course of the simulation, producing continuous stress-strain curves, shown in Figure 2.6, a–b. The obtained stress-strain curves were found to exhibit some dependence on the rate at which $P_{zz}^{\text{target}}$ was changed, especially for positive strains (extension). Therefore, we carried out multiple simulations of both systems at different fixed $P_{zz}^{\text{target}}$ values, monitoring the change of the PF’s extension $L_z$, Figure 2.6 c, and the values of $P_{zz}$.

In Figure 2.7, a–b, we plot the stress and strain in the N and S systems during steady-state fragments of the simulations performed at constant $P_{zz}^{\text{target}}$ values. The curves exhibit interesting non-linearities, notably a marked softening for positive strains. For the N system, we also performed simulations with $L_z$ fixed, thus controlling strain, to verify the stress-controlled results. The starting configurations for the latter simulations
Figure 2.6: (a, b) Stress-strain curves from dynamic stress-controlled simulations of the N (a) and S (b) systems. For each system, the stress was changed at two different rates, 20 bar/ns (black) and 5 bar/ns (red). The 20 bar/ns simulation of the N system was performed starting from the protofilament structure that had not reached the equilibrium extension of 83.9 Å, hence the 20 bar/ns curve does not pass through the origin. (c–f) Raw data from the stress- and strain-controlled simulations of the N and S systems. (c) The $L_2$ length of the N system versus time at various applied stresses. The black curve corresponds to equilibration, i.e. no applied stress. Horizontal lines indicate the mean value attained at the end of each trajectory. (d) The $P_{z z}$ component of the stress tensor in the simulations of the N system versus time at several fixed values of $L_z$. Horizontal lines indicate the mean value attained at the end of each trajectory. (e) The $L_2$ length of the S system versus time at various applied stresses. Horizontal lines indicate the mean value attained at the end of each trajectory. (f) The $L_2$ length of the complete, infinite MT system versus time at various applied stresses. Horizontal lines indicate the mean value attained at the end of each trajectory. In the case of the +10 MPa simulation (blue), equilibrium was not reached within the simulation time scale.

were taken from the dynamically controlled stress simulations. The results of the strain-controlled simulations are shown as filled squares in Figure 2.7 a, and agree well with the stress-controlled data.
The stress-strain data for the N system in Figure 2.7a appears to fall into two or perhaps three distinct domains, each with a different Young’s modulus. For positive strain, or stretching, a linear fit yields a Young’s modulus of 0.3 GPa. For negative strain, or compression, the system is significantly stiffer. When fitting the $-0.01 < \epsilon \leq 0$ and $\epsilon < -0.01$ domains separately, we find Young’s moduli of 1.9 and 1.0 GPa, respectively, while a fit to the whole $\epsilon < 0$ domain yields a value of 1.4 GPa. These values are in broad agreement with experimental estimates [93], as well as MD simulations of the elastic properties of tubulin [99]. A fit to the entire range of strains, meanwhile, yields a Young’s modulus of 0.7 GPa, in very good agreement with the estimate of 0.6 GPa based on fitting to atomic force microscopy radial indentation data [88]. Data for the S system is displayed in Figure 2.7b and shows behavior similar to that of the N system, with calculated negative- and positive-strain Young’s moduli of 1.3 and 0.5 GPa, respectively.

Figure 2.7: Axial stretching and compression simulations. (a) Stress-strain data for the N system. Stress and strain were calculated every 5 ps. Each dot indicates the stress value calculated from the 5-ps average of $P_{zz}$ and the strain value calculated from the instantaneous value of $L_z$. Data were taken from the last 5 ns of the simulation performed at constant $P_{zz}^{\text{target}}$ value. The points are colored alternating shades to distinguish datasets corresponding to different $P_{zz}^{\text{target}}$ values. The mean stress–strain value of each 5 ns data set is shown as an open diamond. The filled squares indicate the mean stress observed in the constant strain simulations. (b) Stress-strain data for the S system. The symbols have the same meaning as in (a). (c) Stress-strain data for the complete MT. The symbols have the same meaning as in (a), but the stress-strain values were taken from the last 2 ns of the simulations at each $P_{zz}^{\text{target}}$ value. The simulation at $P_{zz}^{\text{target}} \approx 10$ MPa did not reach an equilibrium strain within the simulation time of 12 ns.

In order to disentangle monomer deformation and monomer separation in the stretching and compression simulations, we examined the strains of the individual monomers, shown in Figure 2.8. The monomer strain analysis showed that negative strains involve more deformation of the monomers compared to positive strains. This should be expected since the monomers are already in close contact at zero strain. There is a noticeable asymmetry between $\alpha$- and $\beta$-tubulin: $\alpha$-tubulin tends to deform more than $\beta$-tubulin, especially for positive strains. These results are consistent with previous findings showing $\beta$-tubulin to be axially stiffer than $\alpha$-tubulin [99].
Figure 2.8: Monomer strain versus system strain for the N (a–d) and S (e–h) systems. Data were taken from the stress-controlled simulations. The dashed lines in each plot are of slope 1. If the data points lie along this line, then the individual monomers are stretching and compressing to the same degree as the filaments as a whole, implying that the strain of the system is entirely due to strain of the monomers. On the other hand, if the data points fall on the x-axis, then the monomers are not stretching or compressing at all, implying that the strain of the system is due to relative motion of the unstrained monomers, e.g., separation of monomers during stretching. Monomer strain was determined by first fitting select $C^\alpha$ atoms to a reference structure, then calculating the linear regression of the scatter plot $\Delta z_i = z_i - z_i^{\text{ref}}$ versus $z_i^{\text{ref}}$, where $z_i$ and $z_i^{\text{ref}}$ are the $z$-coordinates of the current and reference atom $i$, respectively. The slope calculated is then the strain of the individual monomer. Reference structures were taken as the average of the last 5 ns of equilibration. The residues selected for these calculations had either alpha-helical or beta-sheet secondary structure in the reference coordinates as determined by STRIDE [100].
Construction and equilibration of the MT system

The procedure for constructing the atomic scale model of a complete MT is outlined schematically in Figure 2.9 a. Using the equilibrated conformations of the PFs from the N and S systems, the initial model of the MT system was produced by placing the PFs according to the MT’s geometry. Using harmonic restraints, the interface between the PFs in the MT model was adjusted to reproduce the PFs’ contacts observed in the equilibrated N and S systems. These steps are described in detail in Section B.1.

Following the interface fitting procedure, the system was equilibrated for 14.24 ns in the NpT ensemble with $L_z$ fixed to the crystal value (equilibration-I). As in the case of the 2-PF systems, the $P_{zz}$ component of the stress tensor in the simulation of the MT system was $\sim$100 bar. Using anisotropic pressure control, $P_{zz}^{\text{target}}$ of the Langevin piston was changed from 100 to 1 bar over 5 ns in 1 bar increments, followed by 8.0-ns equilibration with $P_{zz}^{\text{target}}$ set to 1 bar (equilibration-II). Figure 2.9 d shows $P_{zz}$ and $L_z$ during the pressure adjustment and equilibration simulations.

To examine the structural changes during and after the interface fitting procedure, the RMSD of the interface atoms in the MT structure were computed relative to the corresponding coordinates in the equilibrated N and S systems, Figure 2.9 c. For these RMSD plots, atoms within a $2\pi/26$-radian sector centered between PFs were defined as interface atoms, see Figure 2.9 b. The graph shows that the interfaces maintain RMSDs in a tight range below their starting value, indicating that the interface fitting procedure succeeded in producing a stable MT structure. The examined atoms exclude residues 35–60, as inspection of the trajectory showed that a portion of one of the reconstructed loop in $\alpha$-tubulin shifted at $t \sim 19$ ns.

Axial extension and compression of the MT system

Using the anisotropic pressure controls method, we examined the mechanical response of a complete MT to axial compression and extension. Simulations were carried out at four $P_{zz}^{\text{target}}$ values corresponding to axial stress of approximately 10, 5, −5, and −15 MPa.

The stress-strain curve for the complete MT is shown in Figure 2.7 c. The plot shows stress-strain data calculated every 5 ps, as well as averages for each, with the exception of the 10 MPa simulation which had a rising strain value that did not reach a steady state; Figure 2.6 f shows the values of $L_z$ for each simulation. The shape of the stress-strain curve is similar to that seen for the N and S systems, notably the softening seen for positive strains. A linear fit to the averages yields a Young’s modulus of 1.2 GPa, in good agreement with that found for the N and S systems above. It should be mentioned, however, that the system demonstrated very slow relaxation in $L_z$ during equilibration, as seen in Figure 2.9 d, raising the possibility that the stress-controlled simulations did not reach genuine steady states.
Figure 2.9: Construction the MT system. (a) Outline of the procedure. (b) A view of the MT system from the +z direction. The MT structure is shaded according to the proximity of its atoms to the PFs’ interface: the darker the shading, the closer the atom is to the interface. During the interface fitting procedure, atoms shown in black were subject to maximum restraints, while atoms shown in white were unrestrained (see Section B.1). (c) RMSD of the interface’s backbone atoms from their target coordinates during interface fitting, equilibration with $L_z$ fixed, pressure adjustment, and free equilibration simulations, excluding residues 35 to 60 of α-tubulin. Each trace shows the RMSD value of a single interface (13 traces total are shown), defined as a $2\pi/26$-radian sector in (b). The vertical dotted lines in the interface fitting part indicate simulations carried out using different restraint scaling factors (see Section B.1). (d) $L_z$ (solid line) and $P_{zz}$ (dashed line) of the MT system during pressure adjustment and equilibration simulations. The scales for $L_z$ and $P_{zz}$ are shown on the right and left vertical axes, respectively.

Stretching of the MT was found to reduce its radius, as intuition may suggest. For the case of $-20$ MPa stress, we found the average MT radius (measured to the PF centers of mass) to be $111.57 \pm 0.07$ Å, compared to $110.23 \pm 0.07$ Å for the +5 MPa stress simulation. At zero stress, the MT radius was found to be $110.45 \pm 0.06$ Å.
Radial compression of the MT system

To probe the mechanical behavior of a complete MT under radial compression, two forcing surfaces were moved toward one another and the center of the MT at a constant velocity (see Section B.1). Because of the asymmetry of the MT, two compression simulations were performed differing in the directions of the compression force: parallel and perpendicular to the radial position vector of the seam. These simulations are referred to as $S_{\parallel}$ and $S_{\perp}$, respectively.

The strain introduced by the compression force was calculated by fitting an ellipse to the $xy$ coordinates of the PFs’ centers of mass for each trajectory frame. The fit was performed under the following constraints: the ellipse’s center had the same coordinates as at the beginning of the compression simulation; and the major axis of the ellipse was always directed parallel to the forcing surfaces. The initial length of the minor axis ($L_{\text{minor}}$) was taken as a zero strain reference ($\epsilon = 0$), while the strain in the rest of the trajectory was calculated as $\epsilon = (L_{\text{minor,0}} - L_{\text{minor}})/L_{\text{minor,0}}$. Note that this is the negative of the standard definition.

In Figure 2.10a, the average compression force is shown as a function of the strain for both $S_{\perp}$ and $S_{\parallel}$ systems. The inset images show the $S_{\perp}$ and $S_{\parallel}$ systems from the $+z$ direction at a strain of 0.13. The blue slabs represent the forcing surfaces. The simulated MT systems were effectively infinite along the $z$ direction.

The simulated force versus strain dependences have similar shape for both $S_{\perp}$ and $S_{\parallel}$ systems. Both systems exhibited an initial high resistance to deformation, to strains of $\sim 0.01$, followed by a softer response. The regime past $\epsilon \sim 0.2$ ceases to increase monotonically, displaying steps in the response curve for both systems. We hypothesized that the change in resistance to deformation at $\epsilon \sim 0.01$ represents a transition from elastic to plastic deformation. To investigate this possibility, we performed two equilibrium simulations, starting from $\epsilon \approx 0.005$ and 0.055, i.e. before and after the transition, of the $S_{\parallel}$ system. Results of these simulations are shown in the inset of Figure 2.10, which shows strain as a function of time. The first simulation is seen to return to $\epsilon \approx 0$. The second simulation, on the other hand, displays only a slight return toward zero strain, remaining above 0.05, supporting the idea of plastic deformation.

In experiment, the radial indentation of an MT using AFM demonstrated reversible deformation up to 15% deformation [88]. Calculating the effective spring constant of our MT for the strains $0.05 < \epsilon < 0.15$, we find $k \sim 0.3$ N/m, somewhat higher than the value of $\sim 0.07$ N/m found in [88]. However, the deformation mode from a finite AFM tip differs from the compression applied in our simulations, which effectively indents the MT along its entire length. Given this difference, and the factor of $\sim 10^7$ higher loading rate in simulation versus experiment, the approximate agreement in $k$ is satisfying.

We also investigated the deformation mode during compression. Backbone RMSD of the individual PFs
Figure 2.10: Radial compression of a complete, infinite MT. (a) Compression force versus MT strain for the $S^\perp$ (pluses) and $S^\parallel$ (triangles) systems. The average force from the two surfaces is shown. Each point represents a 25-ps-window average. Inset images show the systems from the $+z$ direction at $\epsilon = 0.13$. The seam is at the bottom of the images. (b) Relaxation simulations of the $S^\parallel$ system. The solid trace shows a fragment of the compression simulation, whereas the dashed traces show two equilibration simulations started from different points along the compression trajectory: after 0.6 and 2.0 ns. (c) Center-of-mass circumference of the MT versus strain. This circumference is defined as the sum of the lengths of the line segments connecting the PF centers-of-mass in the $xy$ plane. Symbols are the same as those in (a).

was very low throughout the entire process, remaining below 1.7 Å. Therefore, the deformation is primarily composed of hinge-like motion between PFs, and relative motion of PFs. Analysis of the deformation revealed a constant MT circumference, defined as the sum of straight-line segments between PF centers-of-mass in the $xy$ plane, to strains of $\sim 0.05$, after which it began to rise. This is shown in Figure 2.10 c. This may imply a larger elastic regime more consistent with experiment.

**Twist deformation of the MT system**

To examine the mechanical response of an MT to twist deformation, we constructed a finite-length model of the MT (see Section B.1) that contained three repeats of the $\alpha\beta$-tubulin dimer per protofilament, shown in Figure 2.11 a. Forces were applied to the terminal monomers of the MT, to produce displacement of the MT’s plus and minus ends in opposite circumferential directions (see Section B.1).

Figure 2.11 b plots the total displacement angle versus simulation time for counterclockwise (CCW) and clockwise (CW) twist deformation. The total torque imparted in both cases was $\sim 89$ nN nm, increasing
Figure 2.11: Twist simulations. (a) Finite-length MT system, at the beginning of the twist simulations. (b) Total angular displacement of the microtubule’s terminal monomers in the simulations of the counter-clockwise (solid line) and clockwise (dashed line) twist deformations, as seen from +z direction. The arrows indicate the moment when the twisting forces were set to zero and the structures were allowed to relax.

slightly (<3%) during the course of the simulations due to increased MT radius (see Figure 2.12 a). For deformations <4°, both directions of the torque produced twists of comparable value. Thereafter, twisting the MT clockwise produced a greater twist strain than counterclockwise. After approximately 4 ns of torque application, the forces were switched off, allowing the system to relax. The twist strain was observed to decrease toward zero, although complete recovery of the initial structure was not observed within the time scale of our simulations. The length of the MT fragment under the twist deformation displayed similar behavior in both CCW and CW simulations: the length remained nearly constant, shrinking by 0.5% at high strains in the CW simulation, Figure 2.12 c.

If modeled as an isotropically elastic tube with outer and inner radii of $r_o = 125$ Å and $r_i = 75$ Å, respectively, a torque of $\tau = 89$ nN nm is expected to produce an angular strain of $\Delta \theta \lesssim 8^\circ$, using $\Delta \theta = \tau L/J G$, where $L$ is the length, $J = \frac{\pi}{2}(r_o^4 - r_i^4)$ is the torsion constant, and $G = E/2(1 + \nu)$ is the shear modulus. We have taken the Young’s modulus $E = 1.2$ GPa, and the Poisson ratio $\nu = 0.5$ to set an upper bound on the predicted strain. This prediction is clearly exceeded in Figure 2.11 b, demonstrating the inadequacy of an isotropic treatment of MT elasticity using a thick cylindrical tube model. An MT is softer than predicted by the isotropic tube model, and displays an asymmetry between the two directions of twist deformation.
Figure 2.12: Torque and length of the MT during the twist simulations. (a) Total torque applied during the simulation of CCW (black) and CW (red) twist. (b) Microtubule system used for twist simulations. While most protofilaments have six lateral monomer contacts on both sides, some have only five on one side due to the finiteness of the system. Protofilaments having fewer than six lateral contacts are shown in gray and were not used to compute the average lengths shown in (c) and (d). (c) Distance along the z-axis between the centers of mass of the two end tubulin monomers in the twist simulations, averaged over protofilaments that had six lateral contacts on each side, shown in red in (b). (d) Distance along the z-axis between the centers of mass of the two end tubulin monomers of each protofilament during equilibration, averaged over the same protofilaments.
2.3 Conclusion

In this study, we combined structural data at multiple resolutions to construct an all-atom model of a complete MT. This model should not be considered a substitute for a crystallographic model, and most likely contains small errors, but nevertheless performed very well. Using this model, we were able to determine a detailed stress-strain dependence clearly showing different behavior under extension and compression. Furthermore, the computed Young’s moduli were in very good agreement with published data and estimates. We studied the properties of an MT under radial compression and identified a possible elastic-to-plastic transition. Finally, we applied twisting force and showed a dramatic asymmetry in the MT reaction to the two senses of twist. The twist simulations in particular demonstrate the range of possibilities afforded by the MD method, allowing the computation of quantities well beyond the current reach of experiment. The excellent agreement of the axial strain simulations with available experimental data gives confidence in the other quantities for which experimental data are not available.

This work is one of the first to combine cryo-EM and crystal structures for subsequent all-atom MD simulation. The successful performance of such a model opens the door to the simulation of many other systems whose constituent units are known in atomic detail but whose complete structure is known only at lower resolution. Such simulations allow more biologically relevant length scales to be probed, promising to expand our knowledge of the microscopic mechanisms of life.
Chapter 3

Grid-steered Molecular Dynamics and accelerating DNA translocation


3.1 Introduction

Alpha-hemolysin is a bacterial toxin that self-assembles in a lipid membrane to form a water-filled transmembrane pore [37]. Applying a transmembrane electric potential, solutes of various molecular weights can be admitted into the channel, including rather long (up to 1300 nucleotides) single DNA and RNA strands [38, 39]. The presence of a solute in the channel reduces the transmembrane ionic current below the open pore level, which can be used to identify the solute’s type and concentration [41]. Such ionic current blockades in α-hemolysin have been used to detect small organic analytes [40], metal ions [42], drugs [52], peptides [50], DNA hairpins [44], modified DNA strands [48], proteins [45, 47], and to detect rupture events in nanopore force spectroscopy experiments [54, 55]. In the case of DNA and RNA strands, the level of the ionic current blockades depends on the sequence of the nucleotide fragment confined in the pore constriction [38, 43, 49, 51], which in principle can be used to create a device for high-throughput DNA sequencing [46, 53].

Although α-hemolysin has been successfully employed as a stochastic sensor in a number of studies, the ionic current recordings have limited utility, as the level of the ionic current blockade is determined not only by the atomic-detail structure of the solute, but also by the solute’s microscopic conformation. Hence, to unambiguously interpret the ionic current recordings, the microscopic conformation of the solute in α-hemolysin has to be determined to atomic resolution. Direct experimental visualization of solute conformation in α-hemolysin is not currently possible.

Molecular Dynamics (MD) is a computational method that can provide a realistic account of a transmembrane permeation event [16, 101–110]. In this method, a biomolecular system is represented by an ensemble
of particles that move and interact according to the laws of classical mechanics [111]. Recently, this method was employed to relate the microscopic conformations of DNA in κ-hemolysin to the measured ionic current blockades, revealing that DNA’s global orientation is a significant factor determining the current blockade [112]. Due to the recent dramatic advances in computational technology [23], the ionic conductance of κ-hemolysin can now be accurately computed from all-atom MD simulations [65]. Nevertheless, it is still not feasible to simulate using the “brute-force” approach to the permeation of larger solutes, such as DNA and proteins, since typical permeation times of such solutes range from tens of microseconds to milliseconds.

In this chapter, we present a computational method that, in tens of nanoseconds, can provide a realistic account of a permeation event that would require a millisecond simulation using conventional MD.

One of the methods developed to overcome the timescale limitation of MD is Steered Molecular Dynamics (SMD) [12,13]. In this method, an external force is applied to a solute to facilitate the crossing of free-energy barriers in the channel, thus reducing the permeation timescale. The applied force is usually implemented by attaching a moving harmonic (spring-like) restraint to one or more atoms in the system [12,13,113]. The outcome of many repetitive SMD simulations can be related to the profile of the Potential of Mean Force in the channel through Jarzinsky’s identity [114, 115]. Coarser-scale simulations of membrane transport, although able to describe the general features of the permeation process [116–120], by their very design cannot reveal a permeation mechanism that is sensitive to atomic-scale details.

Although SMD is a popular method to investigate permeation of small solutes through membrane channels [16–18], the application of conventional SMD is troublesome in the case of long linear biopolymers, such as proteins, nucleic acids, or any other biomolecules that are or may transiently become structurally disordered during the permeation process. If a steering force is applied to a terminal of such a solute, this force is likely to distort the solute’s conformation, e.g., overstretch DNA or unfold a protein. An example of such an SMD simulation is shown in Figure 3.1, a–d. Because the timescale of the SMD simulation is too short for relaxation forces to act in response to the strain produced by the steering force, the solute enters the channel in a distorted, physiologically unfeasible conformation. Applying the SMD force to the center of mass of the solute produces similar distortions, as shown in Figure 3.1, e–h.

An alternative method for accelerating the permeation processes is to scale up the transmembrane potential that drives the solutes across the cell membrane. This method was recently employed to simulate the permeation of ions and DNA through biological and artificial membranes [65,112,121]. However, the maximum transmembrane bias that can be applied in such simulations is limited by the stability of the membrane, as a lipid bilayer in an MD simulation is prone to breaking when the transmembrane bias exceeds \(\sim 2\) V. This leads to ion leakage, which subsequently distorts the driving potential and thereby results...
in a simulation of an unrealistic permeation event. Applying a uniform external electrostatic field only to
the solute’s atoms fails to produce a successful translocation, as illustrated in Figure 3.1, i–l.

Here we present a generalization of the SMD methodology tailored to simulations of membrane potential-
driven transport of large solutes. In order to simulate a permeation event, we amplify the three-dimensional
(3-D) electrostatic potential derived from an all-atom MD simulation [65] and apply it to the permeating
solute only. By doing so we facilitate faster translocation of solutes without straining the structure of the
channel or the membrane. Moreover, as the shape of the potential applied to the solutes in our method
exactly matches the transmembrane potential driving the solutes in experiment, the field of forces is very
realistic, and because the forces are distributed over the entire solute, the strain introduced into the solute’s
conformation is much smaller than that introduced by conventional SMD. An example of such a simulation
is shown in Figure 3.1, m–p.

In this chapter, we apply the Grid-SMD (G-SMD) method to simulate the translocation of DNA through
\( \alpha \)-hemolysin. First, we investigate how the translocation rate of a DNA strand depends on the magnitude
of the applied potential and examine the strain introduced into the DNA conformation. We then compare
the results of G-SMD simulations to experiment by measuring the translocation velocity of DNA strands of
different sequences and global orientations.

3.2 Results

After the systems comprising DNA, \( \alpha \)-hemolysin, lipid bilayer membrane, water and ions were assembled
and equilibrated (see Section B.2), external fields of different magnitudes were applied to the DNA, and the
resulting displacements of the DNA in the channel were recorded. All simulations were carried out applying
an electric field equivalent to a 1.2 V transmembrane bias to all atoms in the system [65, 121, 122]. In
addition, the \( z \)-component of the forces derived from the average distribution of the electrostatic potential
were multiplied by the scaling factor \( N \) and applied to DNA atoms only. The steering forces were not applied
in the \( x \)- and \( y \)-directions. Table 3.1 summarizes the results of all MD simulations performed.

First, we investigate how the introduction of a steering potential influences the velocity of DNA translo-
cation. In Figure 3.2 we plot the cumulative number of permeated nucleotides against the simulation time
for four G-SMD runs in which the steering field was applied with the scaling factor \( N \) of 1, 3, 7 and 10. In
the same plot we display the results of a simulation without a steering field, i.e., \( N = 0 \) (a 1.2 V transmem-
brane bias was applied in all simulations). The DNA was observed to permeate the channel faster at larger
steering fields. In the inset to Figure 3.2 we plot the average DNA translocation velocity against the effective
Figure 3.1: SMD simulations of DNA translocation through α-hemolysin. (a–d) SMD force applied to the phosphorous atom of the first nucleotide of the DNA strand. The DNA stretches as it traverses through the pore constriction. This 3 ns simulation was done using an SMD pulling velocity of 43 Å/ns and a spring constant of 500 kcal/mol·Å². (e–h) SMD force applied to the center of mass of the DNA strand. (i–l) DNA permeation driven by a uniform electric field. This simulation was performed by applying constant forces to individual DNA atoms. The magnitude of each force was computed as a product of the atomic charge and the electric field equivalent to a 12 V transmembrane bias. The total simulation time is 4 ns. (m–p) G-SMD simulation of DNA translocation. The transmembrane bias was set to 1.2 V with the scaling factor of $N = 10$ (see Section B.2). The total simulation time is 5 ns.
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Effective bias (V)</th>
<th>Translocation velocity (nuc./ns)</th>
<th>Simulation time (ns)</th>
<th>Translocated nucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>(cis) 3’-dA58-5’ (trans)</td>
<td>1.2</td>
<td>0.021</td>
<td>27.8</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>0.101</td>
<td>19.6</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>4.8</td>
<td>1.10</td>
<td>9.3</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td>9.6</td>
<td>6.94</td>
<td>3.6</td>
<td>24.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.75</td>
<td>3.6</td>
<td>21.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.46</td>
<td>3.3</td>
<td>21.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.21</td>
<td>3.5</td>
<td>21.9</td>
</tr>
<tr>
<td></td>
<td>13.2</td>
<td>9.07</td>
<td>1.2</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.96</td>
<td>1.2</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.13</td>
<td>1.2</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.56</td>
<td>1.1</td>
<td>9.0</td>
</tr>
<tr>
<td>(cis) 5’-dA58-3’ (trans)</td>
<td>1.2</td>
<td>0.071</td>
<td>29.9</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>0.091</td>
<td>18.8</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>4.8</td>
<td>1.25</td>
<td>8.9</td>
<td>10.7</td>
</tr>
<tr>
<td></td>
<td>9.6</td>
<td>6.84</td>
<td>3.4</td>
<td>22.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.66</td>
<td>3.0</td>
<td>22.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.07</td>
<td>2.7</td>
<td>21.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.70</td>
<td>2.8</td>
<td>20.9</td>
</tr>
<tr>
<td></td>
<td>13.2</td>
<td>13.5</td>
<td>1.1</td>
<td>14.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.6</td>
<td>1.2</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.05</td>
<td>1.2</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.5</td>
<td>1.4</td>
<td>14.7</td>
</tr>
<tr>
<td>(cis) 3’-dC58-5’ (trans)</td>
<td>4.8</td>
<td>1.32</td>
<td>9.8</td>
<td>11.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.891</td>
<td>6.6</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>9.6</td>
<td>6.48</td>
<td>4.0</td>
<td>25.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.38</td>
<td>4.0</td>
<td>25.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.89</td>
<td>3.8</td>
<td>30.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.26</td>
<td>4.0</td>
<td>27.9</td>
</tr>
<tr>
<td>(cis) 5’-dC58-3’ (trans)</td>
<td>4.8</td>
<td>1.73</td>
<td>10.8</td>
<td>18.4</td>
</tr>
<tr>
<td></td>
<td>9.6</td>
<td>13.5</td>
<td>1.5</td>
<td>19.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12.2</td>
<td>1.9</td>
<td>20.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12.1</td>
<td>1.6</td>
<td>18.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.6</td>
<td>1.6</td>
<td>17.3</td>
</tr>
<tr>
<td>(cis) 5’-(dAdC)29-3’ (trans)</td>
<td>9.6</td>
<td>10.3</td>
<td>2.2</td>
<td>22.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.25</td>
<td>2.2</td>
<td>20.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.68</td>
<td>2.3</td>
<td>20.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.38</td>
<td>2.2</td>
<td>19.0</td>
</tr>
</tbody>
</table>

Table 3.1: G-SMD simulations of DNA tranlocation through α-hemolysin. The effective bias is computed as \((N + 1) \times 1.2\) V, where \(N\) is the scaling factor applied to the steering potential. The DNA translocation velocities were computed by applying a linear regression fit to the cumulative DNA currents.
bias applied, i.e., \((N + 1) \times 1.2\) V. The DNA translocation velocity was observed to scale superlinearly with the effective bias until \(N = 3\). Note that at \(N = 3\), the translocation velocity is about 1 nucleotide per nanosecond, approximately 1000 times faster than in experiment.

Figure 3.2: Dependence of DNA translocation velocity on steering potential. The total number of nucleotides translocated is plotted against the simulation time. Black triangles, white triangles, diamonds, and squares indicate simulations carried out at a 1.2 V electrostatic field magnified 11, 8, 4 and 2 times, respectively; the circles correspond to the simulation carried out at a 1.2 V bias (no steering force applied). (Inset) Tranlocation velocity versus the effective bias. These simulations were carried out using a poly(dA)_{58} strand in the A5' orientation (5' end of DNA on the trans side; see Figure 3.3a).

In order to assess the conformational strain introduced by our method, we analyzed the number of nucleotides in the vestibule, constriction and stem (see Figure 3.3a) as a function of time. The results of this analysis are shown in Figure 3.4. We found that for \(N = 1\) or 3, the number of nucleotides in the stem remained nearly constant, indicating that virtually no conformational strain was introduced (the rising number of nucleotides in the vestibule results from DNA relaxation; compare the \(N = 0\) case.) The strain remained modest for higher values of \(N\). We thus conclude that \(N = 3\), i.e., an effective bias of 4.8 V, is optimal for G-SMD simulations, as it does not distort the conformation of the DNA in \(\alpha\)-hemolysin while still substantially accelerating its movement. In all simulations, the hydrophobic stacking of the DNA bases is broken in the constriction, as it is too narrow for the DNA strand to fit in in the base-stacked conformation. In the stem and in the vestibule the DNA strand is split into groups of 2–4 nucleotides that preserve the base-stacking pattern. The structure of the lipid bilayer and of the protein is unaffected by the steering potential, as it is applied to atoms of DNA only.

We next investigate whether G-SMD can provide quantitative insights into the mechanics of DNA translocation, despite the dramatic reduction of the translocation timescale. It has been experimentally shown that the global orientation of a DNA strand in the \(\alpha\)-hemolysin channel has a deterministic effect on the velocity
Figure 3.3: The set-up of G-SMD simulations. (a) Microscopic model of the α-hemolysin/DNA system suspended in a lipid bilayer membrane. The DNA atoms are drawn as orange spheres. The channel is drawn as a molecular surface (violet) separating the protein from the membrane and water. This surface is cut by a plane perpendicular to the lipid bilayer, passing through the geometrical center of the protein. The DPPC lipid bilayer is shown in green, water and ions are not shown. The stem and vestibule sections of the protein are indicated; the constriction, the narrowest part of the channel, lies between them. The model comprises of 356,065 atoms. (b) The electrostatic potential map of α-hemolysin. The figure shows a cut through the averaged (over a 5.3 ns MD simulation and the sevenfold symmetry of α-hemolysin) electrostatic potential along the z-axis. A 1.2 V transmembrane bias was applied in this simulation. This 3-D potential is used in our G-SMD simulations to drive translocation of DNA through α-hemolysin.

Figure 3.4: Influence of the driving potential on the conformation of DNA in α-hemolysin. The number of nucleotides in the vestibule (a) and in the stem (b) of the channel are shown as functions of the simulation time. The symbols indicate the scaling of the driving potential: $N = 0$ (circles), $N = 1$ (squares), $N = 3$ (diamonds), $N = 7$ (white triangles) and $N = 10$ (black triangles). Each trace for $N = 7$ or 10 is an average over four trajectories. These simulations correspond to DNA in the A5’ orientation (5’ end of DNA on the trans side; see Figure 3.3 a).
of DNA translocation: when a single DNA strand of adenine nucleotides (poly(dA) strand) is driven by a transmembrane bias [112], the translocation velocity of the DNA strand in the direction of its 3’-end is about 16% higher than that in the direction of its 5’-end. To test whether G-SMD can capture this rather subtle difference of the translocation velocities, we set up two systems in which a poly(dA) strand was threaded through α-hemolysin in two global orientations, (cis) 3’-da₅₈-5’ (trans) (referred to as A5’) and (cis) 5’-da₅₈-3’ (trans) (A3’). Refer to Figure 3.3 for the cis/trans convention used. For each orientation of the poly(dA)₅₈ strand, we carried out four G-SMD simulations applying an effective bias equivalent to 9.6 V, and four simulations at 13.2 V. In Figure 3.5, we plot the cumulative number of translocated nucleotides for the poly(dA)₅₈ systems at a 13.2 V effective bias, indicating by the color of the lines the orientation of the DNA strand in the pore. In the same figure, we plot a linear regression fit to the cumulative current curves averaged over the four simulations (symbols). The slope of the fit yields the average translocation velocity of 11.2 ± 0.4 nucleotides/ns for A3’ and 8.8 ± 0.4 nucleotides/ns for A5’; the error is the 90% confidence interval estimated from the standard deviation of the current traces of the four runs taken together. The same systems at a 9.6 V effective bias produced translocation velocities of 7.4 ± 0.2 nucleotides/ns for A3’ and 6.4 ± 0.2 nucleotides/ns for A5’, or a ratio of 1.16 ± 0.05, in excellent agreement with the experimental ratio of 1.16±0.05 [112]. Hence, we conclude that although the absolute values of the translocation velocities are much higher than the experimental ones, their ratio is in agreement with experiment [112].

![Figure 3.5: Effect of global orientation on DNA translocation velocity. The total number of nucleotides translocated is plotted against the simulation time. The chart plots the results of eight simulations, four for each of the two global orientations of a poly(dA)₅₈ strand. Red and blue lines correspond to A3’ (3’ end of DNA on the trans side; see Figure 3.3 a) and A5’ orientations of the strand; squares and circles indicate a linear regression fit to the above sets of four trajectories. The insets illustrate the conformations of a poly(dA)₅₈ strand in the α-hemolysin pore.](image)

It was suggested in Ref. 112 that tilting of the DNA bases towards the 5’-end of the strand in the
constriction of α-hemolysin is the molecular origin of the observed directionality of the DNA translocation. However, the conventional MD simulation employed in that study could sample only a rather small part of the translocation trajectory, in which the displacement of DNA originated mostly from the relaxation of the DNA conformation in response to the applied electric field. Using G-SMD we could, for the first time, not only simulate the entire translocation of a DNA strand through α-hemolysin but also carry out our simulations multiple times. Visual analysis of the MD trajectories confirmed the original suggestion of Mathé, et al. [112] that reorientation of the DNA bases in the constriction of α-hemolysin is responsible for the observed directionality. In the insets to Figure 3.5 we display conformations of two DNA strands identical in sequence but different in their global orientations. When a DNA strand enters the pore with its 5'-end first, the base stacking is more likely to be disrupted due to the preferential tilt of the DNA bases towards the 5'-end in the constriction.

Another factor that can affect the velocity of DNA translocation is the sequence of the DNA strand. The most probable translocation rates of poly(dA)\textsubscript{100}, poly(dAdC)\textsubscript{50} and poly(dC)\textsubscript{100} single-stranded DNA at 25°C were measured at 1.92, 1.36 and 0.76 μsec/base, respectively, at a 120 mV driving bias [43]. The global orientation of the strands was not resolved in those experiments. To investigate whether G-SMD could capture the effect of the DNA sequence on the translocation velocity, we carried out twelve additional simulations on the system that included a poly(dC)\textsubscript{58} strand threaded through α-hemolysin in two global orientations, (cis)3'-dC\textsubscript{58}-5' (trans) (C5’) and (cis)5'-dC\textsubscript{58}-3' (trans) (C3’), as well as a system of poly(dAdC)\textsubscript{29} in one orientation, (cis)5'-(dAdC)\textsubscript{29}-3' (trans) (AC3’). These simulations were carried out using a steering potential of intermediate strength (\(N = 7\)) corresponding to a 9.6 V effective bias. Four simulations at \(N = 3\) were also conducted. In Figure 3.6 we plot the cumulative currents for each of the three sequences in the same global orientation. We found that the poly(dC)\textsubscript{58} strand permeates the pore faster than the poly(dAdC)\textsubscript{29} strand, which in turn permeates the pore faster than poly(dA)\textsubscript{58}, which is in qualitative agreement with experiment [43]. As the global orientation of poly(dC) strands was not resolved in experiment, direct quantitative comparison between simulation and experiment is not possible. Nevertheless, we note that the ratios of the translocation velocities obtained with G-SMD are systematically smaller than the ratios of the most probable translocation times observed in experiment. For example, in our simulations we observed poly(dC)\textsubscript{58} to move about 1.65 times faster than poly(dA)\textsubscript{58} in the direction of the 3'-end of the strand, whereas the ratio of the most probable translocation times is about 2.5 in experiment [43]. At this time we do not know the exact reason for this quantitative discrepancy. We note, however, that our G-SMD simulations were designed to minimize the interactions between DNA and the cap of α-hemolysin, which can arrest the translocation process (see Figures 3.7 and 3.8.) Also, we observed no anomalously long
translocation events for poly(dA) that, in experiment, result in a very broad distribution of the poly(dA) translocation times.

Figure 3.6: Effect of sequence on DNA translocation velocity. The total number of nucleotides translocated is plotted against the simulation time. The chart plots the results of 12 simulations, 4 for each of poly(dC)$_{58}$, poly(dAdC)$_{29}$ and poly(dA)$_{58}$, all in the (cis) 5’-3’ (trans) orientation (see Figure 3.3 a). Traces of the four trajectories are shown for each sequence, represented by thin lines, as well as linear regression fits indicated by black diamonds, red squares and blue circles, corresponding to poly(dC)$_{58}$, poly(dAdC)$_{29}$ and poly(dA)$_{58}$, respectively.

Figure 3.7: We carried out two simulations of C5’ translocation. In the first simulation (red squares), after $\sim$ 3 ns the DNA strand collided with the cap of $\alpha$-hemolysin, which halted the DNA translocation. Figure 3.8 below illustrates the conformation of the C5’ strand at the end of this simulation. In the second simulation (black circles) we, for a short period of time (200 ps), applied a force to the upper portion of the DNA strand parallel to the lipid bilayer membrane in order to shift the DNA away from the cap, after which the simulation was continued normally.

The ratio of the translocation velocities at different strengths of the steering potential is shown in Figure 3.9. As in the case of poly(dA)$_{58}$, for poly(dC)$_{58}$ we observed faster translocation in the direction of the 3’ end of the strand at both $N = 3$ and $N = 7$. The ratio of the 3’- to 5’-first translocation velocities for poly(dA)$_{58}$ is in the 1.1–1.2 range for $N = 3$, 7 and 10. Overall, the ratio of the translocation velocities was not observed to change dramatically with the scaling of the steering potential for the scaling factors
Figure 3.8: Protein-DNA interaction halts translocation. The image illustrates the final state of a 6.5 ns MD simulation carried out with G-SMD under a 4.8 V effective bias. The translocation of a poly(dC)$_{58}$ strand (yellow) halted after encountering positively charged residues at the cap of $\alpha$-hemolysin: Lys8, Lys21, Lys46, Arg236, Lys 237, and Lys288. The fragment of the DNA strand in the vestibule extends to its full contour length but the bonds between the DNA and the protein persist. The surface of $\alpha$-hemolysin is colored according to the type of the exposed residues: red, blue, green and white correspond to negatively charged, positively charged, polar and nonpolar side chains, respectively.
considered. We note, however, that for very large steering potentials, the hydrodynamic drag force on DNA should become comparable to the friction forces between the stem of α-hemolysin and DNA, which obviously will reduce the influence of the sequence on the translocation velocity.

Figure 3.9: DNA current ratios. Selected DNA current ratios are shown comparing orientation or sequence at different effective biases, e.g., $I_{C3'}/I_{C5'}$ is the ratio of the nucleotide current of the (cis) 5'-dC58-3' (trans) system to that of the (cis) 3'-dC58-5' (trans) system. Error bars indicate 90% confidence interval estimated from the standard deviations of the current traces.

Some of the permeation traces noticeably deviate from a straight line; an exemplary trajectory is presented in Figure 3.10. In some parts of the trajectory, the velocity of DNA translocation changes abruptly. Such “steps” in the DNA translocation, however, do not correspond to the translocation of individual nucleotides, as there are fewer steps in the translocation traces than nucleotides translocated through the pore. To verify that our procedure for computing DNA currents, which involves integration over the entire volume of α-hemolysin, did not conceal some events that could be associated with the translocation of individual nucleotides, we computed the DNA current through the constriction of α-hemolysin only. One such trace in shown Figure 3.10 as a dashed line. The steps in the resulting traces are sharper, but the number of steps remains the same. Visual inspection of the translocation trajectories with VMD [123] revealed that the steps are associated with broken base stacking of the DNA strand entering the constriction, indicated by the arrows in Figure 3.10. Two example base conformations are shown as insets. Thus, our G-SMD simulations suggest that permeation of DNA through the constriction of α-hemolysin takes place in pulses of 2–4 nucleotides that preserve the base stacking pattern of an unconfined strand.
Figure 3.10: Steps in DNA current traces. Exemplary traces of DNA current are shown, both through the entire channel (solid line) as well as through the constriction only (dashed line). Rather than a smooth linear trace, simulations of this system displayed a pronounced step-like current. Inspection with the program VMD [123] revealed the steps to be associated with breaks in the base stacking structure of the DNA entering the protein constriction, causing temporary halts in the translocation followed by surges of DNA current. Times at which the base stacking was broken in the constriction are marked by the arrows. The insets show the conformations of the nucleotides in the constriction (Thr145 to Val149) for the structure breaks at 0 ns (left) and 1.9 ns (right). This simulation corresponds to DNA in the A5' orientation (5' end of DNA on the trans side; see Figure 3.3 a) at a 9.6 V effective bias.

3.3 Discussion and Conclusions

We have developed and tested the extension of the Steered Molecular Dynamics method to simulations of electric field-driven permeation of large solutes through membrane channels. In tens of nanoseconds this method can provide a realistic account of permeation events that would require millisecond simulations with standard MD. The decoupling of the electric field driving the solutes through the channel from the electric field acting on the rest of the system allows us to apply high effective biases without affecting the system's integrity, greatly reducing the permeation timescale. As the SMD forces in our method act on the entire solute via a smooth 3-D potential, the strain introduced into the solute's conformation is small, comparable to the strain observed during a standard MD simulation.

Using the G-SMD method we investigated the mechanics of DNA permeation through α-hemolysin. Our simulations revealed that the base-stacking of nucleotides in single DNA strands has a significant effect on the DNA translocation velocity in α-hemolysin. Our G-SMD results support the molecular mechanism proposed in Ref. 112 to explain the dependence of the DNA translocation velocity on the global orientation of the DNA strands in the channel, i.e., the preferential tilt of the DNA bases towards the 5'-end of the strand near the channel’s constriction. Due to the dramatic reduction of the translocation timescale and small distortions that G-SMD introduces into the solute’s conformation, we could carry out multiple simulations
of the same permeation process and characterize the outcome in statistical terms. The outcome of such simulations can provide the sequence-specific information required for a realistic coarser-scale description of the translocation process [116–120, 124–127].

Our initial motivation for developing the G-SMD method was to enable characterization of the conformations of large solutes during their transport through α-hemolysin. The results of simulations by others in our group demonstrate that the G-SMD method can adequately describe permeation not only of linear DNA strands but also of DNA hairpins and small peptides [22]. The ensemble of conformations resulting from a G-SMD run can be related to the measured ionic current blockages using standard MD [65] or other methods [128, 129]. Next, we investigated if, in addition to screening possible conformations of the solute in the pore, the G-SMD method can adequately describe permeation kinetics and relative translocation velocities. For this purpose we carried out G-SMD simulations on DNA strands of different sequences and global orientations. The obtained ratios of the simulated translocation velocities of DNA strands in different global orientations were in excellent quantitative agreement with experiment. The ratio of the translocation velocities of DNA homopolymers of different sequence were found in semi-quantitative agreement with experiment. Such good agreement with experiment is encouraging given that our method reduces the permeation time 1000-fold.

G-SMD has proven to be an effective technique for probing certain processes that cannot be accurately described using SMD. As with SMD, however, the magnitude of the steering forces must be carefully chosen to ensure that the simulated trajectories correspond to experiment. Ideally, the magnitude of the applied forces and the timescale of the simulation should allow relaxation forces to dissipate the strain introduced by the G-SMD protocol.

The G-SMD method extends the applicability of atomic-scale simulations to processes that were previously off-limits. Among the topics to be addressed in the future are mechanisms of protein translocation through membrane channels, selectivity of ion channels and transporters, simulations of nanopore devices for high-throughput sequencing of DNA and proteins, and, eventually, the transport of biomolecules in nanofluidic systems. In addition to studies of membrane transport, the G-SMD method is expected to find application in multiscale modeling, cryo-electron microscopy data fitting, in developing implicit models of biomaterials, and for simulations of nanoscale hydrodynamics. The G-SMD method will be further developed to enhance random forces producing stochastic motion of the solutes in the channel, to employ a self-consistent electrostatic potential (i.e., one calculated during the G-SMD simulation) to steer permeation of solutes, and to provide methods for extracting the Potential of Mean Force from G-SMD trajectories.

37
Chapter 4

Slowing down DNA translocation


4.1 Introduction

It is well known that interactions between DNA and counterions can profoundly affect its physical properties [130, 131]. Although valuable as first-order approximations, traditional models for polyelectrolyte-counterion interactions, i.e., Manning [130] and Poisson-Boltzmann [132] theory, ignore relevant details such as the discrete nature of charges on DNA, the type of cation, and ion-ion interactions. Here, we examine the results of nanopore experiments using MD simulations to quantitatively reveal the effect of different counterions (K\(^{+}\), Na\(^{+}\), and Li\(^{+}\)) on the charge reduction of a DNA molecule. Surprisingly, we find that the various monovalent ions can have very different effects.

Nanopores have emerged as a versatile tool for the detection and manipulation of charged biomolecules [66, 71, 133–136]. In a typical setup, an external electric field drives a (bio)molecule through a nanometer-size pore in a thin synthetic membrane, producing a characteristic temporary change in the trans-pore ionic current. This can be used for sensitive single-molecule sensing platforms. A major difficulty in experiments to date, however, has been the speed of DNA translocation, which is very fast. The average translocation speed is set by the electrophoretic drive which in turn is determined by the charge on the DNA. A lower charge would result in lower translocation speed and therefore higher read-out accuracy in these types of experiments. In this chapter, we show how the nanopore can be used as a tool to determine the effect of the electrolyte conditions on the charge of DNA.
4.2 Summary of Experimental Results

Our collaborators performed experiments in which double-stranded (ds) DNA transited a 15–20 nm silicon nitride (SiN) nanopore under a modest (\(\sim 0.1 \text{ V}\)) external electric voltage [70]. Figure 4.1a shows the layout of the nanopore experiments, depicting a dsDNA molecule in an ionic solution containing KCl, NaCl, or LiCl. DNA translocation events are evident as a reduction in the ionic current. Our collaborators found that the translocation times increase greatly upon changing the solute from KCl to NaCl to LiCl. They found that for 1 M solutions, the ratios for the experimental translocation times of dsDNA are KCl : NaCl : LiCl = 1 : 1.7 : 4.8 [70]. This is a surprising observation since, a priori, one would expect that the monovalent \(\text{K}^+\), \(\text{Na}^+\), and \(\text{Li}^+\) ions would behave very similarly. Higher salt concentrations of LiCl were found to result in even longer translocation times, with a ratio of 1 : 1.5 : 2 for 1 M : 2 M : 4 M.

![Figure 4.1: (a) Side-view schematic of the experimental device, consisting of a 20 nm thin free-standing silicon nitride window (blue layer) embedded in a silicon wafer. Upon application of an electric field across the nanopore, DNA translocates through the pore. (b) MD simulation system, showing a DNA molecule, a 0.1 M KCl solution and an 8 nm nanopore. Silica is shown as a gray surface, DNA is shown as yellow van der Waals spheres with the phosphates shown in red, chloride ions are shown in blue, and potassium ions are shown in green. Water is not shown.](image)

4.3 MD Results

How can we understand this pronounced and unexpected slowing down of DNA translocation upon merely changing the buffer from KCl to LiCl? To elucidate the microscopic mechanism of the dependence of the DNA translocation velocity on the type and concentration of the electrolyte, we carried out all-atom
MD simulations of several nanopore systems each containing a fragment of dsDNA, a circular nanopore and electrolyte, as depicted in Figure 4.1b. Each system was periodic in all three dimensions, and hence represented an infinite DNA molecule confined in an infinite nanochannel. In the radial direction, the DNA fragment was harmonically restrained to remain in proximity of the geometrical center of the pore. In the axial direction, an additional harmonic potential was enforced between the centers of mass of the DNA fragment and of the nanopore. Subject to an external electric field directed along the nanopore axis, the DNA fragment was observed to move in the direction opposite to the direction of the field until the force of the harmonic potential balanced the effective force exerted by the electric field on the DNA. A detailed description of the simulation methods and protocols is provided in Section B.3.

Using the setup described above, we simulated the dependence of the stall force on the type and concentration of the electrolyte and on the diameter of the nanopore. We chose to simulate the stall force rather than translocation velocity because, for the long DNA molecules used in experiment (λ-DNA), the translocation velocity is determined by the balance of the effective driving force of the electric field in the nanopore and the hydrodynamic drag on the DNA coil outside the nanopore. Assuming the hydrodynamic drag on the coil depends only on the viscosity of the electrolyte, the simulated dependence of the stall force should, ideally, match the measured dependence of the translocation velocity after scaling by the viscosity, even if the DNA translocation velocity varies (increases) as the translocation progresses [137].

Figure 4.2, a–b plots the simulated dependence of the stall force, expressed as translocation time for comparison to experiment. To convert forces to translocation times, we adopt the model of Storm et al. [138], in which the dominant force opposing translocation is the viscous drag of the coil of untranslocated DNA. The persistence length of DNA is independent of ion conditions at the ionic strengths examined here [139], and thus translocation times $\tau$ can be calculated as $\tau = \alpha \eta / F_{\text{stall}}$, where $\eta$ is the solution viscosity [140], $F_{\text{stall}}$ is the stall force determined in the MD simulations, and $\alpha = 62 \mu m^2$ is a parameter chosen to fit the experimental 1.0 M KCl translocation time of 1.7 ms. The simulations were performed for KCl, NaCl, and LiCl at 0.1 M, 0.5 M, 1.0 M, and 4.0 M concentrations using nanopores of 8, 16, and 22 nm in diameter. The results of the simulations are in qualitative agreement with the experimentally observed dependence of the translocation speed on salt species and concentration, as presented in Figure 4.2c. For all pores examined, the stall force in 4.0 M LiCl is lower than in 4.0 M NaCl, which is in turn is lower than in 4.0 M KCl. The simulations also correctly capture the decrease of the stall force as the pore diameter increases, as was measured before [141]. The stall forces are shown in Figure 4.3.

To investigate the microscopic mechanism responsible for the slow translocation of DNA in the presence of LiCl, we computed the average number of ions bound to the DNA surface in the MD trajectories. Direct
Figure 4.2: Molecular dynamics simulations. (a) Translocation time as a function of pore size for 4.0 M ion concentration. LiCl is seen to result in the longest translocation time in all pore sizes, followed by NaCl, and finally by KCl. Lines are linear fits. (b) Translocation time as a function of ion concentration in the 16 nm diameter pore. Translocation times $\tau$ were calculated as $\tau = \alpha \eta / F_{\text{stall}}$, where $\eta$ is the solution viscosity, $F_{\text{stall}}$ is the stall force determined in the MD simulations, and $\alpha = 62 \text{ } \mu\text{m}^2$ is a parameter chosen to fit the experimental 1 M KCl translocation time of 1.7 ms. Lines are linear fits to the data. (c) Experimental translocation times for dsDNA [70] as a function of ion species and concentration (same color coding as in panel b). Solid lines are linear fits to the data.
ion binding to DNA was rarely observed in our MD simulation trajectories. Ion binding in the first solvation shell, however, was prevalent, see Figure 4.4 b for illustration. In Figure 4.4 a, we show the number of bound counterions per base $N$ in the 8 nm pore at 4.0 M ion concentration. The number $N$ is plotted as a function of minimum bond duration (see Section B.3). The plot shows that the instantaneous number of bound ions (minimum bond duration of 0 ps) is independent of ion type. Importantly, however, the bond strengths are not equal: Li$^+$ bonds last longer, on average, than Na$^+$ bonds, which in turn last longer than K$^+$ bonds. For example, in the 4.0 M LiCl system, each base has on average 0.5 lithium ions which have been bound for at least 50 ps, while in the 4.0 M KCl system, each base has less than 0.1 potassium ions which have been bound that long. This difference in bond strength leads to a difference in net force, and thus velocity, for the different solutions. Note that taking into account chloride ions, the instantaneous total charge of bound ions per base approaches 1 as the minimum bond duration approaches 0, see Figure 4.5 a. Chloride ions were found to bind much more weakly to the DNA surface than the cations, Figure 4.5 b.

To demonstrate the effect of bond strength on the force transmitted to the DNA by the ions, we consider a system of electrically driven ions in a periodic potential, shown in Figure 4.6 a. The potential represents a simplified toy model of the charge distribution along the DNA molecule. MD simulations of this system performed for different barrier heights reveal the dependence of the effective force applied by ions to the potential representing the surface of a DNA molecule, as well as the dependence of ion residence time in individual binding sites. Thus, Figure 4.6 b shows $F_{\text{applied}}/qE$, where $F_{\text{applied}}$ is the average magnitude
Figure 4.4: (a) Number of bound counterions per base as a function of minimum bond duration in MD simulation with 4 M ion concentration and an 8 nm diameter pore. Ion binding to DNA takes place through an intermediate water molecule. The instantaneous number of bound ions (i.e., zero minimum bond duration) is seen to be independent of the ion type. However, at higher minimum bond durations, the number of bound lithium ions is seen to be higher than the number of bound sodium and potassium ions. (b) Lithium counterions bound to DNA in MD simulations. DNA is shown in gray, lithium is shown in yellow, and intermediate water is shown in red and white. Only water molecules involved in the bonds are shown. (c) Bond duration $\tau$ at 4 M concentration of LiCl (circles), NaCl (squares), and KCl (diamonds) versus number of bound ions per base $n$ for MD simulations of an 8 nm diameter pore. Each data point represents binding to a single type of DNA atom; the five most popular binding sites are shown and labeled. Atom names are those used in the CHARMM force field. Each binding site is seen to exhibit essentially the same number of bound ions upon changing the ion type. However, the strength of the bonds is highly dependent on ion type, with the most durable bonds are seen in LiCl, followed by NaCl, and finally KCl. Note also that although the phosphate oxygens O1P and O2P exhibit the most bound ions (i.e., large $n$), those bonds are relatively weak (i.e., low $\tau$). The longest-lasting bond type seen is to the O4 oxygen atom in the minor groove. (d) Comparison of bond duration versus number of bound ions per base at different KCl concentrations: 0.1 M (pluses), 1 M (crosses), and 4 M (diamonds). Changing the ion concentration is seen to change the number of bound ions, but, as expected, not the strength of the bonds.
Figure 4.5: (a) Number of bound cations minus the number of bound chloride ions (equivalent to the net charge of the bound ions in units of e) versus minimum bond duration. (b) Number of bound chloride ions per base versus minimum bond duration. Chloride binding events are significantly shorter ($\tau = 17.6, 15.7,$ and $11.6$ ps for LiCl, NaCl, and KCl, respectively) than cation binding events (see Figure 4.4 a).

of the force applied to each ion by the potential (and thus the average force applied by each ion to the potential), $q$ is the charge of the ion, and $E$ is the magnitude of the applied electric field. At a barrier height of $0.25 \ k_B T$, the ions apply little force to the potential, because thermal and electrical energy dominates and the ions do not bind strongly to the binding sites. At a barrier height of $2.25 \ k_B T$, on the other hand, the ions apply nearly the maximum possible force ($qE$), because the ions are strongly bound and spend a relatively long time in each binding site, and therefore transfer the force due to the electric field to the DNA rather than dissipate it to the solution. At intermediate barrier heights, the complete range of possible forces is seen, despite the fact that all ions are bound (i.e. occupy a binding site) at all times. In Figure 4.6 c, we characterize the number of ions that bind to the wells of the potential (compare with Figure 4.4 a) for several intermediate values of the barrier height, illustrating the dependence of bond duration on the latter. The results clearly show the effect of barrier height on the effective force applied by the ions on the DNA, and therefore on the effective charge of DNA in the electric field.

Counterion binding and the dependence on ion type are illuminated further by examining binding to individual sites on DNA. To this end, binding curves such as those shown in Figure 4.4 a were calculated for every binding site seen in the simulations, then fit by an exponential of the form $n e^{-t/\tau}$, where $n$ and $\tau$ are fitting parameters. The five sites with the highest $n$ for the 4.0 M LiCl 8 nm system are plotted in Figure 4.4, c–d. We see in Figure 4.4 c that the number of bound ions per site is largely independent of the ion type, but that the bond duration $\tau$ is highest for lithium, followed by sodium, and finally potassium. Figure 4.4 d shows the dependence on ion concentration: as expected, the ion concentration only affects the
Figure 4.6: The effect of barrier height on force transmission. (a) Model system, in which a sinusoidal potential, representing a toy model for the DNA charge along its length, is applied to ions. Barrier heights of 1 and 2 $k_B T$ are shown. Each period of the potential represents a binding site. Three binding sites are shown; the model system contained 20 binding sites per 6.4 nm length, similar to DNA. (b) Average force per ion applied to the potential representing DNA as a function of barrier height, normalized by the applied electric force. (c) Binding curves similar to those plotted in Figure 4.4a, shown for barrier heights of 1, 2, and 2.25 $k_B T$. Varying the barrier height is seen to affect both the bond lifetime and the effective force applied to the potential representing DNA. Solid lines are exponential fits to the data.

number of bound ions, not the bond duration.

4.4 Discussion

Through a combination of MD simulations and nanopore translocation experiments performed by our collaborators [70], we have characterized the effective charge of a DNA molecule in monovalent electrolytes. While it has been known that the charge of a DNA molecule sensitively depends on the valence of counterions [142,143], here we demonstrate that the size of counterions can considerably influence the effective DNA charge, and hence the average translocation speed through a nanopore.

One may ask whether interactions between DNA and the nanopore surface play a role in the reducing of the average translocation speed. In very small pores (well under 5 nm in diameter) DNA-pore wall interactions were reported to slow down DNA translocation [144,145] as well as to lead to a considerable spread in event characteristics [146]. In our case however, our collaborators used wide pores of $\sim 20$ nm in diameter [70] where such effects have not been reported, and there are no obvious reasons to assume adsorption because of the negative charge of both the membrane surface and DNA. Indeed, our collaborators obtained similarly different translocation speeds for LiCl and KCl for graphene nanopore experiments, confirming that the slowing down is intrinsic to the Li-DNA interactions rather than due to extrinsic reasons such as sticking to the SiN surface.

At the molecular level, we found a partial neutralization of the DNA charge that originates from transient binding of counterions. Although we found the number of bound counterions to be the same for each ion type,
the duration, and hence strength, of those bonds was not. Indeed, lithium bonds to DNA were found to be the longest lived, followed by sodium, and finally potassium. Using a simplified ion system, we demonstrated that such a change in bond strength does indeed affect the force transmitted from ion to DNA, and hence the velocity of the translocating DNA. If compared to direct experimental measurements, the simulations quantitatively underestimate the net force on DNA in the case of KCl. However, the qualitative agreement between simulation and experiment is satisfactory, given the relative simplicity of MD models, the known imperfections of the water model in classical MD simulation (for example, a 3-fold underestimation of its viscosity), and the sensitive dependence of the stall force on the ion-DNA interaction. From a practical perspective, our findings indicate that LiCl offers significant advantages compared to the traditionally used ionic solutions for nanopore applications.
Chapter 5

DNA in graphene nanopores

5.1 Introduction

The idea of using a nanopore to sequence DNA [147] has generated much excitement since DNA translocation through the biological nanopore α-hemolysin was first demonstrated [38]. The principle of nanopore sequencing, illustrated in Figure 5.1, is straightforward: an external electric field is used to drive negatively charged DNA through a nanopore. As DNA transits the pore, its nucleotide sequence is determined by measuring changes in the nanopore ionic current. By reading the sequence directly from genomic DNA, nanopore methods promise to offer single-molecule, label-free DNA sequencing with virtually unlimited read lengths [66,67], overcoming many of the drawbacks of so-called next-generation sequencing platforms [68,69].

Spectacular progress in engineering the biological nanopores α-hemolysin [57,59] and MspA [148,149] for sequencing applications indicate the imminent arrival of nanopores in practical biomedical applications [64]. The key factors enabling DNA sequencing using biological nanopores are the nanometer dimensions of the ion-current-modulating regions of the nanopores and the use of processive enzymes to produce DNA translocation in discrete steps [63,150]. Solid-state nanopores [71] offer a number of practical advantages over their biological counterparts, including superior mechanical properties [151], multiplex detection [152], integration with on-chip electronics [153], and detection modalities other than ionic current [154–161]. However, conventional synthetic membranes are generally over 10 nm thick, meaning the pore is occupied by many DNA bases at the same time, and thus making single-base-sensitive measurements extremely difficult [67].

The advent of graphene [72]—a single layer carbon sheet—has opened a new chapter in the development of solid-state nanopores for DNA sequencing, with several groups already reporting measurements of ionic current blockades produced by transport of double-stranded DNA through graphene nanopores [73–76]. Graphene nanopores share many advantages of conventional solid-state nanopore systems while offering atomically precise control over the nanopore dimensions, in particular the membrane thickness, which can be made smaller than the distance between neighboring DNA nucleotides in a DNA strand. Furthermore, the unique physical properties of graphene membranes offer several possibilities for DNA sequence detection,
including transverse tunneling [162–164], nanoribbon conductance [165,166] and mechanical deformation of DNA [167,168]

In this chapter, we assess the suitability of graphene nanopores for sequencing DNA by measuring ionic current—the method already proven to work in the case of biological nanopores. Using all-atom molecular dynamics (MD) [134], we simulate the translocation of single-stranded DNA (ssDNA) through graphene nanopores and characterize the ionic current blockades produced by DNA nucleotides. We show that graphene nanopores have an essential feature that made biological nanopores amenable to sequencing applications: translocation of ssDNA can occur in single-nucleotide steps. Furthermore, others in our group have shown that the ionic current blockades can be indicative of the type of DNA nucleotides [169].

Figure 5.1: Schematic of a graphene nanopore-based device for sequencing DNA. A graphene sheet containing a nanopore is placed over an aperture in a synthetic membrane. DNA is driven through the pore by a transmembrane electric field. As the DNA transits the nanopore, the ionic current is modulated by the nucleotides in the nanopore, revealing the DNA sequence.
5.2 Results

To simulate the electric field-driven transport of ssDNA through graphene nanopores, we built several all-atom systems, each containing a graphene membrane with a single nanopore in it, a single DNA strand threaded through the nanopore, and 1 M KCl solution. The systems were simulated using the all-atom MD method [134], applying a constant electric field normal to the graphene membrane to produce transport of ssDNA and ions through the nanopore. In such classical simulations, the graphene layer serves merely as a barrier to ion passage, which is an adequate approximation to determine general features of ssDNA and ion transport, but may be not accurate enough to yield precise quantitative insights. A complete description of our simulation protocols is given in Section B.4. A number of pore sizes, pore shapes, membrane thicknesses, and DNA sequences were examined. Table 5.1 lists all of our production simulations.

<table>
<thead>
<tr>
<th>Pore</th>
<th>Layers</th>
<th>Pore radius (Å)</th>
<th>Voltage (V)</th>
<th>$I_0$ (nA)</th>
<th>$\langle I \rangle_{\text{seq}} / I_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
<td>7.7</td>
<td>1.2</td>
<td>5.0</td>
<td>0.12–0.58</td>
</tr>
<tr>
<td>II</td>
<td>1</td>
<td>8.0</td>
<td>0.5</td>
<td>3.3</td>
<td>0.61</td>
</tr>
<tr>
<td>III</td>
<td>2</td>
<td>7.0 × 5.4</td>
<td>2</td>
<td>2.8</td>
<td>0.15</td>
</tr>
<tr>
<td>IV</td>
<td>2</td>
<td>8.0 × 6.2</td>
<td>0.5</td>
<td>1.5</td>
<td>0.37</td>
</tr>
<tr>
<td>V</td>
<td>2</td>
<td>8.0</td>
<td>0.5</td>
<td>2.1</td>
<td>0.63</td>
</tr>
<tr>
<td>VI</td>
<td>3</td>
<td>8.0 × 6.2</td>
<td>0.5</td>
<td>1.1</td>
<td>0.32</td>
</tr>
<tr>
<td>VII</td>
<td>3</td>
<td>8.0</td>
<td>0.5, 0.8</td>
<td>1.9</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Table 5.1: Graphene nanopore systems examined using the all-atom MD method. For elliptical nanopores, major and minor radii are given. For pore VII, all simulations were carried out at 0.5 V, with the exception of the simulation of poly(dG), which was performed also at 0.8 V. The blockade current $\langle I \rangle_{\text{seq}} / I_0$ was averaged over all DNA homopolymer sequences. For pore I, the lowest [poly(dT)$_{10}$] and highest [poly(dA)$_{10}$] blockade currents among the four bases are listed.

Figure 5.2 a illustrates a typical simulation trajectory. In all of our simulations, ssDNA was observed to stick to the surface of the graphene membrane. Although we began our simulations with ssDNA in a conformation extended away from the graphene membrane, within $\sim 25$ ns most DNA bases in most systems were found to be in contact with the graphene membrane. Such adsorption of DNA bases has a hydrophobic origin, as water is excluded from the DNA-graphene contact area [170, 171]. Despite the strong interaction, DNA remained highly mobile at the graphene surface, performing a kind of two-dimensional diffusion. In likely experimental scenarios, the cross sectional area of the graphene nanopore is negligible in comparison to the area of the graphene membrane; therefore, the most probable sequence of events is that the ssDNA first adsorbs to graphene membrane surface from bulk solution and then enters the nanopore by diffusion along the graphene surface.

Our simulations suggest that such 2D diffusion of ssDNA does not limit the rate of ssDNA translocation through the nanopore. In Figure 5.2 b, we plot the diffusivity of DNA bases adsorbed to one of the graphene
surfaces versus the distance from the nanopore, calculated separately for the radial and tangential directions. Near the pore, the tangential diffusivities are seen to be approximately 10 times higher than the radial diffusivities, and both values increase with the distance from the pore, approaching the free-solution value of \( \sim 10 \, \text{Å}^2/\text{ns} \). In the case of the pore featured in Figure 5.2 a, the mean time per base translocation \( \bar{t} \) is \( \sim 16 \, \text{ns} \) at a 500 mV bias. Using a conservative estimate for the 2D diffusion coefficient of a DNA base (4 \( \text{Å}^2/\text{ns} \)), the expected mean-square displacement of the base within the time interval \( \bar{t} \) exceeds the distance between neighboring nucleotides in a DNA strand (\( \sim 5 \, \text{Å} \)).

Figure 5.2: Adhesion of ssDNA to the surface of a graphene membrane. (a) A typical MD simulation of ssDNA translocation through a graphene nanopore. Shown is the trajectory of poly (dT)\(_{20}\) under a 500 mV bias, pore VII (see Table 5.1). Complete translocation data is given in Figure 5.4 d. The snapshots in the top and bottom rows provide the side and top views of the same system. DNA atoms are shown as yellow and red spheres (red spheres indicate the phosphorus atoms); the graphene membrane is shown as a gray semitransparent surface. Water and ions are not shown. The arrows indicate the radial and tangential directions used to compute 2D diffusivities of the DNA bases. (b) Diffusion of bases adsorbed to the surface of one- (green), two- (red), and three-layer (black) membranes. The plots show the diffusivity for radial (solid lines) and tangential (dashed lines) directions.

All pores examined in this study remained filled with water during our MD simulations. Figure 5.3 shows the average water densities along a line passing through the center of the pores normal to the membrane (z-axis in our setup). Some small deviations from the bulk value of 0.1 atoms/Å\(^3\) can be seen in the water density profiles near the center of the graphene membrane (z = 0). However, when DNA atoms are included in the density calculations, the deviations become considerably smaller, indicating that the water deficit is due to displacement by DNA. Thus, we did not observe ion current gating produced by electric field-induced wetting-dewetting transitions that was experimentally observed for larger hydrophobic pores [172,173]. We hypothesize that the presence of a hydrophilic DNA backbone in the nanopore suppresses such dewetting transitions for the pore geometries considered.

In all simulations, ssDNA was observed to either translocate in steps or not to translocate at all. The
Figure 5.3: Water density in pores II (a), V (b), and VII (c). Number density of water atoms alone (dotted line) and of water and DNA atoms together (solid line) are plotted as a function of distance from the center of the graphene membrane along the nanopore axis. Away from the pore and membrane, the number density is 0.1 atoms/Å³. Within the nanopore, some deviation in the density of water can be seen, especially in pore VII (c). However, the density of both water and DNA atoms is close to the bulk value, indicating that water missing from the pore has been displaced by DNA, rather than expelled from the nanopore. All plots are for poly(dT)₂₀ system at a 500 mV bias.
latter may be interpreted as a long step, given the limited timescale of our MD simulations (< 500 ns). For each pore studied, simulations were performed using poly(dA), poly(dC), poly(dG), and poly(dT) homopolymers. In general, the greatest number of nucleotide translocation events were observed for the poly(dT) systems, and the fewest for poly(dG).

The pore diameters explored in our MD simulations ranged from 10.8 to 16 Å. We considered such small pores as they were expected to exhibit deeper ionic current blockades and, possibly, larger sensitivity of the blockade ionic current to the type of DNA nucleotides confined in the pores. However, our smallest pores proved to be so small that ssDNA would not translocate except at a very high bias (∼2 V). In addition to circular pores, we also investigated elliptical pores (III, IV, and VI) to determine whether such pores could enhance the ionic current signal by reducing the pore’s cross-sectional area and control the DNA conformation by breaking the rotational symmetry of the system. By design, the pores were only large enough for bases to transit when properly aligned. However, rather than steering the DNA, the pores tended to jam, severely limiting conformational fluctuations of ssDNA and thus preventing translocation. It is possible that at lower electric biases the DNA would be able to more fully explore conformational space and sample conformations which would allow it to translocate. For larger diameter pores, we expect the strong interactions between the graphene membrane and the bases of ssDNA to remain the limiting factor determining the overall speed of ssDNA translocation.

We found that the transmembrane bias can affect not only the rate of DNA translocation, but also its qualitative character. Higher biases, such as those used in the smallest pores, generally increased the DNA translocation rate. However, increasing the bias was observed to increase the amount of “skips”, i.e. rapid translocation of multiple nucleotides. In the case of elliptical pores, a higher bias may also, somewhat paradoxically, lead to slower translocation due to frustration, by trapping the DNA in a conformation unfavorable for translocation.

Figure 5.4 compares the results of our MD simulations performed using graphene membranes containing one, two or three carbon layers and the same pore size. The figure shows data for poly(dT)20, which is representative of the data obtained using other sequences in pores II, V, and VII; complete translocation data for all systems is provided in Figure 5.5.

In our one-layer systems, Figure 5.4 a, translocation of ssDNA did not reliably occur in single-nucleotide steps, see Figure 5.4 d and Figure 5.5 b. Because the single layer graphene is only one carbon atom thin, the fraction of time that the pore is occupied by a DNA base is quite small (∼36%, see Figure 5.6), which severely limits detection of the DNA sequence by measuring the ionic current. Furthermore, we do not expect that lowering the transmembrane bias would significantly increase the time DNA bases are exposed.
Figure 5.4: DNA translocation through nanopores in one-, two-, and three-layer membranes. (a–c) Snapshots of the pore region of the systems (pores II, V, and VII) featuring one- (a), two- (b), and three-layer (c) membranes. Graphene is shown as a silver molecular surface and DNA is shown as yellow and red van der Waals spheres, where red spheres indicate the phosphorus atoms. (d) Number of translocated nucleotides as a function of simulation time for one- (pore II, green), two- (pore V, red), and three-layer (pore VII, black) systems. The arrow indicates base skipping: multiple bases transit the pore together. Data here are from systems containing poly(dT)\textsubscript{20} under a 500 mV bias. (e–g) Number of translocated nucleotides versus simulation time for pore VII (three-layer membrane) and poly(dA)\textsubscript{20} (e), poly(dC)\textsubscript{20} (f), poly(dG)\textsubscript{20} (g). In panel (g), the vertical dashed line indicates a bias increase from 500 mV to 800 mV.

to the nanopore volume. Instead, lowering the bias is expected to increase the time ssDNA spends between the steps, during which no base occupies the pore.

For a two-layer membrane at same transmembrane bias of 500 mV (Figure 5.4 b), translocation of ssDNA was observed to occur at a higher rate than in the one-layer system (Figure 5.4 d). Although clear single-base steps can be discerned from the DNA translocation trace, there are also many skips. The arrow in Figure 5.4 d indicates one such event: the number of translocated nucleotides rapidly changes from \( \sim 4 \) to \( \sim 6 \) within \( \sim 2 \) ns. In contrast, translocation of ssDNA in the three-layer systems, Figure 5.4 c, was observed to occur predominantly in single-base steps for all four homopolymers sequences studies, Figure 5.4, d–g. In the case of poly(dG) (Figure 5.4 g), translocation was observed only after the transmembrane bias was increased to 800 mV (the translocation was stalled at 500 mV), producing a clearly stepwise translocation pattern in single-base steps. Together, single-base steps accounted for 54\% (20/37) of all translocated nucleotides in the pore VII systems, while in the pore II and pore V systems they accounted for 40\% (4/10) and 29\% (6/21), respectively. Thus, translocation in pore VII was not exclusively in single-base steps; the data for poly(dA) in particular (Figure 5.4 c) show some skipping and continuous translocation. Nevertheless, no other pore that we examined reliably produced single-base steps across all sequences. Overall, the rate of ssDNA translocation was not found to exhibit a simple dependence on the number of carbon layers in a graphene membrane.
Figure 5.5: Number of translocated nucleotides in the MD simulations of nanopores I (a), II (b), III (c), IV (d), V (e), and VI (f). Data for pore VII is shown in Figure 5.4, d–g. Traces are shown for poly(dA) (black), poly(dC) (red), poly(dG) (green), and poly(dT) (blue). In every system, DNA translocation was observed for at least one sequence. Note that pores I and III were simulated at 1.2 and 2 V, respectively, while all other systems were simulated at 500 mV (see Table 5.1).
Figure 5.6: Number of bases within pore II (top), pore V (middle), and pore VII (bottom). A base is considered within the pore if the position of its center of mass in the direction normal to the membrane fulfills the condition $|z| \leq 1.674N\ \text{Å}$, where $N$ is the number of graphene layers composing the membrane. In pore II, no base occupied the pore 64% of the time (averaged over sequences). In pore V, the pore was occupied by a single base 69% of the time, but the pore was still unoccupied a significant fraction of the time. In pore VII, the pore was almost always occupied, most often by a single base (55% average). Two bases were found in pore VII less frequently (38%).

The microscopic mechanics of stepwise translocation can be discerned from Figure 5.4, a–c. Typically, one translocation step involves unbinding of a DNA base from one side of a graphene membrane and subsequent binding of the same base to the other side of the membrane (single-layer graphene), or concerted unbinding/binding of several neighboring nucleotides (two- and three-layer membranes). In both cases, stepwise transport occurs simultaneously with sliding of the DNA strand along the surface of the membrane. For nanopores in two- and three-layer membranes, more often than not only one base occupies the nanopore, see Figure 5.6. Furthermore, due to strong hydrophobic interaction, the base occupying the nanopore is localized near its surface, see Figure 5.7.

The combination of stepwise translocation and localization near the nanopore surface results in a radical reduction of the number of conformations a nucleotide can adopt in the nanopore. To illustrate this point, in Figure 5.8, a–d we plot the $z$ coordinate of the center of mass of each base versus time using the MD trajectories obtained for pore VII. The plots display the hallmark of stepwise translocation we seek: long stable periods during which a measurement can be performed, punctuated by short periods of DNA translocation to advance to the next base. For each DNA sequence, the histograms of the nucleotide coordinates (Figure 5.8, e–h) display one or two peaks near the membrane center ($z = 0$), highlighting repetitive positioning.
of DNA bases within the pore. Thus, single bases residing within pore VII are localized both radially (with respect to the pore axis) and longitudinally (along the pore axis).

Furthermore, the bases adsorbed to the nanopore surface are seen to adopt a limited number of conformations. In Figure 5.8 j we plot a scatter diagram of the angles $\beta$ and $\gamma$, which we use to describe the orientation of a DNA base adsorbed to the nanopore surface. The angle $\beta$ measures the tilt of the base relative to the membrane, while the angle $\gamma$ measures the rotation of the base within the plane defined by $\beta$; these angles are shown in Figure 5.8 k and defined precisely in Figure 5.9. The observed orientations fall into two distinct clusters, centered at $(\beta, \gamma) \approx (80^\circ, 160^\circ)$ and $(65^\circ, 295^\circ)$. The two clusters correspond to conformations that feature the plane of the bases oriented almost parallel to the pore axis (and the nanopore surface) and the two possible flips of the base about the backbone, see Figure 5.8 k. Moreover, the purine and pyrimidine bases were found to display statistically distinct conformations, illustrated in Figure 5.8 l. The clustering of base orientations is indicative of consistent positioning of DNA nucleotides within the nanopore, which is expected to facilitate sequencing due to the dramatic reduction of conformational noise [174].

As coordinates of every ion in the simulation system are known, the ionic current flowing through the nanopore can be easily determined [121, 134]. Typically, we find the open-pore ionic currents $I_0$ to be in the range of several nanoamperes and the average relative blockade current $I/I_0$ to range between 12 and 63%, see Table 5.1. For pores II, V, and VII, which showed consistent DNA translocation, the blockade current
Figure 5.8: Stepwise transport of ssDNA is associated with reduced conformational noise. (a–d) $z$ coordinate of the center of mass of individual bases in poly(dA)$_{20}$ (a), poly(dG)$_{20}$ (b), poly(dC)$_{20}$ (c), and poly(dT)$_{20}$ (d) systems. To make translocation events discernible, coordinates of individual bases are plotted in black, red, green and blue. The span of the $y$ axis ($-5 < z < 5$ Å) corresponds to the thickness of the three-layer membrane. (e–h) Normalized distributions of the bases’ $z$ coordinate. The histogram for poly(dG)$_{20}$ was computed using the MD trajectory at 800 mV (see Figure 5.4 g). (i–l) Orientations of individual bases within pore VII in the simulations of poly(dA)$_{20}$ (black), poly(dC)$_{20}$ (red), poly(dG)$_{20}$ (green), and poly(dT)$_{20}$ (blue). The angles $\beta$ and $\gamma$ describe the tilt and rotation of the bases, respectively; their definitions are shown in panel (k) and in Figure 5.9. (j) Scatter diagram of angles $\beta$ and $\gamma$. (i, l) Normalized distributions of $\beta$ (i) and $\gamma$ (l). The snapshots illustrate typical conformations of purine and pyrimidine bases.

Figure 5.9: Definition of basis vectors used to specify the orientation of the four DNA bases. For purines, $\mathbf{e}_1$ is defined as a unit vector passing from the C4 through the N1 atoms. For pyrimidines, $\mathbf{e}_1$ passes from the C6 through the N3 atoms. To determine $\mathbf{e}_2$, a vector $\mathbf{e}_2^*$ is first defined from C4 to C6 in purines and C6 to C4 in pyrimidines. The vector $\mathbf{e}_2$ is then the unit vector proportional to $\mathbf{e}_2^* - (\mathbf{e}_2^* \cdot \mathbf{e}_1)\mathbf{e}_1$. For all bases, $\mathbf{e}_3$ is defined as $\mathbf{e}_1 \times \mathbf{e}_2$. 

57
was about 60% of the open pore current. The open-pore current, in general, was found to scale with the pore diameter. However, due to the finite size of an ion hydration shell, the effective diameter of the pore for ion transport is smaller than the pore diameter determined using coordinates of carbon atoms.

Figure 5.10: Ionic current traces obtained from all-atom MD simulations of three-layer systems (pore VII, 500 mV bias). (a) Ionic current versus time. Each data point is a 50 ns block average of the instantaneous current. Error bars are 95% confidence intervals. In the low current region of the poly(dC)\textsubscript{20} trace (25 < t < 210 ns), two cytosine nucleotides jam the pore. (b–e) Normalized distributions of blockade ionic current for poly(dA)\textsubscript{20} (b), poly(dC)\textsubscript{20} (c), poly(dT)\textsubscript{20} (d), and poly(dG)\textsubscript{20} (e) systems sampled in 2 ns blocks. Vertical dashed lines indicate the mean current, the numerical value is displayed in the plot. For poly(dG)\textsubscript{20}, the distribution was computed using data obtained at 500 mV bias, see Figure 5.8 b.

Predicting the sequence-specific ionic current directly from all-atom MD simulations requires trajectories that greatly exceed the duration of those described in this work. We illustrate this point in Figure 5.10 by plotting the ion current traces and the normalized current histograms for our MD simulations of pore VII and all four DNA homopolymers. The stochastic variations in the current are too large to draw statistically sound conclusions. Therefore, the sequence-specific difference in ionic current blockades were studied by others in our group [169] using our recently developed atomic-resolution Brownian dynamics (ARBD)
method [175]. The time between single-base steps for DNA at lower transmembrane bias is expected to be in the microsecond range, and ARBD allowed them to readily probe this timescale. They found that ionic current blockades produced by different DNA nucleotides are, in general, indicative of the nucleotide type, but very sensitive to the orientation of the nucleotides in the nanopore [169].

5.3 Discussion

Thus, we have shown that ssDNA translocation through nanopores in graphene membranes may exhibit the same features that have made biological nanopores amenable to DNA sequencing applications. We have found that translocation of ssDNA may proceed in single nucleotide steps, akin to ssDNA translocation controlled by a polymerase motor [63, 150]. Hydrophobic adhesion of nucleobases to graphene reduces the likely ssDNA conformations to a few states, which can be sequence specific. Others in our group showed that the ionic current blockades produced by immobilized DNA nucleotides may be sufficiently different to identify their type via ionic current measurement [169], which may involve measurements under alternating electric field. Our simulations suggest that multi-layer graphene membranes might be better suited to DNA sequence detection via ionic current measurement than single layer graphene, although one might have to use a hidden Markov chain base-calling method [176] to deconvolve the ionic current trace.

Our simulations also revealed potential problems with the approach. The transport of ssDNA is not always stepwise or even unidirectional and the duration of the steps is stochastic, which can lead to insertion and deletion errors in sequence determination. To optimize control of ssDNA stepping through graphene nanopores, one must minimize zero-bias displacements of a DNA strand in the nanopore. Given the exponential dependence of the probability of barrier crossing on applied force, controlled stepwise transport of ssDNA might be realized by pulsing the applied electric field. From the limited number of nucleotide conformations examined, we found the blockade current to depend to the same degree or even more on the conformation of nucleotides as on their chemical structure. Through precise chemical decoration of the nanopore surface [177], one may enhance the resolution of ionic current blockades by engineering sequence-dependent binding of nucleotides in the conformations that enable detection of the bound nucleotides.
Chapter 6

Other applications of Grid-steered Molecular Dynamics

In this chapter, I present ongoing work using G-SMD. In Section 6.1, I describe work to model proteins and other biomolecules using grids rather than an explicit atomic representation. In Section 6.2, I describe an expansion of the G-SMD code to facilitate simulation of thermal gradients in simulation systems.

6.1 Modeling biomolecules

Modeling all atoms in a system maximizes detail and, within the limits of the force field describing the interatomic interactions, also maximizes accuracy. However, explicit treatment of all atoms limits the size and duration of a simulation, and very often yields an unnecessary amount of detail about uninteresting parts of a system. In other words, computational effort is not ideally distributed. Much work has been done on implicit solvents [178] to address this concern, since solvent makes up ∼90% of the atoms in a typical simulation, while the researcher is usually most interested in the solute. On the other hand, much less has been done treating other parts of a system, such as biomolecules and membranes, without explicit atoms or coarse-grained beads. A field-based treatment of biomolecules promises to further focus computational effort, maximizing the useful information derived from a simulation.

G-SMD is the methodological basis for the goal of implicit representation of biomolecules. The work is planned to proceed in stages, first reproducing the hydrophobicity of all-atom surfaces, then protein-protein interactions, and finally transport properties. Work is currently under way to reproduce the hydrophobicity of various all-atom silica surfaces [179]. The hydrophobicity is determined the equilibrium water contact angle of a water droplet on the surface. This project involves reproducing van der Waals interactions with oxygen and hydrogen, as well as electrostatic interactions.

The next stage is to demonstrate the ability to model proteins in grid form. This represents the next level of complexity, as a protein surface is heterogeneous and fluctuates significantly. We will first model a globular protein in solution, and ensure that we can properly capture its interaction with the surrounding solvent. We will then show that a grid-modeled protein can interact realistically with another protein by
Figure 6.1: Replacing the all-atom representation of the membrane protein α-hemolysin (left) with a grid representation (right). By using a grid representation, computational effort may be saved. Additionally, grids may be used to model structures for which no all-atom model exists.

taking a homodimeric protein and modeling one subunit with a grid, keeping the other subunit in all-atom form. This will involve reproducing a multitude of specific interactions.

Finally, we will attempt to reproduce the transport properties of α-hemolysin, modeling the protein with a number of grids while keeping the rest of the system, including the lipid bilayer, in all-atom form. See Figure 6.1 for a representation of this system. Modeling this system and reproducing characteristics such as ionic conductivity will stringently test the limits of grid-based biomolecular representation, as these characteristics are sensitive to protein flexibility in addition to charge distribution.

Others in our group plan to extend this idea to moving grids. Proteins or other biomolecules would be modeled as grids as above, but in addition would be free to move and interact with one another. This will require modification of the integrator to update the positions and orientations of the moving grids, as well as modeling of the grid-grid interactions.

6.2 Thermal gradients

G-SMD may be modified relatively easily to achieve other forms of control in simulations. Others in our lab are currently investigating nanoscale thermal gradients and their effect on transport through a nanopore. With a simple modification of the implementation, I was able to link the G-SMD functionality to a thermostat. This was used to create a heat sink in a region of the solvent, while a Si$_3$N$_4$ membrane was kept at higher temperature. Preliminary data is shown in Figure 6.2. Work on this project is ongoing.
Figure 6.2: Molecular dynamics simulations of heated Si$_3$N$_4$ nanopore. (a) Snapshot of an MD simulation, demonstrating the nanopore (gray), heated region (red) and ions (K$^+$, yellow; Cl$^-$, green). (b) Dependence of the total ionic current obtained in MD simulations on the $T_H$. (c) Typical profile of the temperature in the system on the example of simulations with $T_H = 595$ K. Figure created by Maxim Belkin.
Appendix A

Implementation of Grid-steered Molecular Dynamics

In the G-SMD method, an arbitrary potential is defined on a regular three-dimensional grid $V_{ijk}$. For numerical reasons, we require that the force derived from this potential be continuous across grid cells, and that the method conserves energy within numerical accuracy. These requirements lead naturally to the idea of constructing a $C^1$-continuous interpolating function (interpolant) of the potential $V(x, y, z)$, which can then be analytically differentiated to find the force.

As an interpolant, we follow Lekien and Marsden [180] in choosing a piecewise tricubic polynomial. Polynomials are an obvious choice due to the ease of differentiating them, and it is easily shown that a third degree polynomial is the minimum required to match derivatives at the cell boundaries. The general form of the interpolant is thus:

$$V(x, y, z) = \sum_{i,j,k=0}^{3} a_{ijk} x^i y^j z^k \quad (A.1)$$

The 64 coefficients $a_{ijk}$ in Equation A.1 in general differ between grid cells—hence the interpolant is piecewise—and are determined by boundary conditions. Continuity of $V$ and $\nabla V$ at each of the 8 cell corners provides $8 \times 4 = 32$ coefficients, leaving 32 more to define. These may be chosen in any manner; we choose to emphasize the smoothness of the interpolant, and therefore seek a set of higher derivatives which will also be continuous across cells. These should also be linearly independent of $V$ and $\nabla V$, and invariant under rotation of the axes [180]. This leads to the choice of

$$\left\{ \frac{\partial^2 V}{\partial x \partial y}, \frac{\partial^2 V}{\partial x \partial z}, \frac{\partial^2 V}{\partial y \partial z}, \frac{\partial^3 V}{\partial x \partial y \partial z} \right\}$$

making the full set of continuous quantities

$$\left\{ V, \frac{\partial V}{\partial x}, \frac{\partial V}{\partial y}, \frac{\partial V}{\partial z}, \frac{\partial^2 V}{\partial x \partial y}, \frac{\partial^2 V}{\partial x \partial z}, \frac{\partial^2 V}{\partial y \partial z}, \frac{\partial^3 V}{\partial x \partial y \partial z} \right\}. \quad (A.2)$$

The 64 $a_{ijk}$ in a given cell are then linearly related to the 64 quantities given in (Equation A.2) above.
This relationship is described by a matrix $\mathbf{B}$, which is unique and invertible [180]. In our implementation, the input grid only specifies $V$ at each point, while the rest are calculated from $V$ by finite differences. In principle, however, any or all of the derivatives could be provided as input as well, thus improving the accuracy of the interpolant.

Once the $a_{ijk}$ have been determined for each grid cell, the interpolant is fully defined. The force on a particular atom $i$ is then calculated as

$$f_i = s_i \mathbf{S} \cdot (-q_i \nabla V) \quad (A.3)$$

where $s_i$ is a scaling factor, $q_i$ is the coupling factor (default is the electric charge), and $\mathbf{S}$ is a diagonal scaling tensor, allowing independent scaling of the derived force components. The $s_i$ (and optionally $q_i$) are defined in an input file, allowing an arbitrary set of atoms to be forced.
Appendix B

Methodology


B.1 Grid-steered Molecular Dynamics and microtubule mechanics

MD methods

All simulations were performed using the software package NAMD [23], 1–2–4-fs multiple timestepping, CHARMM27 parameters [5] with CMAP corrections [181], a 10–12 Å cutoff for van der Waals and short-range electrostatic forces, and the Particle Mesh Ewald (PME) method for long-range electrostatics computed over a 1.0 Å-spaced grid. The temperature was maintained at 310 K using the Langevin thermostat with a damping constant of 1.0 ps$^{-1}$. For NpT simulations, the Langevin piston method was used, with a period of 200 fs and a decay of 100 fs. Parameters for GDP and GTP were adapted from those for ADP and ATP, respectively. Flexible fitting was performed using the Grid-Steered Molecular Dynamics (G-SMD) feature of NAMD [22]. A custom anisotropic pressure protocol was implemented in NAMD to independently control the diagonal components of the pressure tensor [92]. Visualization and analysis were performed using VMD [123].
Initial structural data

Coordinates for αβ-tubulin were taken from the Protein Data Bank, code 1JFF. αβ-tubulin was crystallized in antiparallel sheets and originally resolved to 3.7 Å resolution [34] and later refined to 3.5 Å resolution [35]. The structure includes αβ-tubulin, GDP, GTP, and an Mg$^{2+}$ ion in the GTP binding pocket. Taxol, used to stabilize the structure, was removed.

A cryo-electron microscopy (cryo-EM) map of a 13-PF MT at 8 Å resolution [36] was provided by the lab of Kenneth H. Downing. At this resolution, α- and β-tubulin are indistinguishable, and therefore 13-fold helical symmetry was utilized in making the map.

The 1JFF structure is missing coordinates for residues 35 to 60 of α-tubulin. Therefore, we used an alternative tubulin structure, PDB code 1SA1 [182], for residues 35 to 39 and 45 to 60. Coordinates for residues 40 to 44 were missing from the alternative structure as well; however, the remaining gap was near the stretched length of five amino acids, and the coordinates were therefore reconstructed using the Molefacture plugin in VMD [123]. This region is located on the inner microtubule surface and, using the NCBI’s Conserved Domain Database [183], is seen to be poorly conserved across species. Residue 1 of α- and β-tubulin were both missing as well, their coordinates were taken from the structure PDB code 1TUB [34].

Construction of the N and S systems

The N and S systems were built to contain two PFs each in a simulation cell, Figure 2.1 b. Each PF was made of one αβ-tubulin dimer. Through periodic boundary conditions (PBC), the PFs were effectively infinite.

Coordinates for one PF were obtained by rigid fitting the tubulin dimer to a subset of the cryo-EM map using the CoLoRes program [184] with a 2° step and Laplacian correlation. The original map was a 200 × 200 × 96 grid (∼1.7 Å/gridpoint), which was trimmed to the range [54, 148] in x, [14, 59] in y, and [0, 74] in z, for a 95 × 46 × 75 grid. The grid size corresponds, approximately, to the cross section of three PFs, and a height of three monomers. The best fit from CoLoRes was used. Coordinates for a second PF were obtained from the first by geometric transforms consistent with the MT helical lattice, i.e. a rotation of 360°/13 about the map center, and an axial displacement of (3/2) × 81.2 Å/13. This became the N system. The S system was then obtained from the N system by the exchange of α- and β-tubulin, and their associated nucleotides, in the second PF. Due to PBC, this is the equivalent of an axial shift of 40.6 Å or one monomer.

To place water and ions, we constructed two additional systems that had an extra monomer on the minus end of each PF, see Figure B.1. Coordinates for the extra monomers (two β-tubulin’s for the N
system, one α-tubulin and one β-tubulin for the S system) were obtained by shifting the coordinates of the monomers and nucleotides by 81.2 Å in the −z direction. To place water into the internal cavities of the microtubule’s protofilaments the dower sx variation of the program DOWSER was used [185]. As a result of the iterative water placement procedure, 1272 and 1275 water molecules were embedded into the N and S protofilament system, respectively. The protonation states of the histidines were assigned based on the local environment of each histidine residue. Thus, residues 28, 61, 88, 393, and 406 of α-tubulin were assigned the HSE protonation state; 8, 107, 139, 192, 283, and 309 were assigned the HSD state; and HIS:266 was assigned the HSP state. In β-tubulin, residues 6, 37, 139, 192, and 406 were identified as HSE; 28, 107, and 229 as HSD; and 266 and 309 as HSP. Using the SOLVATE program [12], a minimum 3 Å layer of water was created around each protofilament, adding 12,628 and 15,184 water molecules to the N and S system, respectively. Next, the extra monomers and nucleotides were removed; water and ion atoms were trimmed to conform to the 81.2 Å dimension of the system in the z direction. The resulting assembly was embedded in a rectangular volume of pre-equilibrated water using the Solvate plugin in VMD. The final systems contained 80 K⁺ and 10 Cl⁻ ions each (corresponding to a ~190 and ~24 mM concentration of K⁺ and Cl⁻ ions, respectively), measured 134 × 94 × 81.2 Å³, were electrically neutral, and comprised of about ~98,000 atoms.

Figure B.1: Graphical representation of solvation process for the N (top row) and S (bottom row) systems. α-tubulin is shown in red, β-tubulin in blue, and water as translucent light blue. Because DOWSER and SOLVATE do not support periodic boundary conditions, we built intermediate systems with extra monomers, labelled αₓ and βₓ. These systems were then solvated with DOWSER and SOLVATE. The resulting water shell was then trimmed to the periodic length of 81.2 Å in z, and the extra monomers removed, resulting in a system that can tile in the z-dimension without water or protein overlap, depicted on the right with periodic images shown as dashed outlines.
After minimization and 1-ns equilibration in the NVT ensemble, the cryo-EM map of a complete MT was used as an external potential to refine the conformation of the PFs. The original $200 \times 200 \times 96$ gridpoint map was trimmed to the range $[54, 118]$ in $x$, $[14, 59]$ in $y$, and $[0, 47]$ in $z$, producing a $65 \times 46 \times 48$ grid. The grid values were then inverted, shifted and rescaled to the range $[0, 0.1]$:  

$$V_{ijk}^{\text{rescaled}} = 0.1 \text{kcal/}(\text{mol} \cdot \text{amu}) \cdot \left[2 - \left(V_{ijk} - V_{\text{min}}\right)/(V_{\text{max}} - V_{\text{min}})\right] \quad (B.1)$$

The forces derived from the potential were multiplied by the mass of the protein’s or the nucleotide’s atoms. The typical scale of the features in the resulting potential was $\sim k_B T$. Both systems were simulated for 5 ns subject to the potential. Harmonic restraints of 200 kcal/mol-rad were imposed on the dihedral angles of all $\alpha$-helical and $\beta$-sheet fragments of the protein to maintain the secondary structure and avoid overfitting [24]. Subsequent equilibration is described in Section 2.2.

**Construction of the infinite MT system**

The initial model of a complete MT was built by assembling twelve copies of the $\alpha_1$-$\beta_1$ PF from the N system and one copy of the $\alpha_1$-$\beta_1$ PF from the S system along with the solvent into $2\pi/13$ radial sectors of the MT. The atomic coordinates of the PFs were taken from the states attained at the end of equilibration-III, Figure 2.2 (see Section 2.2). Additional water and ions were added, producing a $300 \times 300 \times 81$ Å$^3$ system of $\sim 750,000$ atoms. The MT was effectively infinite due to PBC.

The interface between the PFs was adjusted using the coordinates of the $\alpha$-$\alpha$, $\beta$-$\beta$, and $\alpha$-$\beta$ interfaces in the equilibrated N and S systems as a template. To produce the target coordinates for the interfaces, the N and S systems after equilibration-III were again simulated with G-SMD for 5 ns, using the same parameters as during the flexible fitting stage, ensuring that the angle and axial alignment between the two PFs matched that required for a 13-PF MT. Coordinates of atoms corresponding to a $2\pi/13$ radian sector centered on the center of mass of the two PFs were then taken from each system. The target structure was constructed from twelve copies of the N system interface coordinates and one copy of the S system interface coordinates, shown in Figure 2.9 $b$. The interface fitting was performed using the following harmonic restraint potential applied to all heavy atoms of the PFs:

$$U = -\sum_i k\beta_i (r_i - r_{i,0})^2/2 \quad (B.2)$$

The scaling factor $\beta_i = [1 + \cos (13\theta_i)]/2$, where $\theta_i$ is the radial angle of atom $i$ with respect to the MT seam (visualized in Figure 2.9 $b$), was introduced to accentuate the restraint potential at the interface atoms.
Note that in the initial model of the MT, half of all interface atoms already had the target coordinates. The restraints were applied linearly increasing the spring constant from $k = 0$ to 1 kcal/mol·Å over 5 ns. A value of $k = 1.0$ kcal/mol·Å was then maintained for 1 ns, after which simulations with $k = 0.3$, 0.1, and 0.03 kcal/mol·Å were carried out sequentially over 1, 2, and 2.88 ns, respectively. In total, the interface adjustment simulations lasted 11.88 ns and were performed with $L_z$ of the periodic cell fixed at the value suggested by the crystal structure of tubulin.

**Construction of the finite MT system**

The structure used for the twist simulations was constructed by combining three copies of the unit cell of the infinite MT system with a fourth copy that contained only the solvent, using the set of coordinates obtained at the end of equilibration-II, Figure 2.9. To minimize the extent of the system along the $z$ axis, some monomers were shifted by $3 L_z$ in the $-z$ direction, as shown in Figure 2.11 a. That system was solvated and ionized, producing an electrically neutral system of $\sim 2.83$ million atoms.

**Stress calculation**

Under the anisotropic pressure condition $P_{xx} = P_{yy} \neq P_{zz}$, the difference $\Delta P \equiv P_{zz} - P_{xx}$ must be borne by the PFs [92]. The total force on the PFs is then $\Delta P$ multiplied by the cross-sectional area of the system:

$$f(t) = [P_{zz}(t) - 1 \text{ bar}] \cdot L_x(t)L_y(t)$$  \hspace{1cm} (B.3)

The stress on the PFs is the negative of this divided by the cross-sectional area of the PFs:

$$\sigma(t) = -f(t)/N_{PF}A_{PF} = -[P_{zz}(t) - 1 \text{ bar}] \times L_x(t)L_y(t)/N_{PF}A_{PF}$$  \hspace{1cm} (B.4)

where $N_{PF}$ is the number of PFs, either 2 for the N and S systems or 13 for the MT system, and $A_{PF}$ is the cross-sectional area of a PF, taken as $50 \times 50$ Å$^2$. Strain was computed as $\epsilon(t) = [L_z(t) - L_{z,0}]/L_{z,0}$, where $L_{z,0}$ is the equilibrium length of the periodic cell along the PFs, taken as 83.92 Å for the N and S systems and 83.08 Å for the MT system.

**Radial compression**

To model the radial compression experiments, we defined the compression force acting on atom $i$ as a moving step function directed toward a plane passing through the MT axis: $f_i = -f_0 n(r_i \cdot n)/|r_i \cdot n|$ when
\[ |\mathbf{r}_i \cdot \mathbf{n}| \leq d_0 - vt, \text{ and } f_i = 0 \text{ otherwise. Here } |\mathbf{r}_i \cdot \mathbf{n}| \text{ is the distance of atom } i \text{ from the plane (the MT axis passes through the origin of our coordinate system), } f_0 = 100 \text{ pN, } d_0 = 140 \text{ Å, } v = 12.5 \text{ Å/ns, and } \mathbf{n} \text{ is the unit vector normal to the plane. The force was applied to all heavy atoms of tubulin using the TclBC feature of NAMD [186].}

**Twist**

The twist deformation was produced by forcing the tubulin monomers located at the ends of the finite MT system to rotate in opposite directions. The force on atom \( i \) was computed as:

\[
f_i = C \gamma(\mathbf{r}_i)(\mathbf{k} \times \mathbf{r}_i)m_i/M
\]

where \( C = \pm 1 \text{ kcal/mol} \cdot \text{Å}^2 \), \( \gamma(\mathbf{r}_i) = \pm 1 \) for the atoms located at the plus (+) and minus (−) ends of the MT, \( \mathbf{k} \) was the unit vector directed along the MT axis, \( m_i \) was the \( i \)th atom mass, and \( M \) was the total mass of all forced atoms. To maintain the integrity of the tubulin monomers subject to the external force and whose ends were exposed to solvent, some backbone dihedral angles were harmonically restrained with a spring constant of 400 kcal/mol-rad, using the extrabonds feature of NAMD. Specifically, the dihedral restraints were applied to the atoms of the terminal monomers located closer to the MT ends than the center of mass of the respective monomers.

**B.2 Grid-steered Molecular Dynamics and accelerating DNA translocation**

**Microscopic model of α-hemolysin**

An all-atom model of the α-hemolysin channel assembled with a lipid bilayer was constructed, simulated and tested extensively as described in Ref. 65. Atomic coordinates of α-hemolysin were taken from the Protein Data Bank (entry 7AHL). Coordinates of the atoms missing from the crystallographic structure were reconstructed using the psfgen structure building module of NAMD [23]. All histidine residues were assigned the HSE protonation state (pH 8.0 conditions) [65]. Water molecules were placed in the internal cavities of the protein using the Dowser program [185]. Following that, a 3 Å layer of water was created around the entire protein using the Solvate program [12], which also populated the transmembrane pore and the seven side channels with water. The resulting structure was oriented in space to align the symmetry axis of the transmembrane pore with the \( z \)-axis. Next, the protein was embedded in a patch of a pre-equilibrated...
and solvated DPPC lipid bilayer. All lipid molecules that overlapped with the protein stem were removed, along with all water molecules around the stem of the protein that overlapped with the lipid bilayer. The protein-lipid complex was solvated in a rectangular volume of pre-equilibrated TIP3P [187] water molecules. Corresponding to a solution concentration of 1 M, K\(^+\) and Cl\(^-\) ions were added at random positions located at least 4 Å away from the protein, DNA and membrane, and 3 Å away from each other. The resulting system measured 135 × 137 × 148 Å\(^3\) and included 288,678 atoms. Following 2000 steps of minimization with all protein atoms fixed, the system was equilibrated in the NPT ensemble for 1.3 ns with the backbone of the protein restrained, and for another 3.0 ns without any restraints [65].

**Microscopic models of single-stranded DNA/\(\alpha\)-hemolysin**

Single strands of DNA were threaded through the \(\alpha\)-hemolysin pore using the phantom pore procedures described in detail in Refs. 112 and 186. A 58-base pair double-stranded DNA helix was built from individual base pairs in the geometry suggested by Quanta [188]. Single-stranded DNA was obtained from that structure by removing one of the strands. The remaining strand was then solvated in a pre-equilibrated volume of TIP3P water molecules; K\(^+\) and Cl\(^-\) ions were added, corresponding to a 1 M concentration. The resulting system was equilibrated for 12 ns. The poly(dA)\(_{58}\), poly(dC)\(_{58}\) and poly(dAdC)\(_{29}\) strands were constructed from the equilibrated DNA strand by replacing DNA bases with adenine or cytosine.

In order to thread a DNA strand through the \(\alpha\)-hemolysin pore, KCl electrolyte was built around a DNA strand, conforming to the shape of the \(\alpha\)-hemolysin pore [112]. The shape of the \(\alpha\)-hemolysin pore was represented by a mathematical surface, which we refer to as a phantom pore. Initially, the phantom pore was made 2 nm wider in diameter than the pore of \(\alpha\)-hemolysin, so that the entire 58-nucleotide strand could fit into it. The pore was then gradually shrunk in a 2 ns simulation to the shape of the \(\alpha\)-hemolysin pore. At the same time, 10 pN forces pushed all atoms laying outside of the shrinking surface toward the center of the pore. At the end of the simulation, the DNA strand adopted a straight conformation that conformed to the shape of the \(\alpha\)-hemolysin pore. We carried out two such simulations (in the NPT ensemble) corresponding to the two global orientations of the DNA strand inside the pore. Each DNA strand, as well as the ions found in the stem region of the phantom pore, were merged with the all-atom model of the \(\alpha\)-hemolysin channel. In the resulting structure, water and ions covered the DNA strand completely; one such system is shown in Figure 3.3 a. The final systems measured 135 × 137 × 183 Å\(^3\) and included 356,065 atoms in the case of poly(dA)\(_{58}\), 355,952 atoms in the case of poly(dC)\(_{58}\) and 356,005 atoms in the case of poly(dAdC)\(_{29}\). Following 2000 steps of minimization with all DNA atoms fixed, each system was equilibrated for 2 ns in the NPT ensemble.
MD methods

All our MD simulations were performed using the program NAMD [23], periodic boundary conditions and Particle Mesh Ewald (PME) full electrostatics [189] with a dielectric constant \( \epsilon = 1 \). PME was computed using a grid spacing of \( \sim 1.1 \) Å per grid point in each dimension. Systems comprised entirely of water, ions and nucleic acids were simulated using the AMBER95 [190] force field; the CHARMM27 [5] force field was employed for all other systems. The temperature was kept at 295 K by applying Langevin forces [191] to all heavy atoms; the Langevin damping constant was set to 1 ps\(^{-1}\). The integration timestep chosen was 1 fs. The equilibration in the NPT ensemble was performed using the Nosé-Hoover Langevin piston pressure control [192] at 1 bar. Van der Waals energies were calculated using a smooth (10–12 Å) cutoff. Restraints were imposed by harmonic forces; the force constants were set to 1 kcal/mol·Å\(^2\). All simulations in an external electric field were carried out in the NVT ensemble.

Electrostatic potential

The procedures for determining the average distribution of an electrostatic potential from MD trajectories were described in detail elsewhere [65]. A 5.3 ns simulation of the \( \alpha \)-hemolysin system containing no DNA was carried out while applying a uniform external electric field equivalent to a 1.2 V bias. The simulation produced 5300 snapshots of the system configuration. For every configuration, an instantaneous distribution of the electrostatic potential was computed using the PME electrostatics module of NAMD [23]. The instantaneous potentials were averaged over the entire MD trajectory and over the sevenfold symmetry of \( \alpha \)-hemolysin. A slice through such a potential along the \( xz \)-plane is shown in Figure 3.3b. The electrostatic potential used in our G-SMD simulation was computed over a 96 \( \times \) 96 \( \times \) 96 grid using Gaussians of width \( \beta = 0.395 \) Å\(^{-1}\).

Grid-SMD simulations

To carry out G-SMD simulations, the systems were simulated in the NVT ensemble, applying an external electric field equivalent to a 1.2 V bias. Steering forces derived from the 3-D distribution of the electrostatic potential were applied to all atoms of DNA using either a \texttt{tclforces} script or a custom version of NAMD. These forces were computed from the finite differences of the potential and were scaled with the charge of the DNA atoms and with the scaling factor \( N \). Only the \( z \)-components of the forces were applied. In our G-SMD simulations, we varied the scaling factor \( N \) from 0 to 10, thereby changing the effective bias from 1.2 to 13.2 V in the \( z \)-direction (effective bias is \((N + 1) \times 1.2 \) V). No steering force was applied in the \( xy \)-plane.
To keep the potential aligned with the protein, the latter was restrained through harmonic forces applied to the α-carbon atoms.

**Rate of DNA transport**

To quantitatively characterize translocation of DNA through α-hemolysin, we adapted the method used for computing ionic currents \[65,121,122\]. An instantaneous current of DNA nucleotides was computed as

\[
I_{\text{DNA}}(t) = \frac{1}{\Delta t L M_{\text{DNA}}} \sum_{i} m_{i} [z_{i}(t + \Delta t) - z_{i}(t)],
\]

where \(z_{i}\) and \(m_{i}\) are the z-coordinate and the mass of atom \(i\), respectively, \(L\) is the extension of the system along the direction of the current, and \(M_{\text{DNA}}\) is the mass of one nucleotide. The sum in Equation B.6 runs over all DNA atoms in the volume of interest, which in our case was either the entire channel or the channel’s constriction. To compute the average translocation velocity at a given bias, instantaneous currents of DNA nucleotides \(I_{\text{DNA}}(t)\) were first integrated with respect to time to produce a cumulative current curve; applying a linear regression fit to the cumulative current curve yielded the average translocation velocity.

**B.3 Slowing down DNA translocation**

All simulations were performed using the software package NAMD \[23\], 2–2–6-fs multiple timestepping, CHARMM27 parameters \[193\] with CMAP corrections \[181\], a 7–8 Å cutoff for van der Waals and short-range electrostatic forces, and the Particle Mesh Ewald (PME) method for long-range electrostatics computed over a 1.0 Å-spaced grid. The temperature was maintained at 295 K using the Lowe-Andersen thermostat. For NPT simulations, the Langevin piston method was used, with a period of 2000 fs and a decay of 200 fs. Ions were simulated with parameters produced by recent refinements to Beglov and Roux ion parameters \[194\]. Trajectory frames were saved every 10 ps. Visualization and analysis were performed using VMD \[123\].

The simulation systems consisted of a 6.4 nm thick block of annealed SiO\(_2\) \[179\] containing an 8 nm, 16 nm, or 22 nm diameter circular pore along the short axis of the block, DNA, and ionic solution of 0.1, 0.5, 1 M, or 4 M of LiCl, NaCl, or KCl. Simulations with larger systems, in which the pore was connected to a solution bath, verified that the ion concentrations quoted correspond to bulk concentrations. Systems contained 20 bp (two turns) of a random sequence. To avoid interaction of DNA with the silica surface, the center of mass of DNA phosphate atoms was radially restrained to the center of the pore, with a spring constant of 93.72 kcal mol\(^{-1}\) Å\(^{-2}\). Periodic boundary conditions were used, with DNA covalently bonded over the periodic boundaries, effectively making the simulation system an infinite nanopore containing infinite DNA.
An electric field was applied parallel to the nanopore/DNA axis. Electric potential differences reported refer to the difference over a single 6.4 nm periodic image. No absolute restraints were applied along the dimension of the pore axis of the system, and the zero momentum feature of NAMD was used [195].

Ion binding in the second solvation shell of DNA was computed as follows. An ion was considered bound if it was within 0.34 nm of a water oxygen atom, which in turn was within 0.31 nm of a DNA heavy atom. These distances, which are measured between atom centers, correspond to minima in the applicable radial distribution functions. When calculating the number of bound ions as a function of minimum bond duration, a bond is considered to persist when the same three atoms (ion, water oxygen, and DNA heavy atom) form a bond in adjacent trajectory frames.

The model system measured $2 \times 2 \times 6.4 \text{ nm}^3$. It contained 40 ions ($q = +1$) only; no water was present. Periodic boundary conditions were applied. The system was simulated using NAMD, with a dielectric constant of 80. A sinusoidal potential of the form $(U_0/2) \cos(2\pi mz/L)$ was applied to all ions using Tcl boundary forces [23], where the barrier height $U_0$ was a constant in the range of 0.25–2.25 $k_B T$, $z$ is the position along the long axis of the system, and $m = 20$ is the number of binding sites. Simulations were run for 9.6 ns, and force was output for each ion every 48 fs. A Langevin thermostat was used to control temperature, with a temperature of 295 K and a damping constant of 5.0 ps$^{-1}$. An electric field corresponding to 500 mV was applied along the long axis of the system. Timestep, cutoff, and PME parameters were the same as those used in the all-atom simulations. During each simulation, we recorded the average force the potential applied to each ion. According to the Newtons third law, the magnitude of this force equals the force exerted by each ion on the potential representing the surface of DNA.

## B.4 DNA in graphene nanopores

All MD simulations were performed using NAMD [23], 2–2–6-fs multiple timestepping, CHARMM27 parameters [5] with CMAP corrections [181], a 7–8 Å cutoff for van der Waals and short-range electrostatic forces, and the Particle Mesh Ewald (PME) method for long-range electrostatics computed over a 1.5 Å-spaced grid. The temperature was maintained at 295 K using the Lowe-Andersen thermostat. For NPT simulations, the Langevin piston method was used, with a period and decay of 200 fs. Visualization and analysis were performed using VMD [123].

All-atom models of graphene sheets were generated using the Inorganic Builder plugin [196] of VMD [123]. Pores were generated by removing carbon atoms from the graphene sheets. Carbon atoms were removed if they fulfilled the condition $(x_i/R_1)^2 + (y_i/R_2)^2 < 1$, where $(x_i, y_i)$ is the position of carbon atom $i$ in
the plane of the graphene relative to the pore center, and $R_1$ and $R_2$ are the major and minor radii of the pore, respectively. Dangling atoms connected to the sheets by only one covalent bond were then removed as well. All carbon atoms in graphene were modeled as type CA atoms of the CHARMM force field [5]. In this study, we neglect possible chemical modifications of graphene pore edges [197] and fusion of graphene layers produced by pore fabrication procedures [198]. Single-stranded DNA was then added in the orientation (cis) 5'–dN–3' (trans) (the cis chamber is defined as that with the lower electric potential), and the combined graphene-DNA system was solvated using the Solvate plugin for VMD. Finally, ions were added to achieve a neutral system containing 1.0 M KCl using the Autoionize VMD plugin. Each final system was a hexagonal prism 49.2 Å on side and 100 Å in height; hexagonal periodic boundary conditions were applied in the $xy$-plane. The systems were minimized for 2000 steps, followed by 2 ns of NPT equilibration at 1 atm pressure in which only the system dimension normal to the graphene membrane, $z$, was allowed to change. The final length of each system along this dimension was then taken as the average value over the last 1.6 ns of equilibration. Systems were then equilibrated for 4 ns in the NVT ensemble. Production simulations were performed applying an external electric field in the $z$ direction. The external fields are reported in terms of a transmembrane voltage difference $V = E L_z$, where $E$ is the electric field strength and $L_z$ is the length of the simulation system in the $z$ direction [121]. In all MD simulations, harmonic restraints with a spring constant of 40 kcal/mol·Å$^2$ were applied to a single layer of carbon atoms on the edge of the periodic cell, i.e. the layer of atoms furthest from the pore.

**Number of translocated nucleotides**

To calculate the number of translocated nucleotides reported in Figures 5.4, 5.8 and 5.5, the mass-weighted number of atoms whose centers satisfied the condition $z < 0$ were determined, which was then translated into the number of nucleotides. Each time series $N(t)$ thus obtained was then shifted by an integer $N_0$ such that the $|N(0) - N_0| \leq 0.5$. Note that $z = 0$ corresponds to the center of the membrane in each system.

**Diffusion constants**

To calculate the diffusion constants of nucleotides adsorbed to the graphene surface, the $x$, $y$ and $z$ coordinates of the C1' atom of each base were examined. Bases were considered to be adsorbed if $0 < z < 1.75 N + 3.75$ Å and $\sqrt{x^2 + y^2} > 10$ Å, where $N$ is the number of graphene sheet layers (the origin of the coordinate system is in the center of the pore). For each base, the simulation trajectory was broken into subtrajectories that had the base in question continuously adsorbed to graphene. Subtrajectories less than 100 ps in duration were discarded; the maximum duration of subtrajectories was limited to 2 ns. Each subtra-
jectory was then further divided into fragments of variable length \( t \), and the fragments were used to calculate the square displacement of the base in the \( r \) and \( \phi \) directions, where \( r \equiv \sqrt{x^2 + y^2} \) and \( \phi \equiv \arctan(y/x) \). In order to find the dependence of diffusion on the distance from the nanopore, the average radial distance \( r \) of the base was calculated for each subtrajectory. For a given value of \( r \), square displacements were averaged over the fragments of length \( t \) to yield a plot of the mean square displacement (MSD) \( \langle x^2 \rangle \) versus time \( t \), where \( x \) was either \( r \) or \( r\phi \). The MSD was calculated for \( 0.2 < t < 1 \) ns, and the diffusion constant was determined for each direction using \( 2D = \partial \langle x^2 \rangle / \partial t \), where the slope \( \partial \langle x^2 \rangle / \partial t \) was determined from a linear fit. Note that drift along the \( r \) coordinate (caused by DNA translocation through the nanopore) was not corrected for.

**Nucleotide orientations**

To describe the orientations of DNA bases, a set of basis vectors was defined as follows. For purines (A and G), the vector \( \mathbf{e}_1 \) was defined in the direction of the vector connecting the C4 atom to the N1 atom. The vector \( \mathbf{e}_2 \) was defined by taking the vector from C4 to C6, then subtracting the projection of \( \mathbf{e}_1 \). For pyrimidines (C and T), the vector \( \mathbf{e}_1 \) was defined from C6 to N3, and \( \mathbf{e}_2 \) was defined from C6 to C4 minus the projection of \( \mathbf{e}_1 \). The vector \( \mathbf{e}_3 \) was then defined as \( \mathbf{e}_3 \equiv \mathbf{e}_1 \times \mathbf{e}_2 \). To describe these orientations more succinctly, we used two angles \( \beta \) and \( \gamma \). The angle \( \beta \) is defined as the angle between the vectors \( \mathbf{e}_3 \) and \( \mathbf{z} \), where \( \mathbf{z} \) is normal to the graphene plane. The angle \( \gamma \) is then defined as the angle between the vectors \( \mathbf{e}_2 \) and \( \mathbf{r} \), where \( \mathbf{r} \equiv (\mathbf{e}_3 \times \mathbf{z}) \times \mathbf{e}_3 \). See Figure 5.8 \( k \) and Figure 5.9 for graphical representations of the vectors and angles described here.
Appendix C

Publications


References


