



UPPSALA
UNIVERSITET

1FA598

Examensarbete 30 hp
September 2020

Computational study of single protein sensing using nanopores

Sebastian Cardoch

Masterprogrammet i fysik
Master Programme in Physics



UPPSALA
UNIVERSITET

**Teknisk- naturvetenskaplig fakultet
UTH-enheten**

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
<http://www.teknat.uu.se/student>

Abstract

Computational study of single protein sensing using nanopores

Sebastian Cardoch

Identifying the protein content in a cell in a fast and reliable manner has become a relevant goal in the field of proteomics. This thesis computationally explores the potential for silicon nitride nanopores to sense and distinguish single miniproteins, which are small domains that promise to facilitate the systematic study of larger proteins. Sensing and identification of these biomolecules using nanopores happens by studying modulations in ionic current during translocation. The approach taken in this work was to study two miniproteins of similar geometry, using a cylindrical-shaped pore. I employed molecular mechanics to compare occupied pore currents computed based on the trajectory of ions. I further used density functional theory along with relative surface accessibility values to compute changes in interaction energies for single amino acids and obtain relative dwell times. While the protein remained inside the nanopore, I found no noticeable differences in the occupied pore currents of the two miniproteins for systems subject to 0.5 and 1.0 V bias voltages. Dwell times were estimated based on the translocation time of a protein that exhibits no interaction with the pore walls. I found that both miniproteins feel an attractive force to the pore wall and estimated their relative dwell times to differ by one order of magnitude. This means even in cases where two miniproteins are indistinguishable by magnitude changes in the ionic current, the dwell time might still be used to identify them. This work was an initial investigation that can be further developed to increase the accuracy of the results and be expanded to assess other miniproteins with the goal to aid future experimental work.

Handledare: Ralph Scheicher
Ämnesgranskare: Biplab Sanyal
Examinator: Andreas Korn
000000000000

Populärvetenskaplig sammanfattning

Det här arbetet undersöker huruvida nanoporer kan användas för att studera stora mängder protein. Proteiner är små biomolekyler, vars funktion är att binda till andra biomolekyler inuti alla levande organismer. De mest intressanta proteinerna återfinns i den mänskliga cellen, eftersom kunskap om dessa kan leda till genombrott för nya mediciner och preventiva medicinska åtgärder. En utmaning är att utveckla experimentella metoder som kan användas för att detektera och identifiera protein i stora mängder. En teknik som har potential för detta användningsområde är nanoporer; små porer med endast några få nanometer i diameter. De tillverkas i antingen syntetiska eller biologiska material. I den här studien har en syntetisk nanopor av kiselnitrid med cylindrisk form undersökts. Hur detekteras då protein i en nanopor?

Ett membran, i vilket nanoporen existerar, separerar två regioner som är helt fyllda med en elektrolytisk lösning. Protein placeras i en av dessa regioner och en spänning appliceras över systemet. Laddade lösta partiklar, så som joner och protein, rör sig då i det elektriska fältets riktning genom nanoporen, från ena sidan av membranet till den andra. Jonernas rörelse ger upphov till en elektrisk ström som kan mätas. När ett protein passerar genom poren kallas det translokation, och detta ger upphov till en minskning av strömmen eftersom jonflödet delvis blockeras. Beroende på proteinets storlek och sammansättning minskar jonflödet olika mycket och under olika lång tid. Hypotesen är att man genom att titta på förändringen hos den uppmätta strömmen kan särskilja olika protein från varandra.

Tillvägagångssättet i den här studien var att simulera ett system med en nanopor, ett protein och en jonlösning genom att använda teknikerna "Molecular mechanics" (MM) och "Density Functional Theory" (DFT). MM använder Newtons rörelselagar för att approximera systemet medan DFT tar hänsyn till elektronernas interaktioner och därmed beskriver systemet bättre. Nackdelen med DFT är dock att det är mer beräkningstungt och därför bara kan användas för att simulera ett litet system med några få atomer (hundratals till tusentals).

I den här studien har jag fokuserat på att studera miniproteiner, som är byggstenar i stora protein. Tanken är att studien av miniprotein kan bana väg för att systematiskt kunna studera större och mer komplexa protein. De två valda miniproteinen valdes då deras form och storlek påminner om varandra. På så sätt blir det en större utmaning att skilja dem åt.

Med hjälp av MM simulering beräknades storleken på den ström som uppstår när proteinerna var för sig befann sig i nanoporen. Strömmen mättes för 0.5 och 1.0 V applicerad spänning. Det visade sig att de resulterande strömmarna såg liknande ut för de två proteinerna, för båda spänningarna. DFT användes sedan för att estimerar hur lång tid det tog för proteinerna att ta sig igenom nanoporen, baserat på hur starkt olika aminosyror i proteinerna interagerade med porväggen. Resultaten visade att de båda proteinens relativa uppehållstid skiljde sig med en faktor tio. Dessa resultat indikerade att det skulle kunna vara möjligt att detektera olika proteiner med denna metod. Studien kan användas som en guide till experimentellt arbete inom detta område. Jag presenterar även en simuleringsmetod som kan användas för att vidare studera andra miniprotein.

Contents

1	Introduction	7
2	Physics of sensing proteins with nanopores	10
2.1	From tunable resistive pulse sensors to silicon nitride nanopores	10
2.2	Entropy barrier and physics near the pore	11
2.3	Noise, resistance and detection	13
2.4	Proteins, domains and miniproteins	16
3	Computational theory	19
3.1	Molecular mechanics	19
3.1.1	CHARMM force field	20
3.1.2	Particle mesh Ewald electrostatics	22
3.1.3	Water, ensemble conditions and simulation constraints	23
3.2	Density functional theory	25
3.2.1	Pseudopotential	26
3.2.2	SIESTA	28
4	Molecular mechanics methods	31
4.1	Silicon nitride membrane	31
4.2	Selected miniproteins	32
4.3	Preparing and executing molecular mechanics	33
4.4	Ionic current, potential energy and harmonic restraint	35
5	Quantum mechanics methods	36
5.1	Amino acids and silicon nitride slab	36
5.2	Exchange-correlation functional	38
5.3	Parameters for electronic structure calculations	38
6	Results	39
6.1	Occupied pore ionic current	39
6.2	Estimate of the pore wall and amino acid interaction	45
6.3	Comparative dwell times	47
7	Discussion	50
8	Conclusion	54
9	Acknowledgements	56
	References	57
	Appendix A: Details on proteins and titration	61
	Appendix B: Details on molecular dynamic simulations	63
	Appendix C: Details on electronic structure calculations	72
	Appendix D: Additional results	76

1. Introduction

In this thesis I computationally investigate a method that can potentially detect and categorize proteins using nanopores. To get a deeper understanding of the motivation, aims and assumptions made, I first discuss proteins, their importance and how they are currently being sensed.

I begin with amino acids, which are building blocks for proteins. There are 20 naturally occurring proteinogenic amino acids and these join together covalently to form chains that, depending on length, are called peptides or polypeptides. A protein is constructed by one or more polypeptides and has as its function to bind to other molecules. Figure 1.1 outlines the biological steps from genetic code to the formation of proteins. Proteins can be further understood in terms of structures; a protein's primary structure refers to the amino acid sequence, while local folds such as α helices, β sheets and turns are its secondary structure [1]. Tertiary structures describes the relative position and folds between secondary structures. Finally, if the molecule is made of more than one polypeptide, its quaternary structure describes relative positioning and orientations in these larger sub-units [1]. Proteins have been found to be responsible for different functions in organisms i.e. molecule transport, cell structure, acting as a catalyst for metabolic functions or DNA replication [2].

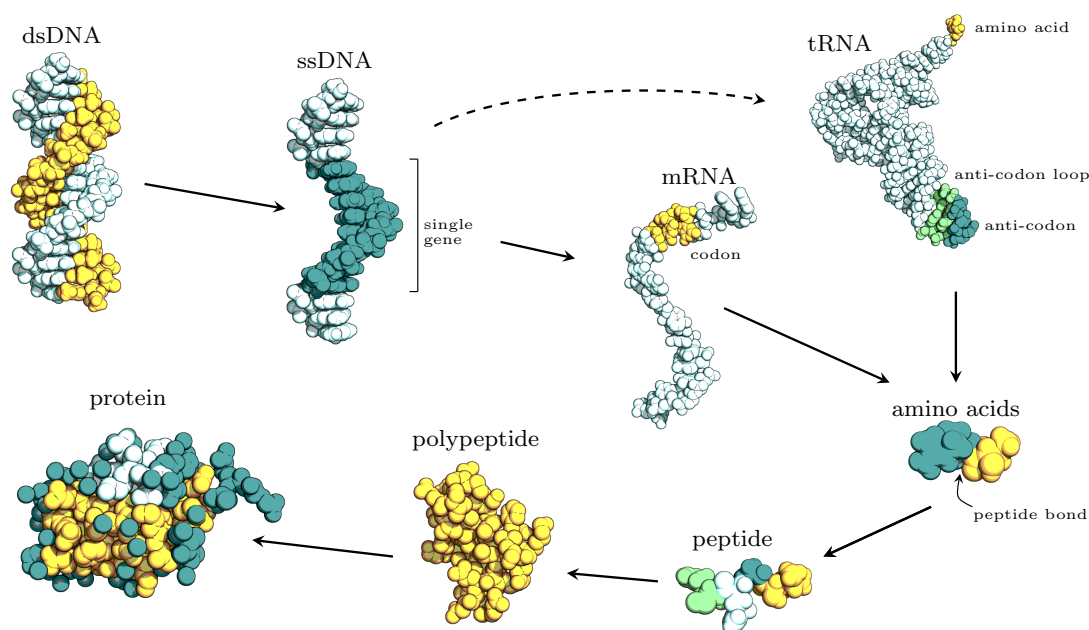


Figure 1.1. Diagram outlining the biological mechanism for the formation of proteins. The process begins when double-stranded DNA (dsDNA) splits into single-stranded DNA (ssDNA). A section of the sequence in ssDNA, or gene, is transcribed into RNA [3]. Each gene in ssDNA codes for a particular RNA i.e. messenger RNA (mRNA) or transfer RNA (tRNA). These two RNA molecules in particular are involved in the formation of proteins; mRNA holds the genetic blueprint in sets of three nucleotides called codons, whereas tRNA holds a single anti-codon and amino acid [3]. Codons bind with anti-codons one at a time. At the same time, at the other end of tRNA, peptide bonds begin to form between neighboring amino acids [3]. This process forms small chains called peptides, larger chains with some folds called polypeptides and eventually proteins.

An emergent area of research and main motivation for this thesis is single cell protein analysis. This field aims to distinguish and quantify the protein content in a single cell to track cell mutations and potentially detect diseases based on a cell's protein content [4]. Different techniques are currently used to study proteins, for example Edman degradation splits the molecule and uses mass spectroscopy to determine a protein's primary structure [1]. Varying degrees of folds can be determined using x-ray crystallography, nuclear magnetic resonance (NMR) [1]. These processes suffer from a number of disadvantages, since they require a bulk number of proteins that in some cases cannot easily be obtained [1]. Additionally, none of these methods give insight into binding, reaction or dynamic processes [1]. Different types of single molecule sensors, such as atomic force microscopy, optical tweezers, fluorosequencing, and Förster resonance energy transfer (FRET) fingerprinting, have emerged as new alternatives to study the dynamics of proteins [1, 5]. Such methods require molecules to be labelled and struggle to detect small energy differences and resolve short time steps [5].

Nanopore sequencing is a unique single-molecule sensing technique at the forefront of innovation that has the potential to circumvent these drawbacks [5]. Oxford Nanopore Technologies, first to offer a commercial nanopore-based DNA sequencer, is a prime example of the possibilities that nanopore sequencing brings [6]. Most recently, nanopores are also being considered as an efficient method to sense proteins to get concentrations and observe dynamic changes. In fact theoretical investigations on folding have already been explored and show that it is possible to obtain distinct ionic signatures for a selected number of proteins in their folded and unfolded state [5]. Nanopore sequencing requires either a solid-state or biological membrane (examples shown in figure 1.2) to be placed in electrolyte solution with a known salt concentration. An electric potential difference applied across the membrane, forces charged particles close to the pore's entrance to electrophoretically translocate through. The flow of ions in the pore is measured as a steady current that will be modulated when a molecule is found inside the pore. Solid-state nanopores are being considered in particular for their great mechanical stability, large degrees of freedom in the shape of the pore and in the case of silicon-based ones their material compatibility with current technology [1].

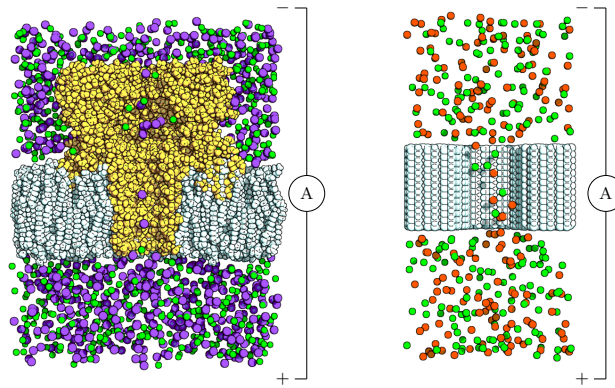


Figure 1.2. Biological and solid-state nanopores. (left) Commonly used biological nanopore α -hemolysin (α -HL) embedded in a lipid bi-layer surrounded by chloride and sodium ions. In solution the nanopore has a 3.6 nm and 2.6 nm diameter at the cap and β -barrel respectively [7]. (right) Solid-state β - Si_3N_4 membrane surrounded by chloride and potassium ions. A pore of the desired diameter can be drilled by electron and ion beams or wet-etching techniques [7].

This thesis thus builds on the potentials of solid-state nanopores over other sequencing methods to sense proteins. I approach the problem from a theoretical perspective by using a combination of classical mechanics and quantum mechanics computational methods to describe a system consisting of silicon nitride membrane, miniprotein (small protein) and solution. The objectives of this thesis are the following:

1. Use force field (FF) molecular dynamics (MD) simulations to compute occupied pore i_{occ} ionic current by placing a miniprotein inside the pore.
2. Use FF MD results to estimate the interaction strength between miniprotein and pore wall, based on changes to the potential energy of the protein while near the membrane and far away from it.
3. Use quantum mechanics (QM) in the form of density functional theory (DFT) calculations to compute the interaction strength between pore wall and individual amino acids that make up the protein.
4. Based on the results above repeated over two miniproteins, we aim to gain insight into the potential of nanopores to classify these and other proteins.

These objectives are based on the following assertions:

1. We consider that the occupied pore current, when the miniprotein is placed at the centre of the pore, corresponds to ionic current amplitudes of typical translocation events.
2. We hypothesize that the interaction strength is proportional to the duration of typical translocation events.
3. We further hypothesize that together the ionic current amplitude and interaction strength can be used as a measure to assess differences in the translocation output signals of different miniproteins.

The implementation of MD and QM methods to address the objectives of this thesis is a compromise between accuracy and speed. Good description of the total energies of the system can be obtained by performing electronic structure calculations. Due to the large size of the system, vast computational power that are not available would be needed to simulate the entire nanopore system. Through MD simulations I can consider a large region of interest using a simplified model based on parametrized forces.

2. Physics of sensing proteins with nanopores

The use of nanopores for protein sensing offers an advantage over other sensing methods, because nanopores have the potential to reach single-protein resolution with high throughput rate. One relevant field discussed in the introduction that could benefit from nanopore technology is single-cell protein analysis, which can be used to study cell evolution through the large-scale study of proteins also known as proteomics [8]. The protein content in a cell is stochastic, therefore cell variations can be spatially and temporally tracked only by significant statistical changes to protein counts [8]. Current protein sensing methods, i.e. cell array and mass spectroscopy, use a bulk number of cells and inherently yield an averaged result with low temporal or spatial resolution [8]. True single-cell resolution has been achieved by different groups with varying degrees of success, but these techniques require labelling (i.e. fluorescence antibody binding) that limits the study to a few selected proteins [8–10]. The use of nanopores to sense proteins would eliminate the need for labels, while achieving true single-cell resolution [11]. Nevertheless, the task remains difficult since small proteins ($M < kDa$) in their native state translocate too quickly and the bandwidth required to detect them yields a signal that is largely obscured by noise [1, 12]. Additionally, the number of proteins in a cell puts a limits on the number of samples that can be used to derive statical differences and protein amplification methods are not currently known [8]. In this chapter, I present background information on solid-state nanopores and miniproteins, which are small protein domains that have the potential for good identification due to their size and low net charge.

2.1 From tunable resistive pulse sensors to silicon nitride nanopores

A tunable resistive pulse sensor (TRPS) is a device that measures a current through an aperture. As briefly described in the introduction, this is done typically by an insulating membrane with a small opening that separates electrolyte solution into two regions *cis* and *trans* [13]. It is possible to apply a bias voltage V_{bias} that causes ions in the solution to feel a force that drives them into a directed motion. This motion is then measured as a steady current by electrodes placed at the end of the two regions. For large enough ion concentrations most of voltage difference occurs across the surface of the membrane, limiting the directed flow of ions inside or near the aperture of the pore. Coulter counters are one of such TRPS devices, typically microns in size and routinely used for cell counting [13]. Coulter counters have served as an inspiration in the development of nanopores, which are a nano-sized version of them ranging from 1-100 nanometers in size [14]. Given the size of the TRPS, the physics near the pore are different. Coulter counters are dominated by bulk behavior, but in the case of nanopores, in addition to electrophoresis, surface charge and electroosmosis play a relevant role [13]. The interplay between these three physical phenomena will be discussed in detail later. A charged polymer such as DNA or protein introduced in the TRPS environment will also translocate from one chamber to the other under the influence of V_{bias} [15]. During the translocation process the flow of ions is restricted, which experimentally was shown correlates with modulations in the measured ionic current [15]. The shape and duration of these modulation can give structural information about the biological molecule i.e. ion-molecule interaction, geometry, size and charge [16].

Advances in nanopore technology have made it possible to create nanopore membranes of biological or solid-state origin. Biological pores, such as α -HL, bacteriophage Φ ,

MspA, have the advantage of being structurally well defined [7]. Their geometry is highly reproducible and can be modified by targeted mutations in the coding of DNA strands [7]. Biological nanopores are typically embedded in lipid bi-layers through self assembly, making their exact position difficult to control [13]. They also exhibit a low capacitance that allows low current detection with a large bandwidth [13]. Solid-state nanopores have gained attraction because, among other advantages, they can be embedded into circuit chips with the help of field-effect transistors [7]. Si_3N_4 , SiO_2 , Al_2O_3 and two-dimensional materials such as graphene and MoS_2 have all been considered as potential candidates [7, 17]. Pores of these materials are drilled in a free-standing membrane using focused ion or electron beams, where the process is usually controlled by photo-lithography [7]. Methods using wet-etching and electric breakdown have also been reported [13]. Such procedures require a clean room, making their commercial preparation a present challenge [13]. Unlike biological nanopores, the position of solid-state pores can be controlled opening the possibility of miniaturization of future devices [13]. Additional coatings on the membrane can also lead to low capacitance and large signal bandwidth, both desirable features for sensing [13].

In particular, silicon nitride Si_3N_4 exhibits thermal, mechanical and chemical stability, characteristic of solid-state membranes [7]. Silicon integrates well with metal oxide semiconductors making it compatible with integrated circuits, opening the possibility for parallelization in the analysis [7]. Pores made of this material have been shown to work well under high electrolyte concentrations and their surface charge can be modified to achieve higher sensitivity [7]. For these reasons Si_3N_4 nanopores have been experimentally and theoretically studied for DNA sequencing in a number of papers (for a review see [13, 15]). Its potential application for the classification of single proteins has also recently been demonstrated in several studies [14]. For example Houghtaling et al. showed that it is possible to classify nine proteins based on their approximate shape, molecular volume and dipole moment [18]. Additionally, Firnkes et al. showed that the direction of transport can depend on the difference in the ζ -potential between the protein and the pore, where diffusive, electrophoretic and electroosmotic forces can cause proteins to move in the opposite direction with respect to the applied electric field [19].

2.2 Entropy barrier and physics near the pore

In its natural state, molecules can take on a number of different conformations N that can be related to a conformational entropy S [15]. This entropy is mainly determined by the stiffness of the backbone and in the case of proteins by its tertiary and quaternary structure [15]. An extreme example of a molecule with high entropy would be single-stranded DNA, which holds no rigidity. Alternatively, a double-stranded DNA forms a characteristic double helix that reduces its entropy. When a polymer is found inside the membrane, the pore walls can also impede its free movement if the radius of the pore R_{pore} is comparable to the size of the polymer R_{protein} [15]. Entropy can be related to the free energy F by

$$F = E - TS = E - T k_B \ln N , \quad (2.1)$$

where E represents the interaction energy between the polymer and solvent, T is the temperature and k_B the Boltzmann constant [15]. Since the pore reduces the entropy, translocation becomes a nucleation process in which the molecule must overcome a free energy barrier as it enters the pore [15]. This mechanism is schematically shown in figure 2.1. For ion concentrations explored in this thesis, the largest voltage difference occurs between the two membrane surfaces. Subject to this environment, biological molecules approach the nanopore aided by Brownian motion and attraction to the surface of the membrane (if any exists). Both of these factors mean the capture and successful translocation of a single molecule happens in large timescales inaccessible to all atomistic simulations methods.

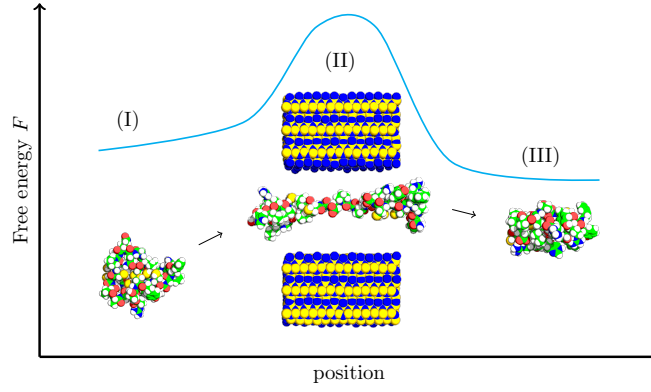


Figure 2.1. Diagram showing a schematic representation of the free energy of a polypeptide at each of the stages during its translocation. The free energy barrier occurs due to the low entropy of the molecule when found inside the pore. Adapted from [15].

Once the biological molecule is near the pore its translocation happens, in addition to random collisions with other particles, through an interplay between electrostatic forces, electrophoresis and electroosmosis [20]. These physical phenomena are a consequence of the electrolyte and the surface charge of the membrane and biological molecule [12]. When the bias voltage is off, ions in solution of opposite polarity form a charged cloud that screen the surface charge of the membrane and biological molecule [21]. The formation of the screening cloud, also referred as a double layer, coupled with the surface charge effectively creates a capacitor at the interface with an electrostatic potential described by

$$\phi(x) = \frac{\sigma_s \lambda_D}{\epsilon_w} e^{-x/\lambda_D} \equiv \zeta e^{-x/\lambda_D}, \quad (2.2)$$

where ϵ_w is the relative permittivity of water, σ_s is the surface charge density of the membrane or molecule and λ_D is the Debye screening length, which describes the thickness of the screening cloud and depends on the ionic concentration [21]. In the equation above I have defined

$$\zeta \equiv \frac{\sigma_s \lambda_D}{\epsilon_w}, \quad (2.3)$$

where ζ is known as the zeta-potential, a useful quantity for modeling the interaction between surface and solution. When the bias voltage is turned on, ions including those making up the double layer feel a Coulomb force that drives them into motion [21]. The flow of ions around a stationary surface such as the membrane is known as osmotic flow. In the case of a freely suspended particle, osmotic flow causes the particle to experience a force from the moving charge in the opposite direction, known as electrophoresis [21]. In the limit of small zeta-potential $\zeta \ll k_B T/e$ and for a symmetric electrolyte such as KCl (meaning the affinity of both ions is the same) these effects can be described as shown in figure 2.2 [21]. Under these conditions, the slipping velocity of the screening cloud can be modeled by

$$\mathbf{u}_s = -\frac{\epsilon_w \zeta}{\eta} \mathbf{E}_{\parallel}, \quad (2.4)$$

where \mathbf{E}_{\parallel} is the z -component of the electric field and η is the fluid's viscosity [21]. The electrophoretic velocity for a freely suspended particle can be modeled in a similar manner by

$$\mathbf{U} = \frac{\epsilon_w \zeta}{\eta} \mathbf{E}, \quad (2.5)$$

where in this case the factor $\epsilon_w \zeta / \eta$ describes the electromobility of the particle [21].

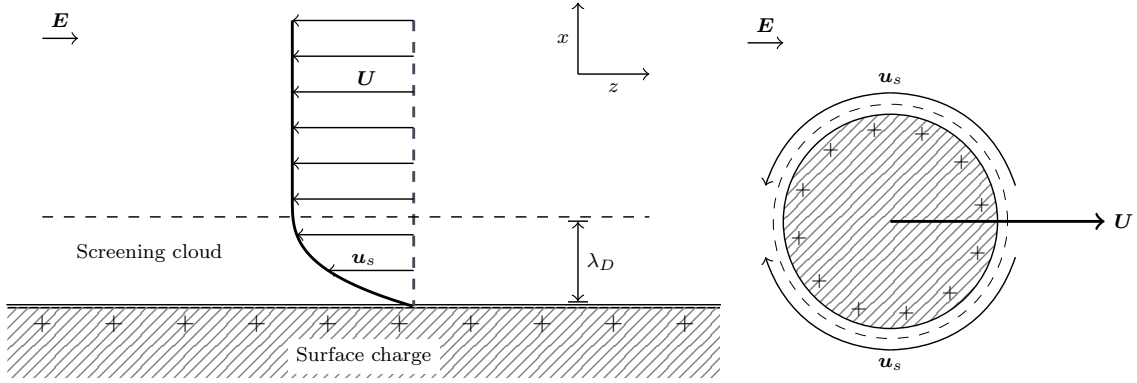


Figure 2.2. Diagram depicting an electric double layer of thickness λ_D forming at the surface of the membrane (left) and at the surface of the biological molecule (right), which under the influence of an external electric field E experience an electroosmotic flow with velocity u_s . In the case of the free-standing molecule, electroosmosis induces a movement on the molecule with velocity U known as electrophoresis. Adapted from [21].

The model considers a membrane and biomolecule with a surface charge that is well distributed and therefore generates a homogeneous zeta-potential. This is an approximation, since defects or regions of excess charge are found in the pore walls and protein. When the double layer is in motion, such defects have been shown to cause eddies and recirculation regions that require non-linear terms to describe the flow [21].

2.3 Noise, resistance and detection

The different physical phenomena described above coupled with the dynamics of the protein make translocation a stochastic process. Distinguishing between proteins then becomes challenging, since a variety of current signatures can be detected for a single polymer. Additionally, a number of sub-events can happen when a molecule translocates or is found near the pore. Besides a smooth transit, a molecule can collide with the pore, make contact with the pore walls, undergo structural changes or even tumble inside the pore [1]. All these sub-events produce distinct current blockades. As a consequence, the repeated sensing of a particular protein will give rise to a current amplitude and dwell time distribution. It has been shown that implicit entropy models can predict the nature of the energy barrier and aid in the design of nanopores, but cannot explain translocation time distributions [15]. For this task, it is necessary to employ an all-atomic model such as the ones offered by molecular dynamics simulations.

The sensitivity and temporal resolution of the nanopore become increasingly relevant to discriminate current modulations between different biological molecules, specially for short-lived translocation events [22]. In this area solid-state nanopores have the disadvantage of producing large noise compared to their biological counterpart [22]. The temporal resolution of solid-state nanopores can be enhanced with a larger bandwidth, but noise also increases with increasing bandwidth [22]. In fact studies have demonstrated that at different frequency regimes distinct noises are the dominant source of degradation in the captured signal. As shown in figure 2.3, dielectric and capacitance noise dominate the high frequency domain. The former can be explained by the fact that non-ideal dielectric materials experience electrical energy loss in the form of heat dissipation, which translates to an increase in thermal noise [22]. The latter refers to a pairing between the noise from the input impedance voltage and the capacitance of the membrane [22]. It has been demonstrated that lowering of the capacitance by applying special coatings to the surface can dampen these noises [22, 23].

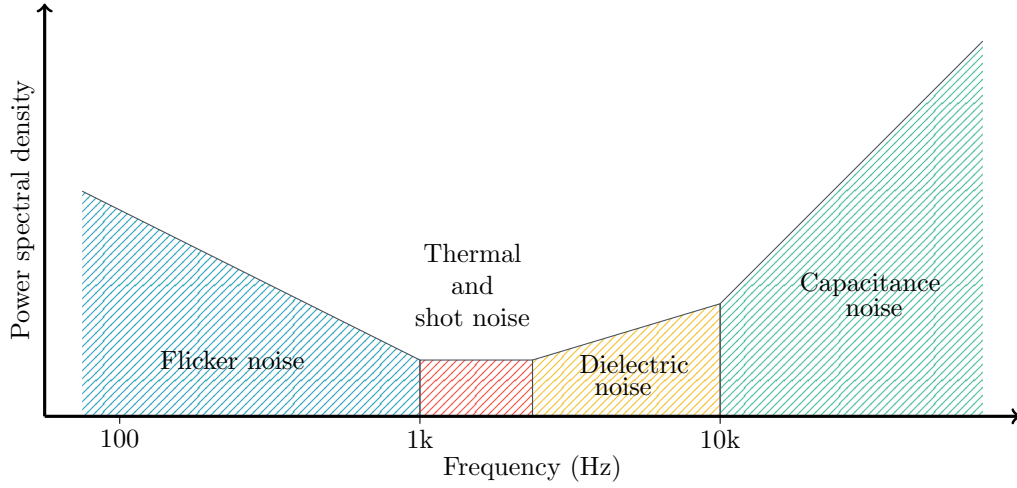


Figure 2.3. Diagram showing the dominant source of noise in a solid-state nanopore at different frequency regimes. Estimates of noise are derived from an equivalent circuit depicting a nanopore connected to a transimpedance amplifier. Adapted from [22, 24].

To get an estimate of the sensing capability of a nanopore and the type of noise that will be most dominant in the system, it is possible to model the system with an equivalent circuit like the one shown in figure 2.4. In such case, an upper limit on the cutoff frequency f_c (frequency at which the signal attenuates) can be found using

$$f_c \leq \frac{1}{4\pi C R_{\text{electrolyte}}} , \quad (2.6)$$

where $R_{\text{electrolyte}}$ the resistance associated with the solution and C is the membrane capacitance [22]. An approximation of the capacitance can be obtained from the geometry of the membrane

$$C = \frac{\epsilon_0 \epsilon_r A}{t} , \quad (2.7)$$

where A is the membrane's area of contact with the electrolyte, t is the thickness of the membrane and ϵ_0 and ϵ_r are the permittivity of free space and membrane respectively [22].

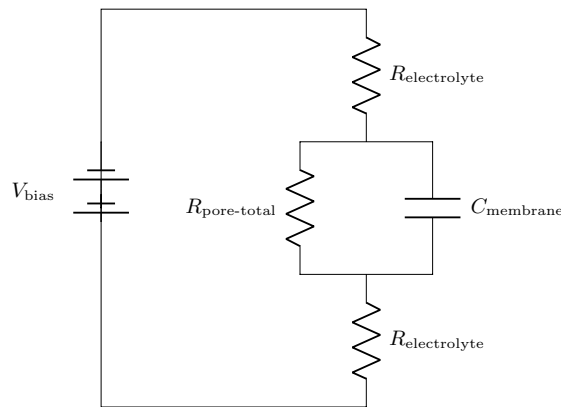


Figure 2.4. Equivalent circuit representing a simplified nanopore system, where $R_{\text{electrolyte}}$ is the resistance associated with the solution and $R_{\text{pore-total}}$ represents a combination of the pore's access resistance and resistance due to ions near its walls [22].

The total pore resistance referred to in figure 2.4 can be divided into an access resistance (AR) and a pore resistance (PR) [13, 16]. AR models the reduction in electrodiffusivity of ions when they converge from the bulk to a confined volume inside the nanopore, which

under low salt concentrations has been shown to be the dominant source of resistance [16]. In a cylindrical pore, AR can be modeled as

$$R_{\text{acc}} = \frac{1}{4\kappa d}, \quad (2.8)$$

where κ is the electrolyte conductivity, and d the diameter of the pore [13]. The PR can be considered to have two main contributors from ions at the surface of the pore walls R_{surf} and ions far from the pore walls R_{bulk} . For a cylindrical-shaped pore, these two quantities can be modeled by

$$R_{\text{surf}} = \frac{4L}{\pi d^2 \kappa} \quad \text{and} \quad R_{\text{bulk}} = \frac{L}{\pi d \sigma_s \mu_i}, \quad (2.9)$$

where L is the pore length and μ_i is the electrophoretic mobility of counter ions i [13]. These two resistance contributions add in parallel to give the pore resistance

$$\frac{1}{R_{\text{pore}}} = \frac{1}{R_{\text{bulk}}} + \frac{1}{R_{\text{surf}}}, \quad (2.10)$$

making the total resistance [13]

$$R_{\text{pore-total}} = 2R_{\text{acc}} + R_{\text{pore}}. \quad (2.11)$$

These models assume for example a constant electric potential, a neutral pore and a homogeneous electrolyte solution [16]. These conditions do not hold exactly true in experimental environments or molecular dynamic simulations. Nevertheless, if the electrolyte is well distributed, the electric field will be approximately constant over the bulk region and change rapidly near the pore, where electrophoretic forces will be present [13]. At some radius r_c the forces associated with Brownian motion and electrophoresis will approximately cancel out [13]. This radius encloses a volume known as the capture region, as shown in figure 2.5.

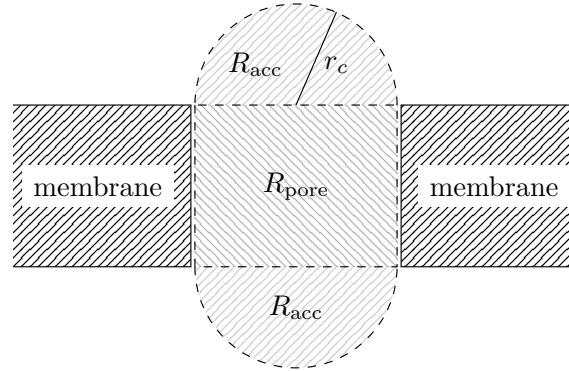


Figure 2.5. Schematic representation of resistance contributions present in a cylindrical nanopore subject to a uniform external field along the latitudinal direction.

Finally, the open pore current i_o for a cylindrical-shaped opening can be approximated by Ohm's law [1]

$$i_o = \frac{V_{\text{bias}}}{R_{\text{pore-total}}}. \quad (2.12)$$

The occupied pore current i_{occ} on the other hand, is proportional to the cross-sectional area of the pore and inversely proportional to the thickness of the pore [20]. This means thin membranes are most desirable to achieve high sensitivity [22]. Additionally, the smaller the pore size, compared to the size of the molecule, the greater signal strength will be

exhibited. This is because it is common to define the modulation in the ionic current as the ratio between the occupied and open pore currents [13]

$$I = \frac{i_{\text{occ}}}{i_0}, \quad (2.13)$$

such that if $\Delta I \equiv i_0 - i_{\text{occ}}$ and the pore size is very large, $i_{\text{occ}} \approx i_0$ and $\Delta I/i_0$ is approximately zero. In the case when the pore is very small $i_{\text{occ}} \approx 0$ and the ratio $\Delta I/i_0$ becomes approximately 1.

2.4 Proteins, domains and miniproteins

As discussed earlier, a protein is built from a sequence of amino acids that assemble into a unit by forming peptide bonds. Each amino acid is composed of a carboxyl group (COOH), an amino group (NH₂) and a radical (R) where the radical in this case is the part that differs for each type of amino acid [2]. As shown in figure 2.6, amino acids join together when the carbon atom from the carboxyl group (COO⁻) shares an electron with the nitrogen from the amino group (NH₃⁺) [2]. The linked carboxyl and amino group make up the backbone of the protein, while the radical becomes a side group [2]. Due to the way these bonds form, two ends emerge; the *N*-terminus and *C*-terminus.

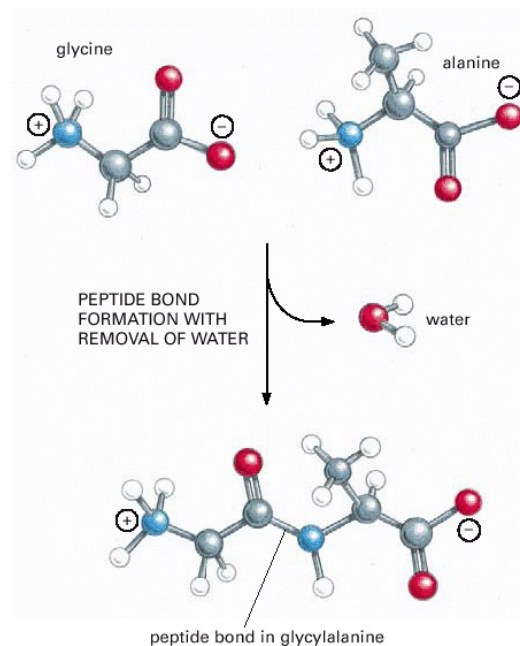


Figure 2.6. Schematic representation of a sequence of amino acids forming a peptide chain. The bonding process releases a water molecule. By convention, the protein sequence is read in the N-C terminus direction [2]. Taken from [2].

Peptide bonds are flexible and allow folds to originate exclusively based on the amino acid sequence of the protein [2]. These folds take on different angles and configurations, but are limited by non-covalent forces between atoms [2]. The entire folded structure of a protein, also called conformation, tries to minimize the energy of the molecule [2]. The stability of the molecule thus depends on its entropy and the enthalpy of the bonds that keep the folds in place [25]. Common fold patterns are α helix or β sheets, as shown in figure 2.7. Based on various folds, proteins tend to form a hydrophobic core, which generally strengthens the protein's conformation and creates binding sites that allow it to interact with other molecules [13].

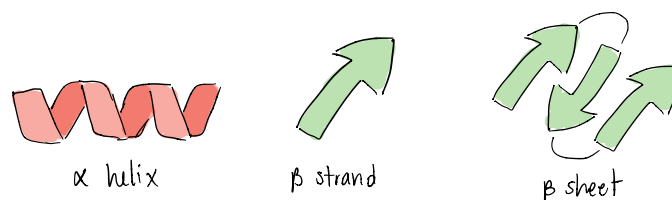


Figure 2.7. Schematic representation of common folds found in proteins.

Besides the primary to quaternary structures discussed in the introduction, a protein can also be subdivided into domains. A protein's domain is any polypeptide section that has the ability to independently fold into a stable structure [2]. It is usually the case that these domains are associated to a particular function of the protein, i.e. regulatory roles, transport or binding to other molecules [2]. Domains make it possible to study protein sensing using nanopores in this thesis. The computational power required for all-atom molecular dynamics and electronic structure calculations places a limitation in the size of the simulation system. For this reason, systems of large proteins are not feasible with the resources that I have available. I therefore turn to the investigation of small domains from proteins that are involved in cellular processes. These domains, typically a single polypeptide with less than 40-50 amino acids long, can be classified as miniproteins [25, 26]. Miniproteins are often less stable due to the fewer number of atoms, which translates to a reduced number of non-covalent bonds and a less hydrophobic core [25]. This also means a single miniprotein contains a limited number of secondary structures, making them suitable to study folds in larger proteins in systematic ways [2, 26].

I selected two miniproteins to study that are involved in cellular processes; the WW domain of FBP11 and AGRP(87-120). General information about these molecules is shown in table 2.1 and their amino acid sequence is shown in appendix A.

Table 2.1. List of miniproteins involved in human cellular processes, selected for their analogous geometry and number of residues. The structure of these proteins has been determined by NMR spectroscopy in aqueous environment. Their structure files were obtained from the Protein Data Bank (PDB) [27].

PDB	protein	structure	length ^a (Å)	mass (Da)	volume ^b (Å ³)	res.	ref.
1ZR7	FBP11 WW	β -sheet, turn, bend	30.7	3587.7	8.47×10^3	30	[28]
1MR0	AGRP(87-120)	β -strand, turn, bend	31.6	3884.5	8.90×10^3	34	[29]

^aLargest distance between any two atomic centers. ^bApproximate volume determined using MDTools *MolVolume* with a probe radius of 2.0 Å and grid size of 0.1 Å.

FBP11 is a human protein that participates in mRNA splicing, a process in which coding regions in RNA are separated from the non-coding regions to form mRNA [28, 30]. There exists two WW domains in FBP11, where each folds into a β sheet consisting of three β strands [31, 32]. The name for this domain is derived from the two tryptophan residues, given the amino acid code W, that form part of its structure [33]. These domains have been shown to bind to the protein huntingtin that is associated with Huntington's disease [28]. A disruption in the normal binding between these two proteins is theorized to contribute to the generation of this disease [31]. The WW domain of FBP11 is shown in figure 2.8.

The agouti-related protein (AGRP) is a regulatory protein that is produced in the hypothalamus, located in a central region of the brain [29]. AGRP regulates feeding behavior and energy balance by acting as an antagonist to other molecules in melanocortin receptors [29]. The C-terminal domain, corresponding to residues (87-132) of the protein, consists mainly of a cysteine knot that has the contact points that are responsible for the protein's interaction with these receptors [29]. To experimentally prove this, Jackson et al. synthesized a miniprotein AGRP(87-120) that only contains the cysteine knot [29]. For

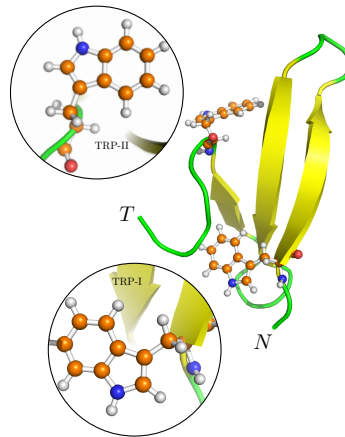


Figure 2.8. Cartoon representation of the structure of the FBP11 WW domain. Beta sheets are shown in yellow, while turns and bends are shown in green. The miniprotein has two tryptophan (W) residues highlighted using balls and sticks, that are 21 amino acids apart. The figure also shows the direction of the sequence from the N-T terminus. Between the two W amino acids, the peptide folds into a beta sheet.

example, experiments in mice have shown that AGRP is linked to weight disorder [29]. Controlling the protein-receptor dynamics can lead to the development of novel treatments for diabetes or substance abuse in humans [29]. The domain AGRP(87-120) is shown in figure 2.9

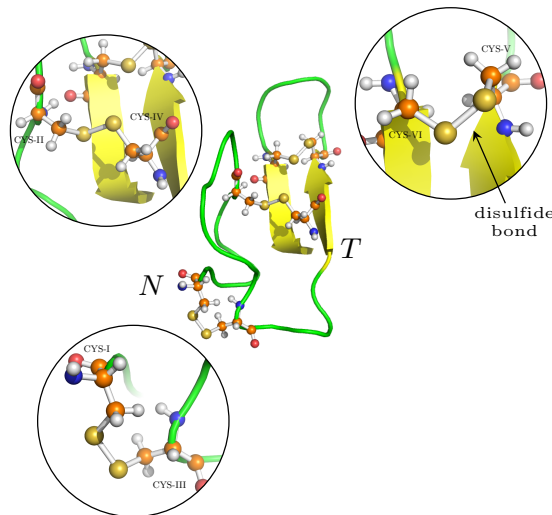


Figure 2.9. Cartoon representation of the structure of AGRP(87-120). Beta strands are represented by yellow arrows, while the turns and bends are represented by green tubes. I highlight cysteine amino acids using balls and sticks (labelled I-V following the miniprotein sequence read in the N-T direction), responsible for cysteine knots formed by disulfide bonds [34].

3. Computational theory

In this section, I present theoretical details on the simulation techniques employed to study the nanopore-miniprotein system. Molecular mechanics can be used to obtain dynamic properties for a large system at the cost of using a simplified model, while a density functional method gives accurate atomic and electronic structures at great computation expense and thus can only be used to study small regions. In this investigation I used the former method implemented in the software NAMD to study transport properties of ions and the latter implemented in the software SIESTA to quantify the interaction strength between pore-wall and miniprotein [35, 36].

3.1 Molecular mechanics

With molecular dynamics (MD) simulations it is possible to model a system made up of individual particles and obtain information about its properties as a function of time by solving Newton's equations of motion. I can start by writing the Hamiltonian of the system as

$$H = H(\mathbf{r}, \mathbf{p}) = U(\mathbf{r}) + K(\mathbf{p}) , \quad (3.1)$$

where $U(\mathbf{r})$ is the potential field under which particles interact, $K(\mathbf{p})$ represents their kinetic energy and variables \mathbf{r} and \mathbf{p} are the set of position and momenta of a single microscopic state of the system in *phase* space $\{\mathbf{r}, \mathbf{p}\}$ [37]. For a single state and for a single atom i with mass m_i , this Hamiltonian can be differentiated

$$\dot{\mathbf{r}}_i = \frac{\partial H(\mathbf{r}, \mathbf{p})}{\partial \mathbf{p}_i} \text{ and } \dot{\mathbf{p}}_i = -\frac{\partial H(\mathbf{r}, \mathbf{p})}{\partial \mathbf{r}_i} , \quad (3.2)$$

to give the equations of motion as two coupled first order differential equations

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} \text{ and } \dot{\mathbf{p}}_i = \mathbf{F}_i , \quad (3.3)$$

where the dot represents a derivative with respect to time and \mathbf{F}_i the net forces acting on the atom [37, 38]

$$\mathbf{F}_i = -\frac{\partial U(\mathbf{r})}{\partial \mathbf{r}_i} . \quad (3.4)$$

Molecular mechanics (MM) is one MD method that expresses $U(\mathbf{r})$ as a mathematical function in the form of an empirical force field (FF) where the nucleus and electrons are combined into inert spherical particles [38]. In this picture the electronic degrees of freedom are decoupled from the motions of the nucleus and are considered as an average. This idea follows from the Born-Oppenheimer approximation where the typical movement of electrons is orders of magnitude faster than the nucleus's due to their mass differences [38, 39].

The equations of motion can be written using a finite difference approach and solved by numerical integration methods. These algorithms have to be efficient, time reversible, accurate (meaning they experience a small short-term energy drift) and symplectic (meaning they experience a small long-term energy drift) [38]. Common integrators are the Verlet

and Leapfrog, where NAMD uses the former coupled with a multiple-time-stepping algorithm to determine momenta and position of particles [35].

From MM simulations, one can readily obtain the dynamic average of an equilibrium or transport property of interest, i.e. for a dynamic variable $A(\mathbf{r}, \mathbf{p})$ the dynamic average is given by

$$\langle A(\mathbf{r}, \mathbf{p}) \rangle_\tau = \frac{1}{\tau} \int_0^\tau dt A(\mathbf{r}(t), \mathbf{p}(t)) , \quad (3.5)$$

where τ is the simulation time [37]. This expression represents the time evolution of a single phase space state, but instead what is generally desired is to obtain the thermodynamic average of this property

$$\langle A(\mathbf{r}, \mathbf{p}) \rangle_Z = \int_V d\mathbf{r} \int_{-\infty}^{\infty} d\mathbf{p} \rho(\mathbf{r}, \mathbf{p}) A(\mathbf{r}, \mathbf{p}) , \quad (3.6)$$

where the first integral is over all momenta, the second over all volume and $\rho(\mathbf{r}, \mathbf{p})$ is the probability of finding the system in a particular state in phase space. For an NVT ensemble, this probability is given by the Boltzmann distribution

$$\rho(\mathbf{r}, \mathbf{p}) = \frac{\exp(-\mathbf{H}(\mathbf{r}, \mathbf{p})/k_B T)}{Z} , \quad (3.7)$$

where k_B is the Boltzmann constant, T the temperature and Z is the partition function [37]. Since Z is an integral over the whole phase space, solving for the thermodynamic average would require to have information of every single microscopic state which is a computationally expensive task [37]. From this point it is assumed that, if the system remains in a stationary state over the length of the simulation, the system is ergodic such that for very long simulation times the thermodynamic average and the dynamic average are approximately the same [37]

$$\lim_{\tau \rightarrow \infty} \langle A(\mathbf{r}, \mathbf{p}) \rangle_\tau \approx \langle A(\mathbf{r}, \mathbf{p}) \rangle_Z . \quad (3.8)$$

In practice these assumptions still require simulations to run for a long time as to obtain reliable dynamic averages, which are determined from the position and velocity of particles [38].

3.1.1 CHARMM force field

There is no unique solution to describe the quantum mechanical energy landscape of a material or molecule using a classical description [39]. To build an empirical FF it is therefore assumed that the energy landscape of the system can be split into contributions that added together yield a good description of the true potential [39]. Furthermore, FF are also assumed to be transferable meaning that their functional form and parameters (with the exception of some cases) are system-independent [39]. Many force fields are available to describe bio-molecules and most split the potential in the general form

$$U(\mathbf{r}) = U(\mathbf{r})_{\text{internal}} + U(\mathbf{r})_{\text{external}} , \quad (3.9)$$

where the internal and external terms model bonded and non-bonded interactions respectively [37]. For this thesis I chose to work with the CHARMM36 FF.

The CHARMM36 FF expands this potential energy function to describe a system like the one shown in figure 3.1 using [40]

$$\begin{aligned}
U(\mathbf{r}) = & \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{UB}} K_{\text{UB}}(S - S_0)^2 + \sum_{\text{angle}} K_\theta(\theta - \theta_0)^2 \\
& + \sum_{\text{dihedrals}} K_\chi(1 + \cos(n\chi - \delta)) + \sum_{\text{impropers}} K_{\text{imp}}(\phi - \phi_0)^2 \\
& + \sum_{\text{residues}} u_{\text{cmap}}(\phi, \psi) + \sum_{\text{non bonded}} \frac{q_i q_j}{\epsilon_1 r_{i,j}} + \epsilon_{i,j} \left[\left(\frac{R_{i,j}^{\text{min}}}{r_{i,j}} \right)^{12} - \left(\frac{R_{i,j}^{\text{min}}}{r_{i,j}} \right)^6 \right]^2. \quad (3.10)
\end{aligned}$$

In this expression K_b , K_{UB} , K_θ , K_χ and K_{imp} are the bond, Urey-Bradley, angle, dihedral angle and improper dihedral angle force constants. The terms b , S , θ , χ and ϕ are the bond length, Urey-Bradley 1,3-distance, bond angle, dihedral angle and improper torsion angle and those marked with the subscript zero represents their respective equilibrium values [40, 41]. The bonded, Urey-Bradley, angle and improper dihedral angle interactions are treated harmonically by a quadratic function, while the dihedral angles have an oscillatory behavior described by a sinusoidal function [41]. For dihedral angles the force constant determines the height of the energy barrier of the rotation whose location and periodicity are defined by the phase δ and the multiplicity n [37]. The u_{cmap} term is a special contribution acting on improper dihedral angles ϕ and ψ found in the backbone of proteins [42, 43]. This term has been added to improve the sampling of certain conformations and folded descriptions in some domains and proteins [42, 43].

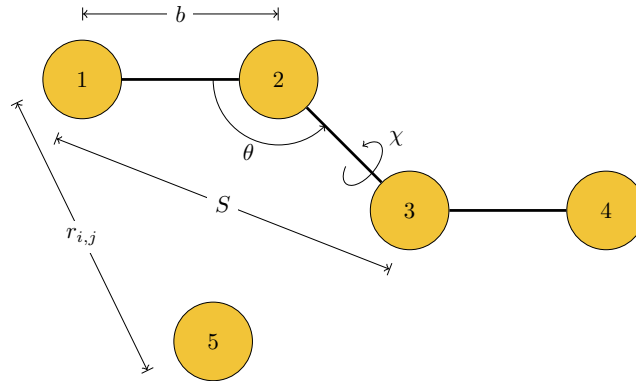


Figure 3.1. Generic system of atoms with bonded and non-bonded interactions. The potential energy of the system can be described by the additive force field CHARMM shown in equation (3.10). The Urey-Bradley term enhances the description of the vibrational spectra of the system by including a bonding term between atoms 1 and 3 and the improper dihedral angle terms treat out-of-plane distortions or prevent the molecule from flipping over to its mirrored image [37]. Adapted from [37].

The non-bonded interactions include the electrostatics between atoms i and j with charge q at a distance $r_{i,j}$ and van der Waals forces modeled by a standard 12-6 Lennard-Jones potential. Parameter ϵ_1 is the effective dielectric constant, which has been taken to be 1 in order not to unbalance the parametrization between forces that hold the molecule together (intramolecular) and forces between molecules (intermolecular) [40]. Along with the potential expression CHARMM also includes best fit parameters that are atom specific and determined either through experiments or quantum mechanical calculations, which all together make the FF [41]. Lennard-Jones parameters involving two different atom types are determined by geometric mean $\epsilon_{i,j} = \sqrt{\epsilon_i \epsilon_j}$ and arithmetic mean $R_{i,j}^{\text{min}} = (R_i^{\text{min}} + R_j^{\text{min}})/2$ [37]. The non-bonded terms have a large influence in the structure of biological molecules such as hydrogen bond formations that determine the stability

of a protein's secondary structure and influence its interaction with the external environment [37]. Based on the assumption presented earlier, equation (3.10) is a simplified form of the true potential energy of the system that is a compromise between chemical accuracy and computational load. Such approximation also limits the range of conditions for which the model remains valid, therefore molecular mechanics simulations must be executed in the vicinity of room temperature [37].

3.1.2 Particle mesh Ewald electrostatics

Because the Coulomb term in the FF falls slowly as $1/r$, electrostatic interactions between particles need to be computed over large distances. Its implementation consumes a large portion of the resources and in the beginning of computer simulations cut-off schemes were introduced to make the task more manageable [41]. Such approach lacks the description of long-range non-bonded interactions that are crucial for modeling biomolecules [41]. Nowadays full-range electrostatics can be obtained using the Ewald summation method, which builds on the premise that an infinite sum of pair-wise electrostatic interactions conditionally converges for a neutrally charged system under periodic boundary conditions [35, 44]. To calculate the electrostatic energy for a simulation box defined by three independent lattice vectors \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 with N particles and volume V , the Ewald sum takes the general form [35]

$$E_{\text{Ewald}} = E_{\text{direct}} + E_{\text{reciprocal}} + E_{\text{self}} + E_{\text{surface}} . \quad (3.11)$$

An infinite series with conditional convergence yields a different result based on the order in which it is carried out. The energy is therefore computed by first summing over each simulation box and later over spheres of boxes with increasing radius [35].

The direct energy contribution in equation (3.11) is computed in real space by pair-wise Coulomb interactions [35]

$$E_{\text{direct}} = \frac{1}{2} \sum_{i,j} q_i q_j \sum_{\mathbf{n}_r} \frac{\text{erfc}(\beta |\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}_r|)}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}_r|} , \quad (3.12)$$

where $\mathbf{n}_r = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3$, n_1 , n_2 and n_3 are integers, q_i is the charge and \mathbf{r}_i is the position of atom i [35]. Ewald summation is an approximate solution to the true electrostatic energy of the system with an error quantified by the complementary error function $\text{erfc}(\beta; \mathbf{r})$ [35]. The weight parameter β determines how much of the electrostatic sum is done in real and reciprocal space, such that after some cutoff radius r_c the real space contribution to the energy becomes negligible [35, 44]. The reciprocal term has the form

$$E_{\text{reciprocal}} = \frac{1}{2\pi V} \sum_{\mathbf{m} \neq 0} \frac{\exp(-\pi |\mathbf{m}|^2 / \beta^2)}{|\mathbf{m}|^2} \left| \sum_i q_i \exp(2\pi i \mathbf{m} \cdot \mathbf{r}_i) \right|^2 , \quad (3.13)$$

where V represents the volume of the simulation box and \mathbf{m} its dimensions in reciprocal space [35]. The last two terms of equation (3.11) are the self energy, which is independent of the size of the simulation box and has the form

$$E_{\text{self}} = -\frac{\beta}{\sqrt{\pi}} \sum_i q_i^2 , \quad (3.14)$$

and the surface energy term, which is proportional to the square of the net dipole moment and has the form [35, 45]

$$E_{\text{surface}} = \frac{2\pi}{(2\epsilon_s + 1)V} \left| \sum_i q_i \mathbf{r}_i \right|^2 . \quad (3.15)$$

This last energy term can be ignored under the so-called *tin-foil* boundary conditions, where the box is assumed to be surrounded by an infinite sphere made of a medium that is perfectly conducting [45]. The dielectric of the surroundings ϵ_s is then considered to be infinitely large $\epsilon_s \rightarrow \infty$ and the surface energy term vanishes [45]. For the system under investigation this approximation holds, since the membrane and protein are surrounded by water with low ion concentration with an approximate dielectric constant $80 \gg 1$.

The Ewald method is commonly replaced by the particle mesh Ewald (PME) method to improve the efficiency of the calculation by evaluating the reciprocal energy contribution using an interpolation scheme [44]. As shown in figure 3.2, a single charge can be distributed over a grid that allows the reciprocal term to be solved using fast Fourier transforms. Additionally, the value of β is readjusted to optimize computational scaling [44].

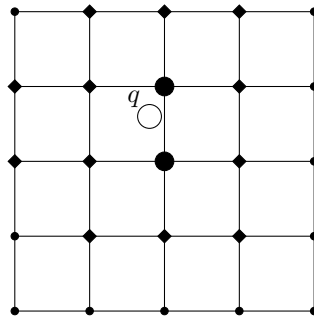


Figure 3.2. Interpolation scheme used by PME to distribute charge q over a two dimensional grid. The weight of the charge distribution on the grid, visually represented by the size of the dots, is determined based on distance. The same distribution can also be extended in three dimensions. Adapted from [35].

3.1.3 Water, ensemble conditions and simulation constraints

To solvate the nanopore-protein system I employed a static TIP3P water, which is a model that has been parametrized to reproduce the structural and energetic properties of gas and liquid water [46]. TIP3P is a 3-site model where two positive charges are placed at the nucleus of each of the hydrogen atoms and a negative charge is placed at the bisector of the H-O-H angle [46]. The OH bonds in the molecule have an equilibrium distance of 0.9572 \AA and the H-O-H angle has an equilibrium value of 104.2° [46, 47]. Relevant computed water properties using this model and experimental values for reference are shown in table 3.1. The TIP3P water model was developed with electrostatic cutoff schemes in mind and the introduction of long-range electrostatic methods such as Ewald summation and PME have resulted in its diffusivity becoming large [41]. The TIP3P water model nevertheless continues to be widely used in MM simulations and the CHARMM36 FF was parametrized to work alongside it. Changing the water model used in the parallelization of the FF can lead to undesired effects in the dynamics of the molecule and is not recommended unless extensive studies are made [41].

Table 3.1. Summary of computed and experimental values for TIP3P and water respectively.

Model	Dipole moment ^b	Dielectric constant ^a	Diffusivity ^b $\times 10^{-9} \text{ [m}^2/\text{s]}$	density ^b $\rho \text{ [g/cm}^3\text{]}$
TIP3P	2.347	82.0	5.6	0.998
Exp.	2.95 ^a	78.4	2.3	0.997

^aTaken from [46]. ^b Taken from [48] at 297.0(0.9) K.

The direct treatment of Newton's equations of motion gives an NVE ensemble, but most experiments are conducted at constant pressure and/or constant temperature [38].

To simulate an environment under these conditions, the equations of motion are modified slightly to model a system that is coupled to a reservoir [35]. In the simulations performed, I employed the Nose-Hoover Langevin (NHL) piston for pressure control and the Lowe-Andersen thermostat for temperature control.

Pressure controlled simulations solve the following problem; it is difficult to know *a priori* the volume of the system that yields reasonable density values in liquids [49]. For simple systems/geometries the volume can be estimated based on bulk experimental values, but in systems with membranes one can expect densities to be different in regions near the nanopore, membrane or protein [49]. In constant pressure simulations densities are therefore determined as a consequence of the force field parameters and the force exerted by a piston of *pseudo-mass* W [49]. Available in NAMD is a modified version of the Nose-Hoover method in which Langevin dynamics are used to control fluctuation in the piston [35]. The Nose-Hoover method in turn is derived from the simpler Anderson method, where the volume of the simulation box is adjusted and the position and momentum of the particles is scaled to keep the pressure constant. Using the NHL piston method, the equations of motion for a single particle i become

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} + \frac{1}{3} \frac{\dot{V}}{V} \mathbf{r}_i, \quad \dot{\mathbf{p}}_i = \mathbf{F}_i - \frac{1}{3} \frac{\dot{V}}{V} \mathbf{p}_i \quad \text{and} \quad (3.16)$$

$$\ddot{V} = \frac{1}{W} [P(t) - P_{\text{ext}}] - \gamma \dot{V} + R(t), \quad (3.17)$$

where V is the volume, P is the instantaneous pressure of the system, P_{ext} the target pressure, γ is the collision frequency and $R(t)$ is a random force taken from a Gaussian distribution with mean zero and variance $\langle R(0)R(t) \rangle = 2\gamma k_B T \delta(t)/W$ [49]. The NHL piston method, through fluctuations in the volume, creates an NPT ensemble [49]. For the set of simulations I performed, the degrees of freedom for this energy exchange are two; volume in the z-direction is allowed to fluctuate freely and in the x-y direction the volume can fluctuate but the ratio of geometry is kept fixed. As a result, the exchange of energy from the piston is slow in typical molecular dynamics timescales and the system must be coupled to a thermostat in order to carry out simulations in an efficient manner [49].

The Lowe-Andersen thermostat is local, Galilean-invariant and conserves momentum; I explain the significance of each of these descriptions briefly. Thermostats can generally be divided into global or local. In global thermostats the scaling factor that controls the temperature is based on properties of the entire system (i.e. looking at the velocity of all particles) and therefore distributes energy evenly over the system [50]. In local thermostats the rescaling factor is determined in a stochastic way that takes effect over a local space only (i.e. velocity reassignment for a single particle) and therefore dissipates the energy in a more realistic way [50, 51]. Galilean invariant means Newton's equations of motion hold true in any inertial reference frame, as a result any motion to the centre of mass of the system will not be considered as a shift in temperature [51]. Finally, the fact that the thermostat conserves momentum means it models hydrodynamics correctly and makes it possible to study transport properties of the system [51].

The Lowe-Andersen thermostat proposes to keep temperatures constant by periodically changing the velocities of pairs of particles that are within a search radius R_T [51, 52]. Every pair of particles in the search radius has a probability $\Gamma \Delta t$ of having a collision with the heat bath, where Γ is the collision frequency and Δt is the simulation time step [51, 52]. In case of a bath collision, the pairs of particles get assigned relative velocities from a Maxwell-Boltzmann distribution in a way that conserves angular momentum [51]. For a pair of particles i and j initially with velocities $\mathbf{v}_i(t)$ and $\mathbf{v}_j(t)$, the criteria for assigning

new velocities $\mathbf{v}_i^*(t)$ and $\mathbf{v}_j^*(t)$ respectively is the following

$$\mathbf{v}_i^*(t) = \begin{cases} \mathbf{v}_i(t) & \Gamma \Delta t < \zeta_1 \\ \mathbf{v}_i + \left(\frac{\mu_{ij}}{m_i}\right) \mathbf{v}' & \Gamma \Delta t \geq \zeta_1 \end{cases}, \quad (3.18)$$

$$\mathbf{v}_j^*(t) = \begin{cases} \mathbf{v}_j(t) & \Gamma \Delta t < \zeta_1 \\ \mathbf{v}_j - \left(\frac{\mu_{ij}}{m_j}\right) \mathbf{v}' & \Gamma \Delta t \geq \zeta_1 \end{cases}, \quad (3.19)$$

where $\mu_{i,j} = m_i m_j / (m_i + m_j)$ is the reduced mass of the pair, $\mathbf{v}' = (\lambda - (\mathbf{v}_i - \mathbf{v}_j) \cdot \hat{\sigma}_{ij}) \hat{\sigma}_{ij}$ is the relative velocity added to the particle pair from the collision with the heat bath and ζ_1 is a number drawn from a Gaussian distribution of unit variance [51]. The relative velocity is determined via a stochastic variable $\lambda = \zeta_2 \sqrt{(k_B T / \mu_{ij})}$, where ζ_2 is also a random number taken from the unit-variance Gaussian distribution, and the unit separation vector is $\hat{\sigma}_{ij} = (\mathbf{r}_i - \mathbf{r}_j) / |\mathbf{r}_i - \mathbf{r}_j|$ [51].

It is desirable to set a high collision rate Γ to keep the temperature of the system constant and achieve a high rate of diffusion. As a result the simulations become more efficient, since the conformation space can be sampled rapidly [51]. This constraint comes at a disadvantage, since the thermostat also disturbs the dynamics by applying a shier stress $\sigma_{xy}(t)$ to the system that contributes with artificial viscosity when compared to a system without a thermostat. This artificial dynamic viscosity is given by the time integral of the stress-stress correlation function

$$\eta^T \sim \lim_{t \rightarrow \infty} \int_0^t dt' \langle \sigma_{xy}(0) \sigma_{xy}(t') \rangle \sim \frac{\pi \rho^2 R_T^5 \Gamma}{75 m}, \quad (3.20)$$

where m and ρ are the mass and density of the material the thermostat is being applied to [51]. The presence of the viscosity was useful in the simulations I carried out, because the TIP3P overestimates the diffusivity of water. In a liquid, momentum is transferred rapidly by the inter-forces between neighboring molecules, whereas the mass transport happens slowly [52]. This effect can be summarized by the Schmidt number S_c , which is the ratio of the kinematic viscosity and diffusivity $S_c = \nu / D$. In liquids the Schmidt number is in the order of $\sim 10^3$ and in gases in the order of ~ 1 [52]. By inducing an artificial viscosity it is possible to adjust S_c such that the hydrodynamics in TIP3P are closer to a liquid; more details are discussed in section 4.

One last aspect relevant to discuss is that additional constraints can be placed on atoms to speed up or control aspects of the simulation environment. One standard practice is to employ a combination of the SHAKE and RATTLE algorithms to reduce vibrational degrees of freedom in the system. The advantage being that larger simulation time steps can be used, since bond vibrations with high frequency that normally require a small time step are suppressed [38]. Rigid bond constraints are commonly placed in water molecules and bonds involving hydrogen atoms. The software I used to carry out MM simulations employs a variant optimized algorithm specifically for water molecules called SETTLE. This constraint was also employed in the parametrization of the CHARMM36 FF.

3.2 Density functional theory

In electronic structure calculations, the aim is to solve for the interactions of a many-body system. In the formalism of density functional theory (DFT), this means solving the single-particle Kohn-Sham (KS) equation

$$(\hat{T} + \hat{V}_{\text{ext}} + \hat{V}_{\text{H}} + \hat{V}_{\text{xc}}) \Psi_k = \epsilon_k \Psi_k. \quad (3.21)$$

where ϵ_k and Ψ_k are the KS eigenvalues and orbitals respectively. The Born-Oppenheimer approximation is used to arrive at this equation, hence the interaction between atomic cores and their kinetic energies can be ignored and the interaction between the electron-cloud and nuclei is considered as an external potential V_{ext} [53]. The operator \hat{T} represents the kinetic energy of the non-interacting electrons

$$\hat{T} = -\frac{1}{2} \sum_i \nabla_i^2 \quad (3.22)$$

and the Hartree potential V_H

$$V_H = \frac{1}{2} \int d^3r d^3r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (3.23)$$

represents the classic electron-electron interaction, where $n(\mathbf{r})$ is the density of the electron gas [53]. The Hartree term yields a non-physical self interaction; additionally, since the electrons are particles their motion will be correlated forming a charge density depletion zone near each electron; finally, due to the exclusion principle there exists a type of exchange contribution [53]. These quantum mechanical effects are casted into the exchange-correlation term \hat{V}_{xc} , which cannot be exactly known and is instead approximated by a functional of the electronic density.

DFT builds on three key theoretical foundations presented by Hohenberg and Khon [54]. The first is *uniqueness*, which tells that a point in energy space can be uniquely determined by a point in density space [53]. Further, the *variational principle* means the ground state energy of the system is exactly determined by ground state electronic density [53]. Finally, *universality* means the electronic kinetic energy and the electron-electron interaction are system-independent [53]. The total energy can therefore be found self-consistently for a system of N electrons by

$$E = \sum_k \epsilon_k - \frac{1}{2} \int d^3r d^3r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - \int d^3r \hat{V}_{\text{ext}}[n] n(\mathbf{r}) + E_{\text{xc}}[n], \quad (3.24)$$

where the exchange-correlation energy functional is $V_{\text{xc}} = \delta E_{\text{xc}}[n] / \delta n(\mathbf{r})$ [53, 55].

3.2.1 Pseudopotential

The idea of a pseudopotential begins by recognizing that there are two distinct regions in the external potential of a single atom, as shown in figure 3.3. In the region close to the core the external potential changes rapidly, while in the region outside the core the potential changes slowly. The first region is composed of tightly bounded core electrons that are mostly unaffected by the presence of other neighboring atoms [56, 57]. On the contrary, the potential in the second region is composed of valence electrons that respond to the presence of other atoms leading to the formation of bonds or rearrangement of the electronic density [56, 57]. The reason for the distinct behavior in these two regions can be explained by the fact that the energy levels for core electrons are deep and their wavefunctions are localized close to the atomic nucleus, compared to valence electrons [57]. This means core electrons are chemically inert and do not *feel* the electrons from other atoms [56].

An additional factor must be considered. The wavefunctions describing the atomic orbitals must be orthogonal to each other, as shown for manganese in figure 3.4, where the 1s orbital has no nodes, the 2s orbital has one node and the 3s orbital has two nodes [56]. To represent all the wiggles in the 3s-wavefunction in real space would require a fine mesh.

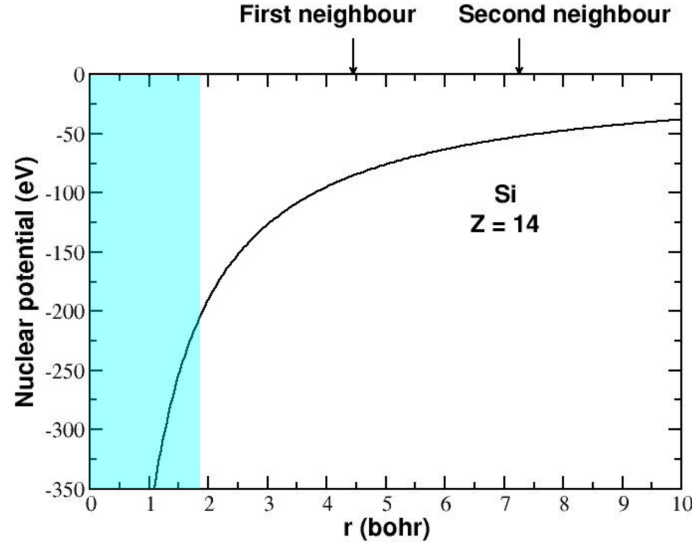


Figure 3.3. Atomic potential for silicon as a function of radial distance from the core. The external potential can be divided into a core (shaded cyan) and valence (unshaded) region. Taken from [58].

Similarly, to represent the wavefunction in reciprocal space would require Fourier components with high frequency and the rapid oscillations in the valence wavefunctions translates to higher kinetic energy [56]. As a consequence higher energy cutoff are needed and the computational cost to handle heavier atoms or systems with many electrons quickly increases.

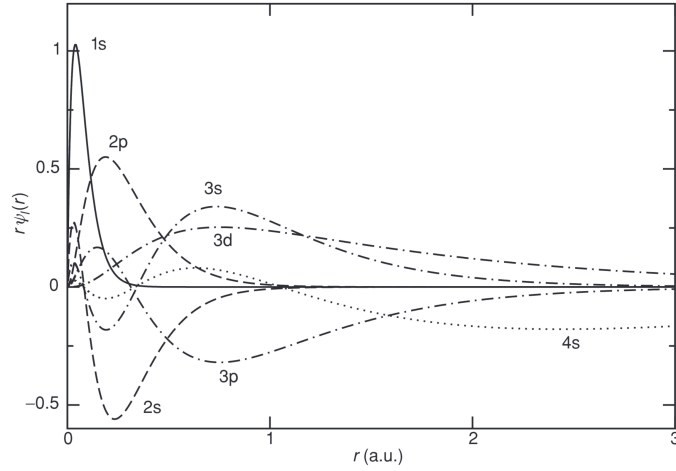


Figure 3.4. Radial component for all the wavefunctions of manganese Mn showing these are orthogonal. SIESTA defines wavefunctions where nodes occur close to the core and the function becomes smooth in the valence region [36]. Higher orbitals have more nodes and therefore more wiggles. Taken from [56].

Pseudopotentials solve this problem by introducing a pseudo-wavefunctions that replaces the original wavefunctions. The pseudowavefunction is defined to be smooth and node-less inside some spherical region with a cutoff radius r_c , where all the wiggles of the all-electron wavefunction are found [56]. This pseudowavefunction modifies the original Hamiltonian and the effective potential is replaced by a pseudopotential [56]. Under this approximation, the core electrons become *frozen* and are assumed to remain unperturbed when atoms in the system are brought together [56]. SIESTA uses a norm conserving pseudopotential, where some additional conditions are imposed. The eigenvalues of the modified Hamiltonian must match the eigenvalues of the all-electron calculation done for a single atom [57]. The pseudowavefunction must also exactly match with the all-electron

wavefunctions in the region outside the cutoff radius [36, 57]. Finally, the pseudowavefunction $\psi_l(r)$ must give a density inside the core region that matches the true electron density, meaning that

$$Q_l = \int_0^{r_c} dr r^2 |\psi_l(r)|^2 , \quad (3.25)$$

where Q_l is the charge of the all-electron wavefunction up to the cutoff radius [36].

3.2.2 SIESTA

The Kohn-Sham orbitals in equation (3.21) are split into two, a radial and angular function

$$\Psi_k = \chi_i(r) Y_{lm}(\theta, \phi) , \quad (3.26)$$

where the angular part is represented in real space using Legendre polynomials

$$Y_{lm}(\theta, \phi) = C_{lm} P_l^m(\cos\theta) , \quad (3.27)$$

C_{lm} is a normalization factor and lm are the quantum numbers [36]. For the radial part, SIESTA uses a linear combination of atomic orbitals

$$\chi_i(\mathbf{r}) = \sum_{\mu} c_{\mu i} \phi_{\mu}(\mathbf{r}) , \quad (3.28)$$

therefore the charge density is given by

$$\rho(\mathbf{r}) = \sum_i n_i |\chi_i(\mathbf{r})|^2 , \quad (3.29)$$

where n_i is the occupation number that follows the Fermi-Dirac distribution [36]. At zero temperature the function is a step function, but at finite temperature the function is smeared around the Fermi level.

In SIESTA the pseudopotential contribution to the Hamiltonian is divided into two parts, a *local* part that acts equally on all electrons independent of their angular momentum and a non-local Kleinman-Bylander part that depends on the angular momentum [36]

$$\hat{V}^{\text{PS}} = V^{\text{local}}(\mathbf{r}) + \hat{V}^{\text{KB}} . \quad (3.30)$$

Far away from the core region, the local part behaves as a Coulomb potential with an attractive force (as seen by a valence electron) given by

$$V^{\text{local}}(\mathbf{r}) \rightarrow -\frac{Z_{\text{val}}}{r} , \quad (3.31)$$

where the valence charge Z_{val} is the net charge of the nucleus plus core electrons [36]. The slow decay constitutes a problem, because calculation of the total energy at any particular point would require to keep track of the contribution from all the core and valence electrons in the system. SIESTA solves this by defining a neutral atom potential V^{NA} for every atom in the system, such that outside some atomic region the attractive long-range local potential is screened by the repulsive potential coming from the Hartree term [36, 59]. I briefly discuss the details of this method.

The electronic density of the system $\rho(\mathbf{r})$ can be given by the sum of the charge density of individual atoms $\rho_{\text{atom}}(\mathbf{r})$ plus some perturbation such that

$$\rho(\mathbf{r}) = \rho_{\text{atom}}(\mathbf{r}) + \delta\rho(\mathbf{r}) , \quad (3.32)$$

where $\delta\rho(\mathbf{r})$ is the change in the electron density due to the creation of bonds and rearrangement of the charge distribution when the individual atoms are brought together to

form the system [59]. Following equation (3.32), the Hartree potential of the system can also be split into contributions from the unperturbed density plus the difference

$$\hat{V}_H(\mathbf{r}) = \hat{V}_H^{\text{atom}}(\mathbf{r}) + \delta\hat{V}_H(\mathbf{r}) , \quad (3.33)$$

where the unperturbed Hartree potential can be expressed as the sum of single-atom potential $V_I^{\text{atom}}(\mathbf{r})$ [59]. SIESTA uses a finite basis set that is localized around the atomic region and, due to this confinement, the single-atom potential are repulsive and decay as Z_{val}/r far away from the core [59]. The neutral atom potential for a single atom I can then be defined as the sum of the local pseudopotential contribution added to the Coulomb repulsion coming from the core and core electrons [36]

$$\hat{V}_I^{\text{NA}} \equiv \hat{V}_I^{\text{local}}(\mathbf{r}) + \hat{V}_I^{\text{atom}}(\mathbf{r}) . \quad (3.34)$$

As seen in figure 3.5, inside the core the neutral atom potential has some smooth shape that is determined by the two terms, but away from the core region the Hartree potential will exactly cancel the local part of the pseudopotential. Using this arrangement, it is no longer necessary to keep track of these two terms for every atom over all space to find the total energy of the system.

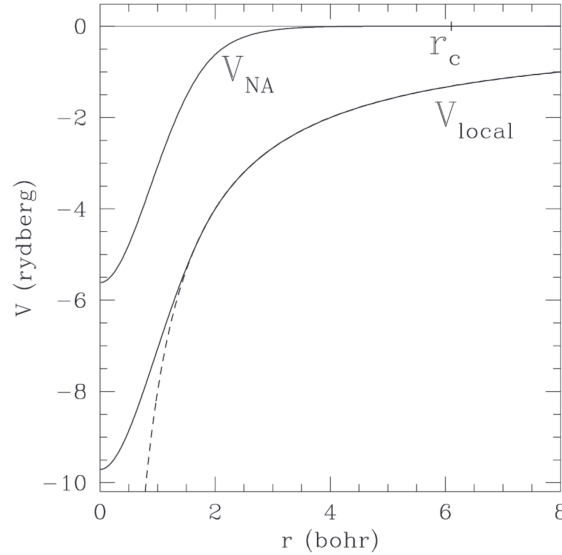


Figure 3.5. Neutral atom potential V_{extNA} for silicon used by SIESTA to accelerate the self-consistent calculations. The local part of pseudopotential that after some radius r_c is screened by the electron charge distribution [36]. Taken from [36].

Following the introduction of the neutral atom potential, the Hamiltonian is then re-expressed in the form [36]

$$\hat{H} = \hat{T} + \sum_I \hat{V}^{\text{KB}} + \sum_I \hat{V}^{\text{NA}}(\mathbf{r}) + \delta\hat{V}_H(\mathbf{r}) + \hat{V}_{\text{xc}}(\mathbf{r}) . \quad (3.35)$$

This modification also speeds up the calculation because the first two terms do not depend on the electronic density and can be computed before the self-consistent loop, which SIESTA does through a two point integral in reciprocal space [36, 59]. The perturbation to the Hartree potential is determined by solving the Poisson equation in Fourier space [59]. To execute such method requires not only to know the density of electrons over all space, but also to impose boundary conditions, which SIESTA fixes as periodic boundary conditions. The rest of the terms in the Hamiltonian are evaluated numerically using a real-space grid inside the self-consistent loop, but these real-space integrals become zero unless there is an overlap between orbitals[59]. As a consequence the computational effort

does not scale with system size, but rather with the addition of more orbitals thus increasing linearly with the number of particles $O(N)$. As shown in figure 3.6, this means the type of system that is studied influences the speed of the computation and it also means vacuum in the system does not significantly increase the cost of the simulations.

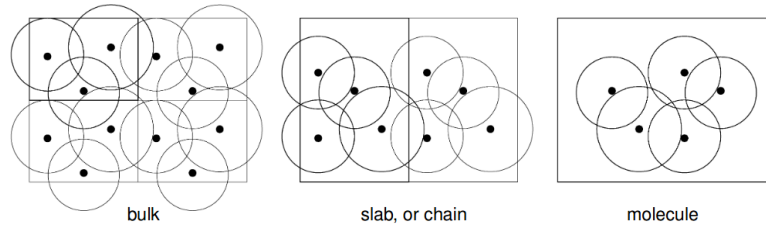


Figure 3.6. Overlap between the extension of the orbitals centered at the core of each atom. Taken from [60].

4. Molecular mechanics methods

In this section I discuss the methods I employed to prepare and carry out MM simulations for a system consisting of silicon nitride nanopore and miniprotein. I built on the methods presented in [61–64] and made use of a modified version of the MSXX FF to describe the membrane and the CHARMM36 FF to model ions, water and proteins [43, 65]. In the process I also optimized the response of the membrane to an external field, adjusted the conductivity of ions to experimental values and modified the protonation states of the miniproteins to match physiological pH conditions.

4.1 Silicon nitride membrane

I used a β -Si₃N₄ unit cell with lattice constants summarized in table 4.1, to build a hexagonal membrane with its surface normal to the z -axis. The shape was chosen to minimize the volume of the membrane, but maximize the separation distance between mirrored images of the system. As a next step I assigned covalent bonds to atoms $< 2 \text{ \AA}$ apart, using periodic boundary conditions along the x - y plane. After, I adjusted the position of the membrane to align its geometric center to the origin of the simulation box and cut a cylindrical pore of radius r_{pore} along the z -axis using

$$\sqrt{x^2 + y^2} < r_{\text{pore}} . \quad (4.1)$$

Unlike the top and bottom faces of the membrane, removing atoms left a pore surface irregular in shape and with nitrogen or silicon atoms exposed in different amounts. This feature of the membrane was not possible to control exactly and created a disadvantage in the model, since exposed atoms influence the interaction strengths between protein and pore wall. I therefore minimized and equilibrated the membrane in vacuum applying a harmonic restraint to all atoms except those close to the pore surface. The procedure allowed surface atoms to find an equilibrium position that reduced the cavities between neighboring atoms and improved the surface ratio of silicon and nitrogen atoms. More details of this procedure can be found in appendix B. Removing atoms from the membrane also modified the ratio of nitrogen and silicon ions and left an excess charge in the membrane. This was corrected by slightly shifting the charge of nitrogen atoms by approximately 0.07%, with the final result shown in figure 4.1.

Table 4.1. Lattice vectors for silicon nitride obtained using x -ray crystallography [61, 66].

Unit cell lattice vectors [\AA]
$\mathbf{a}_1 = 7.595 \times [1, 0, 0]$
$\mathbf{a}_2 = 7.595 \times [1/2, \sqrt{3}/2, 0]$
$\mathbf{a}_3 = 2.902 \times [0, 0, 1]$

The MSXX FF parameters that model the membrane were adapted by Aksimentiev et al. to the CHARMM FF format to be compatible with the program running the MM simulations [61]. The bonded interactions between silicon and nitrogen, which in the MSXX FF are described by the Morse bond stretch potential, were instead fitted around its minimum to a harmonic potential following equation (3.10) using parameter values

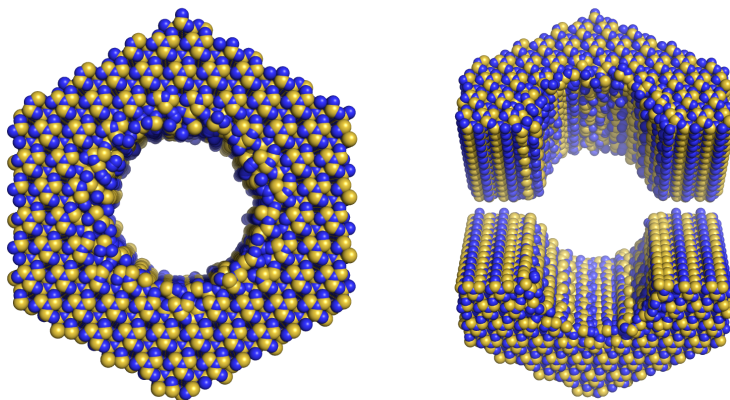


Figure 4.1. Van der Waals representation of Si_3N_4 nanopore in two different orientations. The hexagonal membrane has side length of $s = 37.98 \text{ \AA}$, height of $h = 37.73 \text{ \AA}$ and pore radius of $r_{\text{pore}} = 17.5 \text{ \AA}$. Silicon atoms $r_{\text{Si-VdW}} = 2.10 \text{ \AA}$ are shown in yellow and nitrogen atoms $r_{\text{N-VdW}} = 1.55 \text{ \AA}$ in blue.

$K_b = 5.0 \text{ kcal}/(\text{mol } \text{\AA}^2)$ and $b_0 = 1.777 \text{ \AA}$ [61, 62, 65]. The angular bonds between sets of three atoms and non-bonded interactions, which include electrostatic and van der Waals interactions, were taken directly from the MSXX FF [61]. Values for these non-bonded parameters are shown in table 4.2.

Table 4.2. Summary of the MSXX FF parameters for non-bonded interactions, obtained from Hartree-Fock calculations [65]. To describe interactions between silicon and nitrogen atoms, parameters were determined by geometric and arithmetic mean following CHARMM convention.

Atom	charge (q) [e]	ϵ [kcal/mol]	R^{min} [\AA]
Si	0.767900	0.31	6.270
N	-0.575925	0.19	3.995

Silicon nitride is a dielectric material that polarizes in the presence of an external field. The amount of polarization can be denoted by its relative permittivity constant ϵ_r , experimentally measured to be 7.5 [67]. The default MSXX FF parameters used for the bonded interactions have been optimized to give a dielectric constant equal to vacuum [63, 68]. For this reason, I defined an additional force to all atoms in the membrane by applying a harmonic restraint that allowed me to control the relative permittivity and found a restraint value of $7.0 \text{ kcal}/(\text{mol } \text{\AA}^2)$ to be optimal. Details of the process for obtaining this result are outlined in appendix B.

4.2 Selected miniproteins

The accuracy of molecular dynamics simulations depends largely on the accuracy of the protein model itself. For this reason it is necessary to check the protein structure is complete. In general this means checking that no atoms are missing, filling valence bands with hydrogen and determining the protonation state of titratable amino acid groups based on a specified pH value [69]. The molecular structures for both miniproteins were obtained using NMR spectroscopy which, unlike X-ray spectroscopy, can detect the position of hydrogen atoms. After checking the protein structure, for the other two remaining tasks I employ the open source web server H++ [69–71]. H++ uses the protonated and deprotonated partial charges of each titratable side chain to create titration curves using the AMBER ff99SB FF for proteins [69]. The protonation state of each group is then determined based on the pH and the midpoint of the titration curve pK [69]. Titration curves

represent the protonation state probabilities as a function of pH, therefore if $pK < pH$, the site is deprotonated or protonated otherwise [69]. The output structure file from H++ follows the AMBER FF naming convention, which was translated to CHARMM36 naming convention using the MMTSB tool set [72]. Details on the titration results from H++ are summarized in appendix A.

The charge distribution of the two miniproteins after protonation from H++ are shown in figures 4.3 and 4.2, respectively. These quantities were computed using the adaptive Poisson-Boltzmann solver (APBS) to get an indication of how these molecules will behave under the influence of an external electric field [73].

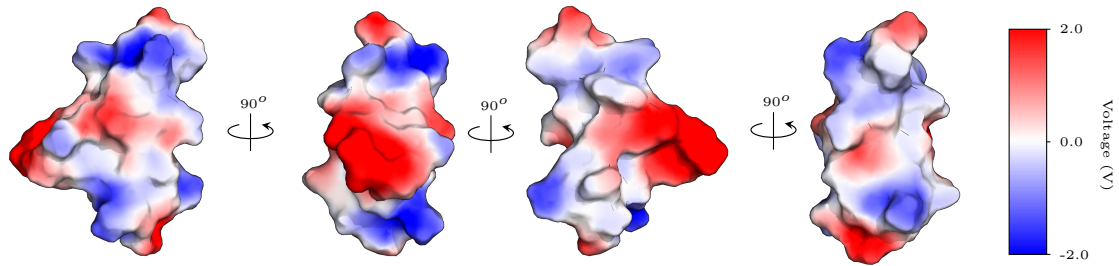


Figure 4.2. Surface render of the FBP11 WW domain miniprotein showing the charge distribution using a dielectric constant of 10 for the protein and 80 for the solvent subject to a 1 M ion concentration.

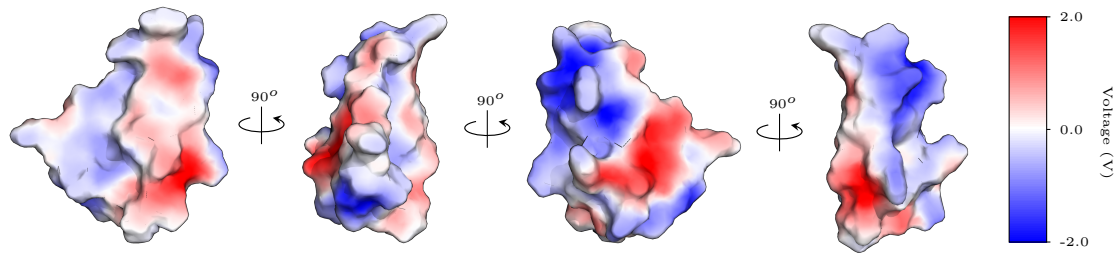


Figure 4.3. Surface render of the AGRP(87-120) miniprotein showing the charge distribution using a dielectric constant of 10 for the protein and 80 for the solvent subject to a 1 M ion concentration.

In section 2.4 I mentioned these two miniproteins were chosen based on their likeness; the volume, shape, sequence length and mass are very similar. Additional properties can be used to further quantify the likeness of the two miniprotein, based on their amino acid sequence. *Identity* is a property that measures the number of amino acids that match exactly in both sequences (excluding gaps). In figure 4.4 I show the optimal alignment of the two protein sequences and find an *identity* value of 14.3%. This low *identity* value means the composition of the two proteins differs significantly, which might also indicate their interaction strength with the nanopore wall (and by extension the dwell time) could differ to a significant degree as well.

AGRP(87-120):	-CVRL <u>H</u> ESCLGQQVPCCDPAATCYCRFFNAFCYCR
FBP11 WW:	GSWTEHKS <u>P</u> DG-----RTYYY---NTETKQSTWEKPD

Figure 4.4. Amino acid sequence represented by single letter residue codes. I underline matching sequences and represent gaps with a dash symbol.

4.3 Preparing and executing molecular mechanics

To prepare the system for molecular dynamics simulations, I used VMD's plugin *solvate* to fill a box with water molecules to a height of 80 Å and removed any excess as to follow

the hexagonal shape of the membrane. A number of potassium K^+ and chloride Cl^- ions were also added using VMD's plugin *autoionize*, equating to a 1 M concentration¹. In cases where a protein is present, I included excess positive or negative charges to keep the total charge of the system zero. According to H++ results, at physiological pH of 7.1 the FBP11 miniprotein has a total charge of $-2e$, while AGRP(87-120) has a total charge of $+1e$. Each protein was placed at the center of the pore in three different orientations. A render of the system after adding water and ions is shown in figure 4.5.

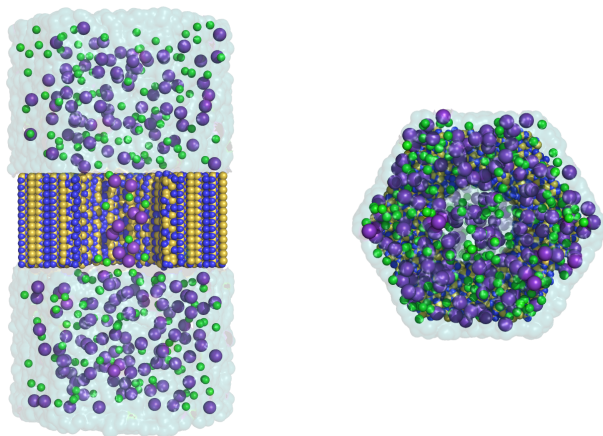


Figure 4.5. Side-view and top-view of the solvated hexagonal silicon nitride membrane with 1 M concentration of potassium chloride ions (open pore system). Water molecules are represented as a transparent surface, potassium ions as purple spheres and chloride ions as green spheres.

I chose MM software NAMD2 to perform simulations with periodic boundary conditions using a time step of 1 femtosecond [35]. Full electrostatics were computed every 4 femtoseconds using particle mesh Ewald (PME) with a grid size of ≤ 1 Å. Local electrostatics and van der Waals interactions were computed every 2 femtoseconds using a switching function with a cutoff distance of 12 Å.

The temperature was fixed at 298 K using a Lowe-Andersen thermostat with a cutoff radius of 3.0 Å and a collision rate of 50/picosecond. Water bonds were set to rigid using NAMD2's optimized algorithm SETTLE [74]. In NPT simulations the Nose-Hoover piston had a target pressure of 1.01325 bar, a Langevin piston period of 200 femtoseconds and a piston decay of 100 femtoseconds. The dimension of the simulation box were allowed to vary in all directions, but I placed constant ratio constraint along the x-y direction.

Pre-production runs

In preparation for production runs the simulation systems were minimized for 0.1 picosecond and equilibrated for a further 1 nanosecond in NPT conditions. From these, an average system box size was computed to be later used in production simulations. During this time the miniprotein was allowed to freely move inside the pore.

Production runs - first set

Production simulations ran for 30 nanoseconds in NVT conditions under a 0.5 V and a 1.0 V voltage bias. To keep the miniprotein inside the pore, atoms that make up the backbone of the protein were harmonically restrained to their initial position. These initial position were determined by the last simulation step from the pre-production equilibration runs.

¹Used molality to keep the concentration of ions independent of the volume of the simulation.

Production runs - second set

I continued to simulate the system for a further 20 nanoseconds by running two simulations 10 nanoseconds in length. This time the harmonic restraint on the protein's backbone was removed, which allowed the protein to freely translocate out of the pore. If the z-coordinate of the center of mass of the protein was found near the edges or outside the pore after the first simulation, I executed the remaining 10 nanoseconds using the initial conditions from the first set of production runs.

Free protein runs

Finally I ran 30 ns-long simulations where miniproteins were simulated *far away* from the pore walls. For example, I imagined a system where the pore radius was much larger than the radius of the miniprotein. In practice it was not realistic to scale up the silicon nitride membrane to a large size. Instead I placed each miniprotein without the presence of the membrane in a bulk solution with the same 1 M ion concentration. In these simulations the bias voltage was scaled proportional to the system size as to mimic as much as possible the environment the protein experiences while inside the pore. I only performed simulations where the bias voltage was scaled from 0.5 V. These simulations were used, along with the second set of production runs, to estimate the interaction strength between the protein and pore wall.

4.4 Ionic current, potential energy and harmonic restraint

The ionic current is based on the trajectory files output from molecular dynamic simulations and is computed by the collective movement of all ions in the system over time t while inside a pore region of length l_z such that

$$I(t) = \frac{1}{\Delta t l_z} \sum_i^N q_i (z_i(t + \Delta t) - z_i(t)) , \quad (4.2)$$

where $z_i(t + \Delta t) - z_i(t)$ is the displacement of a single ion i along the z-direction in an interval of time Δt [61]. The sum runs over all ions that either at time t or $t + \Delta t$ were inside the pore. The transport properties of the system thus becomes relevant, but in MM simulations modeling the movement of ions accurately comes as a challenge, since I have chosen to work with TIP3P known to overestimate the diffusivity of water [75]. To compensate for this effect, I used the artificial viscosity from the Lowe-Andersen thermostat to influence the water diffusivity and obtain a KCl ion conductivity κ , a measure of how ions move in a medium, that is close to theoretically predicted values in bulk solution. Details of this calibration can be found in appendix B.

I employed the VMD plugin NAMD-Energy to compute the bonded and non-bonded interactions of the miniprotein alone, based on trajectory files from simulations. I defined the same bonded and non-bonded configuration settings as those used in simulations, along with the CHARMM36 and MSXX FF parameters.

The harmonic restraint applied to the backbone of the miniproteins based on the position of these atoms ζ at a give simulation time and their initial atomic positions ζ_0 . The harmonic biasing method followed the equation

$$V(\zeta) = \frac{1}{2} k \left(\frac{\zeta - \zeta_0}{\omega_k} \right)^2 , \quad (4.3)$$

where $k = 1 \text{ kcal/mol } \text{\AA}^2$ is the force constant and $\omega_k = 1$ can alter the physical force units of the restraint [76].

5. Quantum mechanics methods

In this section I discuss the methods I used to prepare and carry out electronic structure calculations using the density functional theory package SIESTA-4.0 in order to quantify the interaction strength between miniprotein and pore wall and by extension identify differences in the dwell time [36]. For this I developed a method to determine the interaction strengths from looking at the contribution of individual amino acids to the entire system based on their accessibility.

5.1 Amino acids and silicon nitride slab

Due to the sheer quantity of atoms in the system (approx. 76500 atoms, 79% water, 19% membrane, approx. 1% ions and < 1% biomolecule) and computational resources available, I chose to perform DFT calculations selecting a single amino acid at a time to study the interaction strength between the pore wall and biomolecule. With this approach the question of the interaction strength of the entire miniprotein still remains. To address this I turned to the relative solvent accessibility (RSA), which is a measure of how buried or exposed a residue in the biomolecule is. The RSA has been previously used to describe physical properties and as a predictive measure of evolution rates in proteins [77]. The RSA is a number that for a single amino acid quantifies the accessible surface area (ASA) around its local environment [77]. In the application of this thesis, RSA values identify those amino acids that are most likely to be in contact with the pore walls while the miniprotein translocates. These values are then used to determine the total interaction strength of the miniprotein. There are several ways to compute the maximum solvent accessibility of an amino acid X, which is used as a reference value for RSA. I employed the software Swiss-PDB Viewer that uses the surface area of X when placed between glycine to form a Gly-Gly-X-Gly-Gly pentapeptide in extended conformation as a normalization factor [78]. Glycine in particular is the most simple of the 20 proteinogenic amino acids, since the R-group is composed of a single hydrogen atom. The computed RSA values for FBP11 WW domain and ARGRP(87-120) miniproteins are shown in figures 5.1 and 5.2 respectively.

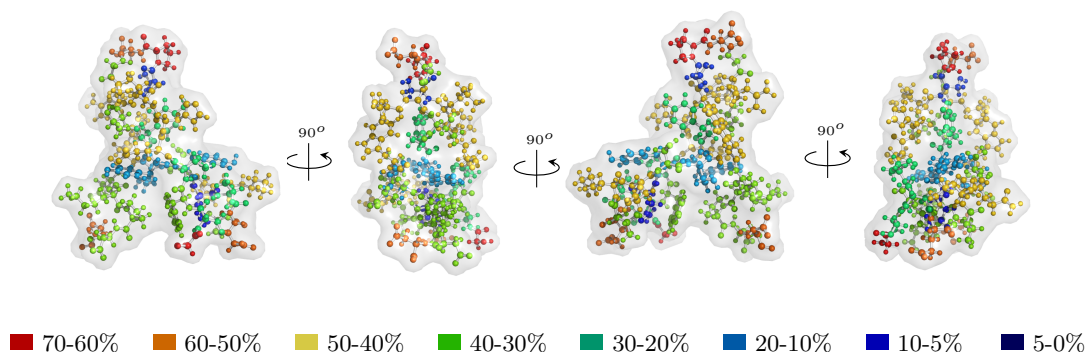


Figure 5.1. FBP11 WW domain miniprotein with its amino acids color coded based on their RSA value. A dark blue color is attributed to completely buried amino acids while red is given to those that have 70-60% of the relative surface exposed.

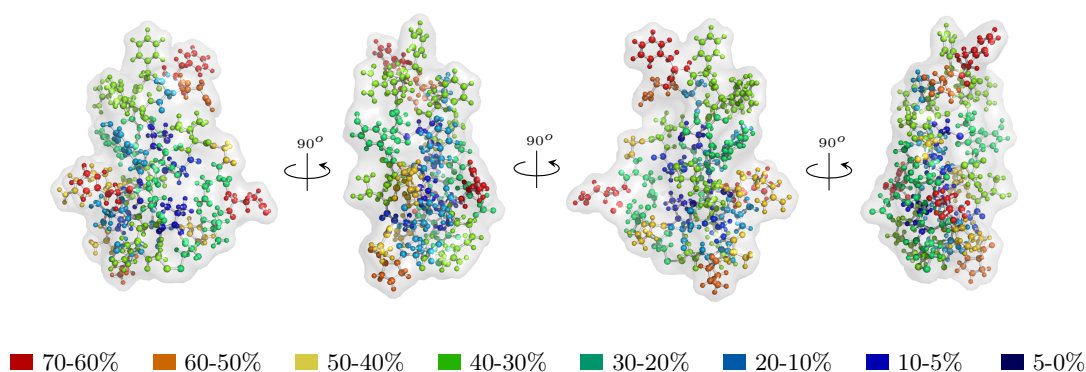


Figure 5.2. AGRP(87-120) miniprotein with its amino acids color coded based on their RSA value. Colors follow the same logic as those presented in the previous figure.

To determine the interaction strength of individual amino acids, I prepared a system with a small triclinic silicon nitride slab with a thickness of approx. 13 Å. The objective is to model the interaction between amino acid inside the pore, I therefore rotated the silicon nitride slab such that the contact surface was as similar to the pore wall as possible. Much like I did in MM simulations I explored minimizing and equilibrating the membrane surface, but observed no changes to the structure because the slab is small and in an extended position. SIESTA simulations are performed in periodic boundary conditions; mirror images of the slab therefore extend continuously along the x-y plane, but were separated by a large distance along the z-direction. Based on the geometry of the slab, mirror images of the protein were also distanced by approx. 14 Å along the x-direction and approx. 15 Å along the y-direction. Depending on the size of the amino acid, mirror images of the silicon nitride slab along the z-direction were at least 39, 46 or 52 Å apart. I determined these distance values were sufficiently large such that periodic images did not affect the total energy of the system, from convergence tests done as a function of simulation box size. Details of these tests are found in appendix C.

The amino acid was initially placed close to the surface of the slab, with the R-group facing the membrane. I took notice to add a hydrogen atom in the carboxyl group or remove a hydrogen atom from the amino group for negatively or positively charged amino acids such as histidine or aspartic acid. I generated snapshots moving the amino acid away from the membrane initially in steps of 0.1 or 0.2 Å and once far away in steps of 1.0 Å. For each of these snapshots I later performed electronic structure calculations to find the ground-state energy of the system after convergence of the electronic density. I show a few of these snapshots for asparagine in figure 5.3.

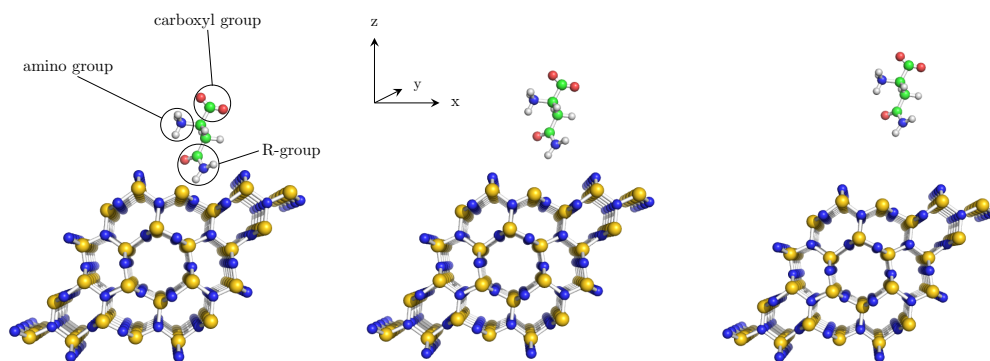


Figure 5.3. Snapshots of asparagine (ASN) placed at different distances away from the triclinic silicon nitride slab. The amino acid is oriented such that the R-group faces the surface of the membrane.

5.2 Exchange-correlation functional

I employed the non-local van der Waals functional DRSLL to describe the exchange correlation part of the Kohn-Sham Hamiltonian. Local and semi-local functional approximations such as LDA and GGA are not well suited to describe long-range interactions present in sparse systems such as biomolecules and soft matter [79, 80]. The VdW functional is expected to yield a better description of the energy landscape when compared to GGA [59]. LDA sometimes yields get correct VdW interactions, but for different reasons that can lead to the wrong interpretation of the physics [59]. VdW binding is a correlation effect between two electrons at locations \mathbf{r} and \mathbf{r}' , which can be determined through the general expression

$$E_{xc} \propto \int d^3r d^3r' n(\mathbf{r}) \phi(\mathbf{r}, \mathbf{r}') n(\mathbf{r}') , \quad (5.1)$$

where $\phi(\mathbf{r}, \mathbf{r}')$ is some function that depends on the electron positions [79]. The exchange part of DRSLL is taken directly from the *revPBE* functional, a conscious choice by Dion et al. that makes it possible to express van der Waals binding exclusively from the correlation term [80]. The exchange choice was made by testing the predictive power of GGA functionals, using the exchange contribution alone, to compute the binding energy of different dimers [80]. In the case of *revPBE*, the exchange lacks description of this binding energy [79, 80].

5.3 Parameters for electronic structure calculations

For SIESTA calculations I employed a single k-point along the Brillouin zone, since the system dimensions were large enough that its reciprocal space was very small. I used a double- ζ polarized basis set and a Pulay mixing scheme with a mixing coefficient of 0.02 and a self-consistent cycle history of 5 to reach convergence. From trial and error, I found this small mixing value guaranteed convergence for all systems. I fixed a mesh cutoff energy of 300 Ry according to convergence runs I performed in appendix C. SIESTA employs pseudopotential to speed up simulations; I selected Abinit's GGA-PB non-relativistic pseudopotential for all elements in the system. The electronic temperature was fixed at 300 K and the tolerance for convergence of the density matrix was kept at 1×10^{-4} eV.

6. Results

I sectioned the results in two parts; in the first section I studied occupied pore currents i_{occ} from results obtain exclusively from MM simulations. In the second section I studied the interaction between pore wall and miniprotein to estimate the dwell time, where results were obtained from both MM and electronic structure calculations.

6.1 Occupied pore ionic current

I explored three different random orientations for each miniprotein and looked at the occupied pore ionic current for the systems under a 0.5 and 1.0 V voltage drops. I arrived at each of these orientations by rotating the protein a few degrees along the x-y-z axes, such that their positioning differed significantly with respect to one another. To keep the results concise, I mostly present results from simulations with a single bias voltage and place remaining findings in appendix D.

Harmonically restrained protein simulations

In an ideal experiment the pore has been submerged in solution and the bias voltage has been left on *long enough* such that when the protein reaches the access region, the system is in a steady state. Conditions in the simulated environment were different, since the electric field was turned on when the protein was already inside the pore. As a first step I determined the simulation time when the system reached a state that closely resembled experimental conditions, while the protein remained fixed. I conjectured that in a well-equilibrated system, the ion count inside the pore stays roughly constant. I decided then to monitor the number of ions inside the pore at any given time in the simulation, which I achieved by saving the position of all atoms at equally spaced intervals. The results for both miniproteins are shown in figures 6.1 and 6.2.

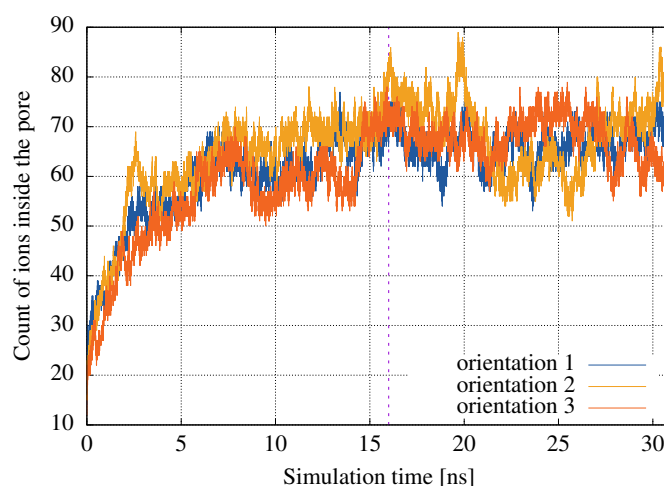


Figure 6.1. Ion count as a function of simulation time for pore with protein FBP11 WW domain under a 0.5 V voltage drop. Ion count is shown for all three orientations. The system was minimized, equilibrated and after 1 nanosecond the field was turned on and the protein was harmonically fixed. Ion positions were stored every 1 picosecond. The dashed purple line marks when experimental conditions were reached.

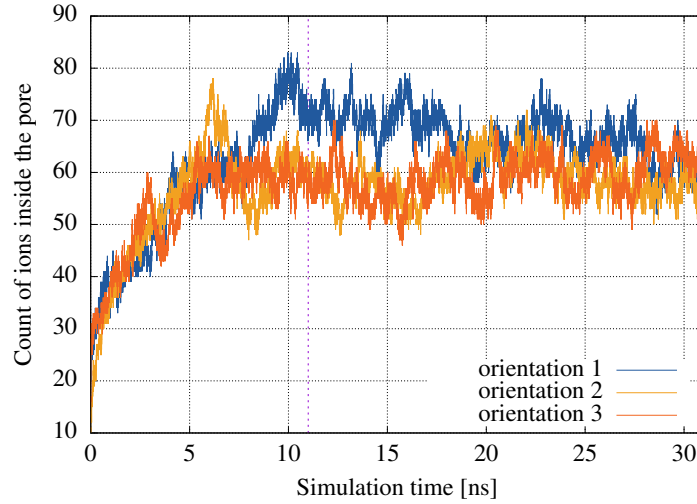


Figure 6.2. Ion count as a function of simulation time for pore with protein AGRP(87-120) under a 0.5 V voltage drop. Ion count is shown for all three orientations. The system was minimized, equilibrated and after 1 nanosecond the field was turned on and the protein was harmonically fixed. Ion positions were stored every 1 picosecond. The dashed purple line marks when experimental conditions were reached.

Based on the results above, the number of ions inside the pore reached a steady state after 16 nanoseconds for FBP11-WW and 11 nanoseconds for AGRP(87-120). From the remaining 15 and 20 nanoseconds respectively, I computed the occupied pore current as a function of time based on the trajectory of the ions and equation (4.2). Protein sensing experiments using a silicon nitride nanopores such as those performed by Rosenstein et al. utilize a patch-clamp amplifier that can capture ionic currents in the range of MHz [81, 82]. Typical ionic current measurements at this frequency have large noise from both the system and instrumentation that attenuate the signal and make it difficult to retrieve meaningful information [1, 12]. Noise in experiments can be removed using a low pass filter (LPF), i.e. Bessel filter, with a pass-band in the order of 1kHz-100kHz [1, 12]. This range of frequencies cannot easily be accessed in simulations, since they translate to timescales in the order of microseconds. Fortunately, instrumentation noise was not present in simulations and I applied a Bessel LPF with a cutoff frequency of 5 GHz. For illustration, I show the raw and filtered i_{occ} for a single orientation of the FBP11 WW domain in figure 6.3. The LPF results for all the orientations and for each miniprotein, are shown in figures 6.4 and 6.5.

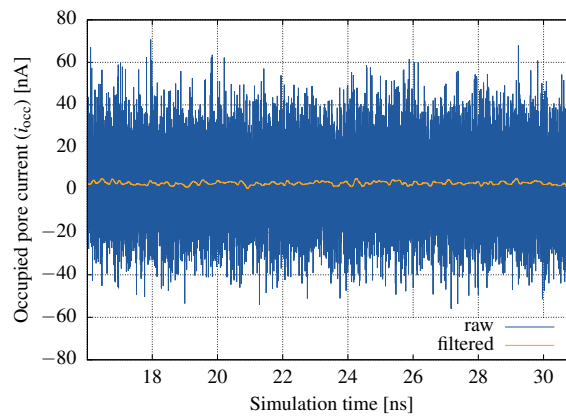


Figure 6.3. Raw and filtered occupied pore current computed for protein FBP11 WW domain as a function of simulation time. The block average was taken using a window size of nanoseconds. The system was subject to 0.5 V voltage drop along the z-direction. Ion positions were stored every 1 picosecond.

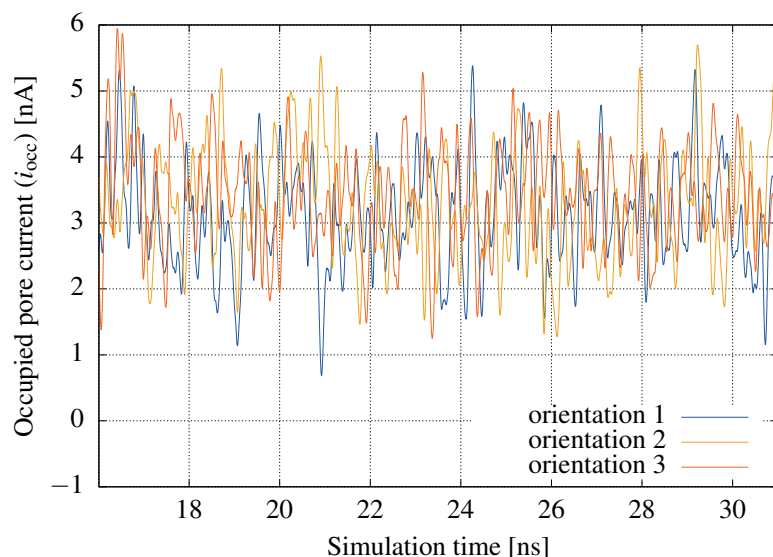


Figure 6.4. LPFd occupied pore current computed for protein FBP11 WW domain as a function of simulation time for all three orientations subject to a 0.5 V bias voltage. The protein's backbone remained harmonically restrained during the simulation.

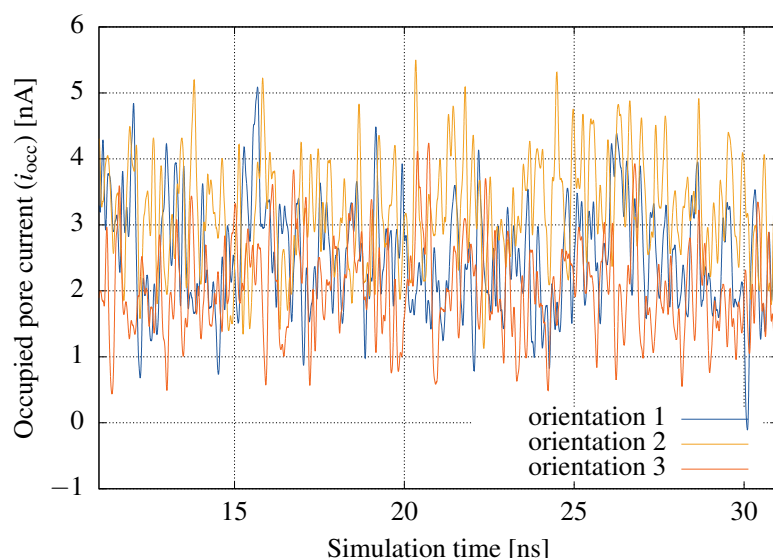


Figure 6.5. LPFd occupied pore current computed for protein AGRP(87-120) as a function of simulation time for all three orientations subject to a 0.5 V bias voltage. The protein's backbone remained harmonically restrained during the simulation.

From the LPFd results above it was possible to obtain a histogram, which I then used to compare the occupied pore currents that could be expected to be captured in typical translocation events of these two miniproteins. The histogram in figure 6.6 shows the combined results for all orientations for the two proteins, when the system were subject of a 0.5 V voltage bias. Similarly, I obtained a histogram of the recorded occupied pore current signatures after applying a low pass filter and aggregating the results from the three different orientations when the system was subject to a 1.0 V voltage bias. I show these results in figure 6.7.

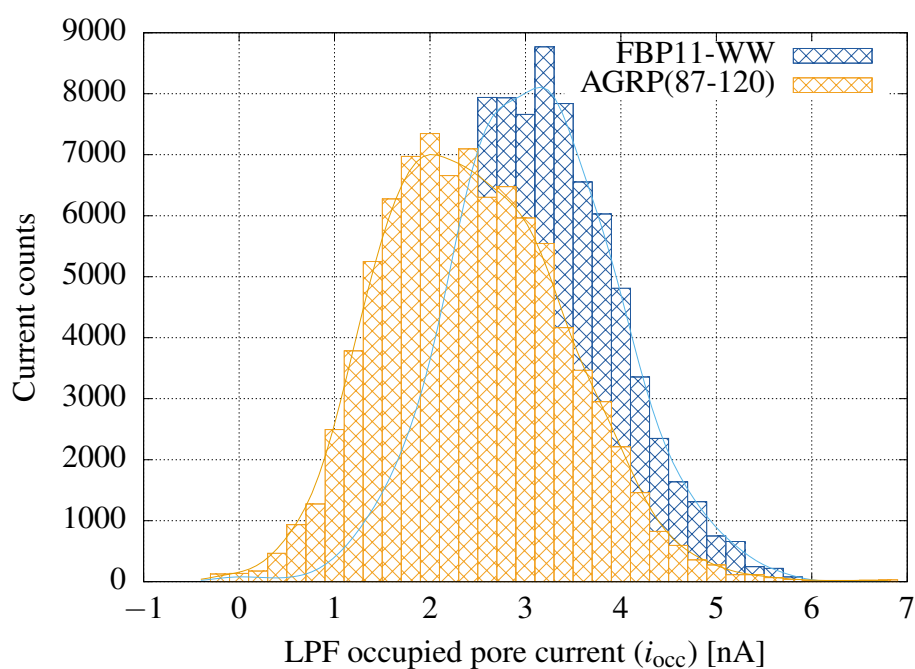


Figure 6.6. Histogram of LPFd occupied pore currents for both miniproteins when the system was subject to a 0.5 V bias voltage. I used a bin width of 0.2 nA and combined results from all orientations. The protein's backbone remained harmonically restrained during the simulation.

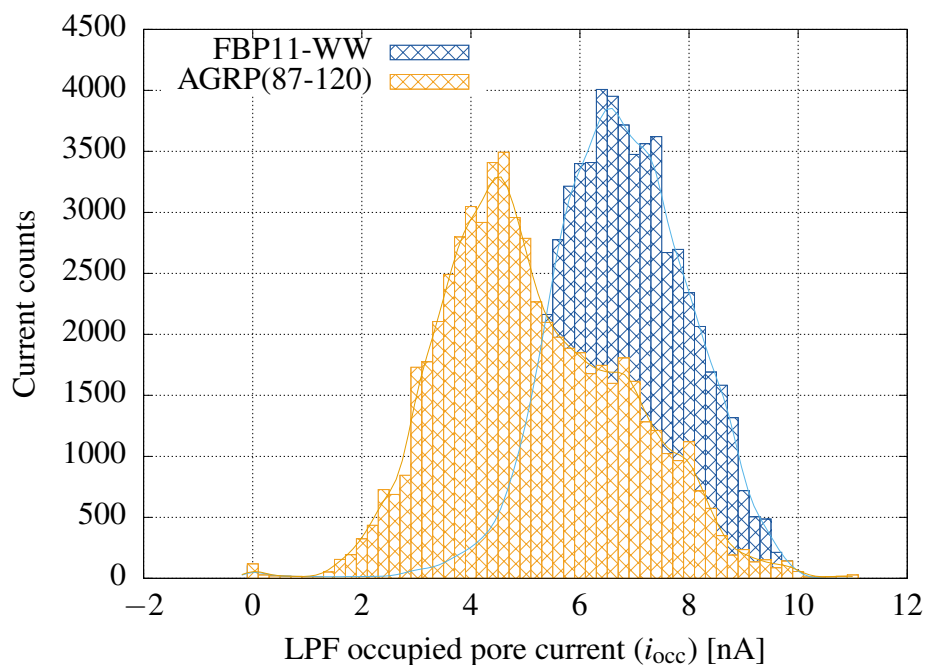


Figure 6.7. Histogram of LPF occupied pore currents for both miniproteins when the system was subject to a 1.0 V bias voltage difference. I used a bin width of 0.2 nA and combined results from all orientations. The protein's backbone remained harmonically restrained during the simulation.

Unrestrained protein simulations

From the last phase state of the harmonic-restraint simulations I ran two 10 nanosecond-long simulations for each protein orientation, where the restraint on the backbone of the protein was removed. During these simulation, it was possible for the miniproteins to leave the pore. If after the first unrestrained simulation the protein was on the edge or outside the pore, I executed the remaining 10 nanoseconds restarting from the last phase state of the harmonic-restraint simulation, otherwise I continued the unrestrained simulations for an additional 10 nanoseconds. In figures 6.8 and 6.9 I show the resulting LPFd occupied pore currents from these simulations when the system was subject to a 0.5 V voltage bias. The LPF needs a short period of adjustment, I therefore omitted results for the initial 0.5 nanoseconds.

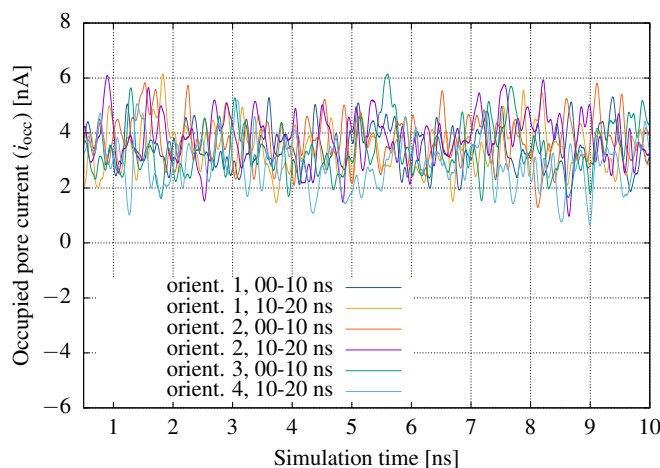


Figure 6.8. LPFd occupied pore current computed for protein FBP11 WW domain as a function of simulation time for all three orientations subject to a 0.5 V voltage bias. No external restraints were applied to the protein.

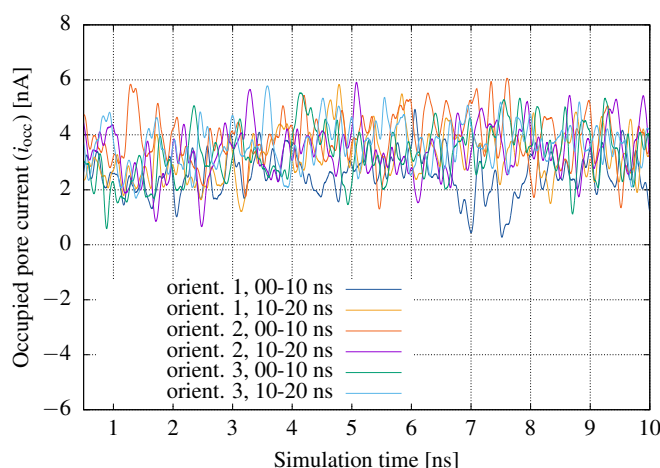


Figure 6.9. LPFd occupied pore current computed for protein AGRP(87-120) as a function of simulation time for all three orientations subject to a 0.5 V voltage bias. No external restraints were applied to the protein.

Again, I compared these occupied pore ionic currents results using a histogram. Results from a 0.5 V voltage bias are shown in figure 6.10 and from a 1.0 V voltage bias are shown in figure 6.11.

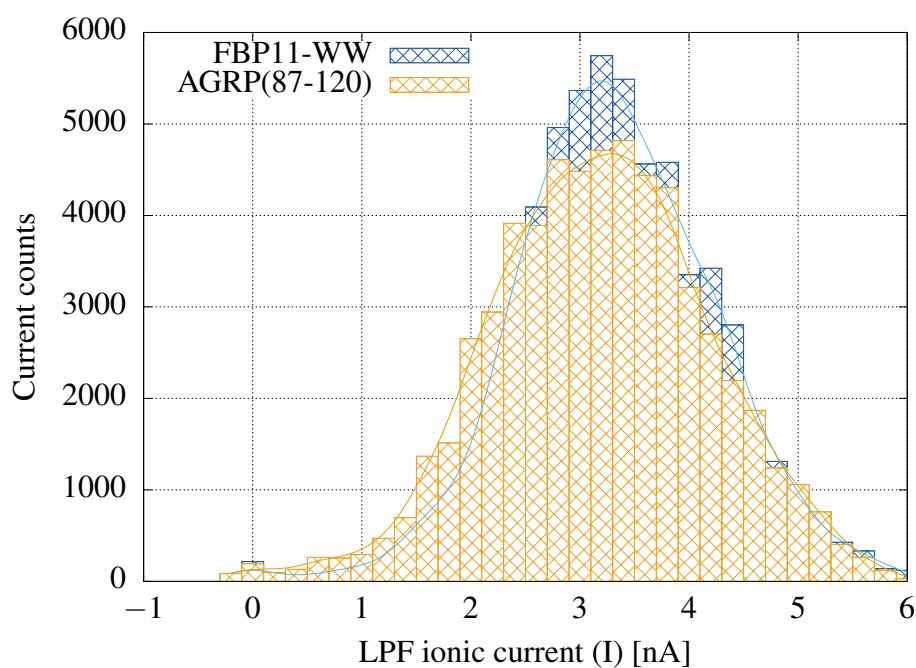


Figure 6.10. Histogram of LPF occupied pore ionic currents for both miniproteins when the system was subject to a 0.5 V voltage bias. No external restraints were applied to the protein. I used a bin width of 0.2 nA and combined results from all orientations.

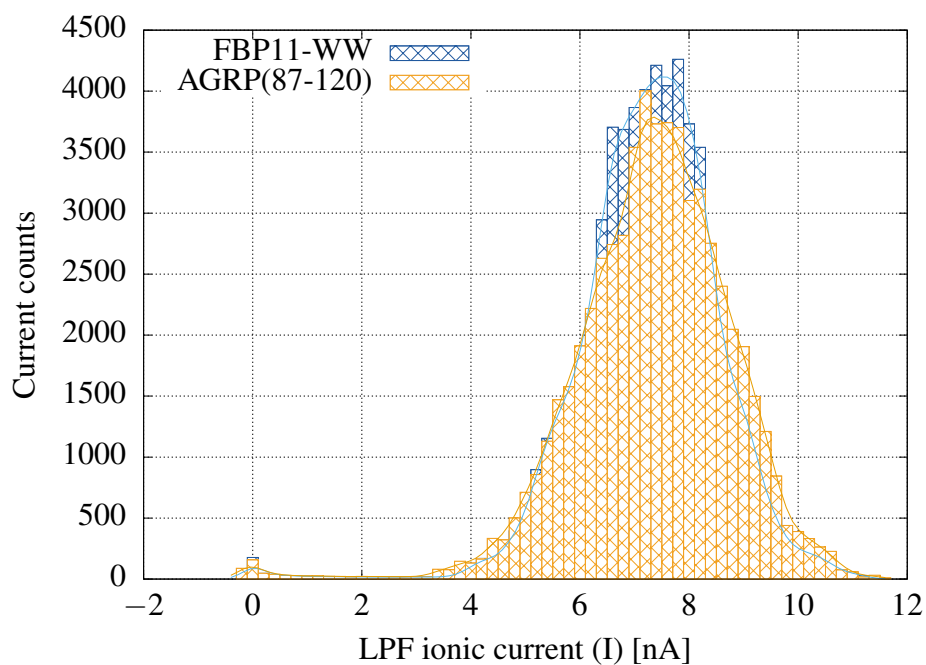


Figure 6.11. Histogram of LPF occupied pore currents for both miniproteins when the system was subject to a 1.0 V voltage bias. No external restraints were applied to the protein. I used a bin width of 0.2 nA and combined results from all orientations.

6.2 Estimate of the pore wall and amino acid interaction

I estimated dwell-time differences in the translocation of the two miniproteins, based on MM simulations carried out using the CHARMM36 FF model. I computed the total potential energy of each miniprotein when placed inside the nanopore and when placed far away from it. We reasoned that changes to the internal energy of the miniproteins in these two environments gave an indication on the interactions between the biomolecule and the pore wall. Simulations when the protein was *far away* from the pore ran for 30 nanoseconds, but I used the last 20 nanoseconds for the energy analysis. Calculation of the potential energy included all bonded and non-bonded interactions specified in equation (3.10). For illustration I first show in figure 6.12 the potential energies for the FBP11 WW domain in these two environments. Due to the large noise in the computed energies I applied a Bessel LPF to the data with a 5 GHz cutoff frequency. Results when the protein is near the pore wall are an average resulting from all three different orientations.

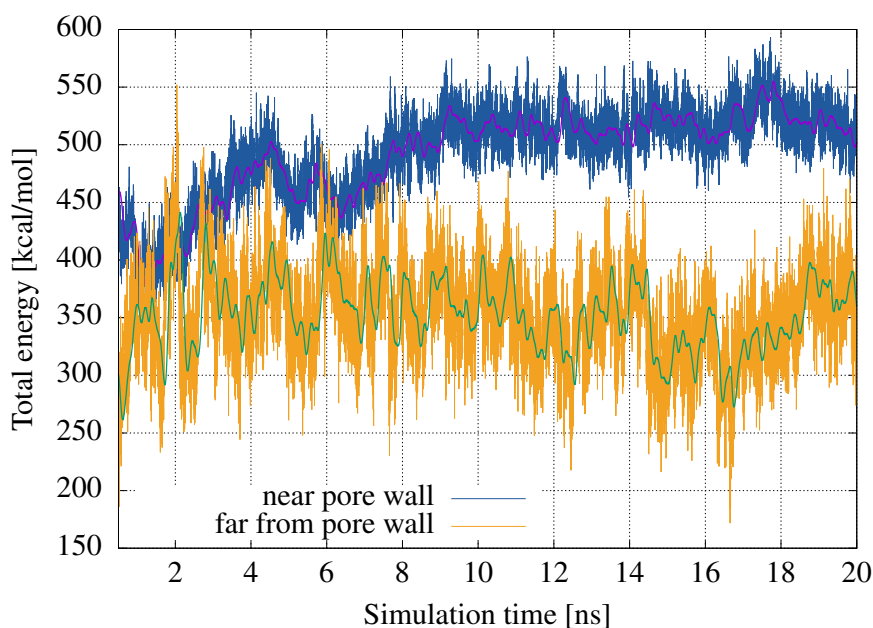


Figure 6.12. Raw and filtered total potential energy for FBP11-WW domain as a function of simulated time. The potential energy is computed from trajectory files every 1 picosecond and includes bonded and non-bonded interaction. I ignored the first 0.5 nanoseconds to allow the LPF time to adjust.

I used these LPF energies in these two environments to compute potential energy differences for both miniproteins and show the results in figure 6.13. Additionally, I took these energy differences and plotted them in a histogram where it could be easy to visualize changes between the two proteins. I show the histogram results in figure 6.14.

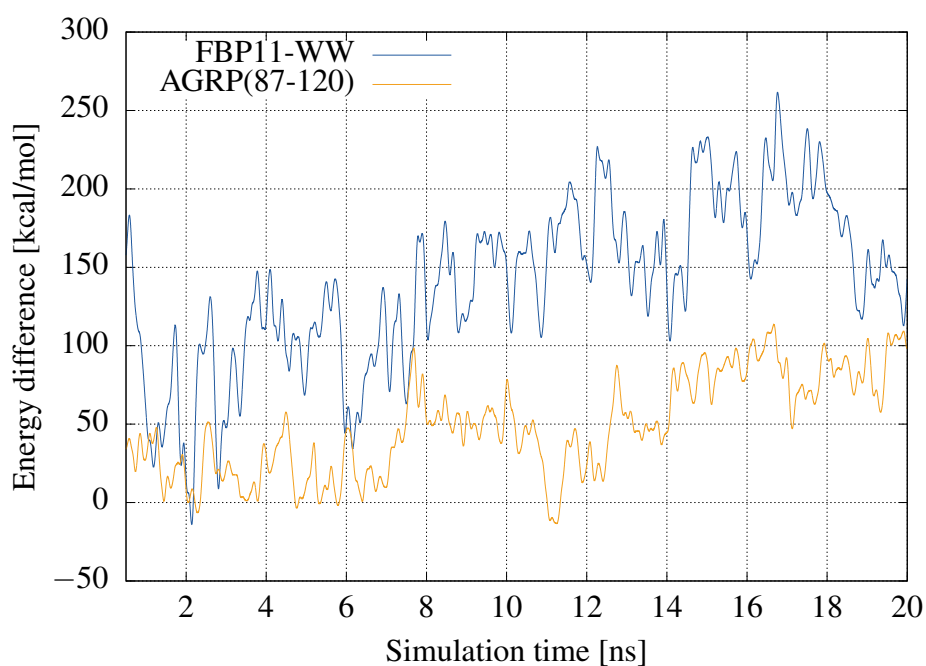


Figure 6.13. Potential energy difference when miniprotein is inside the pore and far away from the pore. The differences were taken from LPF data of the bonded and non-bonded interactions.

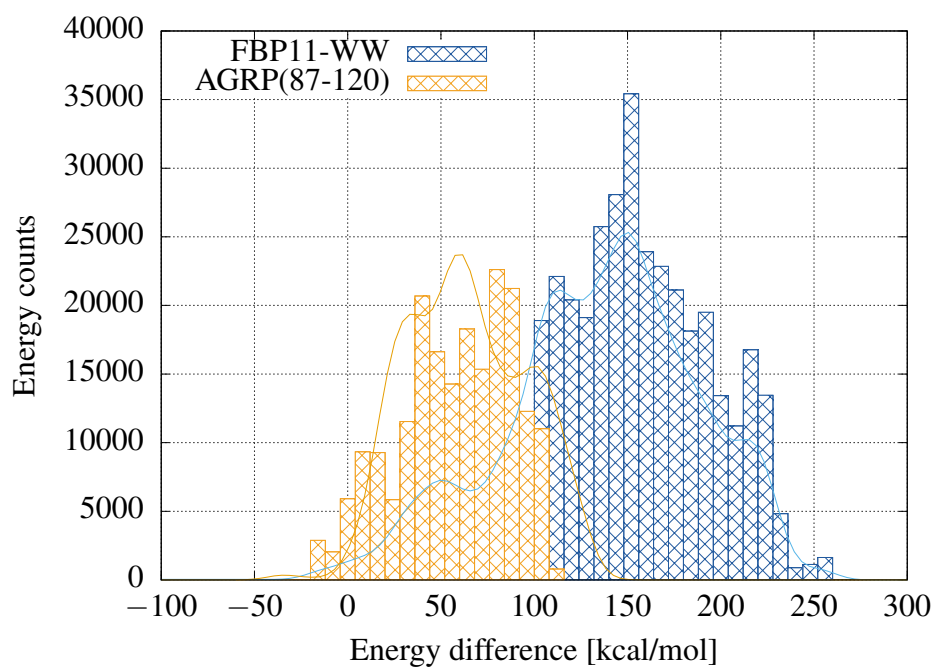


Figure 6.14. Histogram of the potential energy difference when the miniproteins were inside the pore and far away from the pore.

6.3 Comparative dwell times

To compute the interaction energy between miniprotein and pore wall explicitly and with increased accuracy, I carried out DFT simulations for each of the amino acids found in the miniprotein's sequence.

Total energy difference

The interaction strength was determined based on the difference in the total energy of the membrane and amino acid at their equilibrium distance and at the asymptotic limit. To achieve this I placed the amino acid close to the membrane surface and gradually increased their separation distance. At each step I determined the total energy, once self-consistency was reached. To get a good estimate of the asymptotic limit, the distance between mirrored images of the simulated environment had to be large. I studied the energy dependence on the simulation box size and show the results in appendix C. For a few selected amino acids I show the total energy results in figure 6.15.

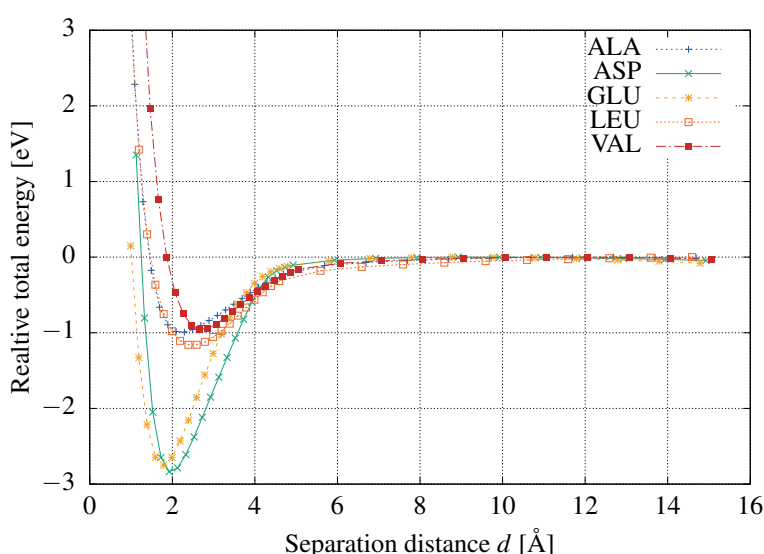


Figure 6.15. Total energy as a function of separation distance for selected amino acids. The energy was computed after the calculations reached self-consistency. The separation distance was defined between a line at the surface of the membrane and the atom making up the amino acid with the lowest z -position. To compare results between amino acids, I shifted the energies such that the total energy approached zero at far away separation distances.

In the results above, it is possible to see a slight lowering of the energy at very large distances. The energy difference was therefore taken at the minima of the potential well and the maxima in the asymptotic limit. It was not possible to attain energies for every separation space imaginable. To solve this I fitted a Morse potential function of the form

$$M(d) = A(e^{-2B(d-C)} - 2e^{-B(d-C)}) + D \quad (6.1)$$

where A , B , C and D are parameters to be fitted, to the total energy profiles for each amino acid to obtain a smoother energy description. To achieve the best possible fit, I modeled individual Morse potentials to data points computed near the potential-well and those found at the asymptotic limit. In figure 6.16 I demonstrate how the Morse function describes the behavior of the energy. In table 6.1 I tabulate the computed changes in energy with respect to the energy of the system when the molecule was far away from the pore wall.

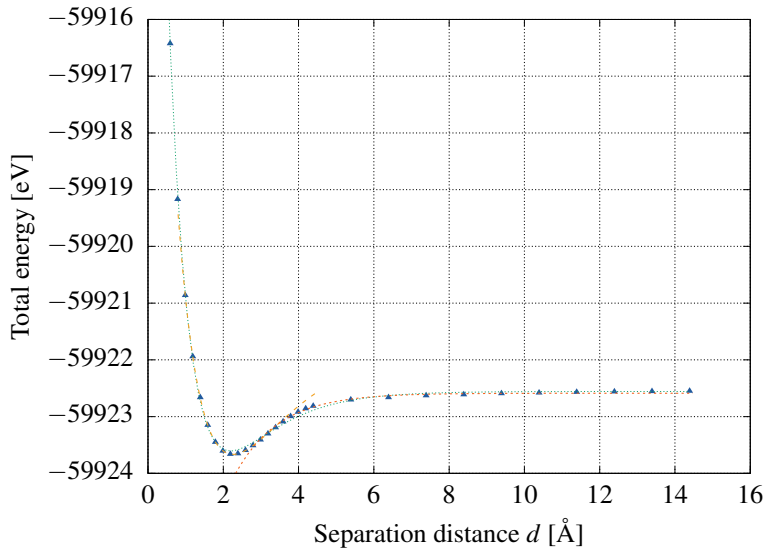


Figure 6.16. Total energy of tryptophan (TRP) and pore wall as a function of separation distance. Results from DFT calculations are represented by blue triangles and are fitted to a Morse potential represented by a blue dashed line. The Morse potential models the general trend of the energies as a function of distance, but sometimes can do poorly fitting the potential well or asymptotic limit. I therefore fitted individual Morse potentials for short (dashed orange line) and large (dashed red line) separation distances.

Table 6.1. Calculated energy differences found at the equilibrium point and at large separation distances for each amino acid. The energies at these two regions are computed by separate Morse potential best-fits. Amino acids are organized based on the charge of the side chain determined at physiological pH [83]. Uncertainties are 1σ deviation from the mean fit given by best-fits using all combinations of fitted parameters.

Side-chain	Amino acid	Energy min. [eV]	Energy max. [eV]	Energy diff. [eV]
positively charged	HIS	-59367.11 ± 0.09	-59366.241 ± 0.006	-0.9 ± 0.1
	ARG	-59723.53 ± 0.07	-59720.9559165	-2.57 ± 0.07
	LYS	-59179.86 ± 0.07	-59176.10 ± 0.05	-2.9 ± 0.1
negatively charged	ASP	-59354.01 ± 0.08	-59351.170 ± 0.005	-2.84 ± 0.09
	GLU	-59540.80 ± 0.05	-59538.042 ± 0.006	-2.76 ± 0.06
uncharged	SER	-58761.39 ± 0.03	-58759.887 ± 0.009	-1.50 ± 0.04
	THR	-58949.98 ± 0.03	-58948.990 ± 0.008	-0.99 ± 0.04
	ASN	-59205.56 ± 0.04	-59204.303 ± 0.007	-1.26 ± 0.05
	GLN	-59393.88 ± 0.08	-59392.913 ± 0.008	-0.97 ± 0.09
hydrophobic	ALA	-58324.64 ± 0.02	-58323.651 ± 0.005	-1.02 ± 0.02
	LEU	-58891.89 ± 0.04	-58890.767 ± 0.009	-1.12 ± 0.05
	PHE	-58667.731 ± 0.009	-58666.844 ± 0.005	-0.89 ± 0.01
	TRP	-59923.61 ± 0.05	-59922.575 ± 0.010	-1.04 ± 0.06
	TYR	-59760.93 ± 0.03	-59759.682 ± 0.010	-1.24 ± 0.04
	VAL	-58702.39 ± 0.03	-58701.467 ± 0.005	-0.92 ± 0.03
other	CYS	-58643.94 ± 0.02	-58642.477 ± 0.009	-1.47 ± 0.03
	GLY	-58136.11 ± 0.09	-58134.62 ± 0.02	-1.5 ± 0.1
	PRO	-58669.27 ± 0.01	-58668.183 ± 0.007	-1.09 ± 0.02

Dwell time estimate

From the relative surface accessibility and changes in energies of each amino acid, I estimated a dwell-time τ for the two miniproteins. We considered the dwell time to be proportional to the transition between two states; when the protein was interacting with the pore wall s_1 and when it was far away from it s_2 . The transition between these two states is determined by the ratio of their probabilities which is given by a Boltzmann distribution function, such that

$$\tau \propto \frac{p_{s_1}}{p_{s_2}} = \frac{e^{-\beta E_{\text{interacting}}}}{e^{-\beta E_{\text{far-away}}}} = e^{-\beta \Delta E_{\text{protein}}}, \quad (6.2)$$

where $\Delta E_{\text{protein}} = E_{\text{interacting}} - E_{\text{far-away}}$ is the total interaction energy difference of the miniprotein and β represents a characteristic energy for the transitions. In equilibrium conditions β is defined as a factor that depends on the Boltzmann constant and temperature, but in our case the leading contribution for the transitions came from the bias voltage applied across the membrane. If we represent the miniprotein as an object of unit charge, when subject to a 1.0 V bias voltage, the characteristic energy β can be estimated as 1 eV. Equation (6.2) allowed us to directly compare transition states to dwell times using an arbitrary value such that, if the interaction between pore wall and miniprotein was very small or none existent, the measured dwell time approached τ_0 . In other words,

$$\text{if } \Delta E_{\text{protein}} \approx 0, \quad \tau = \tau_0 e^{-\beta \Delta E_{\text{protein}}} \approx \tau_0.$$

In the case where interaction energy between these two was repulsive,

$$\Delta E_{\text{protein}} > 0, \quad \tau = \tau_0 e^{-\beta \Delta E_{\text{protein}}} \rightarrow \text{small}, \quad \tau < \tau_0$$

and the protein on average was expected to translocate in less time than the dwell time τ_0 . Finally if the interaction was attractive

$$\Delta E_{\text{protein}} < 0, \quad \tau = \tau_0 e^{-\beta \Delta E_{\text{protein}}} \rightarrow \text{large}, \quad \tau > \tau_0$$

and the protein's translocation was expected to take longer.

To determine $\Delta E_{\text{protein}}$, I conjectured that at any given time in a typical translocation event at most N amino acids in the sequence found themselves at their equilibrium distance from the pore wall and only those had a leading contribution on $\Delta E_{\text{protein}}$. It remained to determine, which amino acids? And how many? To answer the first question, I decided to select individual amino acids from a weighted random choice based on the computed RSA values (see section 5.1 for more details). I argued that the probability for any amino acid to be close to the pore wall depended on how solvent-accessible they were. To determine the number of amino acids, we were motivated by experimental dwell time measurements. Investigations by Lu et al. on tau peptides fragments using biological nanopores and Larkin et al. on large proteins using solid-state nanopores, revealed that dwell times for a single type of protein ranged between 2 or 3 orders of magnitudes [84, 85]. Based on these observations and the dwell times computed by considering different number of amino acids, we estimated that 4 amino acids contributed to $\Delta E_{\text{protein}}$.

From the method described above with amino acids selections repeated 100'000 times, I determined the energy changes for each miniprotein to be

$$\begin{aligned} \Delta E_{\text{AGRP}(87-120)} &= -5.39 \pm 1 \text{ eV} \text{ and} \\ \Delta E_{\text{FBP11-WW}} &= -7.46 \pm 2 \text{ eV.} \end{aligned}$$

Based on these results, one can expect both translocations to be slower than τ_0 , but their dwell times on average to be one order of magnitude apart

$$\begin{aligned} \tau_{\text{AGRP}(87-120)} &= (219 \pm 41) \tau_0 \text{ and} \\ \tau_{\text{FBP11-WW}} &= (1737 \pm 470) \tau_0. \end{aligned}$$

7. Discussion

Occupied pore ionic current

The MM simulations served mainly to model the ionic currents that might be detected during typical translocation events for the two miniproteins in an experiment. Initially, we explored the possibility of restraining the protein at its center of mass, which allowed the protein the freedom to rotate around this point and also to reach different conformations. With this solution, we found that even at low voltages the protein denatured. One potential cause could be attributed to the directed movement of ions and water in the pore region, which forced the ends of the protein to slowly unfold. Another could be that since the protein is polar, after alignment with the external field, the side chains with most net charge slowly were forced into an unfolded configuration. We also explored the possibility of defining a protein-specific potential-well around the pore region with barriers large enough that prevented the protein from leaving the pore. This solution experienced the same drawbacks as the first solution. Restraining the backbone of the protein and later allowing the molecule to be unrestrained, was a compromise between a model that is close to experimental conditions but also unphysical due to the artificial forces that allowed the protein to remain unfolded for long simulation times. I discuss the results obtained using this latter method.

When the backbones were harmonically restrained, simulation showed noticeable differences in the occupied ionic currents between the two miniproteins. Larger currents were computed for FBP11-WW domain than for AGRP(87-120). When the system was subject to 0.5 V voltage drop, the differences in ionic currents were in the order of approx. 1.5 nA and their distributions were roughly similar in shape. When the system was subject to 1.0 V, the measured occupied currents for both proteins increased as expected. Their distributions showed relative differences between them also increased to approx. 3.0 nA. The shape of the distribution of FBP11-WW domain remained the same, but AGRP(87-120) developed a tail towards larger ionic currents with similar magnitudes as those of FBP11-WW.

More realistic experimental conditions were achieved by letting the protein be unrestrained. Now it became possible for the protein to move entirely out of the pore, due to a combination of Brownian motion and electrophoretic/electroosmotic forces. I therefore decided to run shorter simulations in intervals of 10 nanoseconds. After the first set of simulations, I continued to evolve the system for a further 10 nanoseconds if the protein was still near the center of the pore. Otherwise if the protein was close to the edge of the pore or left the pore, I restarted the simulations for 10 nanoseconds from the last step of the harmonic-restraint. When the system was exposed to 0.5 V, both proteins remained inside the pore for the total 20 nanoseconds. But AGRP(87-120) reached the edge of the nanopore after 10 nanoseconds, when the system was exposed to 1.0 V. These observations suggested that differences could exist in the interaction strength between the two miniproteins when in contact with the pore wall. It also suggested that for large voltages, the translocation time for AGRP(87-120) could be too short-lived to be detected in real experiments considering limitations created by the noise at these short time scales.

Unrestrained simulations showed no differences in the occupied ionic currents between the miniproteins subject to either bias voltage. These results suggested that experiments might not be able to differentiate between these two miniproteins based on the modulation of the ionic current alone. Still it is unclear whether differences could exist in the occupied pore currents when these proteins enter or leave the pore. The simulations I carried

out (membrane has zero net change) show that sections of the proteins felt an attractive force to the pore wall. A possibility also exists that if the silicon nitride surface charge could be modified by changing the charge of the membrane, the interaction between pore wall and miniprotein might change to different extents. This modification might show as differences in blocked ionic currents as well. In the introduction I mentioned that a number of sub-events can happen in typical translocations. It is also possible, due to the geometry or interaction with the membrane, that bias exist in the sub-events for one of the miniproteins that could further influence recorded current modulations.

MM simulations rely on the accuracy of the force field to be able to describe an environment that closely resembles reality. I went to length to model the transport properties of ions using the TIP3P water model. By introducing an artificial viscosity via the thermostat I was able to adjust the conductivity of KCl close to experimental conditions. Such modifications also have a disadvantage, since large viscosities could lead to undesired effects in the simulation such as stiffening of the protein. Improvements could be made by introducing a different water model that has a water diffusivity closer to experimental values. A systematic study needs to be made along with the CHARMM36 FF to check a different water model ie. describe well conformation in the protein. I also made an effort to model the response of the silicon nitride membrane to the external field as to obtain a field profile inside the pore that resembles experiments. One improvement that can be made to the simulated environment is to perform a wettability test on the silicon nitride membrane to improve the description of transport properties. Finally, the CHARMM36 and MSXX FF were fitted independently, so the interaction between the nanopore and miniprotein largely depends on the non-bonded description between these two being *good enough*.

Dwell time - MM simulations

Aware of the shortcomings of the FF-based results, I explored changes in the potential energy of the miniproteins, when they were located close to the pore wall and far away from it. This energy difference, when compared against the two miniproteins could be also related to dwell time differences in their typical translocations. The results summarized in a histogram when the system is subject to 0.5 V, showed significant differences in the way their potential energy changed. The results suggested that the time these miniproteins spend inside the nanopore could be different, but it was not possible to relate these results to an absolute measure of the dwell time. Nevertheless the results were encouraging enough to move forward with a more accurate simulation method.

Dwell time - DFT simulations

To improve the description of the interaction between pore wall and miniprotein, I performed electronic structure calculations using DFT. For this I created a smaller system and decided to investigate each amino acid in the sequence of these two miniproteins individually. The pore wall in the hexagonal membrane was replaced by a smaller silicon nitride slab that was rotated such that the surface facing the protein resembled the pore surface as much as possible. In this new system, the curvature of the pore wall was absent. To speed up the calculations, water and ions were also ignored. This is a crude approximation that could have a significant influence in the interaction results, as suggested by the investigation done by de Freitas et al. on the interaction of DNA nucleotides in graphene [86]. The total energy difference for each miniprotein in these two environments depended on the relative surface accessibility. RSA values were computed over a single snapshot of the miniproteins. An improvement on the model would be to capture conformation snapshots from MM simulations, while the protein remains inside the pore and compute RSA values for each of these to obtain an average value.

Intuitively, the total energy of the system when the amino acid is far from the membrane should match the total energy of the system computed from independent simulations of the membrane and the amino acid alone. In reality, the energy of these two arrangements

do not match exactly due to what is referred as the basis set superposition error. For a finite set, the more basis functions are used the better the description of the energy. When the protein and membrane are brought together in the same system some of these bases overlap and the number of available states for those atoms increases. This consequently improves the description of the energy when compared to those two systems simulated individually, bringing about a numerical error in the solution. I therefore obtained energy changes from simulations where the separation distance between amino acid and pore wall were gradually increased.

Generally, amino acids sharing the same type of side chain have similar energy differences as shown in table 6.1 . One notable exception was histidine (HIS), which had a much lower energy compared to arginine (ARG) and lysine (LYS). One possible explanation is that histidine is a rather short amino acid compared to the other two negatively charged side chain amino acids, therefore the amino and carboxyl group are closer to the surface of the membrane and likely contribute in a significant way to the total energy. Additionally the *R* group of histidine, rather than being a linear carbon chain like in most other amino acids, is a carbon (imidazole) ring. Proline (PRO), phenylalanine (PHE), tryptophan (TRP) and tyrosine (TYR) also have a carbon rings of various types and similar to histidine their energy difference hovered around 1 eV. Disparities between computed energies for amino acids could also be attributed to the fact that it was not possible to control the exact position of the biomolecule with respect to distance of all silicon nitride atoms along the x-y plane.

Larger changes in energies for charged amino acids compared to those computed for uncharged ones made sense, since silicon nitride is polar. We expected positively charged side group to interact strongly with nanopore surface atoms. One improvement that could be made in this regard is to also take the polar influence of the applied bias voltage on the silicon nitride membrane. I expect such change will reduce the polarization effect from the amino acid on the nanopore wall.

We computed a larger energy difference for FBP11-WW domain, compared to the energy difference found for AGRP(87-120). These finding qualitatively agree also with the previously computed energy differences derived from MM simulations. The results suggest FBP11-WW interacts with the pore walls more strongly therefore its dwell-time, when compared to the dwell-time of AGRP(87-120) is expected to be longer.

Answering the thesis

We considered that occupied pore currents distribution results from MM simulations yield ionic current modulations recoded from typical translocation events. We also considered that relative changes in the total energy of the protein-nanopore system when close to the pore wall and far away from it are proportional to the dwell-time of a typical translocation event. These two factors were used to assess if nanopores could be employed to sense and identify differences between them. We found that while the occupied pore currents were roughly equal for both molecules, their dwell times can be expected to be different. These results thus show promise in the experimental detection of these miniproteins.

Dwell time in particular plays an important role in the potential detection of these miniproteins. As stated by Di et al., recorded measurements due to noise are typically filtered to frequencies of 1kHz-100kHz, but Rosenstein et al. have also demonstrated high signal to noise ratio at detection frequencies of 1MHz [1, 12, 82]. From this investigation it is not possible to determine the exact dwell time, but observations in the movement of the proteins inside the pore from MM simulations suggested that AGRP(87-120) might experience translocation times in the order of tens to hundreds nanoseconds under a 1.0 V bias voltage. Sub-microsecond dwell times means translocation events might be very difficult, if not impossible, to detect under these conditions for this miniprotein. The situation could improve significantly if much lower voltages are used and the instrumentation noise can be kept low without the need for filtering. Surface charge modification, salt concen-

tration gradients could also slow down translocation times. Specially encouraging are experimental results from Di et al., Nir et al. and Huang et al. that successfully demonstrated solid-state and biological nanopore sensing capabilities for kDa proteins [12, 87, 88]. We were discouraged from employing low voltages typically used in experimental settings (ie. 0.4-0.1 V) for MM simulation, because it meant lower statistics since the number of ions passing through the pore is reduced. Limitation on the statistics can be mitigated by accessing longer simulation times, which can be achieved in a number of ways. For example, the time-step for recoding ionic positions can be made larger, since recording atomic positions takes memory and computational time. In fact measured occupied ionic currents from the MM simulations I carried out showed that ion positions did not need to be recorded every 1 picosecond, but rather could be increased by one or two orders of magnitude. This is evident by the large noise captured from the measured ionic currents. Another example is taking advantage of a simplified water model. I used a static water model that opens the possibility to use larger time-step ie. 2 femtoseconds, that also increases simulated times at reduced computational costs.

8. Conclusion

Molecular mechanics and density functional theory results combined carried out in this project indicate that silicon nitride nanopores show potential in the detection and identification of miniproteins. The two miniproteins I computationally studied showed similar occupied pore currents, but their dwell-times were estimated to differ by one order of magnitude. These results, although involving only two miniproteins very similar in shape, demonstrate miniproteins have the potential to exhibit distinct-enough changes in their structure, shape or composition such that their detection might be possible in an experiment of similar characteristics. To make more concrete statements about the detection of miniproteins in experiments, the work I have presented in this thesis can be expanded in several ways.

One possibility is to examine a larger number of different miniproteins. An obvious choice is to continue with miniproteins that involved in cellular processes ie. P53 tetramerization domain, fragment(1-31) of human parathyroid hormone and the third extracellular loop of cholecystokinin-A receptor. Larger polymers or even full proteins are also relevant, but the advantage of using domains is that they can potentially make the study of very larger proteins systematic.

MM and DFT simulations can be further developed, such that both results are much more interconnected with each other. For example, I could use snapshots obtained directly from MM simulations that include water molecules and ions to carry out DFT calculations of individual amino acids. The inclusion of water and ions changes the interaction energies between biomolecule and pore wall; computing and averaging the interaction energy over several conformations can yield more accurate results. The positioning of each individual amino acid with respect to the structure of the miniprotein can also be considered relevant. I take cysteine as a concrete example, which in AGRP(87-120) makes cysteine-knots at various places in the miniprotein. DFT simulations in this work place the sulfur atom facing towards the surface of the pore wall, but in reality the R-group combines to form disulfide bond that face the structure. One last consideration can be made to the influence of the electric field on the pore wall and amino acid. Amino acids are polar and silicon nitride is a dielectric material. In the calculations I carried out, the amino acid polarizes the silicon nitride which influences their interaction strength. In reality the silicon nitride is polarized mainly by the external bias voltage applied across the system. Such external influence should also be taken into consideration in DFT calculations.

The connection between MM and DFT simulations can also be strengthened in the opposite direction, by utilizing DFT calculations to obtain a better description between the pore wall and amino acid for MD simulations. At the moment I utilized the CHARMM36 non-bonded parameters to describe the interaction between individual atoms found in the pore wall and miniprotein. Carrying out DFT for small sections of the miniprotein and silicon nitride slab, I could develop a custom force field that could act independently of the CHARMM36 and MSXX FF to get a better description of these two objects interacting together.

The current method for estimating relative dwell-time differences between miniproteins could also be improved if RSA values are computed over difference conformations of the protein. We determined the number of ions that have a leading contribution on the interaction strength of the biomolecules, based on the range of experimentally observed dwell times for a single type of protein. A study on its own could be to use a more sophisticated method to model the number of amino acids with leading contribution. It is

possible to even expand the model to consider secondary and tertiary contributions, based on the fact that at large distances we expect the total energy to fall approximately as $1/r^6$.

A different approach all together can also be taken to investigate miniprotein sensing using nanopores. Originally we had explored the possibility of using a quantum-mechanics/molecular-mechanics (QM/MM) method to model the membrane-protein interactions. This method splits the entire system into two regions; the QM and MM regions. The QM is a small region of interest where the protein and pore wall are most proximate to each other. The MM region is then everything else not encompassed in the QM region. DFT simulations are carried out in the QM region considering the electrostatic influence of the MM region as an external potential in the Hamiltonian. Dividing the system this way means that covalent bonds in the membrane and the miniprotein must be broken. One method to deal with broken bonds is the link-atom approach, where dangling bonds in the QM region are saturated with hydrogens.

9. Acknowledgements

NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign [89]. The computations were performed on resources provided by SNIC through HPC cluster Tetralith and Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) and under Project SNIC 2020/5-193. Some images of miniproteins and/or nanopores were generated using the molecular visualization program VMD [90]. Additional images of biomolecules and/or nanopores were generated using the open-source version of PyMOL molecular visualization system. Relative accessibility values were computed using the modeling environment Swiss PDB-Viewer [78]. Density functional theory simulations were carried out by the computer program implementation SIESTA [36].

I would like to take the opportunity to thank my supervisor Ralph Scheicher. Ralph, you have helped me tremendously over the years, always taking the time to discuss topics, giving me the opportunity to contribute, letting me be involved from day one on your research and encouraging me to follow what I found interesting. It has been three years that have really helped me grow academically, so thank you and I hope that we can continue collaborating in the future. Ralph is a great supervisor, super knowledge-full, always gives valuable feedback, always took the time to understand what I had to say and explain information in an easy and enthusiastic manner. I can strongly recommend Ralph to anyone who is interested in this field of physics.

Thank you to Prof. Shi-Li Zhang and Dr. Chenyu Wen for giving me an insight into the experimental world of nanopores. Thanks to Dr. Jeffrey Comer, Dr. Ernane de Freitas Martins, Jacob and Ibbe for discussion and helping me on the methods that I used for this thesis. Thanks to my colleagues at the MT department for the nice talks around the corridors.

Malin, family and friends: I thank you for filling this time with great memories and for being there every step of the way. Malin, familia y amigos: les agradezco por llenar este tiempo con grandes memorias y por estar presentes en cada momento.

References

- [1] Nitinun Varongchayakul, Jiayi Song, Amit Meller, and Mark W Grinstaff. Single-molecule protein sensing in a nanopore: a tutorial. *Chemical Society Reviews*, 47(23):8512–8524, 2018.
- [2] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. Chapter 3 proteins. In *Molecular Biology of the Cell. 4th edition*. Garland Science, 2002.
- [3] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. Chapter 6 from dna to rna. In *Molecular Biology of the Cell. 4th edition*. Garland Science, 2002.
- [4] Vivien Marx. A dream of single-cell proteomics. *Nature methods*, 16(9):809–812, 2019.
- [5] Wei Si and Aleksei Aksimentiev. Nanopore sensing of protein folding. *ACS nano*, 11(7):7091–7100, 2017.
- [6] Michael Eisenstein. An ace in the hole for dna sequencing, 2017.
- [7] Yanxiao Feng, Yuechuan Zhang, Cuifeng Ying, Deqiang Wang, and Chunlei Du. Nanopore-based fourth-generation dna sequencing technology. *Genomics, proteomics & bioinformatics*, 13(1):4–16, 2015.
- [8] Daojing Wang and Steven Bodovitz. Single cell analysis: the new frontier in ‘omics’. *Trends in biotechnology*, 28(6):281–290, 2010.
- [9] Bo Huang, Hongkai Wu, Devaki Bhaya, Arthur Grossman, Sebastien Granier, Brian K Kobilka, and Richard N Zare. Counting low-copy number proteins in a single cell. *Science*, 315(5808):81–84, 2007.
- [10] Rong Fan, Ophir Vermesh, Alok Srivastava, Brian KH Yen, Lidong Qin, Habib Ahmad, Gabriel A Kwong, Chao-Chao Liu, Juliane Gould, Leroy Hood, et al. Integrated barcode chips for rapid, multiplexed analysis of proteins in microliter quantities of blood. *Nature biotechnology*, 26(12):1373, 2008.
- [11] Yuhui He, Makusu Tsutsui, Chun Fan, Masateru Taniguchi, and Tomoji Kawai. Controlling dna translocation through gate modulation of nanopore wall surface charges. *ACS nano*, 5(7):5509–5518, 2011.
- [12] Nicolas Di Fiori, Allison Squires, Daniel Bar, Tal Gilboa, Theodore D Moustakas, and Amit Meller. Optoelectronic control of surface charge and translocation dynamics in solid-state nanopores. *Nature nanotechnology*, 8(12):946–951, 2013.
- [13] Tim Albrecht. Single-molecule analysis with solid-state nanopores. *Annual Review of Analytical Chemistry*, 12, 2019.
- [14] Taoli Ding, Antony K Chen, and Zuhong Lu. The applications of nanopores in studies of proteins. *Science Bulletin*, 2019.
- [15] Murugappan Muthukumar. Mechanism of dna transport through pores. *Annu. Rev. Biophys. Biomol. Struct.*, 36:435–450, 2007.
- [16] Marcel Aguilera-Arzo and Vicente M Aguilera. Access resistance in protein nanopores. a structure-based computational approach. *Bioelectrochemistry*, 131:107371, 2020.
- [17] Yuhuan Luo, Linlin Wu, Jing Tu, and Zuhong Lu. Application of solid-state nanopore in protein detection. *International Journal of Molecular Sciences*, 21(8):2808, 2020.
- [18] Jared Houghtaling, Cuifeng Ying, Olivia M Eggenberger, Aziz Fennouri, Santoshi Nandivada, Mitu Acharjee, Jiali Li, Adam R Hall, and Michael Mayer. Estimation of shape, volume, and dipole moment of individual proteins freely transiting a synthetic nanopore. *ACS nano*, 2019.
- [19] Matthias Firnkes, Daniel Pedone, Jelena Knezevic, Markus Doblinger, and Ulrich Rant. Electrically facilitated translocations of proteins through silicon nitride nanopores: conjoint and competitive action of diffusion, electrophoresis, and electroosmosis. *Nano letters*, 10(6):2162–2167, 2010.
- [20] Jiali Li, Marc Gershow, Derek Stein, Eric Brandin, and Jene Andrew Golovchenko. Dna molecules and configurations in a solid-state nanopore microscope. *Nature materials*, 2(9):611–615, 2003.
- [21] Todd M Squires and Martin Z Bazant. Induced-charge electro-osmosis. *Journal of Fluid Mechanics*, 509:217–252, 2004.
- [22] Vincent Tabard-Cossa. Instrumentation for low-noise high-bandwidth nanopore recording. In *Engineered Nanopores for Bioanalytical Applications*, pages 59–93. Elsevier, 2013.

- [23] Vincent Tabard-Cossa, Dhruti Trivedi, Matthew Wiggin, Nahid N Jetha, and Andre Marziali. Noise analysis and reduction in solid-state nanopores. *Nanotechnology*, 18(30):305505, 2007.
- [24] Meni Wanunu. Nanopores: A journey towards dna sequencing. *Physics of life reviews*, 9(2):125–158, 2012.
- [25] Emily G Baker, Gail J Bartlett, Kathryn L Porter Goff, and Derek N Woolfson. Miniprotein design: Past, present, and prospects. *Accounts of chemical research*, 50(9):2085–2092, 2017.
- [26] Derek N Woolfson, Emily G Baker, and Gail J Bartlett. How do miniproteins fold? *Science*, 357(6347):133–134, 2017.
- [27] Helen Berman, Kim Henrick, and Haruki Nakamura. Announcing the worldwide protein data bank. *Nature Structural & Molecular Biology*, 10(12):980–980, 2003.
- [28] Yusuke Kato, Yumi Hino, Koji Nagata, and Masaru Tanokura. Solution structure and binding specificity of fbp11/hypa ww domain as group-ii/iii. *Proteins: Structure, Function, and Bioinformatics*, 63(1):227–234, 2006.
- [29] Pilgrim J Jackson, Joseph C McNulty, Ying-Kui Yang, Darren A Thompson, Biaoxin Chai, Ira Gantz, Gregory S Barsh, and Glenn L Millhauser. Design, pharmacology, and nmr structure of a minimized cystine knot with agouti-related protein activity. *Biochemistry*, 41(24):7565–7572, 2002.
- [30] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. Isolating, cloning, and sequencing dna. In *Molecular Biology of the Cell. 4th edition*. Garland Science, 2002.
- [31] Mark Allen, Assaf Friedler, Oliver Schon, and Mark Bycroft. The structure of an ff domain from human hypa/fbp11. *Journal of molecular biology*, 323(3):411–416, 2002.
- [32] Kai-Ti Lin, Ruei-Min Lu, and Woan-Yuh Tarn. The ww domain-containing proteins interact with the early spliceosome and participate in pre-mrna splicing in vivo. *Molecular and cellular biology*, 24(20):9176–9185, 2004.
- [33] Ralph A Bradshaw and Edward A Dennis. *Handbook of cell signaling*. Academic press, 2009.
- [34] Shalini Iyer and K Ravi Acharya. Tying the knot: The cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines. *The FEBS journal*, 278(22):4304–4322, 2011.
- [35] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16):1781–1802, 2005.
- [36] José M Soler, Emilio Artacho, Julian D Gale, Alberto García, Javier Junquera, Pablo Ordejón, and Daniel Sánchez-Portal. The siesta method for ab initio order-n materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002.
- [37] Oren M Becker, Alexander D MacKerell Jr, Benoît Roux, and Masakatsu Watanabe. *Computational biochemistry and biophysics*. Crc Press, 2001.
- [38] Susanna Hug. Classical molecular dynamics in a nutshell. In *Biomolecular Simulations*, pages 127–152. Springer, 2013.
- [39] Luca Monticelli and D Peter Tieleman. Force fields for classical molecular dynamics. In *Biomolecular simulations*, pages 197–213. Springer, 2013.
- [40] Alex D MacKerell Jr, Donald Bashford, MLDR Bellott, Roland Leslie Dunbrack Jr, Jeffrey D Evanseck, Martin J Field, Stefan Fischer, Jiali Gao, H Guo, Sookhee Ha, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, 102(18):3586–3616, 1998.
- [41] Alexander D MacKerell Jr. Empirical force fields for biological macromolecules: overview and issues. *Journal of computational chemistry*, 25(13):1584–1604, 2004.
- [42] Alexander D Mackerell Jr, Michael Feig, and Charles L Brooks III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of computational chemistry*, 25(11):1400–1415, 2004.
- [43] Robert B Best, Xiao Zhu, Jihyun Shim, Pedro EM Lopes, Jeetain Mittal, Michael Feig, and Alexander D MacKerell Jr. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *Journal of chemical theory and computation*, 8(9):3257–3273, 2012.
- [44] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log(n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

- [45] Simon W de Leeuw, John William Perram, and Edgar Roderick Smith. Simulation of electrostatic systems in periodic boundary conditions. i. lattice sums and dielectric constants. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 373(1752):27–56, 1980.
- [46] Timothy A Isgro, Marcos Sotomayor, and Eduardo Cruz-Chu. Case study: Water and ice, 2011.
- [47] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.
- [48] Pekka Mark and Lennart Nilsson. Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k. *The Journal of Physical Chemistry A*, 105(43):9954–9960, 2001.
- [49] Scott E Feller, Yuhong Zhang, Richard W Pastor, and Bernard R Brooks. Constant pressure molecular dynamics simulation: the langevin piston method. *The Journal of chemical physics*, 103(11):4613–4621, 1995.
- [50] Xin Yong and Lucy T Zhang. Thermostats and thermostat strategies for molecular dynamics simulations of nanofluidics. *The Journal of chemical physics*, 138(8):084503, 2013.
- [51] EA Koopman and CP Lowe. Advantages of a lowe-andersen thermostat in molecular dynamics simulations. *The Journal of chemical physics*, 124(20):204103, 2006.
- [52] CP Lowe. An alternative approach to dissipative particle dynamics. *EPL (Europhysics Letters)*, 47(2):145, 1999.
- [53] John M Wills, Mebarek Alouani, Per Andersson, Anna Delin, Olle Eriksson, and Oleksiy Grechnev. *Full-Potential Electronic Structure Method: energy and force calculations with density functional and dynamical mean field theory*, volume 167. Springer Science & Business Media, 2010.
- [54] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.
- [55] Richard M Martin. *Electronic structure: basic theory and practical methods*. Cambridge university press, 2020.
- [56] Jorge Kohanoff. *Electronic structure calculations for solids and molecules: theory and computational methods*. Cambridge University Press, 2006.
- [57] Javier Junquera. Norm-conserving pseudopotentials in electronic structure calculations. http://video.tau.ac.il/events/index.php?option=com_k2&view=item&id=5127:norm-conserving-pseudopotentials-in-electronic-structure-calculations&Itemid=552, September 2014.
- [58] Javier Junquera. Pseudopotentials. https://personales.unican.es/junqueraj/JavierJunquera_files/Talks/Pseudopotentials.pdf, 2015.
- [59] Pablo Ordejon. Siesta internal details. http://video.tau.ac.il/events/index.php?option=com_k2&view=item&id=5129:siesta-internal-details-h-s-grid-k-points&Itemid=552, September 2014.
- [60] Andrei Postnikov. Siesta caminando tutorial. *Universit e Paul Verlaine–Metz*, 2009.
- [61] Aleksij Aksimentiev, Jiunn B Heng, Gregory Timp, and Klaus Schulten. Microscopic kinetics of dna translocation through synthetic nanopores. *Biophysical journal*, 87(3):2086–2097, 2004.
- [62] Jeffrey R Comer, David B Wells, and Aleksei Aksimentiev. Modeling nanopores for sequencing dna. In *DNA Nanotechnology*, pages 317–358. Springer, 2011.
- [63] Alek Aksimentiev and Jeffrey Comer. Bionanotechnology tutorial. In *Proc. Comput. Biophys. Workshop*, pages 1–69, 2011.
- [64] Aleksij Aksimentiev and Klaus Schulten. Imaging α -hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophysical journal*, 88(6):3745–3761, 2005.
- [65] John A Wendel and William A Goddard III. The hessian biased force field for silicon nitride ceramics: Predictions of thermodynamic and mechanical properties for α - and β -si₃n₄. *The Journal of chemical physics*, 97(7):5048–5062, 1992.
- [66] Reinard Grün. The crystal structure of β -si₃n₄: structural and stability considerations between α - and β -si₃n₄. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, 35(4):800–804, 1979.
- [67] Simon M Sze and Kwok K Ng. *Physics of semiconductor devices*. John wiley & sons, 2006.
- [68] David J Griffiths. *Introduction to electrodynamics*. Cambridge university press, fourth edition edition, 2017.
- [69] Ramu Anandakrishnan, Boris Aguilar, and Alexey V Onufriev. H++ 3.0: automating p k prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations.

- Nucleic acids research*, 40(W1):W537–W541, 2012.
- [70] Jonathan Myers, Greg Grothaus, Shivaram Narayanan, and Alexey Onufriev. A simple clustering algorithm can be accurate enough for use in calculations of pks in macromolecules. *Proteins: Structure, Function, and Bioinformatics*, 63(4):928–938, 2006.
 - [71] John C Gordon, Jonathan B Myers, Timothy Folta, Valia Shoja, Lenwood S Heath, and Alexey Onufriev. H++: a server for estimating p k as and adding missing hydrogens to macromolecules. *Nucleic acids research*, 33(suppl_2):W368–W371, 2005.
 - [72] David A Case, Thomas E Cheatham III, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M Merz Jr, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J Woods. The amber biomolecular simulation programs. *Journal of computational chemistry*, 26(16):1668–1688, 2005.
 - [73] Elizabeth Jurrus, Dave Engel, Keith Star, Kyle Monson, Juan Brandi, Lisa E Felberg, David H Brookes, Leighton Wilson, Jiahui Chen, Karina Liles, et al. Improvements to the apbs biomolecular solvation software suite. *Protein Science*, 27(1):112–128, 2018.
 - [74] James Phillips, Timothy Isgro, Marcos Sotomayor, and Elizabeth Villa. Namd tutorial. 2003.
 - [75] Guillaume Lamoureux, Alexander D MacKerell Jr, and Benoit Roux. A simple polarizable model of water based on classical drude oscillators. *The Journal of chemical physics*, 119(10):5185–5197, 2003.
 - [76] M Bhandarkar, R Brunner, C Chipot, A Dalke, S Dixit, P Grayson, J Gullingsrud, A Gursoy, D Hardy, W Humphrey, et al. Namd user guide, April 2012. NAMD version 2.9.
 - [77] Matthew Z Tien, Austin G Meyer, Dariya K Sydykova, Stephanie J Spielman, and Claus O Wilke. Maximum allowed solvent accessibilities of residues in proteins. *PloS one*, 8(11):e80635, 2013.
 - [78] Nicolas Guex and Manuel C Peitsch. Swiss-model and the swiss-pdb viewer: an environment for comparative protein modeling. *electrophoresis*, 18(15):2714–2723, 1997.
 - [79] David C Langreth, Max Dion, Henrik Rydberg, Elsebeth Schröder, Per Hyldgaard, and Bengt I Lundqvist. Van der waals density functional theory with applications. *International journal of quantum chemistry*, 101(5):599–610, 2005.
 - [80] Max Dion, Henrik Rydberg, Elsebeth Schröder, David C Langreth, and Bengt I Lundqvist. Van der waals density functional for general geometries. *Physical review letters*, 92(24):246401, 2004.
 - [81] Molecular devices. *Axon Axopatch 200B Microelectrode Amplifier*, 2012.
 - [82] Jacob K Rosenstein, Meni Wanunu, Christopher A Merchant, Marija Drndic, and Kenneth L Shepard. Integrated nanopore sensing platform with sub-microsecond temporal resolution. *Nature methods*, 9(5):487, 2012.
 - [83] Dan Cojocari. Twenty-one amino acids. web, 2010. pKa data from CRC Handbook of Chemistry, v.2010.
 - [84] Joseph Larkin, Robert Y Henley, Murugappan Muthukumar, Jacob K Rosenstein, and Meni Wanunu. High-bandwidth protein analysis using solid-state nanopores. *Biophysical journal*, 106(3):696–704, 2014.
 - [85] Si-Min Lu, Xue-Yuan Wu, Meng-Yin Li, Yi-Lun Ying, and Yi-Tao Long. Diversified exploitation of aerolysin nanopore in single-molecule sensing and protein sequencing. *View*, page 20200006.
 - [86] Ernane de Freitas Martins, Rodrigo G Amorim, Gustavo Troiano Feliciano, Ralph Hendrik Scheicher, and Alexandre Reily Rocha. The role of water on the electronic transport in graphene nanogap devices designed for dna sequencing. *Carbon*, 158:314–319, 2020.
 - [87] Gang Huang, Kherim Willems, Misha Soskine, Carsten Wloka, and Giovanni Maglia. Electro-osmotic capture and ionic discrimination of peptide and protein biomarkers with frac nanopores. *Nature communications*, 8(1):1–11, 2017.
 - [88] Iftach Nir, Diana Huttner, and Amit Meller. Direct sensing and discrimination among ubiquitin and ubiquitin chains using solid-state nanopores. *Biophysical journal*, 108(9):2340–2349, 2015.
 - [89] James C Phillips, David J Hardy, Julio DC Maia, John E Stone, João V Ribeiro, Rafael C Bernardi, Ronak Buch, Giacomo Fiorin, Jérôme Hénin, Wei Jiang, et al. Scalable molecular dynamics on cpu and gpu architectures with namd. *The Journal of Chemical Physics*, 153(4):044130, 2020.
 - [90] William Humphrey, Andrew Dalke, Klaus Schulten, et al. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.
 - [91] Dong Xu, James Christopher Phillips, and Klaus Schulten. Protein response to external electric fields: relaxation, hysteresis, and echo. *The Journal of Physical Chemistry*, 100(29):12108–12121, 1996.
 - [92] Lou Coury. Conductance measurements part 1: theory. *Current Separations*, 18(3):92, 1999.

Appendix A.

Details on proteins and titration

I show the amino sequence (reads from top to bottom from left to right) for FBP11 WW domain. Titrable amino acid residues are marked by a star along with their pK value determined by H++.

GLY:10	PRO:18	ASN:26	TRP:34
SER:11	ASP:19* 3.7	THR:27	GLU:35* 4.9
RRP:12	GLY:20	GLU:28* 4.0	LYS:36* 11.0
THR:13	ARG:21* > 12.0	THR:29	PRO:37
GLU:14* 4.0	THR:22	LYS:30* 11.9	ASP:38* 3.8
HSE:15* 6.6	TYR:23 10.7	GLN:31	ASP:39* 1.8
LYS:16* 10.3	TYR:24* > 12.0	SER:32	
SER:17	TYR:25* 10.4	THR:33	

I also show the amino sequence for AGRP(87-120).

CYS:87	GLY:96	ALA:105	ASN:114
VAL:88	GLN:97	ALA:106	ALA:115
ARG:89* >12.0	GLN:98	TYR:107	PHE:116
LEU:90	VAL:99	CYS:108	CYS:117
HIS:91* 6.6	PRO:100	TYR:109* 10.5	TYR:118* 10.3
GLU:92* 3.9	CYS:101	CYS:110	CYS:119
SER:93	CYS:102	ARG:111* 11.6	ARG:120* >12.0
CYS:94	ASP:103* 3.0	PHE:112	
LEU:95	PRO:104	PHE:113	

Titration curves for some selected sites are shown in figures 9.1 and 9.2 for each miniprotein. The parameters used in H++ are summarized in table 9.1.

Table 9.1. Summary of H++ parameters used for both proteins. The protonation state of side chains depends on their electrostatic interaction with the environment [69]. Different results are resolved depending on the dielectric constants assumed for solvent and protein. Since I am most interested in surface interactions, I use a high value for the protein's dielectric constant. I also use a pH value within the range of intracellular pH.

Parameter	value
solvent dielectric	80
protein dielectric	10
ionic concentration	1 M
target pH	7.1

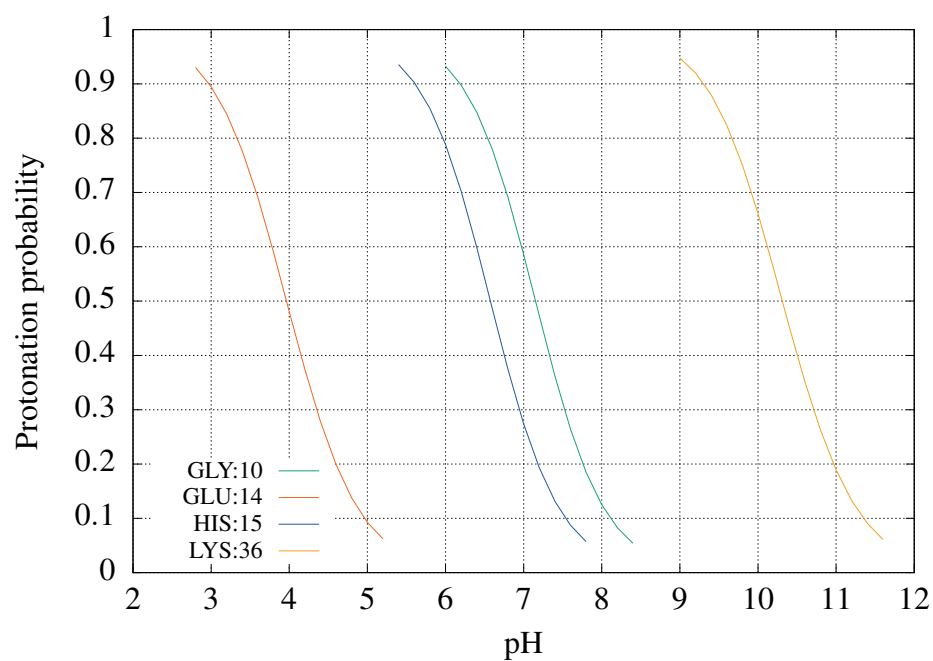


Figure 9.1. Titration probability as a function of pH for selected titrable amino acids in FBP11 WW domain miniprotein. The pK value is determined by the pH when the titration probability is half.

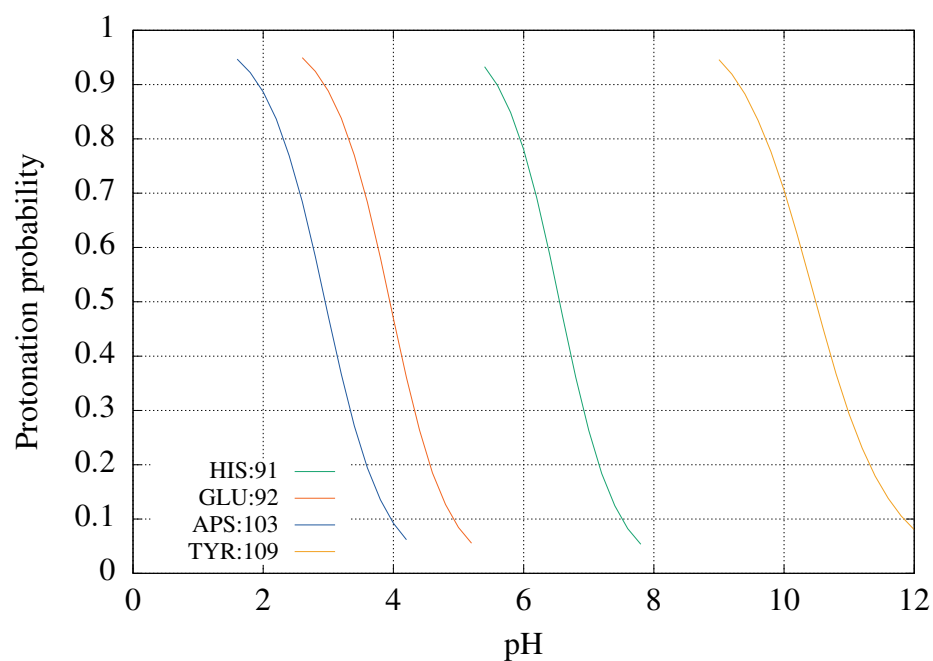


Figure 9.2. Same information as in figure 9.1 but for AGRP(87-120) miniprotein.

Appendix B.

Details on molecular dynamic simulations

In this section I show the procedures I used to improve the description of surface atoms in the silicon membrane pore, correctly model the movement of ions in a TIP3P environment using a momentum conserving thermostat and study the response of the membrane to an external field. In the last two procedures I was guided by experimental results and found parameters for MM configuration files, referenced in section 4, that best described the system.

Silicon nitride relative permittivity

The motion of ions and protein also depends on the response of the membrane to an applied potential difference such that the electric field distribution near and inside the pore region resembles experimental conditions. One approach is to apply an additional constraint on the membrane atoms to modify the relative permittivity of the material. A silicon nitride membrane is made of charged particles that induce an internal dipole moment \mathbf{D} that can be computed using the multi-pole expansion

$$\mathbf{D} = \sum_i^N (q_i - q_0) \mathbf{r}_i, \quad (9.1)$$

where N is the number of particles in the membrane with charge q_i at position \mathbf{r}_i [68]. The term q_0 is the monopole term of the expansion given by

$$q_0 = \sum_i^N q_i, \quad (9.2)$$

which accounts for the fact that dipole moment should be independent of the measurement coordinates [68]. When an external electric field is applied, the dipole moment tends to line up in the direction of the field and the material becomes polarized with its polarization \mathbf{P} described as the dipole moment per unit volume [68]. In the case of MM simulations, the polarization is instead determined by a time average

$$\mathbf{P} = \frac{\langle \Delta \mathbf{D} \rangle_t}{V}, \quad (9.3)$$

where $\Delta \mathbf{D}$ is the dipole moment change induced by applying an external field to the material and V the volume [91]. If I now consider a silicon nitride sphere that is uniform, the volume charge density of the membrane ρ_b can be neglected and the electric field inside the sphere depends only on its surface charge density σ_b [68]. It follows from this approximation that the electric field inside the sphere can be expressed as

$$\mathbf{E} = -\frac{1}{3\epsilon_0} \mathbf{P}, \quad (9.4)$$

where ϵ_0 is the permittivity of free space [68]. If I apply an uniform external field \mathbf{E}_{ext} to the dielectric material, I find the field inside the sphere can also be determined by

$$\mathbf{E} = \frac{3}{\epsilon_r + 2} \mathbf{E}_{\text{ext}}, \quad (9.5)$$

where ϵ_r is the relative permittivity of silicon nitride [68]. I can rearrange equation (9.5) to compute ϵ_r based on a time average of the changes to the dipole moment inside the sphere when an external field is present and absent in the system. A diagram of these two scenarios is shown in figure 9.3.

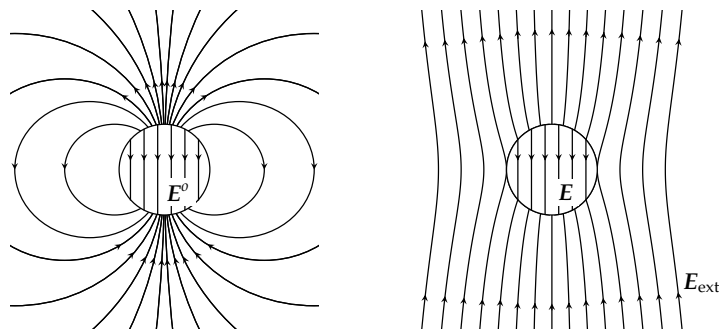


Figure 9.3. Diagram showing electric field lines for a homogenous linear dielectric sphere (left) in vacuum and (right) subject to an uniform external electric field E_{ext} . Adapted from [68].

The polarization of the sphere in MM simulations can be controlled by applying a harmonic restraint to all membrane atoms, which acts as a fictional force that restricts the alignment of individual dipole moments in response to the external field. Changes to the magnitude of the harmonic restraint therefore translate to adjustments in the relative permittivity, which can be used to match the experimental value for silicon nitride of 7.5.

To accomplish this, I ran two sets of simulations systematically adjusting the harmonic restraints to explore the space. The first set of simulations ran for 1 nanosecond and the second set, which spanned over a smaller region of harmonic restraints, ran for 1.6 nanoseconds. The membrane was placed in a vacuum cube 18 nanometers in length and the temperature was kept constant at 298 K using a Lowe-Andersen thermostat with the optimized R_T value determined in the previous section. Before production simulations, the system was minimized and further equilibrated for 1.0 picoseconds. In the cases when the sphere was subject to an external field, I applied a 1 V voltage drop along the z -direction. The results from these simulations are shown in figure 9.4.

The behavior of the relative permittivity in figure 9.4 agrees with my understanding of the FF parameters used, since the ϵ_r approaches a value of 1 as the magnitude of the harmonic restraint was lowered. I fitted a logarithmic function to these results and determined a restraint of 7.0 kcal/(molÅ²) gave reasonable agreements with experiments. The computed ϵ_r values proved to be very sensitive to the volume of the membrane, for example a small change in the van der Waals radius of the surface atoms, which caused a 2 Å increase in the sphere radius, caused a shift in permittivity values of ~ 1.5 . The uncertainty in the figure is derived from fluctuations in the electric field inside the membrane, which was determined by computing the time average of the dipole moment change following equation (9.3).

The dipole moment and volume of the sphere were calculated based on the trajectory files from the simulation, which were stored every 0.1 picoseconds. An example of these two values as a function of time is shown in figure 9.5. The volume of the sphere in particular was determined by the distance between the geometric center of the sphere and the atomic center of all surface atoms plus the average van der Waals radius of the surface atoms.

It is also relevant to look at the dependence of the relative permittivity as a function of the applied external field. For this I ran 1 nanosecond long simulations under the influence of electric fields produced by voltage differences of 0.75, 1.0, 1.5, 2.0 and 5.0 V. The uncertainty grows large for lower voltages, I therefore ran additional 2 nanosecond long simulations at applied external voltage differences of 0.5, 0.4, and 0.3 V. A summary of the results is shown in figure 9.6. These suggested that either there is no dependence be-

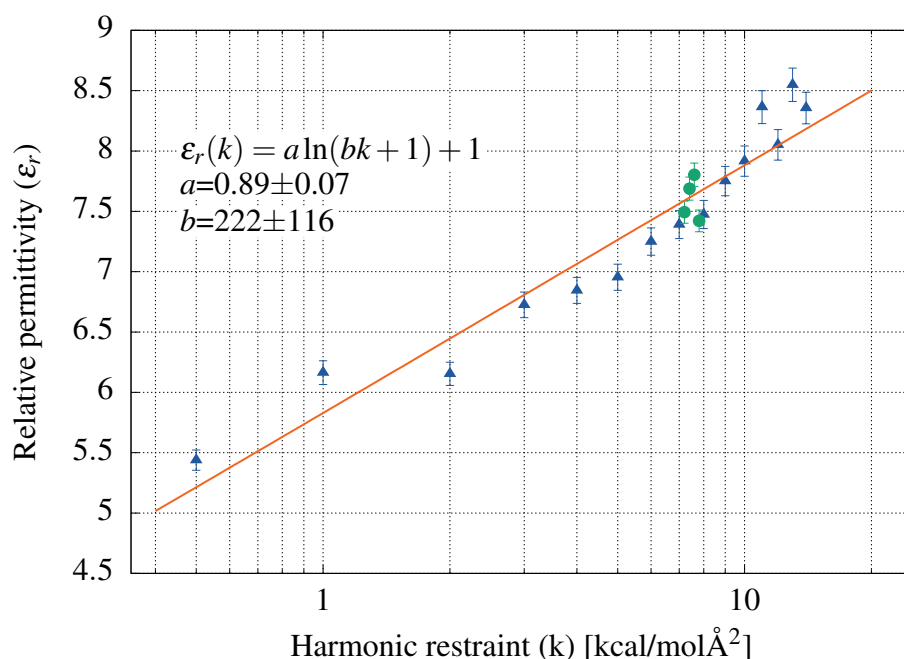


Figure 9.4. Relative permittivity as a function of applied harmonic restraint for a spherical silicon nitride membrane of radius $r = 36.5$ Å. The blue triangles and green circles represent ϵ_r results from 1 and 1.6 nanosecond-long simulations respectively. The solid red line represents a weighted logarithmic best fit of the data.

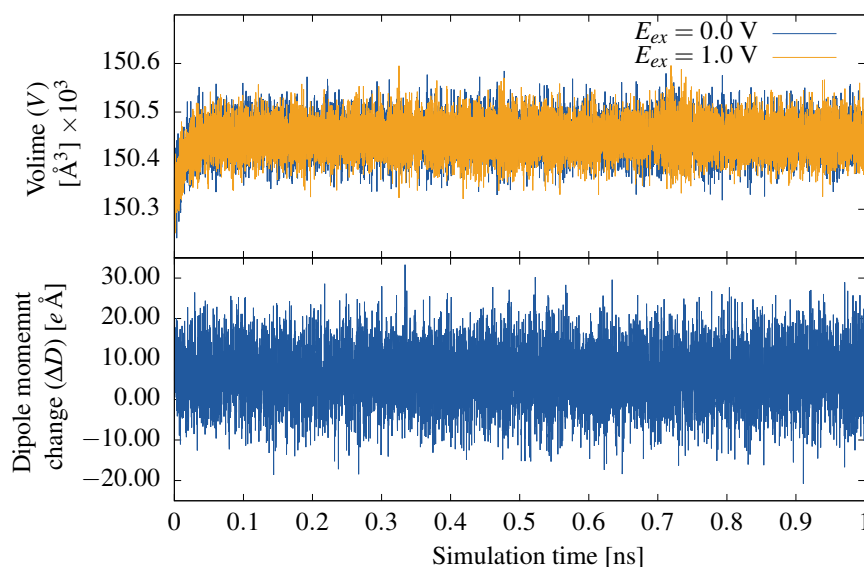


Figure 9.5. Computed dipole moment change and volume for a silicon nitride sphere captured every 0.1 picoseconds for a total of 1 nanosecond, when a 1 V voltage drop is applied and when system is simulated in vacuum. The sphere experience harmonic restraint of $7.0 \text{ kcal}/(\text{molÅ}^2)$.

tween the external field and ϵ_r in the simulated environment or that the dependence is too small to be captured from the simulation results. For both cases the relative permittivity stays within uncertainty near the experimental value. Due to the large uncertainty, the results also show that it not advisable to execute simulation with lower than 0.3 V voltage difference unless even longer calibration simulations are executed.

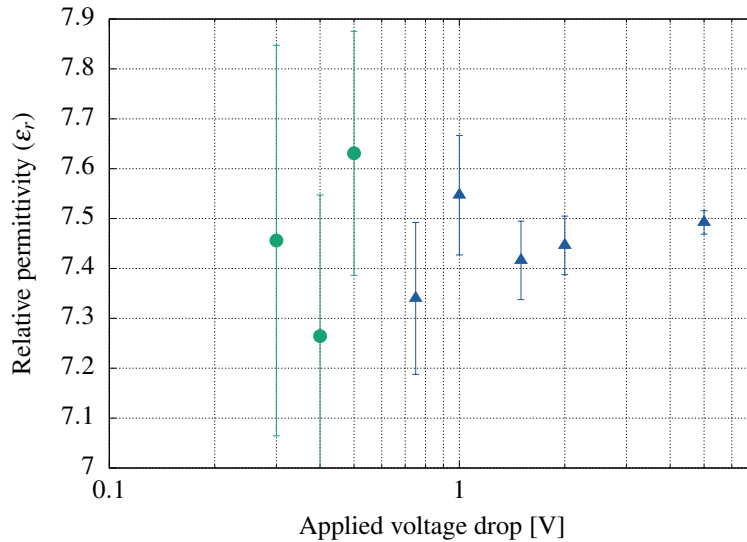


Figure 9.6. Sphere relative permittivity as a function of external voltage drop. Blue triangles show results from 1 nanosecond long simulations and green circles from the 2 nanosecond long simulation. Errors in the permittivity values are derived from fluctuations in the computed polarization.

Figure 9.7 shows a render of the sphere used to carry out the MM simulations. The sphere was formed from a rectangular block of repeating $\beta - \text{Si}_3\text{N}_4$ unit cells. To make the sphere of a volume that matched the one of the hexagonal membrane that is later used in nanopore simulations, I removed atoms outside a radius of $r = 36.5 \text{ \AA}$. The resulting structure had an irregular surface that exposed regions with excess positive or negative charges and did not preserve the 3:4 ratio of silicon to nitrogen atoms. The system charge was brought back to zero by slightly shifting the charges of all atoms to obtain a neutral sphere.

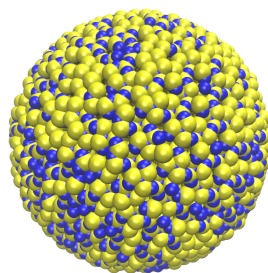


Figure 9.7. Van der Waals render of silicon nitride sphere of approximate radius of 36.5 \AA . Silicon atoms are depicted in yellow and nitrogen atom in blue.

Derivation of the electric field inside the sphere considers the silicon nitride sphere to be a uniform dielectric material. This approximation can be brought to question, since the sphere is made of discrete charged particles and the surface has regions with excess silicon or nitrogen atoms. To know more about the charge distribution of the sphere, I evaluated the electrostatic potential using VMD's PMEpot plugin. The plugin approximates each charged particle as a spherical Gaussian with a width determined by a parameter called the Ewald factor [64]. A single slice of the electrostatic potential when no external field

is applied is shown in figure 9.8. Electrostatics were evaluated from trajectory file every 10 picoseconds from a 1 nanosecond long simulation.

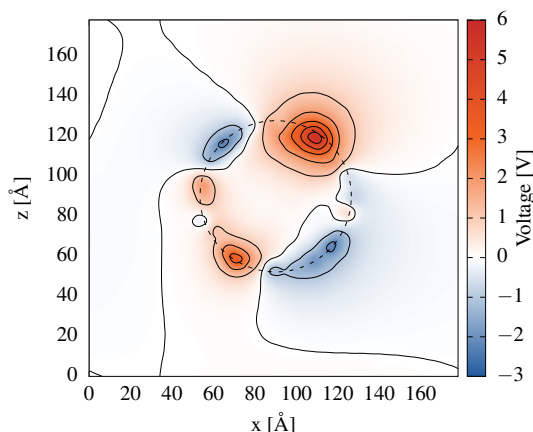


Figure 9.8. Single x - z electrostatic potential slice at $y = 90$ Å for silicon nitride sphere. Values were computed using 1 Å resolution with an Ewald factor of 0.25, every 10 picoseconds and averaged over the entire 1 nanosecond simulation time. The separation between each contour line corresponds to a 1.0 V difference. The dashed line represents the boundaries of the sphere determined by the centre of surface atoms.

Silicon nitride pore surface

Pore surface atoms are the closest membrane atoms that interact with water, protein and ions and thus play a role in the duration and magnitude of ionic current signatures obtained in MM simulations. Because the membrane was formed by repeating triclinic unit cells and the pore was carved-out by removing a cylindrical-shaped group of atoms with cutoff radius r_c , the pore surface was left with unequal amounts of nitrogen and silicon exposed. The ratio of surface silicon and nitrogen atoms fluctuates entirely based on the value of r_c , which was undesirable since a slightly larger or smaller radius could give different results and became a motivation to perturb their crystalline arrangement to obtain a more uniform surface.

For this task, I minimized the nanopore in vacuum for 0.01 nanoseconds, equilibrated for 0.15 nanoseconds and further minimized the system for 0.01 nanoseconds. The equilibration ran in NVT conditions at a temperature of 298 K using a Lowe-Andersen thermostat. In all simulations the membrane atoms were harmonically restrained except for those near the surface that were enclosed within a radius of 25 Å from the centre. In figure 9.11 I show the atoms that were selected to be restrained. I used the modified MMSX FF described in section 4 to define bonded and non-bonded interaction in the membrane.

From the simulation trajectories, recorded every 0.1 picoseconds, I monitored the ratio of silicon to nitrogen atoms at the pore surface. The results shown in figure 9.9 and also in figure 9.11 brought silicon atoms closer to regions with excess nitrogen and vice versa, thus reducing the effects of choosing a particular pore radius. The number of surface atoms as a function of the simulation time is also considered in figure 9.10, which suggest that the pore slightly increased in diameter. I additionally explored annealing the surface by slowly increasing the temperature of the membrane near or above melting temperature and rapidly cooling the membrane. The annealing procedure did not make a difference in the disorder of surface atoms, when compared to the procedure discussed above. For large temperature to make a difference, a FF that allows for the formation and breaking of bonds is needed.

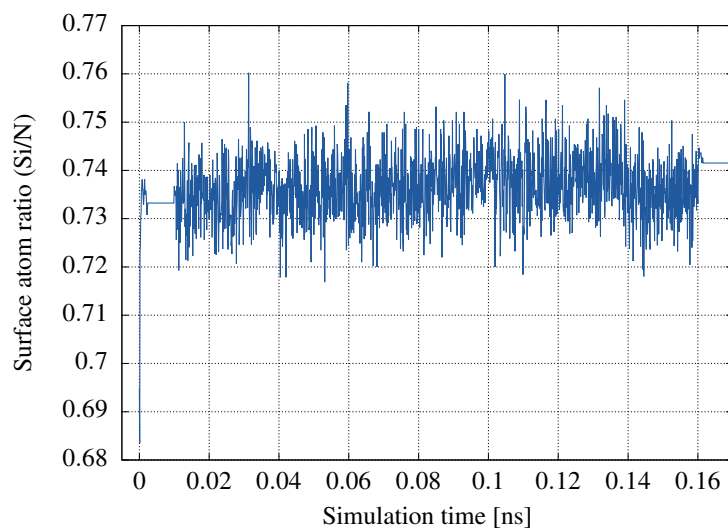


Figure 9.9.

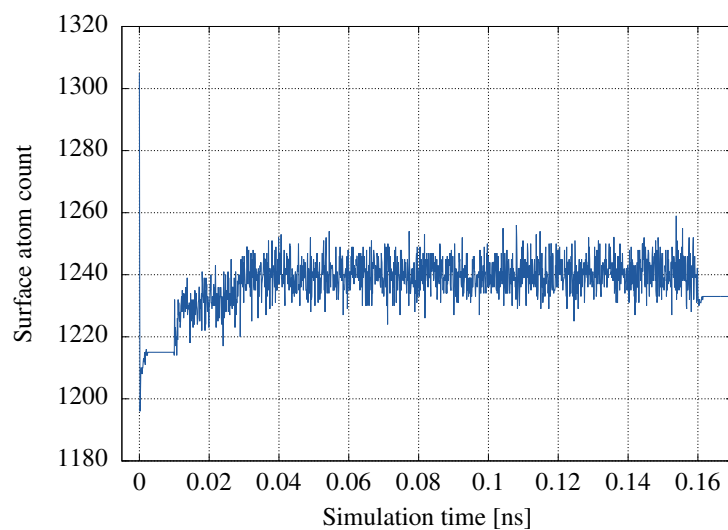


Figure 9.10.

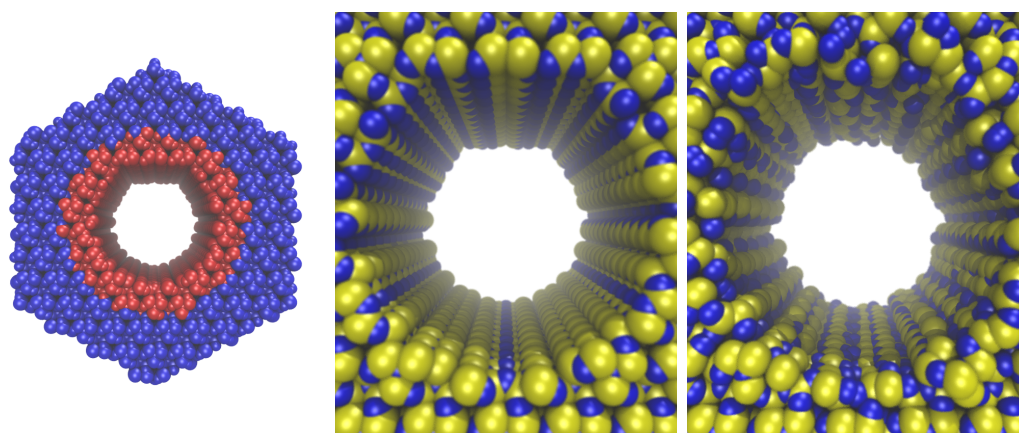


Figure 9.11. Renders of silicon nitride membrane pore surface. (left) Selection of silicon nitride atoms to apply a harmonic restraint, with blue atoms being selected and red atoms left unrestrained. (middle) Pore surface after removing atoms to create the pore. (right) Pore surface after minimization, equilibration at 298 K for 0.15 nanoseconds and minimized again.

Ion conductivity

To validate the open and occupied pore ionic current results obtained from MM simulations, it is necessary to accurately model the motion of potassium and chloride ions in solution. One technique is to study the ion conductivity, which is a measure of the solution's ability to conduct electricity when subject to an external field. If a potential difference is applied across an electrolyte solution by means of two electrodes, a current I is generated by the movement of ions with a magnitude described by Ohm's law

$$V = IR . \quad (9.6)$$

The resistance R is determined by a number of factors such surface characteristics of the electrodes, transport properties of the solvent, the types of ions and their concentration [92]. Its reciprocal, conductance S , can be decoupled into two contributions

$$S = \kappa \phi , \quad (9.7)$$

where the ion conductivity κ holds intrinsic property of the ions and the cell constant ϕ , defined as the ratio between the surface area of the electrodes and their separation distance, holds system-specific information. The conductivity still depends on the ion concentration C , but can be written in a more abstract way by introducing the molar conductivity Λ

$$\kappa = C\Lambda . \quad (9.8)$$

The molar conductivity of a single electrolyte solution can be computed by

$$\Lambda = \nu_+ \lambda_+ + \nu_- \lambda_- , \quad (9.9)$$

where ν_{\pm} represent stoichiometric coefficients and λ_{\pm} are ionic molar conductivities [92].

The molar conductivity Λ and by extension the ion conductivity κ are proportional to the mobility of the ion which in turn is inversely proportional to the viscosity of the solution. As a consequence, limitations in the water model will impact the mobility of ions. In fact it is well known that water diffusivity in the TIP3P model is significantly larger compared to experimental measurements and therefore not surprisingly ion conductivity is overestimated in MM simulations when no additional constraints are applied [75, 92]. Other water models i.e. TIP4P, SPC and SWM4 have been shown to yield better diffusivity values among other properties, nevertheless the CHARMM36 FF was parametrized with TIP3P and the use of a different model can lead to undesired effects or bias in the dynamics of the proteins under investigation [41].

To improve the description of κ using the TIP3P model, I made use of the artificial viscosity introduced by the Lowe-Andersen thermostat η^T . A similar approach has already been introduced by Aksimentiev et. al. using a different thermostat [64]. Two parameters in equation (3.20) can be controlled to modify the solution viscosity; the rate of particle collision with the heat bath Γ and the interaction radius of the thermostat R_T [51]. It is most desirable if the rate of particle collisions is maximized to ensure the target temperature stays constant, thus leaving R_T to be freely adjusted [51].

I determined the theoretical conductivity of $\kappa = 14.985$ S/m for a 1 M KCl electrolyte solution from literature values summarized in table 9.2. To explore the space, I ran simulations for a solvated cube box 4 nanometers in length and 1 M KCl concentration for a range of different Lowe-Andersen radii values. After minimization and a 0.05 nanosecond equilibration period, each simulation with unique R_T was exposed to voltage drop ranges of 0.1-5.0 V applied along the z-direction for 1 nanosecond. In all cases, the water box had periodic boundary conditions and the target temperature was set to 298 K. The equilibration simulations were carried out in NPT conditions at 1 atm where the volume was allowed to fluctuate along the z-direction. Production simulations were done in NVT conditions where the system volume was determined by the average volume from the

Table 9.2. Literature ion molar conductivity values for potassium and chloride ions at 25° C. [92].

Ion	λ [S m ² /mol]
K ⁺	7.350×10^5
C ⁻	7.635×10^5

last 0.04 nanoseconds of the equilibration. Ionic currents were calculated according to equation (4.2) over the length of the simulation box using unwrapped trajectories.

Conductivity results from simulations where the LA radius remained constant, but the system was subject to different voltage drop magnitudes is shown in figure 9.12. The ionic conductivity κ was computed from the slope of the weighted best fit.

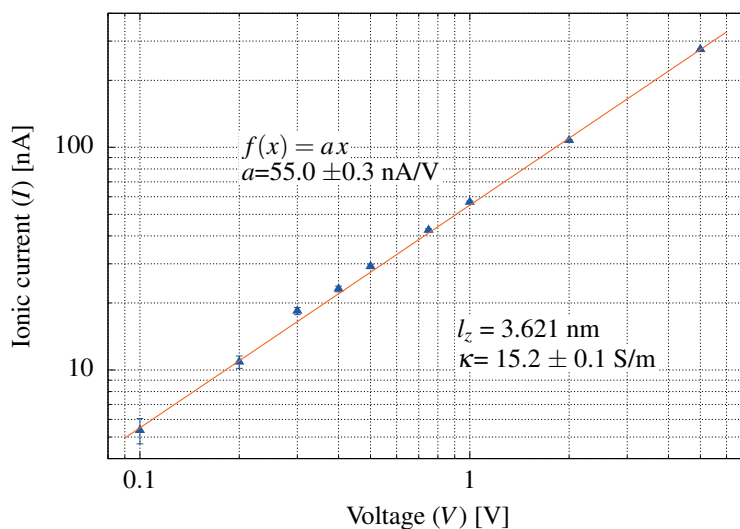


Figure 9.12. Ion conductivity as a function of applied voltage for a 1 nanosecond simulation with a 1 M ion concentration and LA radius $R_T = 3.0$ Å. The uncertainty was determined from fluctuations in the current the over time. The best fit is shown by a solid red line.

The process, repeated for different LA radius values, yielded the results shown in figure 9.13. As expected the ion conductivity decreased with increasing R_T , since the viscosity of the solvent was systematically increased by the thermostat. I fitted the trend to a power function, which showed the target ion conductivity was reached by defining a LA radius of $R_T \sim 3.0$ Å. Finally I plot the ionic current ratio in figure 9.14 and compare it with the ration obtained using literature ionic molar conductivities for low ion concentrations. The results showed good agreement between simulation and experimental values.

In all simulations I employed a static TIP3P water model using NAMD's SETTLE algorithm, in accordance with the parametrization of the CHARMM36 FF. If water molecules were allowed to vibrate, elastic collision between water and ions would further hinder the mobility of ions lowering their conductivity. Such configuration would add a level of accuracy to the simulations, but at the cost of increased computational power. Ultimately the choice of restricting the water vibrations and use of the LA thermostat instead of other common thermostats (i.e. Langevin thermostat) were justified by the ability of the system to conserve momentum and yield reasonable ion conductivity values without a very large increase in R_T values. Any residual systematic bias in the conductivity with respect to experimental values were later removed from calculations, since I was interested in the ratio between open and occupied pore currents. It is worth noting that the viscosity of the water to an extend also influences the dynamics of the protein.

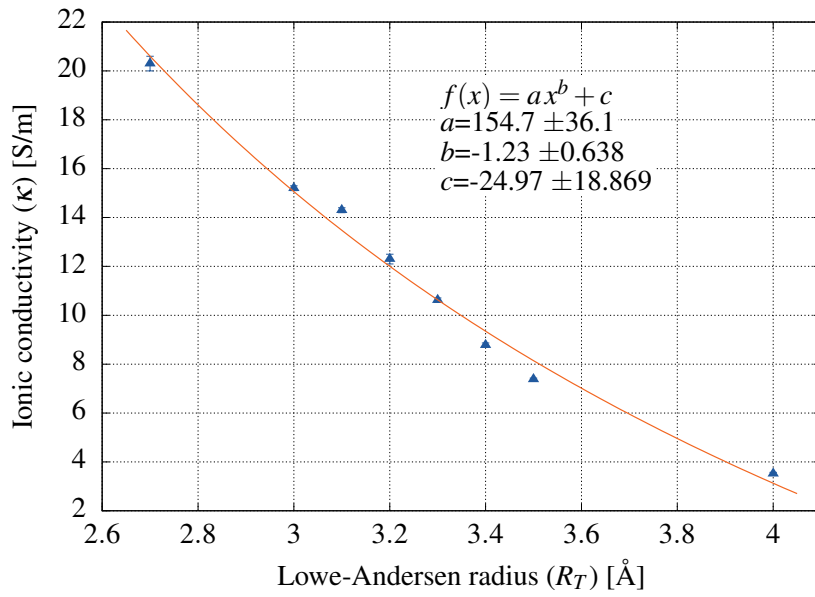


Figure 9.13. Ion conductivities as a function of the Lowe-Andersen radius. Conductivity values are obtained from a weighted best fit of the ionic current as a function of applied voltage from 1 nanosecond simulations.

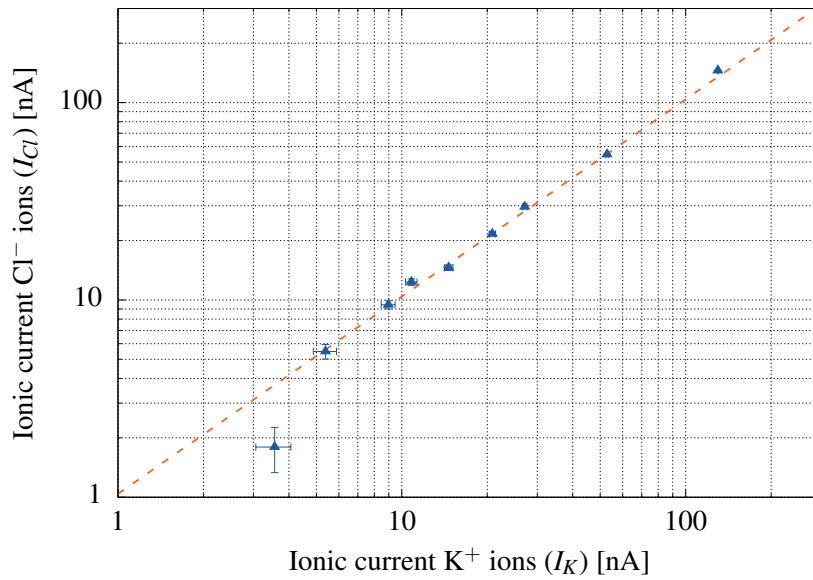


Figure 9.14. Cl^- and K^+ current ratio from a 1 nanosecond simulations using $R_T = 3.0$ Å. Errors come from fluctuations in the ionic current over time and the dashed red line shows the ratio of literature values of Λ .

Appendix C.

Details on electronic structure calculations

Ground state energy converge

I am interested in the simulation cell dimensions, energy mesh cutoff and k-points that yield a well-converged ground-state energy for the silicon nitride slab, the amino acid and both objects combined. In all simulations I employed a double- ζ polarized basis set with the van der Waals functional DRSLL. Convergence was achieved with the Pulay convergence accelerator with a 0.02 mixing factor (0.1 for simulations of the amino acid alone) and a history of 5 self-consistent cycles.

For the silicon nitride slab, made of 280 atoms, I first ran simulations varying the separation distance along the z-direction between periodic images. The results are shown in figure 9.15. Later I systematically increase the energy cutoff to reduce contributions from the *egg-box* effect on the final energy. Results are shown in figure 9.16. I did not study the ground-state energy of the system with increasing k-points, because the simulation box size is large enough such that in reciprocal space is very small and computations can be done along a single k-point (usually the γ -point). Based on the results, I find parameter values for separation distance of ~ 13.0 Å and energy mesh cutoff of 300 Ry are best suited.

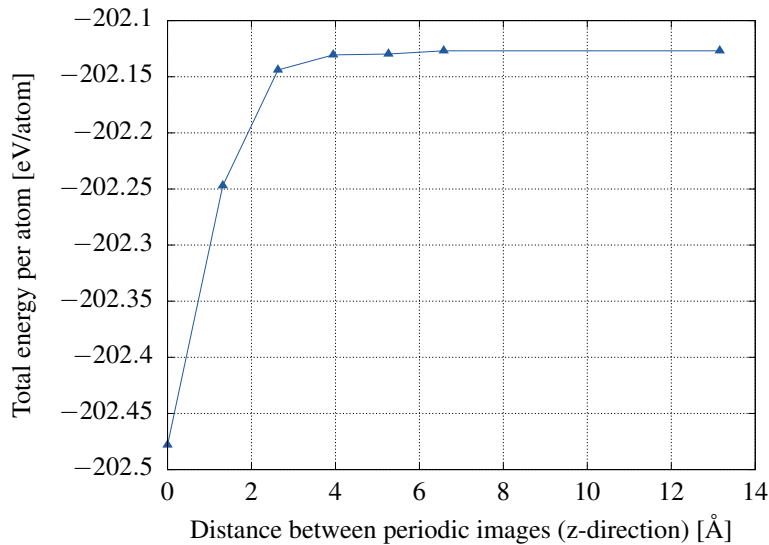


Figure 9.15. Ground state energy of the silicon nitride slab as a function of z-direction separation between layers on silicon nitride slabs that are periodic along the x-y direction.

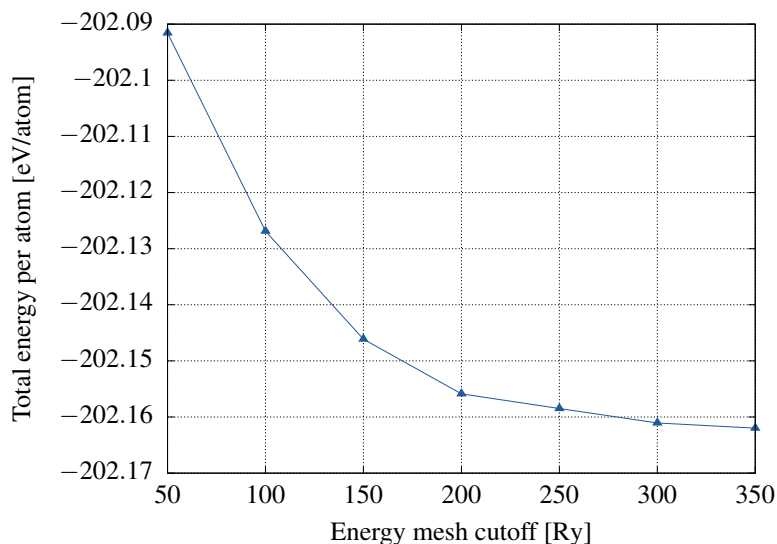


Figure 9.16. Ground state energy of the silicon nitride slab as a function of energy cutoff.

In a similar manner as above, I studied the ground state energy converge for a single amino acid glutamine made of 20 atoms as a function of the simulation box length and cutoff energy. The results are shown in figures 9.17 and 9.18 respectively. In this case, the simulation box might be small enough such that the number of k-points could have an influence in the total energy of the system. Results of the k-point convergence are shown in figure 9.19. Based on the results and taking into consideration the simulation wall time, I find parameter values for separation distance of ~ 16 Å, energy mesh cutoff of 300 Ry and k-points of 1 are best suited.

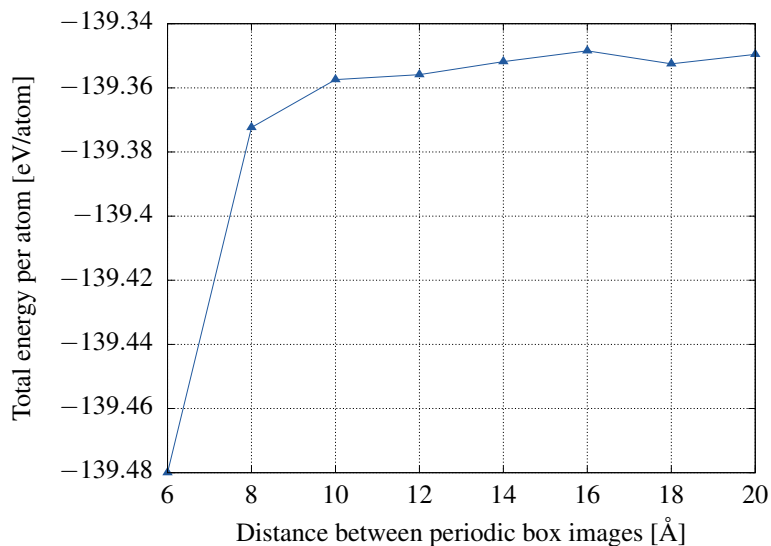


Figure 9.17. Ground state energy of amino acid glutamine as a function of simulation box length.

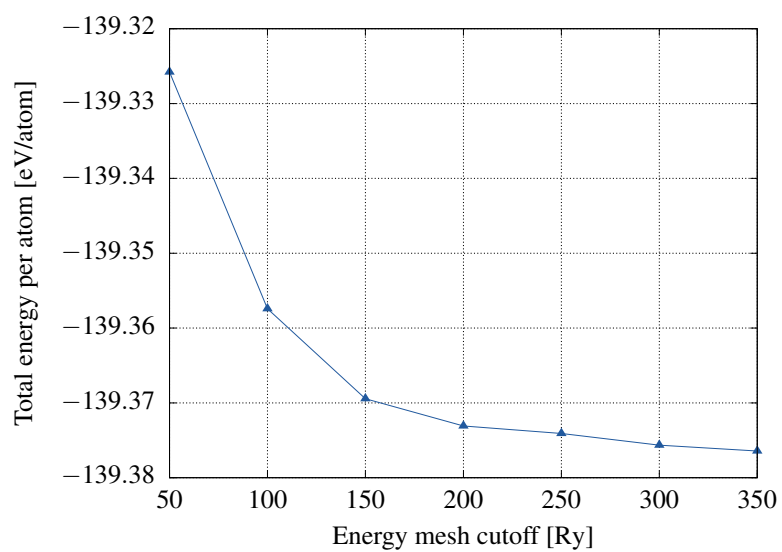


Figure 9.18. Ground state energy of amino acid glutamine as a function of energy cutoff.

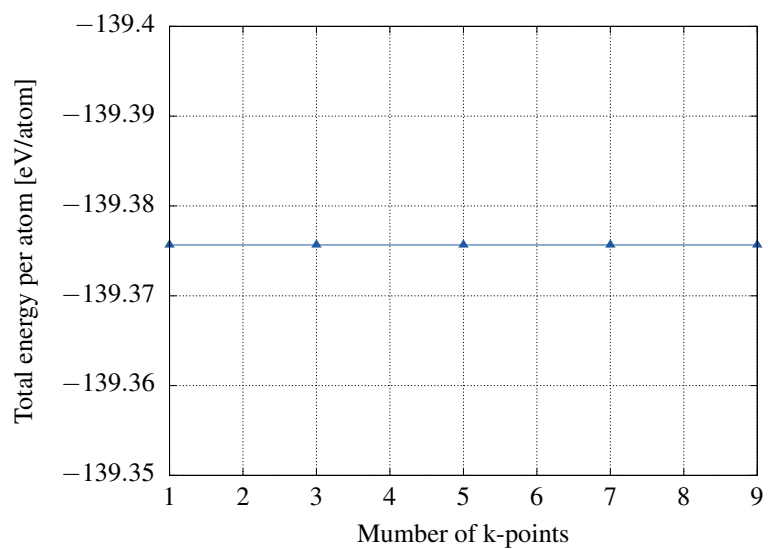


Figure 9.19. Ground state energy of amino acid glutamine as a function of k-points.

Finally, I studied the ground state converge of the silicon nitride slab and amino acid glutamine combined into a single system (300 atoms in total). Energy results for increasing z-direction separation distance between periodic images of the membrane and mesh cutoff energy are shown in figures 9.20 and 9.21 respectively. Based on the results and taking into consideration the simulation wall time, I find parameter values for separation distance of ~ 26 Å and energy mesh cutoff of 300 Ry are best suited.

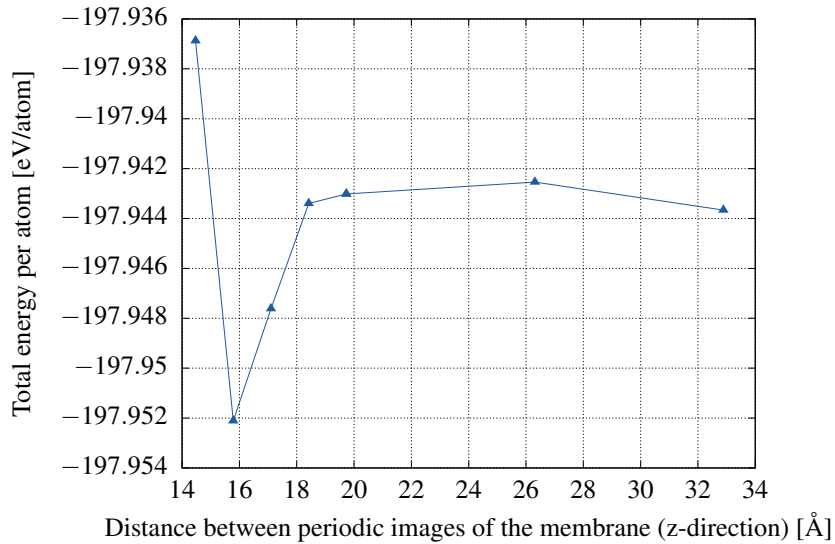


Figure 9.20. Ground state energy of silicon nitride slab with amino acid glutamine as a function of simulation box length.

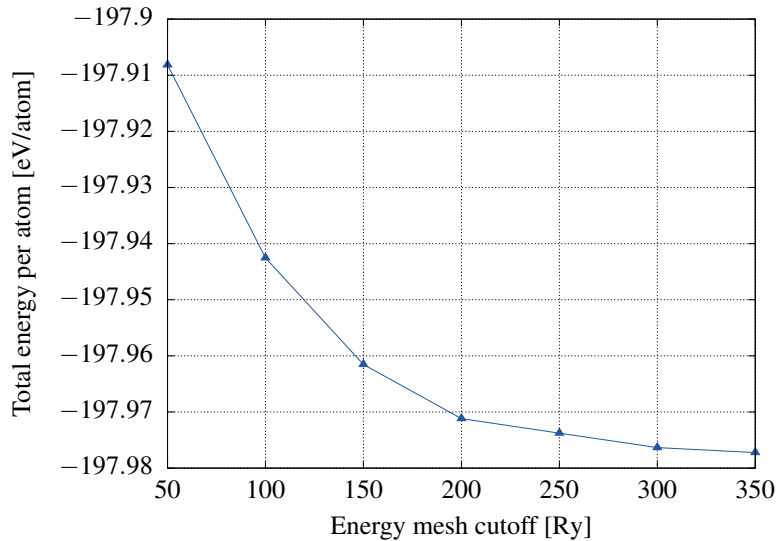


Figure 9.21. Ground state energy of silicon nitride slab with amino acid glutamine as a function of energy cutoff.

Appendix D.

Additional results

Occupied pore ionic currents

Ion counts for occupied pore simulations subject to a 1.0 V voltage bias are shown in figures 9.22 and 9.23. Based on these results, the number of ions inside the pore reached a steady state after 8 nanoseconds for both proteins.

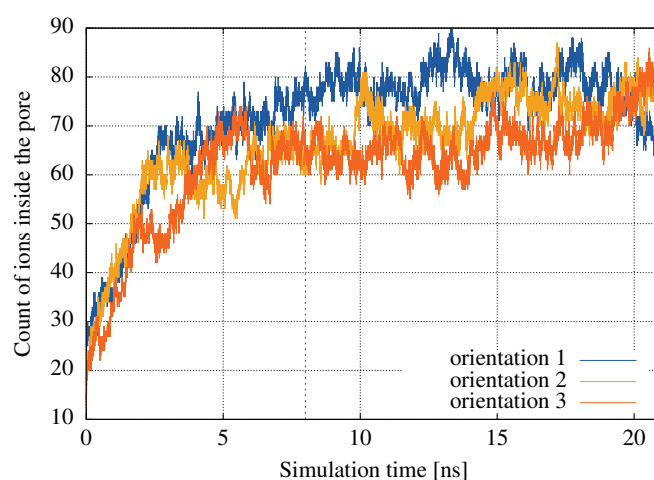


Figure 9.22. Ion count as a function of simulation time for pore with protein FBP11-WW under a 1.0 V voltage bias. Ion count is shown for all three orientations of the protein. The system is first minimized, equilibrated and after 1 nanosecond the field is turned on and the protein is harmonically fixed. Ion positions were stored every 1 picosecond. The dashed purple line marks when experimental conditions were reached.

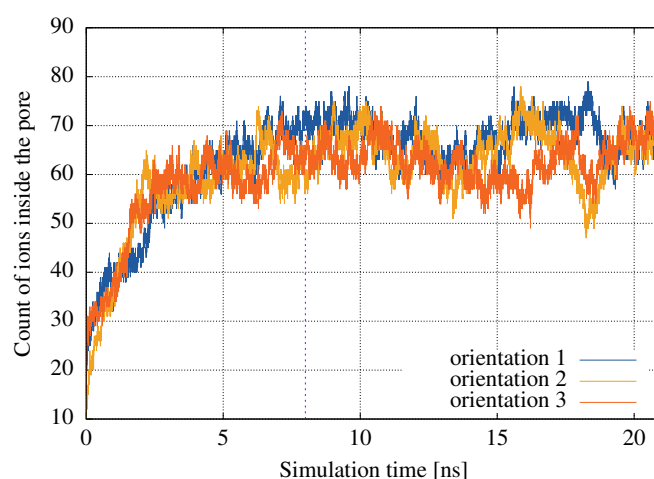


Figure 9.23. Ion count as a function of simulation time for pore with protein AGRP(87-120) under a 1.0 V voltage bias. Ion count is shown for all three orientations of the protein. The system is first minimized, equilibrated and after 1 nanosecond the field is turned on and the protein is harmonically fixed. Ion positions were stored every 1 picosecond. The dashed purple line marks when experimental conditions were reached.

I used the remaining 13 nanoseconds of the simulation to compute the ion current as a function of time based on the trajectory of the ions and I applied a Bessel LPF with a cutoff frequency of 5 GHz. The results are shown in figures 9.24 and 9.25.

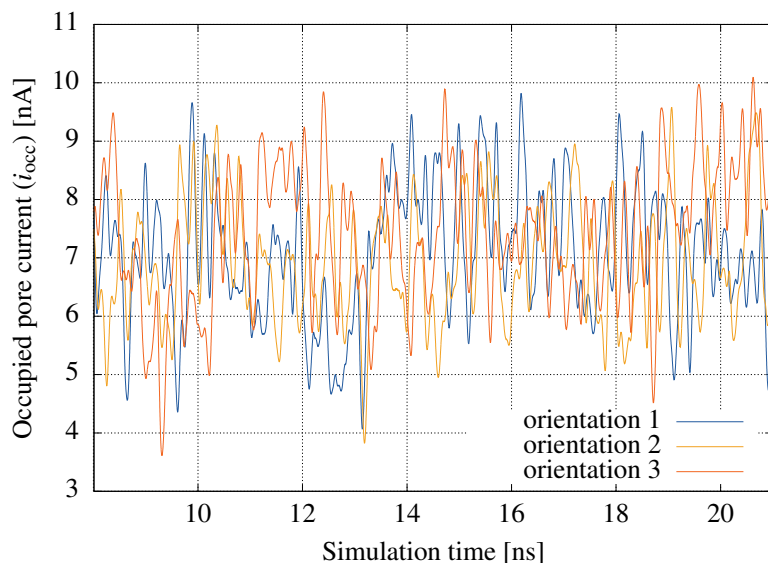


Figure 9.24. LPFd occupied pore current computed for protein FBP11 WW domain as a function of simulation time for all three orientations, while the system is subject to a 1.0 V voltage bias. The protein is harmonically restrained.

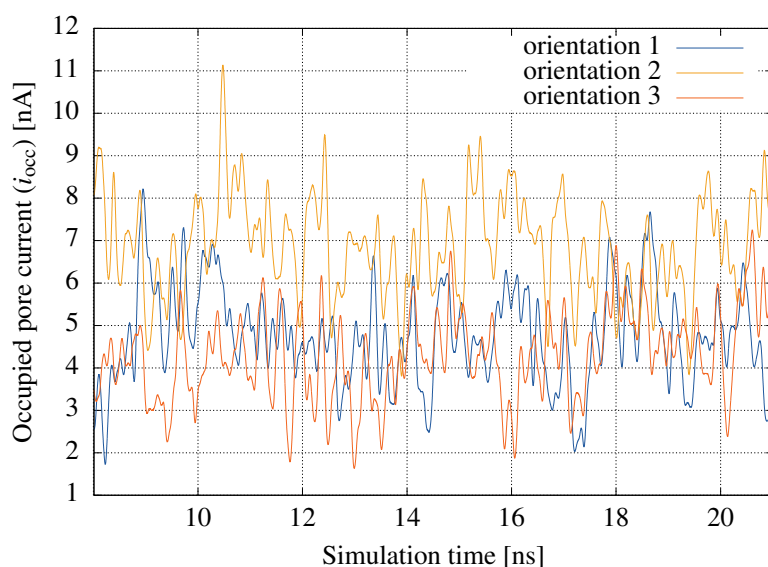


Figure 9.25. LPFd occupied pore ionic current computed for protein AGRP(87-120) domain as a function of simulation time for all three orientations, while the system is subject to a 1.0 V voltage bias. The protein is harmonically restrained.

I followed the same treatment for occupied pore currents results from simulations where the harmonic restrain on protein's backbone is removed. The LPFd occupied pore current results when the system is subject to a 1.0 V voltage bias are shown in figure 9.26 and 9.27.

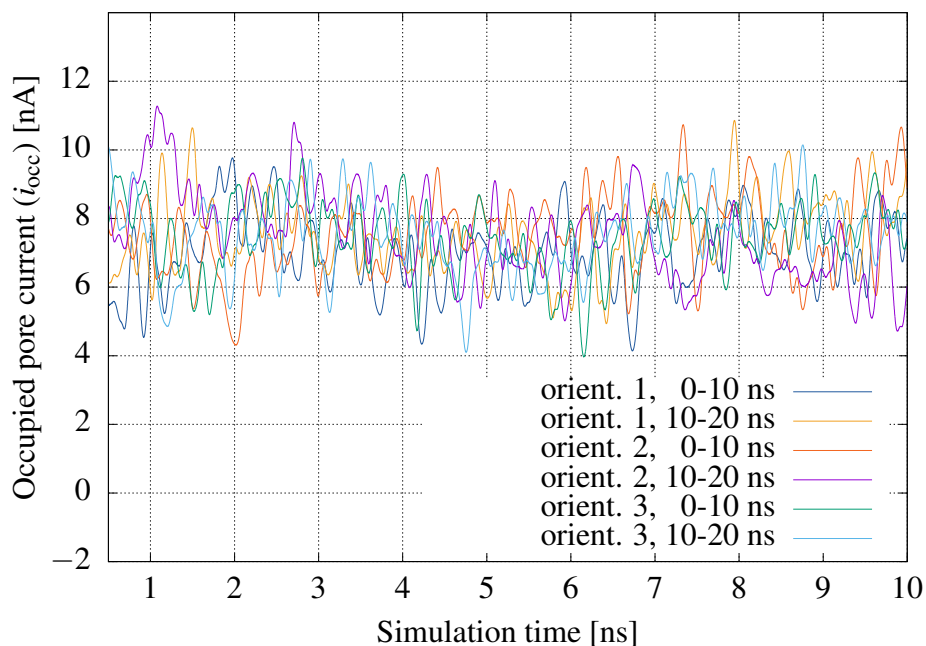


Figure 9.26. LPFd occupied pore current computed for protein FBP11 WW domain as a function of simulation time for all three orientations, while the system is subject to a 1.0 V voltage bias.

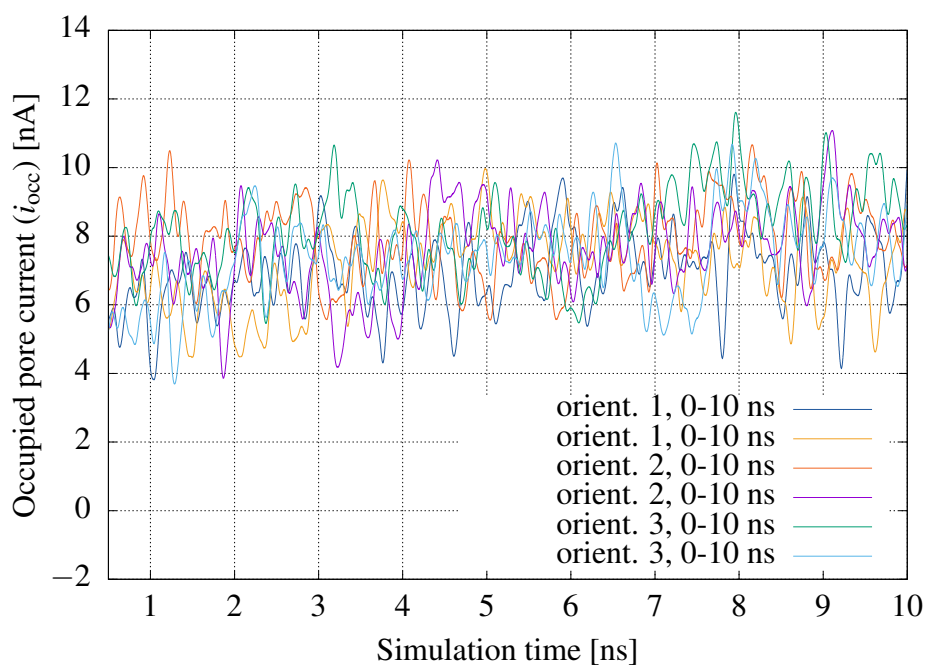


Figure 9.27. LPFd occupied pore ionic current computed for protein AGRP(87-120) domain as a function of simulation time for all three orientations, while the system is subject to a 1.0 V voltage bias.