UNIVERSITY OF CENTRAL OKLAHOMA

MASTER'S THESIS

Molecular Dynamics Simulation of DNA Translocation Through Solid-State Nanopores

Author:
Micah Watson

Supervisor:
Dr. Benjamin TAYO

A thesis submitted in fulfillment of the requirements for the degree of MASTER OF SCIENCE

in the

School of Engineering

December 6, 2024

Molecular Dynamics Simulation of DNA Translocation Through Solid-State Nanopores

By Micah Watson

11/26/2024

Jackson College of Graduate Studies at the University of Central Oklahoma

A THESIS APPROVED FOR

Micah Watson

By

Committee Chairperson

Committee Member

Committee Member

ABSTRACT

This thesis explores the use of molecular dynamics (MD) simulations to model double-stranded DNA (dsDNA) translocation through nanopores, aiming to optimize nanopore-based DNA sequencing technologies. Using both LAMMPS and NAMD, the study investigates the potential for MD simulations to provide insights into nanopore sequencing mechanisms and system behaviors. Insights gained from LAMMPS tutorials by Simon Gravelle informed the understanding of force fields, system minimization, heating, and equilibrating processes necessary for accurate MD simulations. The atom pull method presented in these tutorials was foundational in understanding force application within molecular systems, analogous to grid-steered molecular dynamics (G-SMD) techniques later applied to DNA sequencing.

Following these foundational steps, NAMD simulations based on protocols by Jeffrey R. Comer were performed to simulate dsDNA translocation through two types of nanopores: an alpha-hemolysin biological nanopore and a Silicon Nitride (Si_3N_4) solid-state nanopore. The alpha-hemolysin simulation allowed for an initial examination of DNA behavior in a biological nanopore, while the Si_3N_4 nanopore simulation provided detailed ionic current signatures critical for sequencing analysis. Protocols from Comer's nanopore modeling guide were closely followed to accurately construct these systems, and no structural modifications were made. Results indicate that nanopore geometry, such as the hourglass shape in Si_3N_4, provides a stable pathway for DNA translocation and yields consistent current signals. The simulations utilized a higher-than-normal voltage to accelerate translocation, which provided faster insights but deviated from experimental conditions, suggesting areas for methodological refinement. The high than normal voltage is due to this work following steps described 15 years ago before GPU-computing which led to simplifications in the system. These findings were compared with both another Si_3N_4 simulation and real-world experimental data using Si_3N_4.

The findings suggest that MD simulations can offer valuable insights into nanopore design and sequencing efficiency, providing a foundation for further optimization of real-world DNA sequencing applications. Future work could focus on refining computational methods to simulate translocation under more realistic voltages, ultimately bridging the gap between simulation and experimental results. This study contributes to the evolving field of nanopore sequencing, suggesting avenues for improvements in nanopore structure and operational conditions to enhance sequencing accuracy and reliability.

ACKNOWLEDGMENTS:

I would like to express my deepest gratitude to Dr. Benjamin Tayo for the opportunity to be a part of his research team throughout the course of this project. His unwavering support and invaluable guidance have been vital to the success of our research. Dr. Tayo has not only played a critical role in the success of this project but has also greatly contributed to my academic, professional, and personal development by sharing his expertise and offering thoughtful mentorship.

I am also immensely grateful to Dr. Jiang for his academic guidance and unwavering support throughout the master's program. His assistance with navigating academic schedules, addressing class-related queries, and providing valuable life advice has been instrumental in shaping my academic and career trajectory.

Additionally, I would like to extend my sincere thanks to Dr. Xu for his insightful contributions to my project. His perspectives have added depth to my understanding and have significantly enhanced the quality of my work.

Most importantly, I would like to express my heartfelt gratitude to my fiancée, Sophia, for her unwavering support and patience throughout the challenges of crafting and revising this document. Her encouragement and belief in me have been a constant source of strength.

Finally, I am profoundly thankful to my parents—Nikki, Jimmy, Cory, and Vickey—as well as the rest of my family, friends, and colleagues for their encouragement and support. Their love and kindness have been a source of strength and inspiration throughout this journey.

Sincerely,

Micah Watson

Contents

LIST OF FIGURES AND TABLES

CHAPTER 1: INTRODUCTION

1.1: BACKGROUND AND MOTIVATION:

Current research has found several ways to determine the sequence of Deoxyribonucleic acid (DNA) [2]. The current goal of DNA research is to make the detection quicker, cheaper, and more accurate. Sequencing DNA offers numerous benefits across various fields, from medicine to agriculture and beyond [3]. Some of the benefits include the ability to identify genetic disorders, predictive genetic testing, crop improvement, livestock breeding and more. In the past two decades research has advanced to sequence DNA through nanopores by measuring electric readouts, ionic currents, or tunneling current voltage characteristics [4].

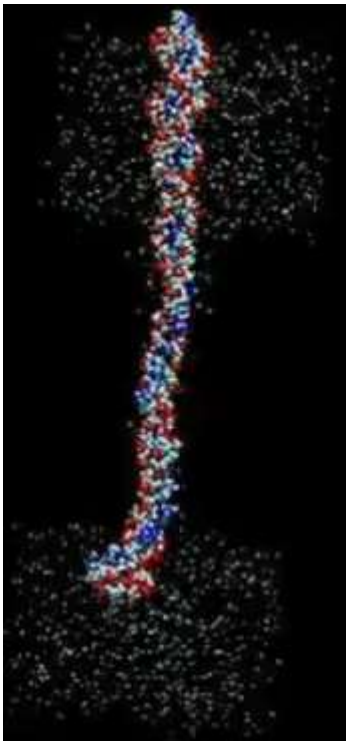

Figure 1: Simulation of DNA translocating through a Silicon Nitride (Si_3N_4) nanopore.

1.2: CHALLENGES IN DNA SEQUENCING

DNA sequencing faces several challenges that highlight the need for more accurate and efficient methods. One of the primary issues is accuracy, even advanced techniques like nanopore sequencing and next-generation sequencing (NGS) can have high error rates. Short read lengths in many sequencing platforms, like Illumina, limit the ability to resolve long repetitive sequences or structural variations, which are crucial for understanding complex diseases and certain genomic regions [5]. Additionally, sequencing methods often struggle with GC-rich or repetitive regions, which can cause biases in data and coverage gaps. Time and cost are also significant challenges, especially for large-scale projects like whole-genome sequencing, where lengthy and expensive processes limit accessibility. Another major issue is the handling of degraded or low-quality DNA samples, which are common in forensic, ancient, or clinical contexts, and often fail in traditional sequencing workflows. Finally, as the volume of sequencing data grows, data storage, processing, and interpretation become overwhelming, requiring more advanced bioinformatics tools and infrastructure.

1.3: ADVANTAGES OF NANOPORE SEQUENCING

Nanopore sequencing offers several advantages over traditional methods like Illumina and Sanger sequencing. Its ability to produce long reads (up to hundreds of kilobases) makes it ideal for assembling complex genomes and identifying structural variants. Unlike other methods, nanopore sequencing provides real-time results without the need for Polymerase chain reaction (PCR) amplification, reducing biases and errors. It's also highly portable, with devices like MinION, making it accessible for fieldwork and rapid diagnostics [6]. Additionally, nanopore technology can directly detect epigenetic modifications, sequence a variety of molecules like DNA, RNA, and proteins, and is cost-effective with lower infrastructure needs. While it has higher raw error rates, ongoing improvements in base-calling are enhancing its accuracy, further solidifying its place as a flexible and scalable sequencing method.

1.4: MOLECULAR DYNAMICS SIMULATIONS TO OPTIMIZE NANOPORE-BASED SEQUENCING

Regarding nanopore sequencing researchers have ran into occasional errors when identifying DNA bases [3]. To determine the best parameters to use to obtain the most identifiable DNA researchers have turned to molecular dynamics (MD) software to simulate DNA translocating through a nanopore. MD simulations can improve the DNA sequencing technology by providing information about molecular-level interactions, enhancing the accuracy of detection, and reducing experimental costs by allowing researchers to test their system before putting it into action and accelerating the development of more efficient sequencing methods. There are many parameters that one can adjust when working with MD. The parameters include: nanopore material, nanopore size, nanopore opening, potential current applied, force parameters, and molarity of system [7].

1.5: THESIS OBJECTIVE:

This thesis uses MD to provide insight to what parameters would provide the most recognizable result with DNA sequencing. The history of the first, second and third generation of DNA sequencing will be explained. Then nanopore sequencing will be described and its advantages over traditional methods of DNA sequencing. Afterwards the molecular dynamics simulation software will be explored and the equations and parameters needed to run the software to simulate DNA translocating through a nanopore. The software used in this research is Nanoscale Molecular Dynamics (NAMD 3) on the OU Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma (OU) and Visual Molecular Dynamics (VMD) on the University of Central Oklahoma’s supercomputer, "Buddy"

Molecular dynamics (MD) simulations of DNA translocating through nanopores are essential for understanding the mechanisms of nanopore sequencing and optimizing experimental conditions. However, current MD simulations often fail to accurately replicate experimental results due to the complexity of DNA-nanopore interactions, limitations in force field accuracy, and the computational intensity of such simulations. There is a critical need to optimize the simulation parameters, including force fields, time steps, and environmental conditions, to improve the reliability and predictive power of these simulations. Using VMD and NAMD 3, advanced software for molecular visualization and dynamics, presents an opportunity to fine-tune these parameters and better align simulated outcomes with real-world experiments. Addressing these challenges will not only enhance the understanding of DNA translocation through nanopores but also provide more accurate guidance for experimental designs, improving the efficiency and effectiveness of nanopore-based sequencing technologies.

1.6: STRUCTURE OF THE DOCUMENT

Chapter 1 introduces the background, motivation, and objectives of the study.

Chapter 2 reviews the literature on DNA sequencing and nanopore technologies.

Chapter 3 outlines the theoretical foundation of molecular dynamics simulations.

Chapter 4 describes the methods used in setting up and running the simulations.

Chapter 5 presents the results of the MD simulations.

Chapter 6 discusses the implications and conclusions of the research.

CHAPTER 2: LITERATURE REVIEW

2.1: DNA REVIEW

DNA (Deoxyribonucleic Acid) is the molecule responsible for storing genetic information in almost all living organisms. It consists of two long chains of nucleotides that form a double helix. Each nucleotide is made up of three components: phosphate group, deoxyribose sugar and a nitrogenous base. The nitrogenous base can be Adenine (A), Thymine (T), Cytosine (C), Guanine (G). The two strands are held together by hydrogen bonds between complementary bases (A pairs with T, and C pairs with G). The sequence of these bases encodes genetic instructions [8].

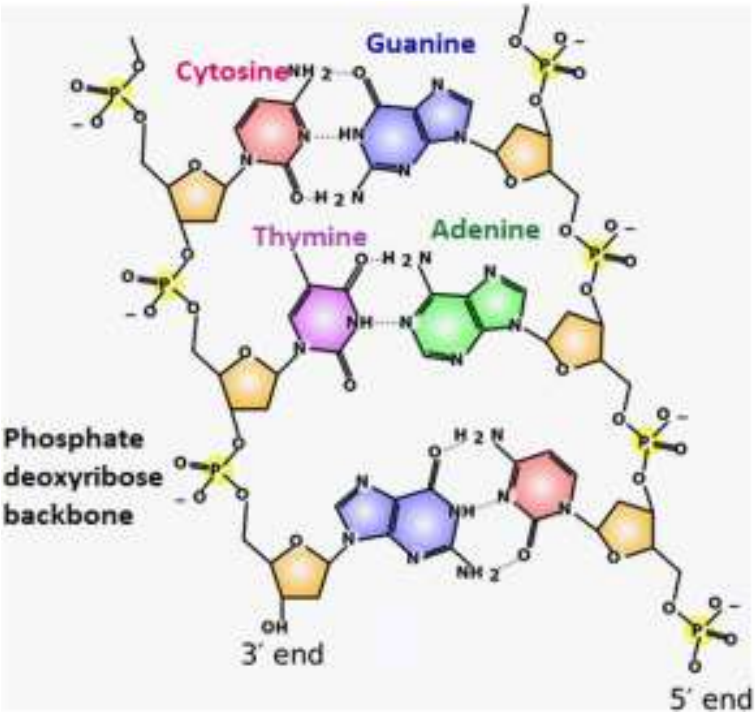

Figure 2: Structure of DNA [9].

2.2: DNA SEQUENCING HISTORY

The history of DNA sequencing can be categorized into three major generations: first-generation, second-generation (Next-Generation Sequencing or NGS), and third-generation technologies. The first-generation method, known as Sanger sequencing, was developed in 1977 by Frederick Sanger. This method uses chain-terminating nucleotides (ddNTPs) to produce DNA fragments of varying lengths, which are then separated by electrophoresis to determine the sequence [10]. Sanger sequencing was highly accurate but slow and labor-intensive, making it impractical for large-scale sequencing projects. Despite this, it was crucial for early breakthroughs, such as the first viral genome sequence (bacteriophage ϕX174) and the Human Genome Project.

In the mid-2000s, second-generation sequencing, or Next-Generation Sequencing (NGS), emerged, revolutionizing the field with massively parallel sequencing capabilities. Technologies such as Illumina sequencing allowed millions of short DNA fragments to be sequenced simultaneously, drastically reducing costs and increasing throughput [11]. NGS made it possible to undertake large-scale projects like whole-genome and exome sequencing, providing powerful insights into genetics and biology [12]. However, these methods typically produced short reads (100–300 base pairs), making it challenging to sequence repetitive or complex regions of genomes.

The development of third-generation sequencing in the 2010s addressed some of the limitations of NGS, particularly the need for long-read sequencing. Unlike previous methods, third-generation sequencing reads single molecules of DNA without the need for amplification. Technologies such as Oxford Nanopore and Pacific Biosciences (PacBio) allow the sequencing of entire DNA molecules in real-time [13]. Nanopore sequencing can produce ultra-long reads, while PacBio's Single Molecule Real-Time (SMRT) sequencing generates long, high-fidelity reads. These advancements improve genome assembly and are particularly useful for studying complex structural variants and sequencing repetitive regions. While third-generation technologies initially had higher error rates, improvements in accuracy, especially with PacBio's HiFi reads, have made them increasingly reliable. As a result, these technologies have become essential in areas ranging from clinical diagnostics to real-time sequencing in the field.

2.3: NANOPORE SEQUENCING

Nanopore sequencing allows for the direct sequencing of long DNA or RNA strands by passing them through tiny pores called nanopores. Developed by Oxford Nanopore Technologies (ONT), it stands out from earlier sequencing technologies by providing real-time sequencing capabilities and the ability to read very long stretches of DNA, sometimes up to megabases in length [14].

Nanopore sequencing works by threading a single-stranded DNA or RNA molecule through a protein nanopore embedded in a synthetic membrane. As the nucleotides pass through the nanopore, each base (A, T, C, G, or Uracil (U)) causes a characteristic disruption in the electric current flowing through the pore. These changes in current are recorded in real time and analyzed by software to determine the sequence of the nucleotides [14]. The technique allows for direct sequencing of native DNA or RNA, without the need for amplification, providing a high level of flexibility and speed.

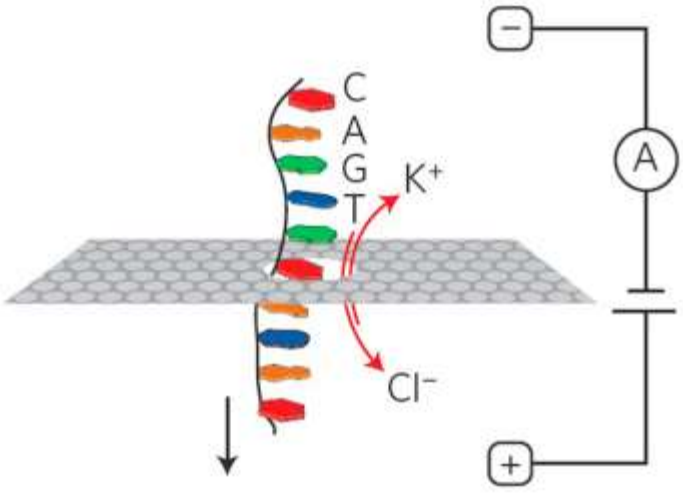

Figure 3: Schematic of ssDNA translocating through a nanopore [9].

2.4: NANOPORE SEQUENCING TECHNOLOGIES

Nanopore sequencing technologies, developed by Oxford Nanopore Technologies (ONT), offer a range of devices designed to meet different sequencing needs, from portable, small-scale experiments to large, high-throughput projects. The MinION is a highly portable, USB-powered device capable of generating up to 50 gigabases (Gb) of data in a single run, making it ideal for fieldwork, clinical diagnostics, and small research projects [15]. For mid-scale applications, ONT offers the GridION, a benchtop device that can run multiple MinION flow cells simultaneously, allowing for higher throughput in laboratory settings. At the high end, the PromethION is designed for large-scale genomic studies and can produce terabytes of data, making it suitable for sequencing large genomes or running many samples concurrently. For smaller-scale, cost-effective experiments, the Flongle adapter is available for use with MinION or GridION, providing low-cost sequencing for simpler or rapid tests. These devices make nanopore sequencing highly adaptable, providing options for everything from portable, real-time analysis to large-scale genomics projects.

2.5: OXFORD NANOPORE, BIOLOGICAL VS. SOLID-STATE NANOPORES

2.5.1: BIOLOGICAL NANOPORE

Oxford Nanopore Technologies (ONT) is a leader in biological nanopore sequencing, while solid-state nanopores are an emerging alternative technology. Biological nanopores, such as the commonly used α-hemolysin or MspA, are protein-based pores embedded in a membrane [16]. One of the major advantages of biological nanopores is their high sensitivity and ability to sequence long fragments of DNA or RNA in real-time, as well as their capability to detect epigenetic modifications without additional steps. However, biological nanopores are delicate and have limited stability, which requires careful handling and having a finite lifespan [17].

2.5.2: SOLID-STATE NANOPORE

In contrast, solid-state nanopores are synthetic, made from materials like silicon nitride or graphene, and offer greater durability and robustness. They are tunable, meaning their properties can be engineered for specific applications, and they can operate under a wider range of conditions than biological nanopores [18]. Despite

these benefits, solid-state nanopores currently suffer from lower accuracy and uniformity, making them less commercially viable at present. While solid-state nanopores are still in the research phase, biological nanopores are more advanced and widely used in real-world sequencing applications.

CHAPTER 3: THEORY (MOLECULAR DYNAMICS SIMULATIONS)

3.1: BASIC IDEA

Molecular dynamics (MD) is a computational simulation technique that models the physical movements of atoms and molecules over time, providing a detailed microscopic view of dynamic processes in materials and biological systems [19]. By numerically solving Newton's equations of motion for a collection of interacting particles, MD calculates the trajectories of atoms and molecules. Each particle in the system is assigned an initial position and velocity, and their interactions are described using force fields. Force fields are mathematical functions that represent the potential energy of the system, including both bonded interactions (like bond stretching and angle bending) and non-bonded interactions (such as van der Waals forces and electrostatic interactions). The simulation advances in small time steps, continually updating the positions and velocities of particles based on the forces acting upon them. This approach allows researchers to investigate structural, thermodynamic, and kinetic properties at the atomic scale, offering insights into phenomena like protein folding, chemical reactions, material deformation, and molecular diffusion [19].

3.2 HIERARCHIES IN MOLECULAR DYNAMICS METHODS

The relationship among these MD methods is inherently hierarchical, allowing researchers to choose an approach based on the required level of detail and computational constraints [21]. QM methods sit at the top for accuracy but are limited in scope to small systems. Semi-empirical methods balance accuracy and efficiency, making them suitable for medium-sized systems. MM methods, at the lower end of the hierarchy, offer the broadest scalability, ideal for large systems but without the ability to capture electronic details. The QM/MM hybrid approach leverages QM accuracy within an MM environment, making it ideal for studying systems where both electronic and large-scale effects are significant. Together, these methods provide a flexible framework to simulate a wide range of molecular systems and processes.

3.2.1: QUANTUM MECHANICS (QM) METHODS

Quantum mechanics (QM) methods, including ab initio techniques like Density Functional Theory (DFT) and Hartree-Fock, involve solving Schrödinger’s equation to accurately describe the electronic structures and energy profiles of molecular systems [20]. Due to their high computational demand, these methods are generally limited to small molecules or specific regions within larger systems, such as the active sites of enzymes. The precision of QM methods makes them essential for validating and parameterizing lower-level models, such as semi-empirical and molecular mechanics (MM) methods. When accurate electronic detail is necessary and computational resources allow, QM methods are a reliable choice for high-level molecular dynamics (MD) simulations.

3.2.2: SEMI-EMPIRICAL METHODS

Semi-empirical methods offer a computationally efficient approximation of QM, where certain integrals are precomputed or approximated, reducing the time and power required for calculations [21]. Examples of semi-empirical methods include AM1 and PM3, which provide sufficient electronic detail for medium-sized systems or larger molecules that would otherwise be too costly to simulate with full QM approaches [22]. By balancing

electronic detail with efficiency, semi-empirical methods act as a bridge between QM and MM, capturing critical electronic properties while maintaining better scalability. These methods are especially useful when electronic effects are essential to the simulation, but computational resources are limited.

3.2.3: MOLECULAR MECHANICS (MM) METHODS

Molecular mechanics (MM) methods treat atomic interactions using classical mechanics principles, using empirical force fields like Chemistry at Harvard Molecular Mechanics (CHARMM), Assisted Model Building with Energy Refinement (AMBER), and Optimized Potentials for Liquid Simulations (OPLS) to model systems at a molecular level. Atoms are treated as point charges, and bonds, angles, and torsions are defined with parameters derived from empirical data. MM methods are highly efficient, making them well-suited for large biomolecular systems like proteins, membranes, and even entire cellular components [21]. However, MM models lack the ability to describe changes in electronic structure, such as bond formation or breaking, limiting their use in studying chemical reactions. Despite this, MM remains a powerful tool for large-scale MD simulations where the primary focus is on structural and dynamic behavior rather than electronic detail.

3.2.4: COMBINED QM/MM METHODS

To bridge the gap between QM and MM, combined QM/MM methods are used, especially for systems where both electronic precision and large-scale modeling are essential. In QM/MM methods, the QM region usually includes the reactive center or active site of interest, where electronic changes are most relevant [21]. The surrounding environment is treated using MM, capturing the larger system's dynamics without excessive computational cost. This hybrid approach is particularly valuable in studies of biochemical reactions and large molecular complexes, where capturing both the quantum-level details and the broader molecular environment is necessary.

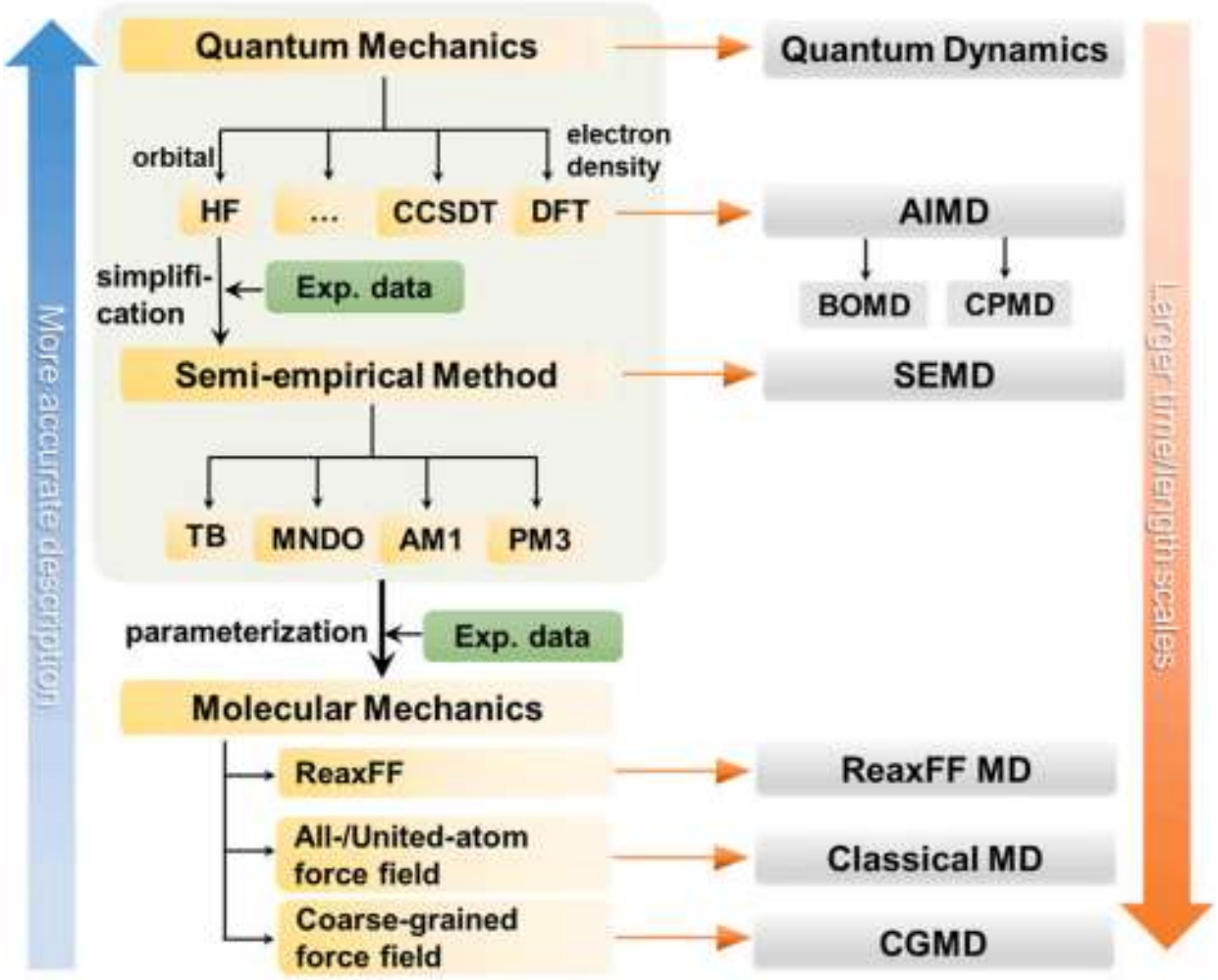

Figure 4: The hierarchical relationship in Molecular dynamics methods [21].

3.3: MOLECULAR DYNAMICS: EQUATIONS OF MOTION

3.3.1: MAIN EQUATIONS

1. **Newton's Equation of Motion**

Newton's second law governs the motion of each particle:

$$\vec{F}_i = m_i \frac{\partial^2 \vec{r}_i}{\partial t^2}. \quad (1)$$

- m_i is the mass of particle i.
- $\vec{r}_i$ is the position vector of particle i .
- $\vec{F}_i$ is the net force acting on particle i.

2. **Force as the negative gradient of potential energy.**

$$\overrightarrow{F_i} = -\frac{\partial V}{\partial \vec{r}_i} \quad (2)$$

- V is the potential energy acting on an. The potential energy from interatomic interactions is determined by several parameters.

3. **Basic summary of potential energies calculated**

$$V(\vec{r}) = V_{str} + V_{bend} + V_{tors} + V_{cross} + V_{el} + V_{vdW} \quad (3)$$

- $\boldsymbol{V_{str}}$(Stretching terms) considers the sum of all 1-2 interactions (atoms directly bonded to each other) in a molecule
- $\boldsymbol{V_{bend}}$ (Bending terms) considers the sum of all 1-3 interactions (Interactions between atoms separated by two covalent bonds i.e. three atoms in the chain)
- $\boldsymbol{V_{tors}}$(Torsional terms) considers all 1-4 interactions (Interactions between atoms separated by three covalent bonds.)
- $\boldsymbol{V_{cross}}$(Cross terms) considers interactions between Stretching, Bending and Torsional terms.
- $\boldsymbol{V_{el}}$(Electrostatic interaction) considers the fluctuation of the electronic cloud generating positively and negatively charged regions.
- $\boldsymbol{V_{vdW}}$(Van der Waals interaction) considers the Van der Waals attractions and repulsions in atoms that are not directly bonded

4. **Solving Newton's equation of motion for $\boldsymbol{a}$ to determine position.**

$$\vec{a}_i = \frac{\vec{F}_i}{m_i} \quad (4)$$

- By solving the equation, the positions can be updated.

5. **Ion current computation**

To minimize the uncertainty in the current estimate, we compute the current only over the region within the pore. Because the silicon nitrate membrane has a thickness of ~ 100 A, we compute the current over $l/2 \le z \le l/2$, where l= 90 A. The current I at time $t + \Delta t/2t$ is computed using the formula:

$$I(t + \Delta t/2) = \frac{1}{\Delta t_{\mathrm{i}}} \sum_i^N q_i \left(\zeta_i(t + \Delta t) - \zeta_i(t)\right) \quad (5)$$

Where,

$$\zeta_i(t) = \begin{cases} z_i(t) \; if \;\; |z_i(t)| \le \ell/2 \\ -\ell/2 \; if \;\; z_i(t) < -\ell/2 \\ \ell/2 \; if \;\; z_i(t) > \ell/2 \end{cases} \quad (6)$$

- q_i is the charge of the i-th ion.
- ζ_i represents the z-coordinate position of each ion within the pore region at times t and $t + \Delta t$.
- z_i represents the **z**-coordinate position of the i-th ion at a given time t.
- $z_i(t)$ indicates the location of each ion along the length of the nanopore at time t.

The position $z_i(t)$ is used to determine whether each ion is inside the region of interest within the nanopore (between $-l/2$ and $l/2$). If an ion's $z_i(t)$ value falls outside this region, it is adjusted to either $-l/2$ or $l/2$,

depending on whether it is below or above this range, as outlined in Equation (6). This ensures that the current calculation focuses only on the ions within the pore region, minimizing the effects of ions outside this area on the current estimate. This approach allows for a more accurate estimation of the ion current by focusing on the region most affected by translocation events.

3.3.2: VELOCITY VERLET ALGORITHM

Method used to integrate Newton's equations of motion in molecular dynamics simulations

1. **First Half-Step Velocity Update:**

$$\vec{v}\left(t+\frac{1}{2}\Delta t\right)=\vec{v}(t)+\frac{1}{2}\vec{a}(t)\Delta t$$

Calculates the velocity halfway through the time step using the acceleration $\vec{a}(t)$

2. **Position Update:**

$$\vec{x}(t+\Delta t)=\vec{x}(t)+\vec{v}\left(t+\frac{1}{2}\Delta t\right)\Delta t$$

Updates the position $\vec{x}(t)$ to $\vec{x}(t+\Delta t)$ using the half-step velocity calculated previously

3. **New Acceleration Calculation:**

$$\vec{a}(t+\Delta t)=\vec{f}\left(\vec{x}(t+\Delta t)\right)$$

New acceleration $a(t+\Delta t)$ is calculated based on the updated position $x(t+\Delta t)$, using the force function $f(x)$ that depends on position.

4. **Second Half-Step Velocity Update:**

$$\vec{v}(t+\Delta t)=\vec{v}\left(t+\frac{1}{2}\Delta t\right)+\frac{1}{2}\vec{a}(t+\Delta t)\Delta t$$

Updates the velocity to the full-time step $\vec{v}(t+\Delta t)$, using new acceleration

The use of this algorithm results in an error of $\sim \mathrm{O(N^2)}$. To lessen the error from the algorithm NAMD uses the Particle Mesh Ewald (PME) algorithm which takes the full electrostatic interactions into account. This algorithm reduces the computational complexity of electrostatic force evaluation from $\sim O(N^2)$ to $\sim O(N \log N)$ [23]

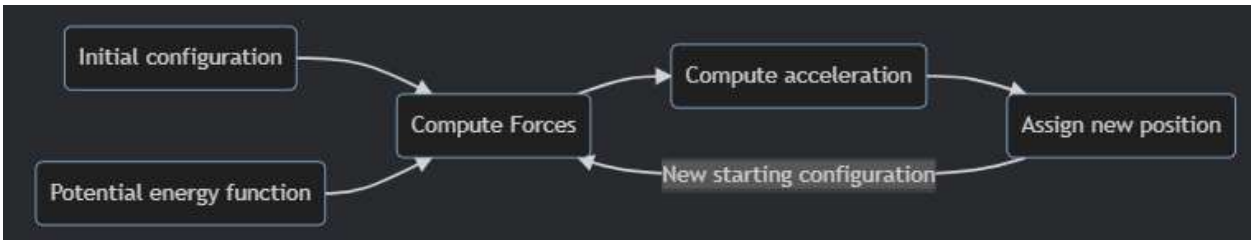

Figure 5: Provides an overview of molecular dynamics process [24].

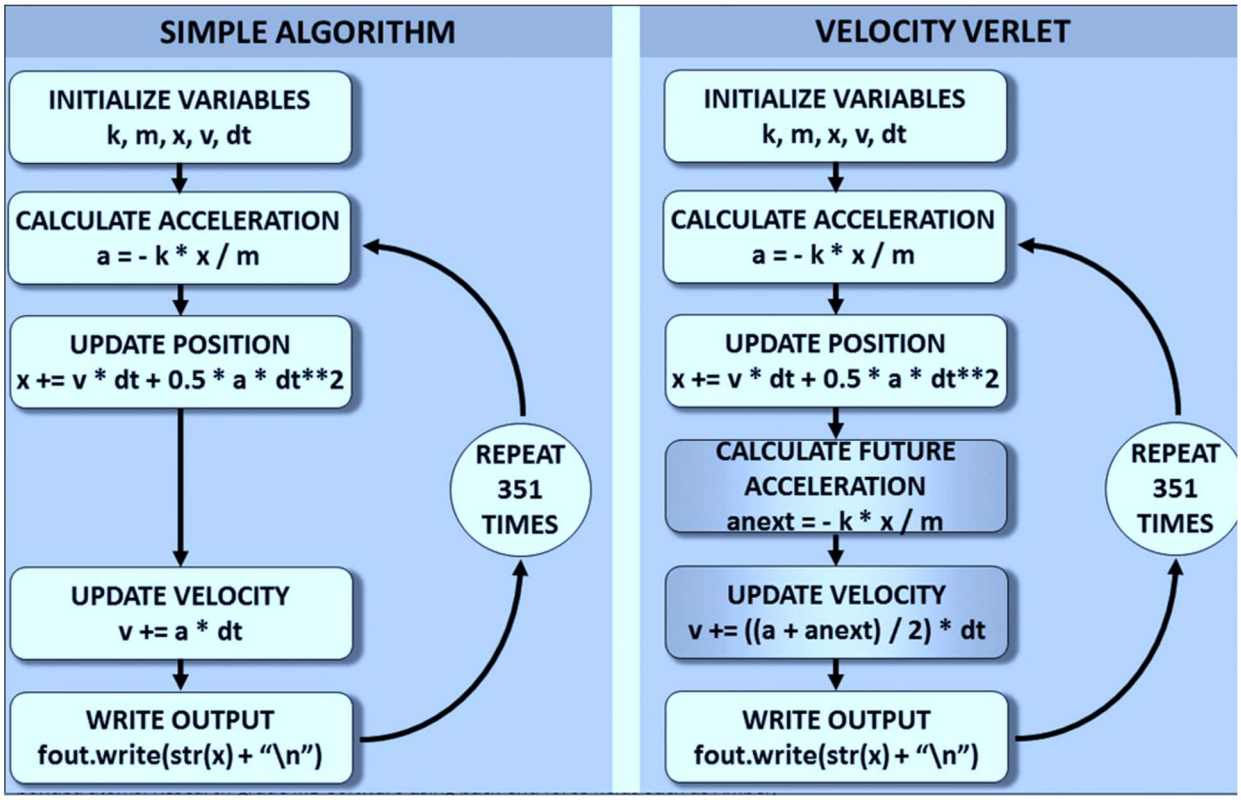

Figure 6: Compares a simple algorithm to the Velocity Verlet algorithm [25].

3.4: MD SOFTWARE:

3.4.1 NAMD

Strengths: Optimized for large-scale simulations of biomolecular systems and excellent scaling across many processors.

Key Features: Good for high-performance computing, often used for large biomolecular systems such as membranes or proteins.

Input Files:

- **Protein Structure File** (.psf): Defines the molecular structure (atoms, bonds, etc.).
- **Coordinate File** (.pdb, .coor): Initial positions of atoms.
- **Parameter File** (.prm, .par): Force field parameters.
- **Configuration File** (.namd): Contains the instructions and simulation settings.

Website: https://www.ks.uiuc.edu/Research/namd/

3.4.2 LAMMPS

Strengths: Very flexible, capable of simulating a wide range of materials (biomolecules, polymers, solids, etc.).

Key Features: Handles atomic, mesoscopic, and coarse-grained simulations, and supports non-biological simulations well.

Input Files:

- **Input Script File** (.in): Contains commands and parameters to set up and run the simulation.
- **Data File** (.data): Includes atomic coordinates, bonds, angles, and other molecular or atomic details.
- **Force Field Files** (.txt, .table, or .in): Custom potential or force field parameter files.

Website: https://lammps.sandia.gov/

3.4.3 AMBER

Strengths: Specialized in biomolecular simulations, particularly for nucleic acids and proteins.

Key Features: AMBER force fields are widely used, and it includes tools for molecular dynamics as well as analysis.

Input Files:

- **Topology File** (.prmtop): Contains atom types, bonds, and force field parameters.
- **Coordinate File** (.inpcrd, .rst7): Atomic coordinates and, optionally, velocities.
- **Control File** (.in): Contains the simulation parameters.
- **Parameter Files** (.frcmod): Used for specifying additional force field parameters.

Website: https://ambermd.org/

3.4.4 GROMACS

Strengths: Highly efficient, widely used for biomolecular simulations (proteins, lipids, nucleic acids).

Key Features: Excellent parallelization, support for large systems, and strong community support.

Input Files:

- **Topology File** (top): Defines the molecular structure, including atoms, bonds, angles, and dihedrals.
- **Coordinate File** (.gro, .pdb): Contains the coordinates of all the atoms.
- **Molecular Topology File** (.itp): Used for defining specific force fields or molecular fragments.

- **Parameter File** (.mdp): Contains simulation parameters (e.g., temperature, time step, pressure coupling).
- **Run Input File** (.tpr): A binary file that includes the topology, coordinates, and parameters required to run the simulation.

Website: https://www.gromacs.org/

3.5: GRAPHICAL INTERFACES FOR MOLECULAR DYNAMICS PREPROCESSING

When setting up molecular dynamics (MD) simulations, preprocessing steps such as building molecular models, assigning force fields, and preparing simulation environments are crucial. Graphical user interfaces (GUIs) simplify these tasks by providing intuitive visual tools. One of the most widely used GUIs in MD is VMD (Visual Molecular Dynamics).

3.5.1: VMD (VISUAL MOLECULAR DYNAMICS)

VMD is a powerful molecular visualization and analysis tool designed to display, animate, and analyze large biomolecular systems. While primarily known for visualization, VMD also offers extensive preprocessing capabilities for MD simulations. It allows users to build and edit molecular structures, including proteins, nucleic acids, and small molecules [30]. VMD facilitates system preparation by enabling the addition of solvent boxes, insertion of ions, and setting up periodic boundary conditions. It supports the assignment of force field parameters to molecules, which is essential for accurate simulations. Furthermore, VMD provides a Tcl/Python scripting interface that allows automation of tasks and customization of functionalities. It integrates seamlessly with MD software like NAMD, making it possible to generate input files compatible with these engines.

3.5.2: OTHER PREPROCESSING GRAPHICAL INTERFACES

Other graphical interfaces assist in the preprocessing of MD simulations. Chimera and ChimeraX provide advanced molecular visualization and editing tools, integrating with AmberTools for force field assignment and simulation setup [31]. PyMOL is used for visualizing and modifying molecular structures, supporting exports compatible with MD software [32],[33]. CHARMM-GUI automates the setup of complex systems like membranes and protein–ligand complexes, generating input files for MD engines such as CHARMM, NAMD, Groningen Molecular Simulation (GROMACS), and AMBER [34]. Molecular Operating Environment (MOE) combines modeling and simulation preparation, automating tasks like fixing missing atoms and assigning protonation states [35]. Avogadro is an open-source molecular builder that supports force field assignment and interfaces with MD packages [36]. Discovery Studio Visualizer offers tools for modeling and simulation setup with a user-friendly GUI [37]. Packmol generates initial structures by packing molecules in defined regions, suitable for complex mixtures and solvated systems [38]. GaussView, primarily used with Gaussian for quantum chemistry, prepares small molecules for MD simulations, exporting geometries to MD-compatible formats [39].

3.5.3: ROLE OF GRAPHICAL INTERFACES IN MD PREPROCESSING

Graphical interfaces like VMD and the others mentioned play a vital role in the preprocessing stage of MD simulations. They simplify complex tasks through intuitive menus and visual aids, reducing errors by providing immediate visual feedback that helps identify and correct issues before running simulations. These interfaces

allow customization and scripting of repetitive tasks, tailoring functionalities to specific needs. They facilitate seamless workflows by integrating with various MD engines and file formats, making MD simulations more accessible to researchers across disciplines [40]. Additionally, they serve as educational tools for teaching concepts in molecular modeling and simulation.

3.6: IMPORTANCE OF SOLVENT IN MD

Solvents are essential in molecular dynamics simulations because they provide a realistic environment in which most molecular interactions naturally occur, particularly in biological and chemical systems. Including solvent molecules, such as water, allows for accurate modeling of solute-solvent interactions like hydrogen bonding, electrostatic screening, and hydrophobic effects, all of which significantly influence the structure, stability, and dynamics of molecules [41]. The presence of a solvent affects thermodynamic and kinetic properties by stabilizing certain molecular conformations, facilitating conformational changes, and influencing reaction mechanisms and rates. Solvents also play a crucial role in processes like protein folding, enzyme activity, and ligand binding by affecting the accessibility and behavior of binding sites. Although incorporating solvents increases the computational complexity of simulations, their inclusion is vital for capturing the true behavior of molecular systems under realistic conditions, making them indispensable for accurate and meaningful molecular dynamics studies.

3.7: LIMITATIONS IN SIMULATION TIME SCALES, ACCURACY, AND FORCE-FIELDS

Molecular dynamics (MD) simulations are powerful tools but have limitations related to simulation time scales, accuracy, and force fields.

3.7.1: COMPUTATIONAL COST

MD simulations of large systems or long durations require significant computational resources. The high demand for processing power and memory can make extensive simulations computationally expensive and time-consuming [21].

3.7.2: TIME SCALE LIMITATIONS

Standard MD simulations typically cover timescales from nanoseconds to microseconds. This short time scale may not capture slower processes like protein folding or diffusion in solids, which occur over milliseconds to seconds [21].

3.7.3: FORCE FIELD ACCURACY

Force fields are based on empirical parameters and may not accurately represent all interactions, especially in unusual environments or novel materials. This limitation affects the reliability of simulations when dealing with systems outside the scope of existing force fields [42].

3.7.4: SAMPLING ISSUES

Systems can become trapped in local energy minima, making it difficult to explore the full phase space. These sampling issues prevent the simulation from capturing all possible configurations and behaviors of the system [43].

3.7.5: CLASSICAL MECHANICS APPROXIMATION

Standard MD simulations rely on classical mechanics and ignore quantum mechanical phenomena, which can be significant in some systems. Ignoring quantum effects limits the accuracy of simulations involving electronic transitions or reactions at the atomic scale [44].

3.8: FORCE FIELDS

3.8.1: DEFINITION

A force field is a set of mathematical equations and parameters used in molecular dynamics simulations to calculate the potential energy of a system of particles. It defines how particles interact by incorporating parameters derived from experimental data and quantum mechanical calculations [19].

3.8.2: COMPONENTS

Force fields consist of bonded and non-bonded terms. Bonded terms account for interactions between atoms that are directly bonded and include bond stretching, angle bending, dihedral torsion, and improper torsion to maintain molecular geometry. Non-bonded terms describe interactions between atoms that are not directly bonded, such as van der Waals interactions and electrostatic interactions [21]. These interactions are formulated so that a simulation will match experimental data.

3.8.3: COMMON FORCE FIELDS

- **AMBER**: Suitable for proteins, nucleic acids, and organic molecules.
- **CHARMM**: Broad applicability, including lipids and carbohydrates.
- **OPLS-AA**: Used for organic and biomolecular systems.
- **Condensed-phase Optimized Molecular Potentials for Atomistic Simulation Studies (COMPASS)**: Designed for accurate simulation of condensed-phase materials.

3.8.4: PARAMETERIZATION

Parameters in force fields are derived by fitting experimental data and high-level quantum mechanical calculations [21]. A key goal is transferability, ensuring that these parameters can be applied across different molecules and systems without significant reparameterization.

3.9: WAYS TO ACCELERATE SIMULATION TIME WITH ENHANCED SAMPLING TECHNIQUES

Accelerating simulation time in molecular dynamics (MD) is crucial for studying complex systems over longer timescales or larger sizes. Various strategies have been developed to improve computational efficiency and enable the exploration of phenomena that occur over extended periods.

3.9.1: METADYNAMICS

Metadynamics accelerates MD simulations by adding a history-dependent bias potential to the system. This bias helps overcome energy barriers, allowing the simulation to explore free energy surfaces more efficiently and sample rare events that would otherwise require long simulation times [45].

3.9.2: UMBRELLA SAMPLING

Umbrella sampling applies biasing potentials to constrain the system in specific regions of phase space. By focusing on rarely visited states and reweighing the results, it enables accurate calculation of free energy differences between states, enhancing the sampling efficiency [46].

3.9.3: REPLICA EXCHANGE MOLECULAR DYNAMICS (REMD)

REMD involves running multiple simulations (replicas) at different temperatures or Hamiltonians. Periodic exchanges between these replicas allow the system to overcome energy barriers more readily, providing a more comprehensive exploration of conformational space [47].

3.9.4: COARSE-GRAINED MODELS

Coarse-grained models simplify the system by grouping atoms into larger pseudo-atoms or beads, reducing the number of particles. This simplification lowers computational demands, allowing for the simulation of larger systems over longer timescales while capturing essential system behaviors [48].

3.9.5: GPU ACCELERATION

Utilizing Graphics Processing Units (GPUs) can significantly speed up MD simulations through parallel computing. GPUs are designed for efficient parallel computations, and many MD software packages (e.g., GROMACS, AMBER, LAMMPS, NAMD) offer GPU-accelerated versions that take advantage of this hardware to perform longer simulations in less time [49].

3.9.6: PARALLEL COMPUTING

Parallel computing distributes computational workloads across multiple processors or computing nodes. By splitting simulation tasks, it drastically reduces simulation time for large systems. Efficient parallelization and scalability are key to maximizing the benefits of this approach [50].

3.9.7: MACHINE LEARNING POTENTIALS

Machine learning potentials use algorithms like neural networks to create accurate and computationally efficient representations of potential energy surfaces. Models such as DeepMD and SchNet capture complex interactions with lower computational costs, enabling longer and larger-scale simulations without sacrificing accuracy [51].

3.9.8: MULTI-SCALE MODELING

Multi-scale modeling combines different levels of theory within a single simulation to balance accuracy and computational efficiency. Quantum Mechanics/Molecular Mechanics (QM/MM) simulations treat critical parts of the system quantum mechanically while using molecular mechanics for the rest. Adaptive resolution schemes dynamically adjust the level of detail in different regions as needed [52].

3.9.9: TIME STEP OPTIMIZATION

Time step optimization techniques allow the use of larger time steps to accelerate simulations. Multiple time stepping applies larger time steps to slower degrees of freedom and smaller steps to faster motions. Constraint algorithms fix high-frequency vibrations (e.g., bonds involving hydrogen), permitting larger overall time steps and reducing the number of steps required [53].

3.10: STUDIES ON MD SIMULATIONS OF DNA TRANSLOCATION THROUGH NANOPORES.

One relevant study, "Imaging α-Hemolysin with Molecular Dynamics: Ionic Conductance, Osmotic Permeability, and the Electrostatic Potential Map" by Aksimentiev and Schulten (2005), used MD simulations to investigate α-hemolysin, focusing on ionic conductance and permeability [54]. Their work provided a detailed atomic-scale view of ion and water transport through α-hemolysin, revealing how electric fields influence ionic currents and how electrostatic interactions affect DNA translocation dynamics. This paper aligns well with this research, as it demonstrates the use of MD to simulate α-hemolysin nanopores

The study by Di Marino et al. (2015), titled *"All-Atom Molecular Dynamics Simulation of Protein Translocation through an α-Hemolysin Nanopore"* in *The Journal of Physical Chemistry Letters*, investigates protein translocation dynamics, focusing on thioredoxin (Trx), a small redox protein essential for cellular redox balance in many organisms [55]. Through molecular dynamics simulations, the researchers modeled the forces and unfolding events Trx experiences as it translocates, identifying specific barriers within the nanopore, particularly at constriction points that affect translocation rates and unfolding behavior. This work provides insights into the structural transitions of molecules moving through nanopores, contributing valuable knowledge to molecular translocation mechanisms at the nanoscale.

The review by Heerema and Dekker (2016), titled *"Graphene Nanodevices for DNA Sequencing"*, explores different methods for employing graphene nanopores in DNA sequencing, emphasizing graphene's atomically thin structure, which potentially allows single-nucleotide resolution [56]. The paper discusses key factors influencing DNA translocation, such as ionic currents, hydrophobic interactions, and the challenges posed by noise and rapid translocation speeds. These insights are directly relevant to understanding DNA dynamics within nanopores, contributing valuable perspectives on optimizing nanopore-based sequencing technologies

3.11: CHALLENGES IN MD SIMULATIONS OF DNA

3.11.1: STEEP LEARNING CURVE

Creating molecular dynamics simulations of DNA translocating through a nanopore presents a steep learning curve due to the intricate complexities of both biological macromolecules and advanced computational methods. This challenge stems from the need to model large, complex systems consisting of thousands to millions of atoms, including DNA, the nanopore, solvent molecules, and ions. It requires proficiency in specialized simulation software, high-performance computing to manage substantial computational resources, and a solid grasp of interdisciplinary concepts from molecular biology, physics, and chemistry. Furthermore, setting up such simulations involves selecting and parameterizing appropriate force fields, applying external electric fields to drive DNA translocation, and using advanced sampling methods to overcome the timescale limitations inherent in molecular dynamics. Together, these factors contribute to a significant learning curve, requiring extensive knowledge and experience to effectively simulate and analyze DNA translocation through nanopores [57].

3.11.2: DNA FLEXIBILITY AND CONFORMATION

DNA is a flexible molecule whose conformation can change significantly during translocation. Capturing this flexibility requires a detailed all-atom representation, which increases computational demands. The DNA may bend, stretch, or undergo conformational transitions as it interacts with the nanopore, and these changes can influence the translocation dynamics [55]. Simplifying the DNA model may overlook important structural features, while fully detailed models are computationally expensive. In "All-atom molecular dynamics simulation of protein translocation through an α-hemolysin nanopore" [56] it explains the unfolding of Trx in the biological nanopore. In a solid-state nanopore it would require setting up the system so the DNA would not fold.

3.11.3: TIMESCALE LIMITATIONS

MD simulations are typically limited to nanosecond to microsecond timescales due to computational constraints. However, DNA translocation through a nanopore can occur over much longer timescales, ranging from microseconds to milliseconds or even seconds. Capturing the complete translocation process requires simulating durations that are often beyond the reach of standard MD techniques. This limitation makes it challenging to observe the full event and necessitates the use of enhanced sampling methods or coarse-grained models to extend the accessible timescales [19].

CHAPTER 4: METHODS

Original Plan

Step 1: Create Water Box
- Simulation box size: 5.80nm X 5.80nm X 10nm
- VMD uses angstroms 58A X 58A X 100A
- Saturate Water Box With Proper Ions

Complete

Step 2: Using TopoTools transfer file into script that can be read by LAMMPS

Requirements
- Understanding TopoTools
- Being able to check file when finished

IN PROGRESS
- Need to email creator of TopoTools to check our understanding of the process
- Trouble obtaining output file to view in VMD

Step 3: Run LAMMPS input file
- Check all program conditions are correct
- Make sure a dump file can be extracted

IN PROGRESS

Step 4: View in VMD

IN PROGRESS

Step 5: Add A Force Field To simulation
- TopoTools→LAMMPS→VMD
- See if the effect from the forcefield in VMD

Step 6: Add A Graphene Nanopore to the water box VMD file to create one file
- TopoTools→LAMMPS→VMD
- See the output in VMD

Step 7: Add Current Reading Across Nanopore
- More Research Required for this

Step 8: Add DNA to Simulation
- TopoTools→LAMMPS→VMD

Should obtain data to create a current vs time graph

Figure 7: Original plan

Reworked Plan

LAMMPS Experiments

Step 1: Create Water Box in VMD
- Simulation box size: 5.80nm X 5.80nm X 10nm
- VMD uses angstroms 58A X 58A X 100A
- Saturate Water Box with 1.0 mol/kg KCl

Complete

Step 2: Using TopoTools transfer file into script that can be read by LAMMPS
- Understanding TopoTools
- Being able to check file when finished

Complete

Step 3: Run several LAMMPS input files
- Check all program conditions are correct
- Make sure a dump file can be extracted

Complete

Step 4: View output in VMD and analyze data
- Analyze water box
- Analyze PEG molecule solvated in water
- Analyze PEG molecule being stretched

Complete

NAMD Experiments

Step 5: Create Input files in VMD
- Solvate and ionize DNA with 1.0 mol/kg KCl
- Create and merge α-hemolysin with DNA and Silicon Nitride with DNA
- Minimize, Heat and equilibrate systems

Complete

Step 6: Run α-hemolysin with DNA simulation
- Obtain simple output

Complete

Step 7: Run Silicon Nitride with DNA simulation
- Obtain simple output

Complete

Step 8: Compare output
- With Simulation data
- With Experimental data

Complete

Step 8: Future Works

Complete

Figure 8: Reworked Plan

4.1: SOFTWARE USED

4.1.1: VMD

Visual Molecular Dynamics (VMD): Used for system set up

The research started with VMD due to its ability to handle complex biomolecular systems and its extensive support for commonly used molecular dynamics file formats. VMD efficiently utilizes PSF (Protein Structure File) and PDB (Protein Data Bank) files to define molecular structures, including atoms, bonds, and coordinates, making it an excellent tool for preparing systems. While VMD's TopoTools [58] plugin was used to convert PSF and PDB files into the specific LAMMPS data file format required for LAMMPS simulations, this process proved to be complex due to the need for additional formatting steps. As the research progressed, the workflow shifted toward using NAMD, as its input files directly support PSF and PDB formats. This made VMD even more effective when paired with NAMD, streamlining the simulation setup process and eliminating the need for format conversion, ultimately making NAMD a more convenient choice for the simulations.

4.1.2: LAMMPS

Large-scale Atomic/Molecular Massively Parallel Simulator: Used for MD simulation

LAMMPS was initially selected as the primary software for this research, which focused on replicating and expanding the simulations discussed in the paper "DNA Detection with Single-Layer Ti3C2 MXene Nanopore"

[58]. Throughout the process, the research significantly enhanced the author's knowledge of molecular dynamics, culminating in the successful simulation of a PEG molecule solvated in ionized water. This experience provided valuable insights into MD techniques and their application to nanopore-based systems.

4.1.3: LINUX

Linux is an open-source operating system (OS) widely used as an alternative to proprietary systems like macOS and Microsoft Windows. It is known for its high performance, flexibility, and extensive customization options. In this research, Linux was utilized to connect to the OU supercomputer. The Windows Subsystem for Linux (WSL) was employed, allowing Linux to run seamlessly on a Windows environment without the need for a separate virtual machine or dual-booting.

4.1.4: NAMD

Nanoscale Molecular Dynamics (NAMD) was used for molecular dynamics simulations in this research. It was downloaded onto the OU supercomputer, where tests were performed to simulate DNA translocation through a α-hemolysin nanopore. The research was guided by the cited paper, "Modeling Nanopores for Sequencing DNA,"[60] which provided a foundational approach for these simulations. Through this process, valuable data was obtained, along with a deeper understanding of the mechanisms involved in DNA translocation through nanopores.

4.2: SUPERCOMPUTERS USED

4.2.1: UCO BUDDY SUPERCOMPUTER

Buddy is a 37-node supercomputer funded by the National Science Foundation (NSF) and designed for high-performance computing tasks. Detailed information about Buddy can be found on the NSF website. In this research, Buddy was used to run molecular dynamics simulations using VMD for visualization and pre-processing, along with LAMMPS for running the simulations themselves. The combination of VMD and LAMMPS on Buddy allowed for efficient handling of complex molecular systems and large-scale computations essential for simulating molecular interactions and dynamics.

4.2.2: OU SCHOONER SUPERCOMPUTER

The OU Schooner supercomputer, equipped with over 10,000 CPU cores, 23TB of RAM, and 450TB of usable storage, provides the computational power necessary to handle even the most demanding tasks. In this research, Schooner was used to run NAMD simulations, leveraging its high-performance architecture to efficiently model molecular dynamics. Schooner’s extensive resources made it an ideal platform for complex simulations, enabling the study of intricate biomolecular systems. For more details about Schooner, you can visit the OU supercomputing website.

4.3: ACCESSING VMD AND NAMD

4.3.1: VMD ON BUDDY

Faculty members or students working on a project at UCO, or collaborating with a UCO faculty member, can request access to Buddy by sending an email to hpc@uco.edu to initiate account setup. Once the account is

created, users can access Buddy from any location via a web browser by visiting https://buddy.uco.edu. The platform utilizes Open OnDemand software, which simplifies job submission and offers interactive graphical user interfaces for managing computational tasks. VMD is also available through the Interactive Apps section, allowing users to conveniently work with molecular systems in an intuitive environment.

4.3.2: LAMMPS ON BUDDY

LAMMPS is accessible on UCO's Buddy super computer through running the sbatch (located in appendix A) file in terminal.

4.3.3: NAMD ON THE OU SUPERCOMPUTER

To use NAMD on the OU supercomputer, the process begins by requesting an account through OU OSCER Account Request. After receiving account approval, users can connect to the supercomputer, named Schooner, via SSH by entering "ssh yourusername@sooner.oscer.ou.edu" in a terminal, where "yourusername" is replaced with their OSCER credentials. With OU transitioning from CentOS 7 to Rocky 9, the machine name for this project is sooner.oscer.ou.edu. The build that was used for NAMD on Sooner was an easy build configuration for NAMD 3.0 from https://github.com/easybuilders/easybuild-easyconfigs/pull/20026. A sample batch script for NAMD can is in appendix A

4.4: SIMULATION DETAILS

4.4.1: FORCE FIELDS USED

The LAMMPS simulations utilized the GROMOS 54A7 force field for the PEG molecule and SPC/Fw for water (H_2O). Force fields are derived from experimental data and must be carefully selected based on the specific properties or behaviors you aim to study in a given molecule. In the following section, a comparison of water force fields with experimental data will illustrate how certain force fields excel in specific areas while underperforming in others, highlighting the importance of choosing an appropriate force field for your simulation objectives.

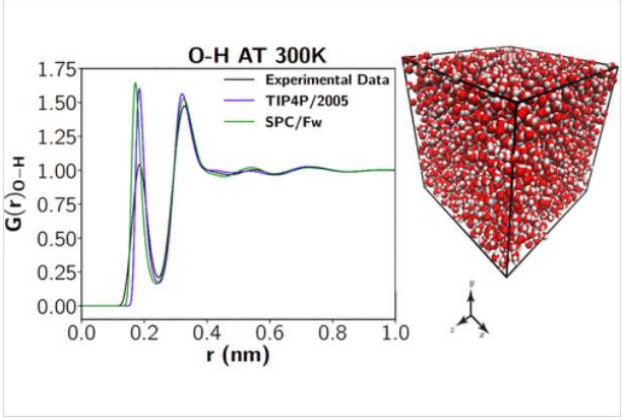

Figure 9: Comparison of water force fields to experimental data [61].

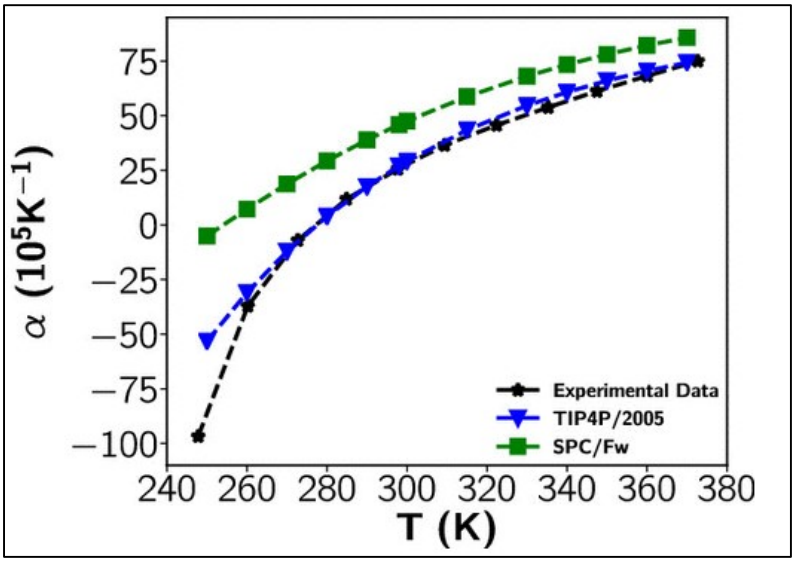

Figure 10: Thermal expansion coefficient at atmospheric pressure [61].

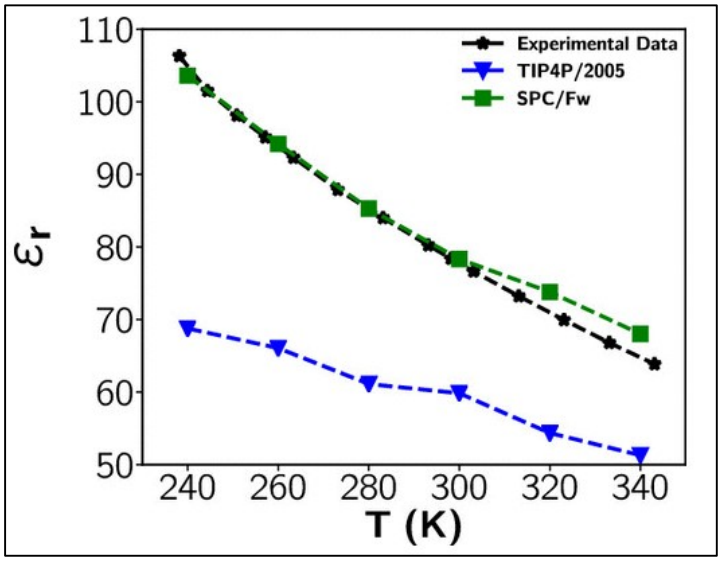

Figure 11: Calculated dielectric constant at several temperatures at a density of 1 g cm–3 and P = 1 atm [61].

The NAMD simulation uses Toppar files with c32b1 release of CHARMM. Includes the cross-term map correction (CMAP) extension to the CHARMM22 protein force field and various extensions of the force fields. The files were obtained from https://mackerell.umaryland.edu/charmm_ff.shtml and use the CHARMM topology files top_all27_na.rtf and top_all27_prot_lipid.rtf as well as the CHARMM parameter files par_all27_na.prm and par_all27_prot_lipid.prm are used in this work.

4.4.2: UNITS USED

In the LAMMPS simulations, the command units real are used. With this setting, masses are measured in grams per mole, distances in Angstroms, time in femtoseconds, and energies in kcal/mol.

In NAMD, the standard units include Angstroms for length, kcal/mol for energy, Kelvin for temperature, bar for pressure, and femtoseconds for time.

4.4.3: MINIMIZATION, HEATING, AND EQUILIBRATION

Minimization, heating, and equilibration are essential preparatory steps in molecular dynamics simulations to ensure system stability and accurate results. Minimization reduces the potential energy of the system by eliminating any initial steric clashes or unfavorable interactions between atoms, providing a stable starting structure. Heating then gradually increases the system's temperature to the desired level, allowing the molecules to adjust without causing abrupt energy spikes [62]. Finally, equilibration ensures that the system reaches and maintains a steady state, where temperature, pressure, and other properties stabilize, allowing for reliable data collection during the production phase of the simulation. Together, these steps prepare the system for accurate and physically meaningful molecular dynamics simulations.

4.5: NANOPORE DNA SEQUENCING PROCESS

4.5.1: NANOPORE STRUCTURE

A nanopore can be biological (such as α-hemolysin or MspA proteins) or synthetic (such as silicon-based materials). The nanopore is embedded in a membrane that separates two chambers filled with a conductive solution [63]. An electric field is applied across the membrane, creating an ionic current through the nanopore.

4.5.2: DNA TRANSLOCATION

A single-stranded DNA (ssDNA) molecule is guided into the nanopore under the influence of the applied electric field. As the DNA moves through the pore, each nucleotide sequentially blocks or alters the flow of ions. The degree to which the current is disrupted depends on the size and shape of each base (A, T, C, or G).

4.5.3: IONIC CURRENT MODULATION

As the DNA strand translocates through the nanopore, the characteristic change in the ionic current is measured. Different nucleotides cause distinct current blockages due to their unique molecular structures, allowing the system to distinguish between them.

4.5.4: BASE CALLING

The data collected from the ionic current changes are processed by algorithms that "read" the sequence of nucleotides based on the unique current signatures corresponding to each base. This step, known as base calling, allows the real-time identification of DNA sequences.

4.6: STRUCTURE AND DYNAMICS OF DNA AS IT INTERACTS WITH THE NANOPORE

The structure and dynamics of DNA as it interacts with a nanopore are critical in determining how effectively translocation occurs, impacting speed, accuracy, and resolution. DNA conformation is important; as DNA enters and interacts with the nanopore, it often adopts specific conformations based on the pore's dimensions and surface characteristics. For example, single-stranded DNA (ssDNA) typically has greater flexibility than double-stranded DNA (dsDNA), which affects how each type of DNA strand passes through and interacts with

the pore's inner surfaces. ssDNA's flexibility makes identifying individual discrete structures difficult, if not impossible [64].

4.6.1: FACTORS AFFECTING TRANSLOCATION SPEED, ACCURACY, AND RESOLUTION

Several factors affect the translocation speed, accuracy, and resolution of DNA moving through a nanopore. Electric fields are essential in driving DNA through the nanopore, as the electrophoretic force propels DNA across the channel. The field strength directly affects translocation speed; higher fields increase speed but may reduce resolution, as DNA bases move too quickly to capture precise current readings. In contrast, lower electric fields slow down translocation, potentially enhancing accuracy and resolution for identifying bases, though this requires more time [65].

Nanopore size and shape are further critical factors. Pore diameter and shape directly impact DNA's ability to pass through. Narrower nanopores can enhance resolution by restricting DNA movement and improving current signal differentiation for individual bases, whereas larger pores facilitate faster translocation but may reduce the precision of base readings. Thermal fluctuations also affect DNA's motion, as random movement at nanoscale dimensions contributes to noise in current readings. Controlling temperature helps ensure consistent DNA behavior within the pore, which can improve resolution.

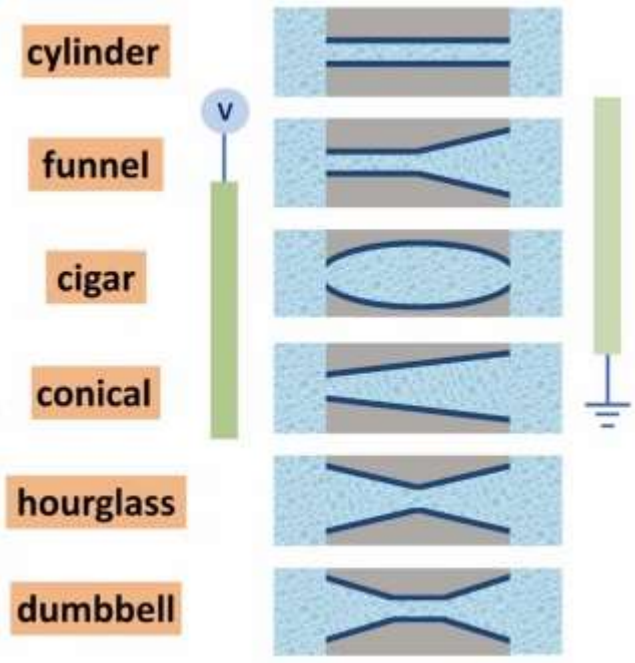

Figure 12: Depicts different shapes of nanopores that have been used in experiments [66].

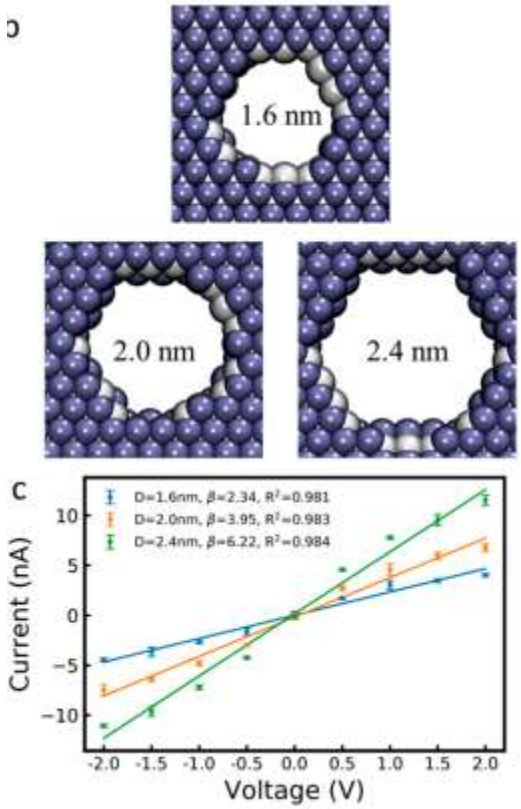

Figure 13: Shows the nanopore of 1.6 nm diameter has a greater resistance to the flow of ionic current, which is consistent with experimental data and previous results [59].

4.7: SYSTEM SETUP

Boundary conditions: Periodic boundary conditions are applied to avoid edge effects and maintain a consistent ionic distribution around the nanopore, creating a more realistic, unbounded simulation environment.

Timesteps: A multiple time-stepping approach is utilized. Bonded forces are calculated at intervals of 1 fs, Lennard-Jones and short-range electrostatic forces are calculated every 2 fs, and long-range electrostatic forces are computed every 4 fs. Electrostatic forces are calculated using the particle mesh Ewald algorithm, with grid spacing set to less than 1.5 Å [60].

Ionic Concentration and Type: The system includes a KCl solution with a molality of 1.0 mol/kg, added using the "add ions" command in VMD.

Nanopore models: Both alpha-hemolysin (biological nanopore) and Silicon Nitride (solid-state nanopore) models are simulated.

DNA configuration: ssDNA and dsDNA are used in the alpha-hemolysin and silicon nitrate simulations respectively.

Solvent model: The TIP3P water model is used to represent the solvent, employing three interaction points that correspond to the three atoms of the water molecule, as set by the VMD solvent command.

CHAPTER 5: RESULTS

5.1: CREATION OF IONIZED WATER BOX IN VMD

First a water box is created in VMD with dimensions of 50Å x 50Å x 50Å using the solvate function.

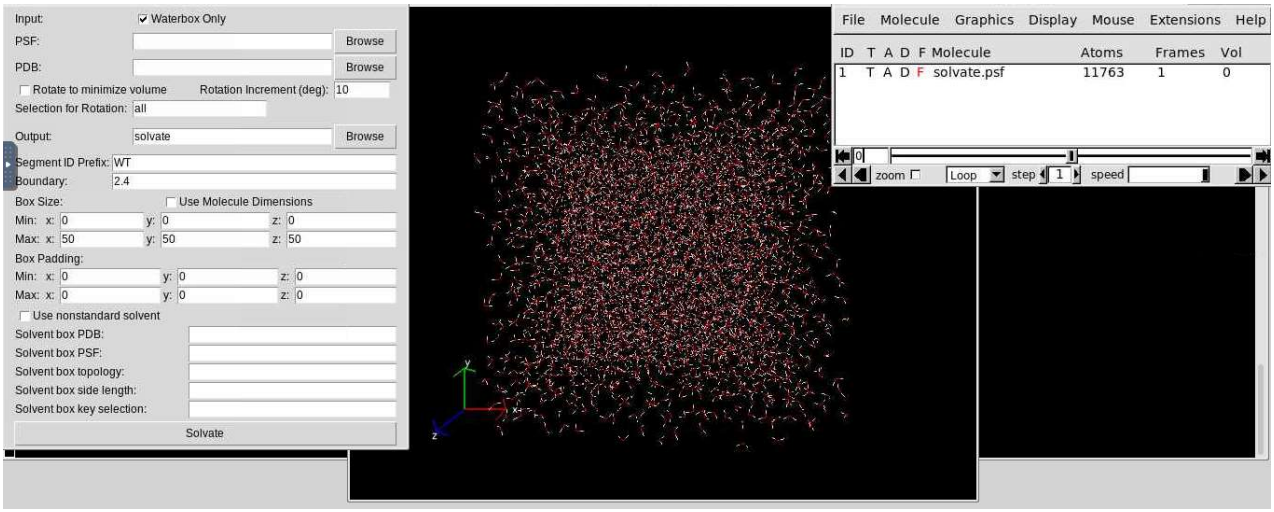

Figure 14: Water box

Next the water box is ionized using the autoionize function to saturate a system with 1.0 mol/kg of KCl ions.

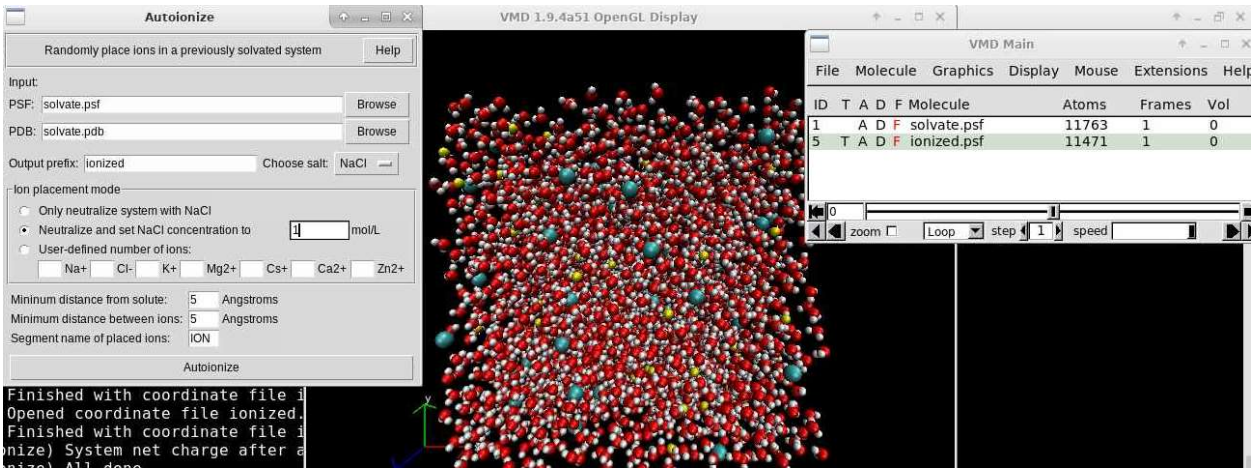

Figure 15: System with NaCl concentration set to 1 mol/L

5.2: LAMMPS SIMULATIONS

5.2.1: SIMULATION OF WATER BOX IN LAMMPS

The water box is first prepared without the PEG molecule present. A rectangular box of water is created and equilibrated at ambient temperature and ambient pressure. The water forcefield used is the SPC/Fw model. The code for water box is located with explanation is in Appendix B.

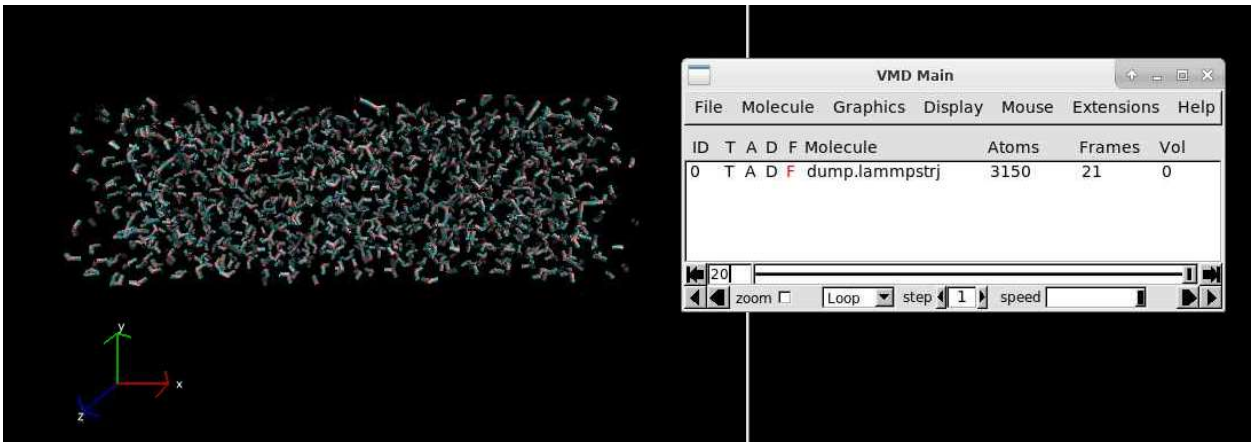

Figure 16: Dump file from simulation

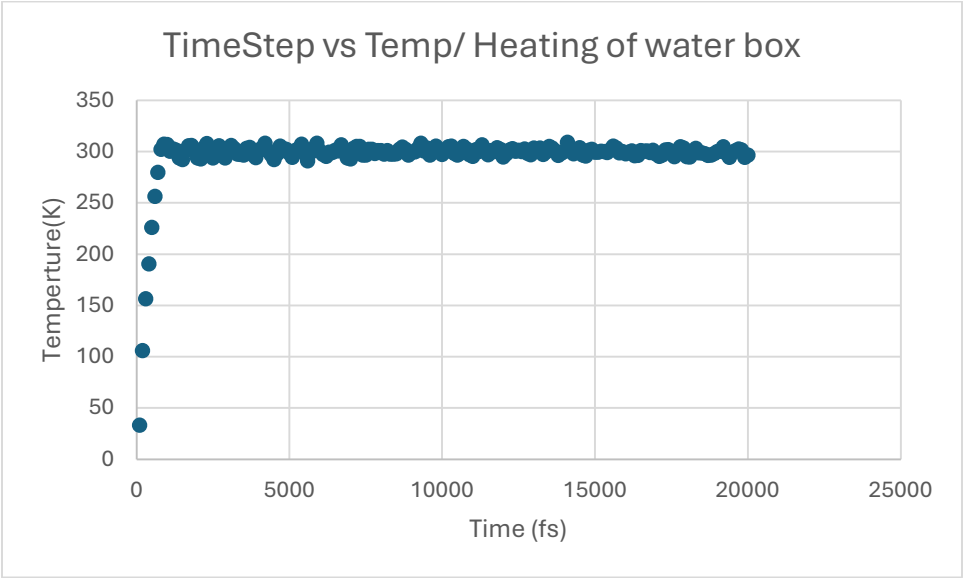

Figure 17: Heating up water box system

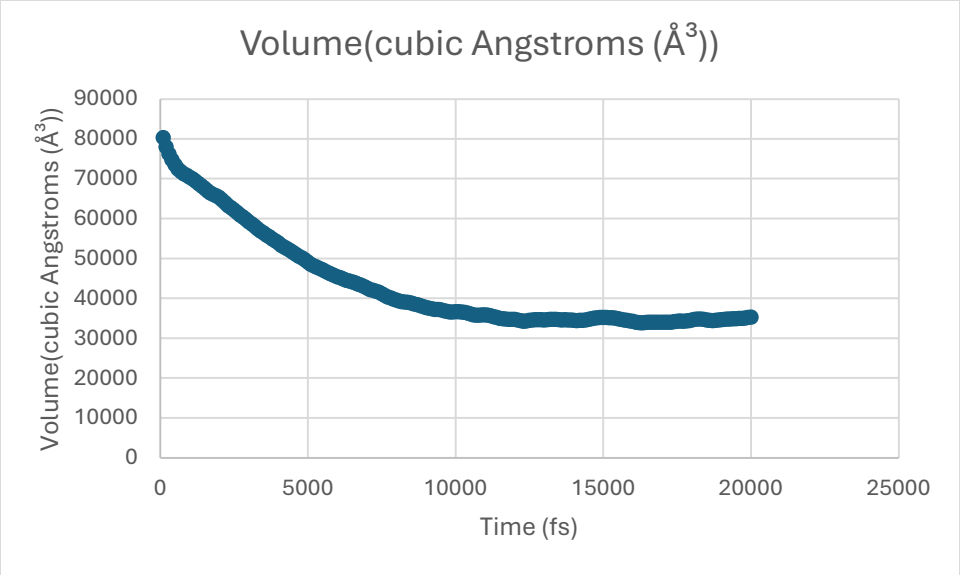

Figure 18: Volume as system heats up

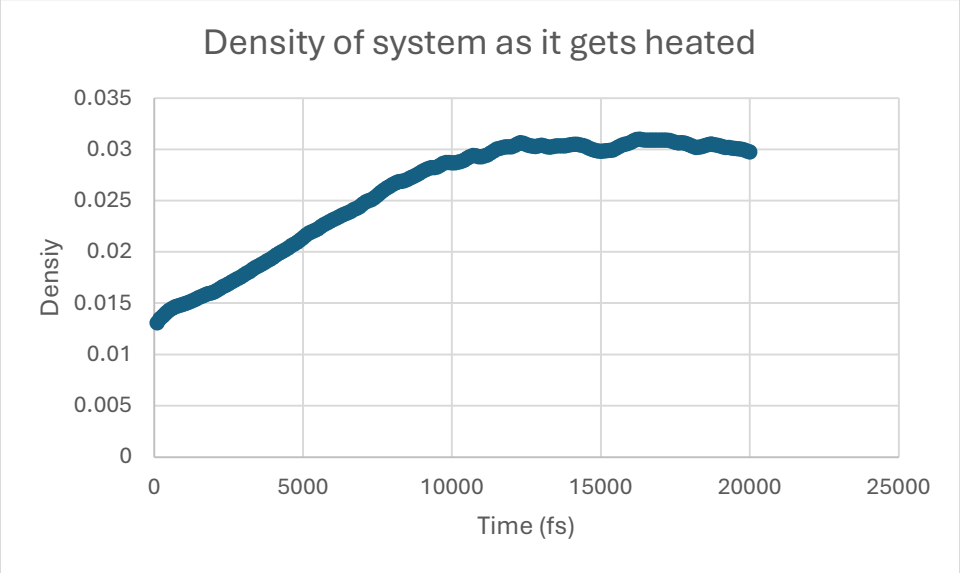

Figure 19: Density as system heats up. Shows the density reaches equilibrium around 12ps.

5.2.2: SIMULATION OF PEG MOLECULE MERGED WITH WATER BOX

The PEG molecule has a formula $C_{28}H_{58}O_{15}$, and uses GROMOS 54A7 force field [67]. The molecule is placed in the center of the box but there is some overlap with the atoms, so we delete the H2O atoms within a 2-angstrom distance. The code is run, and the temperature is set to 300K and pressure is set to 1 atmosphere. The Code for Simulation of PEG molecule solvated in water using LAMMPS is located with explanation in Appendix B.

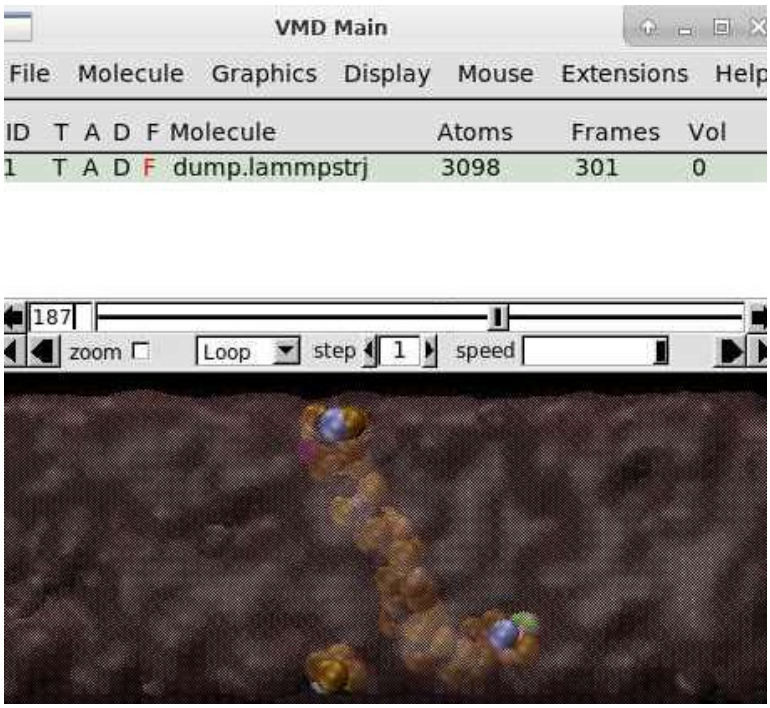

Figure 20: Visualization of system is uploaded at https://youtu.be/Q0mcDjEwcSU

5.2.3: PEG MOLECULE BEING STRETCHED

The final test run in LAMMPS stretches the PEG molecule by applying a force on the Oxygen atoms at each end. Understanding how forces can be applied to a system will be helpful for future testing. The code for water box is located with explanation is in Appendix B

Visualized at: https://youtu.be/8zmbSJq3PEo

Figure 21: Before being stretched

Figure 22: After being stretched

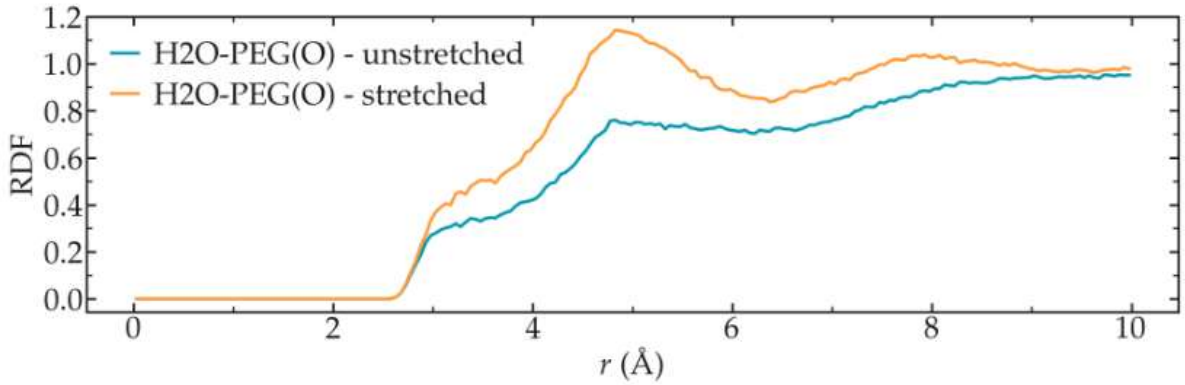

Figure 23: Radial distribution function between the oxygen atoms of water and the oxygen atoms of the PEG molecule. The peak in the unstretched function represents where particles are located and because it is greater than one it suggests that the molecule has structural ordering.

5.3: CREATION AND SIMULATION OF A-HEMOLYSIN-DNA SYSTEM IN NAMD

5.3.1: SSDNA SOLVATED

Obtained a DNA molecule online and by using VMD was able to solvate the DNA and add ions to the system. Also created a ssDNA file separately by deleting one chain of atoms. In this process VMD generates a psf file with the pdb file we have provided. It completes this process by using a guess algorithm. Next 1.0 mol/kg of KCl ions will be added to the simulation by using the VMD autoionize tool. To cancel the charge of DNA an additional 39K^+ ions are added to neutralize the system.

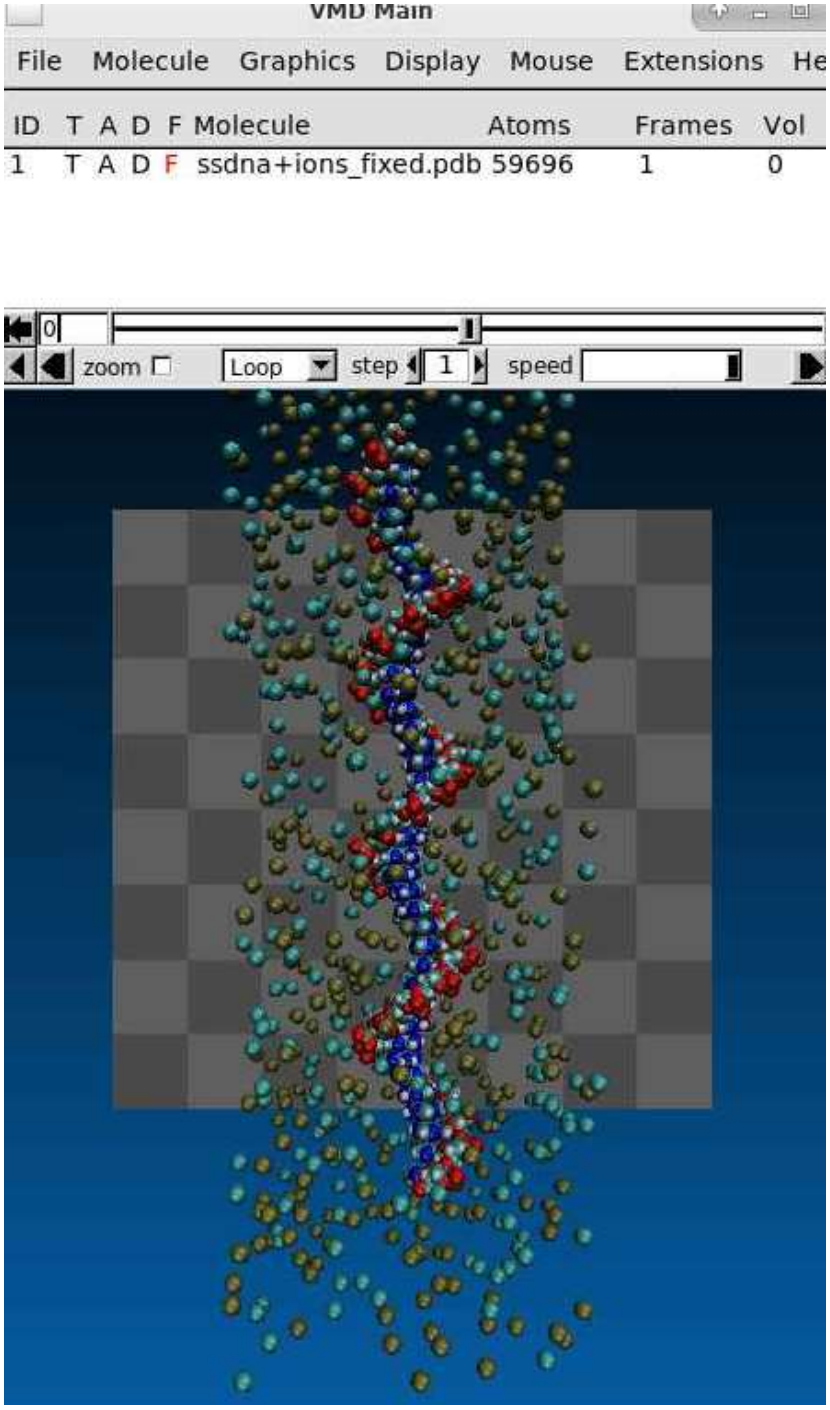

Figure 24: VMD visualization of solvated ssDNA (water not viewed).

5.3.2: MINIMIZATION, HEATING, AND EQUILIBRATION OF SYSTEM IN NAMD

The importance of Minimization, Heating, and Equilibration is described in 4.4.3: Minimization, Heating, and Equilibration. The code for MINIMIZATION, HEATING, AND EQUILIBRATION OF SYSTEM is in appendix D.

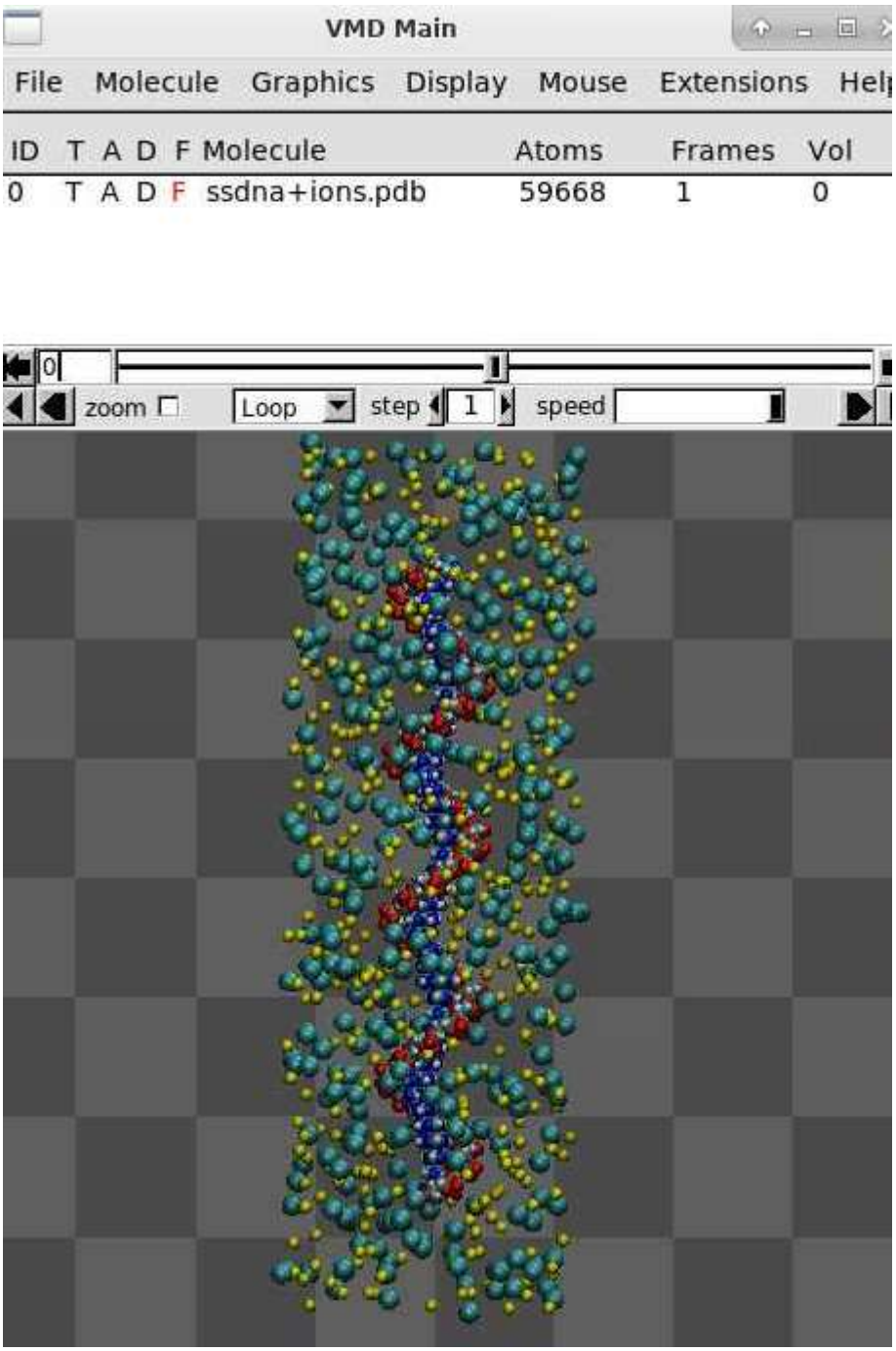

Figure 25: The ssDNA system after Minimization, Heating, and Equilibration

5.3.3: CONSTRUCTION OF A-HEMOLYSIN NANOPORE

α-hemolysin was obtained from protein data bank using PDB code 7AHL. Then the protein is solvated creating a 3-angstrom thick shell of water. The next step is adding the lipid bilayer membrane in which α-hemolysin will sit. To do this the “Membrane” plugin for VMD is used. The protein and lipid membrane are combined into one psf and one pdb file. After that the entire system is solvated to get the final box size. Now we use the autoionize

function to add 1.0 mol/kg of KCl ions to the system. Finally, minimization, heating and equilibration are performed on the system.

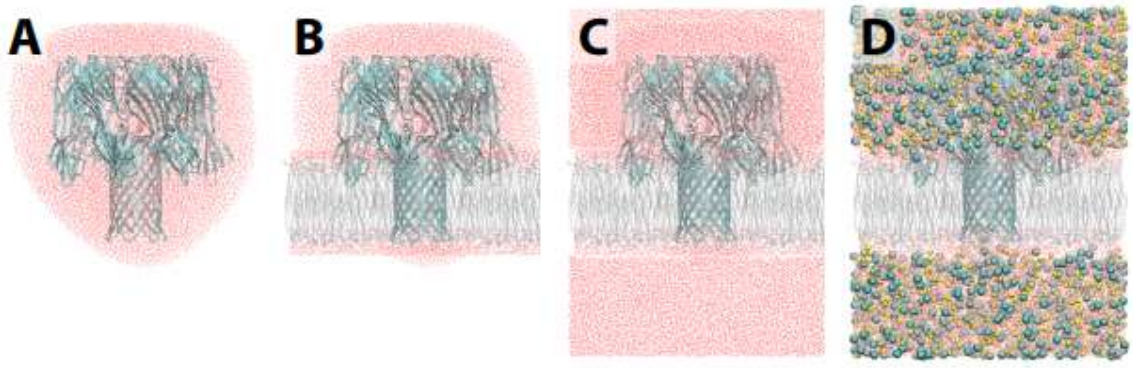

Figure 26: (A) α-hemolysin with a 3-angstrom thick shell of water surrounding it. (B) α-hemolysin combined with the lipid bilayer membrane. (C) System after it is solvated. (D) Final α-hemolysin system with KCl ions added.

5.3.4: A-HEMOLYSIN-DNA SYSTEM

To make sure that the ssDNA will fit through the nanopore we must apply the phantom pore method to retrieve a conformation of the ssDNA that will fit through the nanopore without steric clashes. This process uses a method in NAMD that will push the DNA in α-hemolysin into a five angstrom-radius cylinder. With the outside being confined in a cylinder with a 15-angstrom radius.

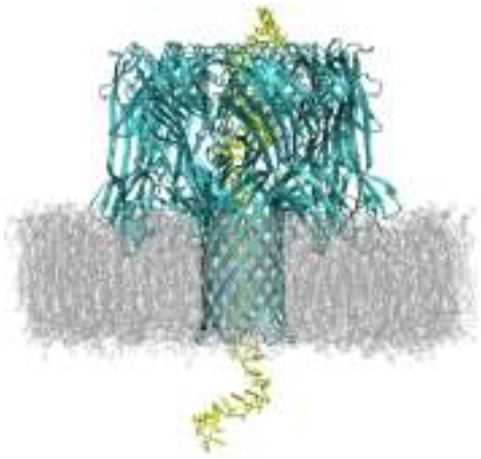

Figure 27: Snapshot of ssDNA through α-hemolysin

5.3.5: SIMULATING A-HEMOLYSIN USING GRID-STEERED MOLECULAR DYNAMICS

The Guided Steered Molecular Dynamics (G-SMD) feature in NAMD applies directional forces to guide molecular systems, such as DNA, along a predefined path, simulating the forces driving DNA through biological or synthetic nanopores. In this simulation, the electrostatic potential used was derived from the study “Modeling Nanopores for Sequencing DNA,” with a transmembrane bias of 1.2 V. This setup accelerates

DNA translocation while maintaining a realistic potential, resulting in a permeation event that closely resembles experimental conditions [68].

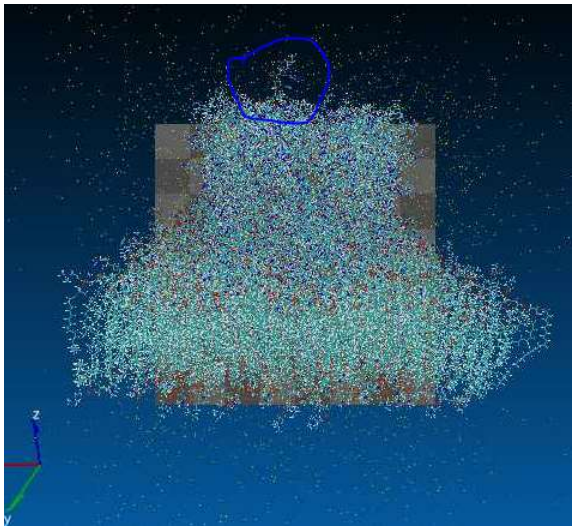

Figure 28: Shows that ssDNA starts inside alpha-hemolysin

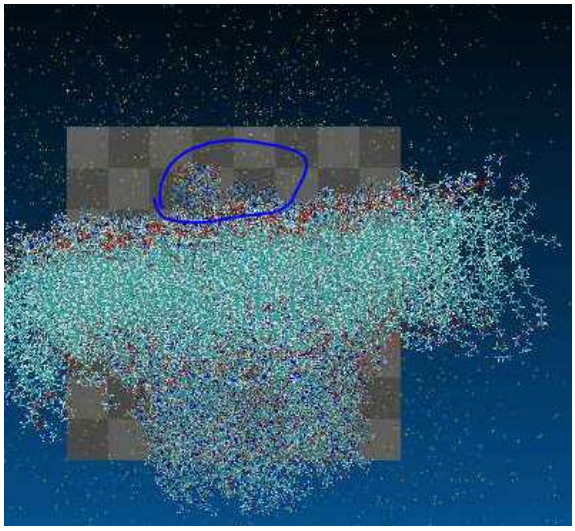

Figure 29: Shows that ssDNA sticking to the surface after translocating.

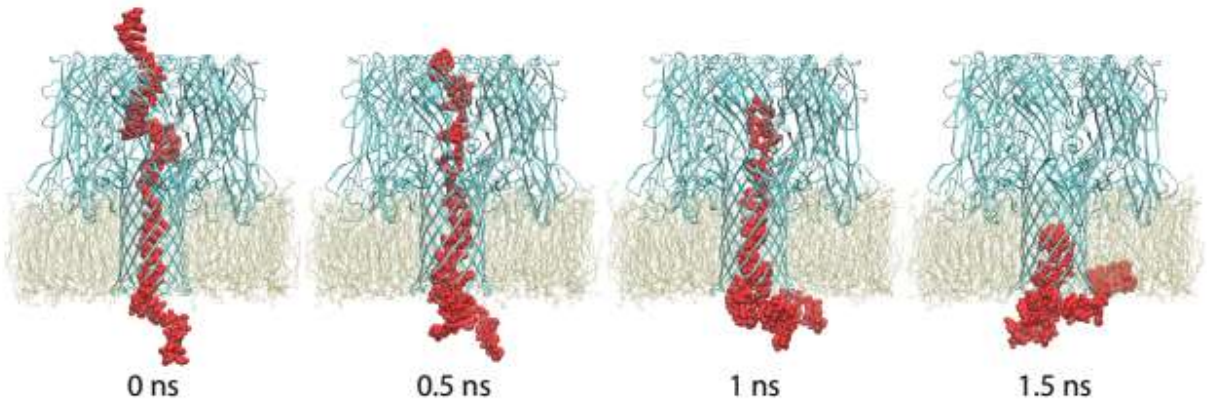

Figure 30: ssDNA translocating through the nanopore [60].

5.4: ANALYSIS OF DNA TRANSLOCATING THROUGH A BIO NANOPORE

In the simulation, DNA successfully translocated through the biological nanopore; however, ionic current calculations have not yet been performed due to key limitations in the setup. A major factor is that the research builds upon work conducted 15 years ago, prior to the advent of GPU computing, which imposed constraints on simulation complexity and resolution. One issue is that the ssDNA was initially placed inside the nanopore, preventing a complete representation of the translocation process. Additionally, the ssDNA adhered to the nanopore walls, further complicating the calculation of ionic currents. Future work will focus on refining the simulation to leverage modern computational advancements and provide a more comprehensive analysis of the translocation process through the α-hemolysin nanopore.

5.5: DSDNA THROUGH AN SI_3N_4 NANOPORE

A simulation of a dsDNA molecule translocating through a Silicon Nitride nanopore was conducted after designing the nanopore. Notable characteristics of the nanopore include its hourglass shape and hexagonal outline, as described in Section 4.6.1: Factors Affecting Translocation Speed, Accuracy, and Resolution. These structural elements mirror the geometry frequently observed in experimental setups.

Simulations were performed both with and without DNA present in the pore to evaluate the difference in ionic current between the two states. An external voltage of 20 V was applied during the simulations, a significantly higher value than typical experimental conditions. This elevated voltage was used to visualize the translocation process efficiently and is reflective of the constraints of the past research, which builds upon work conducted 15 years ago, prior to the advent of GPU computing. The limitations of computational resources at that time necessitated faster DNA translocation to fit within the simulation timescales, leading to the use of such high voltages. These results set the stage for future simulations leveraging modern computational capabilities to achieve more realistic conditions.

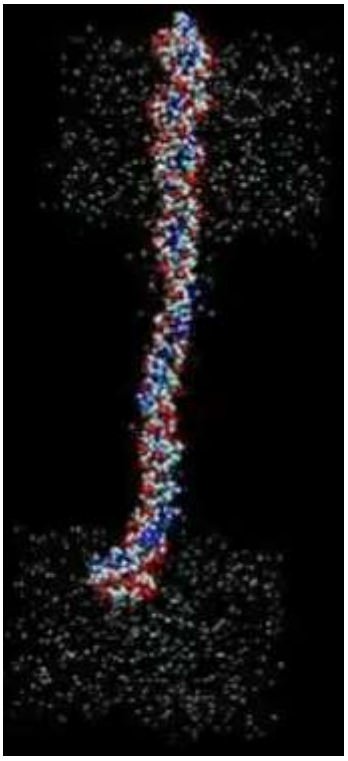

Figure 31: Snapshot of a dsDNA translocating through a Si_3N_4 nanopore. Video can be seen at: https://youtu.be/fuir2Aa9mPQ?si=hSbJdkFiGuvwsTCV

5.6: ANALYSIS OF DSDNA THROUGH AN SI_3N_4 NANOPORE

5.6.1: TIME TAKEN FOR DSDNA TO TRAVEL

Data extracted from this simulation is nucleotides permeated over time, and current(nA) vs time(ns)

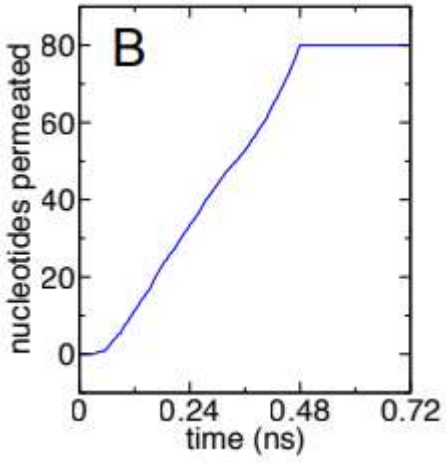

Figure 32: Reveals the time the dsDNA is in the nanopore.

We observe that DNA begins permeating the nanopore at around 0.06 ns and completes its translocation by approximately 0.48 ns. During this period, the number of nucleotides passing through the nanopore steadily

increases, reaching around 80 nucleotides by 0.48 ns, after which the permeation plateaus, indicating that the DNA has fully exited the nanopore.

5.6.2: IONIC CURRENT SIGNATURE

With this graph one can calculate the difference between the open nanopore and the nanopore with the DNA. From the nucleotides permeated vs time we can see that the DNA is translocating from approximately 0.8ns to 0.48ns.

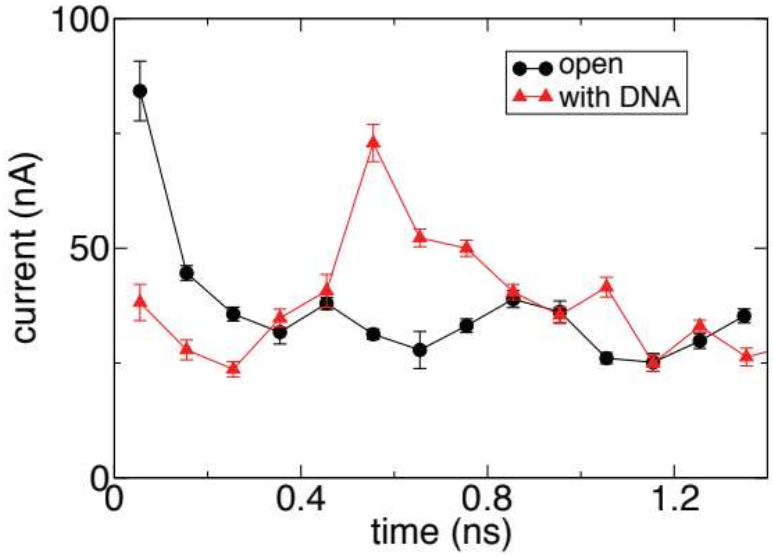

Figure 33: The Difference between the ionic current with and without DNA.

Analyzing the ionic current data during the same time frame (from 0 ns to 0.48 ns) using the current signature graph:

1. At 0 ns: The initial current difference between the open nanopore and the DNA-containing nanopore is approximately 37.5 nA (with DNA) versus 80 nA (open). With a current difference of 42.5 nA.

2. At 0.2 ns: The current difference decreases slightly, with the open nanopore at 40 nA and the DNA-containing pore at 25 nA. This reduction in the difference could be attributed to the DNA progressively occupying the nanopore, creating a partial blockage that alters the ionic flow. With a current difference of 15nA.

3. At 0.4 ns: The current for the open nanopore is around 30 nA, while with DNA it reaches 37.5 nA. This significant increase in the DNA-containing current suggests that the DNA presence is heavily influencing the ionic current, likely due to more nucleotides leaving the nanopore, allowing more ions to enter. With a current difference of -7.5 nA.

4. By 0.48 ns: The translocation process is complete. As the DNA exits the nanopore, a large influx of ions occurs, causing a sharp peak in the current. From this point onward, the current measurements stabilize, reflecting the post-translocation state where the pore is no longer obstructed by DNA.

5.6.3: COMPARISON TO ALTERNATE SIMULATION OF SI_3N_4

An alternative simulation of Si_3N_4 nanopore translocation provides a valuable comparison, showcasing a current versus time graph that includes the open pore current. In this simulation, the current difference between the open pore and the DNA-occupied pore is 7.5 nA, closely aligning with the results of our study. However, significant differences emerge in the translocation time and applied potential. The alternative simulation uses a potential of 1.4 V and achieves a translocation time of 48.9 ns, which is 100 times slower than the 0.48 ns translocation observed in our work. This difference arises because our simulation parameters are based on research conducted 15 years ago, before advancements in GPU computing that now allow for longer and more realistic simulations. Consequently, our experiment employed a higher voltage of 20 V and shorter time scales to effectively visualize the translocation process within the computational constraints of that era.

Additionally, the alternative simulation demonstrates how more up-to-date computational methods and smaller time steps for a longer period of time can more accurately model the dynamics of DNA translocation. However, challenges remain consistent across both studies, such as DNA bases adhering to the nanopore surface, which complicates the translocation process and affects the system's dynamics. This comparison highlights the progress in molecular dynamics simulations, showcasing the evolution of methodologies and computational power that enable more precise and realistic experiments.

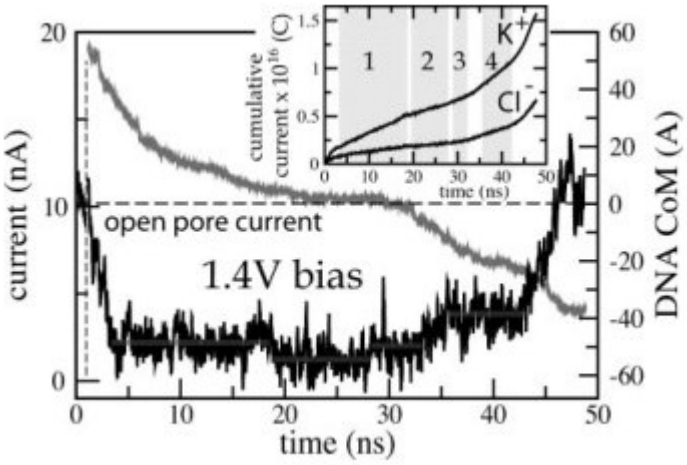

Figure 34: Current vs time graph with open pore current being shown [69].

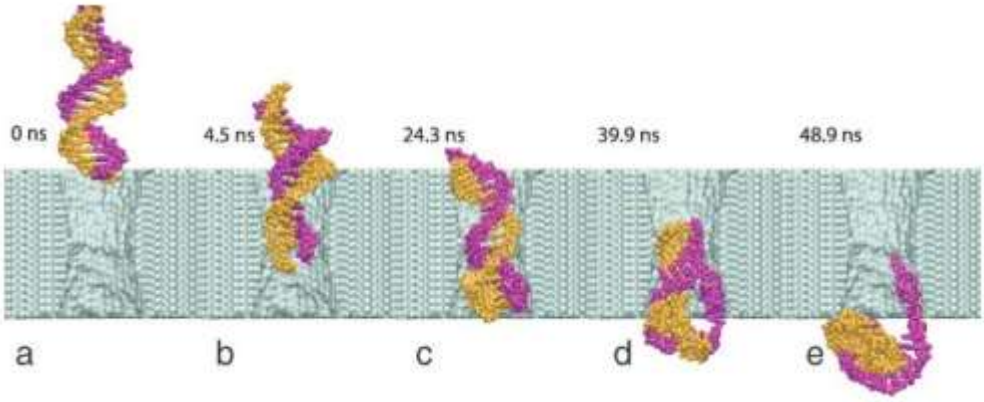

Figure 35: (a-e): Shows the translocation process; (d) specifically depicts one DNA base at the end firmly attached to the nanopore surface [69].

5.6.4: COMPARISON OF SIMULATION RESULTS WITH AVAILABLE EXPERIMENTAL DATA

Below shows experimental data of ssDNA translocating through an α-hemolysin nanopore.

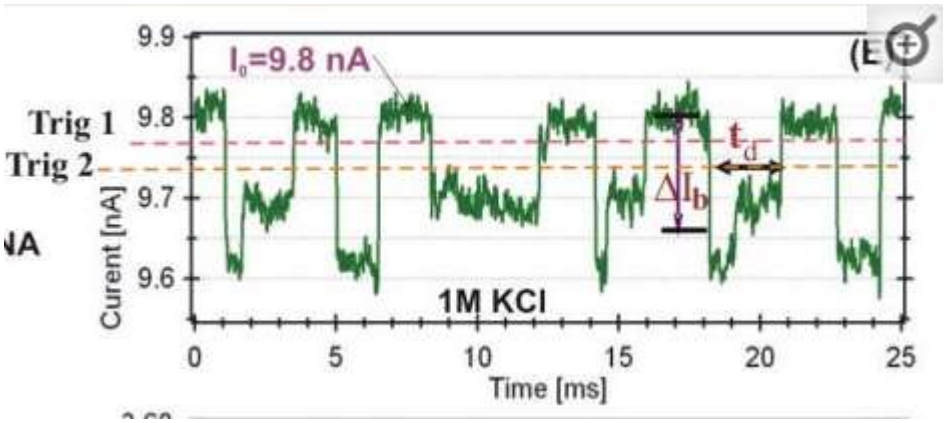

Figure 36: Ionic current vs time [70].

The difference between the first peak and the first minimum is 0.2 nA, a result that differs from the simulation outcomes for several reasons. In the simulation, DNA translocation occurs at a significantly faster speed—within 0.48 ns—due to the high voltage applied and the computational constraints of the system. In contrast, the real-world experiment reports translocation over 110 ± 5 μs, a much longer time frame that allows for more accurate measurements and reduced error.

This stark difference between the simulation and experimental time scales highlights the need for more efficient and advanced simulation methods to better capture the dynamics of the translocation process. Parameters such as pressure, voltage, and solution concentration are adjusted by researchers in experimental setups to observe the translocation process under various conditions and to obtain accurate results. Incorporating similar adjustments and leveraging modern computational advancements in molecular dynamics could enable simulations to more closely mimic real-world experiments, ultimately improving the reliability and applicability of simulation data.

CHAPTER 6 DISCUSSION AND CONCLUSION

6.1: IMPACT OF FINDINGS

This research provides valuable insights into the design and application of nanopore-based DNA sequencing using molecular dynamics (MD) simulations. By comparing DNA translocation through a Silicon Nitride (Si_3N_4) nanopore to other simulations and real-world experiments, we have validated key findings and highlighted the impact of parameters such as applied potential on translocation speed and current signature accuracy.

A significant difference between the simulations and real-world experiments lies in the computational constraints. Simulations required a higher voltage (20 V) and shorter time scales (0.48 ns) due to the limitations of older computational systems, which restricted their ability to replicate millisecond-scale translocations observed experimentally. These constraints underscore the need for more efficient computational methods to bridge the gap between simulations and real-world conditions.

While the findings align with established principles, this study emphasizes the importance of optimizing parameters such as voltage, pore geometry, and solution conditions to achieve more realistic and accurate simulations. Such advancements can deepen our understanding of DNA translocation dynamics and inform the development of improved nanopore designs for DNA sequencing applications.

6.2: PRACTICAL APPLICATIONS AND DESIGN RECOMMENDATIONS

This study provides a proof of concept for the use of molecular dynamics simulations in evaluating and optimizing nanopore-based DNA sequencing systems. The hourglass-shaped Silicon Nitride (Si_3N_4) nanopore demonstrated effective translocation of double-stranded DNA (dsDNA), offering a stable pathway and consistent current readings. Its structural durability and suitability for solid-state applications underline its potential as a reliable material for sequencing systems, laying a solid groundwork for future advancements.

Although no significant advantage was observed for alpha-hemolysin nanopores due to the lack of current vs. time data, this study establishes a framework for further exploration and refinement. The findings emphasize the importance of nanopore geometry and material selection in enhancing sequencing performance. This work sets the stage for future research to build upon, using modern computational tools to simulate more realistic conditions and expand the practical applications of nanopore-based sequencing technologies.

6.3: SIMULATION METHODOLOGY AND CHALLENGES

Simulating DNA translocation presented several challenges, particularly in mastering the setup and execution of complex simulations. Achieving higher accuracy in force fields and implementing smaller time steps could significantly improve the realism of the results. In this study, a high voltage was applied to accelerate DNA translocation, resulting in speeds much faster than those observed experimentally. This approach was necessary to accommodate computational constraints, as lower voltages would have required prohibitively long simulation times given the available resources.

Despite these limitations, this study serves as a proof of concept and highlights the need for optimization in computational setups. Future work could focus on integrating enhanced sampling techniques, as discussed in Section 3.9, to enable simulations at lower voltages and longer time scales, bridging the gap between simulated and experimental conditions. Such advancements would further improve the reliability and applicability of molecular dynamics simulations for nanopore DNA translocation studies.

6.4: FUTURE RESEARCH DIRECTIONS AND PROSPECTS

Future research should focus on optimizing simulation environments to operate at lower, more experimentally relevant potential differences. This would yield translocation times and current signatures that better align with experimental observations. For example, calculating the ionic current vs. time for the alpha-hemolysin nanopore system would allow for direct comparisons with Silicon Nitride (Si_3N_4) nanopores, offering insights into the efficiency and effectiveness of each structure in DNA sequencing. Comparing metrics such as current vs. residence time could also enhance base identification accuracy and provide further validation of simulation results against experimental data.

Looking ahead, molecular dynamics (MD) simulations are poised to play a pivotal role in advancing nanopore technology and biosensing applications. As computational power and simulation techniques improve, MD simulations will provide increasingly detailed insights into optimizing nanopore efficiency, stability, and accuracy. These advancements could enable nanopores to function effectively across diverse environments, enhance base-calling accuracy, and facilitate the sequencing of longer DNA strands. Beyond DNA sequencing, this study lays the groundwork for future research into biomolecular translocation, aiding in the design of biological, synthetic, and solid-state nanopores for applications like protein analysis, drug delivery, and environmental sensing.

By combining refined simulation methodologies with the evolving capabilities of modern computing, future research will continue to bridge the gap between simulated and experimental conditions, unlocking the full potential of nanopore-based technologies.

6.5 CONCLUSION

This study advances the field of DNA sequencing by providing insights into DNA translocation dynamics through nanopores, highlighting the significance of nanopore design and simulation parameters. The findings from molecular dynamics (MD) simulations, combined with comparisons to experimental data, establish a solid foundation for improving nanopore-based DNA sequencing.

By addressing simulation limitations and identifying critical design considerations, this research offers practical guidance for developing more robust, accurate, and versatile nanopore systems. These systems hold the potential to revolutionize applications in genomics, biosensing, and molecular diagnostics, paving the way for faster, more reliable, and cost-effective sequencing technologies.

BIBIOLOGRAPHY

1. Kumawat, Rameshwar L., and Biswarup Pathak. "Identifying Single-Stranded DNA by Tuning the Graphene Nanogap Size: An Ionic Current Approach." The Journal of Physical Chemistry B, vol. 126, no. 6, 2022, pp. 1178-1187. https://doi.org/10.1021/acs.jpcb.1c09266.

2. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics. 2016 Jan;107(1):1-8. doi: 10.1016/j.ygeno.2015.11.003. Epub 2015 Nov 10. PMID: 26554401; PMCID: PMC4727787.

3. Azimi, M., Schmaus, K., Greger, V., Neitzel, D., Rochelle, R. and Dinh, T. (2016), Carrier screening by next-generation sequencing: health benefits and cost effectiveness. Mol Genet Genomic Med, 4: 292-302. https://doi.org/10.1002/mgg3.204

4. Identifying Single-Stranded DNA by Tuning the Graphene Nanogap Size: An Ionic Current Approach Rameshwar L. Kumawat and Biswarup Pathak The Journal of Physical Chemistry B 2022 126 (6), 1178-1187 DOI: 10.1021/acs.jpcb.1c09266

5. Nicholas Stoler, Anton Nekrutenko, Sequencing error profiles of Illumina sequencing instruments, NAR Genomics and Bioinformatics, Volume 3, Issue 1, March 2021, lqab019, https://doi.org/10.1093/nargab/lqab019

6. Jain, M., Olsen, H.E., Paten, B. et al. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17, 239 (2016). https://doi.org/10.1186/s13059-016-1103-0

7. Mlýnský V, Kührová P, Stadlbauer P, Krepl M, Otyepka M, Banáš P, Šponer J. Simple Adjustment of Intranucleotide Base-Phosphate Interaction in the OL3 AMBER Force Field Improves RNA Simulations. J Chem Theory Comput. 2023 Nov 28;19(22):8423-8433. doi: 10.1021/acs.jctc.3c00990. Epub 2023 Nov 9. PMID: 37944118; PMCID: PMC10687871.

8. “Deoxyribonucleic Acid (DNA) Fact Sheet.” Genome.Gov, NIH, www.genome.gov/about-genomics/fact-sheets/Deoxyribonucleic-Acid-Fact-Sheet#:~:text=Deoxyribonucleic%20acid%20(DNA)%20is%20a,to%20their%20offspring%20during%20reproduction. Accessed 14 Oct. 2024.

9. Asma Wasfi, Falah Awwad, Ahmad I. Ayesh, Graphene-based nanopore approaches for DNA sequencing: A literature review, Biosensors and Bioelectronics, Volume 119, 2018, Pages 191-203, ISSN 0956-5663, https://doi.org/10.1016/j.bios.2018.07.072.

10. Totomoch-Serra, Armando et al. “Sanger sequencing as a first-line approach for molecular diagnosis of Andersen-Tawil syndrome.” F1000Research vol. 6 1016. 28 Jun. 2017, doi:10.12688/f1000research.11610.1

11. Satam, Heena et al. “Next-Generation Sequencing Technology: Current Trends and Advancements.” Biology vol. 12,7 997. 13 Jul. 2023, doi:10.3390/biology12070997

12. Satam, Heena et al. “Next-Generation Sequencing Technology: Current Trends and Advancements.” Biology vol. 12,7 997. 13 Jul. 2023, doi:10.3390/biology12070997

13. Udaondo, Zulema et al. “Comparative Analysis of PacBio and Oxford Nanopore Sequencing Technologies for Transcriptomic Landscape Identification of Penaeus monodon.” Life (Basel, Switzerland) vol. 11,8 862. 23 Aug. 2021, doi:10.3390/life11080862

14. Oxford Nanopore Technologies. "Technology." Oxford Nanopore Technologies, nanoporetech.com/platform/technology. Accessed 25 Oct. 2024.

15. “Minion 10 Year Promo Packs.” Store.Nanoporetech.Com, Oxford Nanopore Technologies, store.nanoporetech.com/us/.

16. (Farzin Haque, Jinghong Li, Hai-Chen Wu, Xing-Jie Liang, Peixuan Guo, Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA, Nano Today, Volume 8, Issue 1, 2013, Pages 56-74, ISSN 1748-0132, https://doi.org/10.1016/j.nantod.2012.12.008. (https://www.sciencedirect.com/science/article/pii/S1748013212001454))

17. Hu, Rui et al. “Four Aspects about Solid-State Nanopores for Protein Sensing: Fabrication, Sensitivity, Selectivity, and Durability.” Advanced healthcare materials vol. 9,17 (2020): e2000933. doi:10.1002/adhm.202000933

18. Chen, Qi, and Zewen Liu. “Fabrication and Applications of Solid-State Nanopores.” Sensors (Basel, Switzerland) vol. 19,8 1886. 20 Apr. 2019, doi:10.3390/s19081886

19. Hollingsworth, Scott A., and Ron O. Dror. "Molecular Dynamics Simulation for All." Neuron, vol. 99, no. 6, 2018, pp. 1129-1143. https://doi.org/10.1016/j.neuron.2018.08.011.

20. Tomáš Bučko et al. “Improved Description of the Structure of Molecular and Layered Crystals: Ab Initio DFT Calculations with van der Waals Corrections”. In: The Journal of Physical Chemistry A 114.43 (Oct. 2010), pp. 11814–11824. DOI: 10.1021/jp106469x. URL: https://doi.org/10.1021%2Fjp106469x

21. Qian Mao, Muye Feng, Xi Zhuo Jiang, Yihua Ren, Kai H. Luo, Adri C.T. van Duin, Classical and reactive molecular dynamics: Principles and applications in combustion and energy systems, Progress in Energy and Combustion Science, Volume 97, 2023, 101084, ISSN 0360-1285, https://doi.org/10.1016/j.pecs.2023.101084.

22. Kahn, Kalju. Introduction to Semiempirical Methods. Department of Chemistry and Biochemistry, University of California, Santa Barbara. Available at: https://people.chem.ucsb.edu/kahn/kalju/chem226/public/semiemp_intro.html. Accessed 27 Oct. 2024.

23. Phillips, J. C., et al. NAMD User's Guide, Version 2.9. University of Illinois at Urbana-Champaign, Theoretical and Computational Biophysics Group, 2012. Available at: https://www.ks.uiuc.edu/Research/namd/2.9/ug/node5.html.

24. Computational Chemistry Resources. Molecular Dynamics: Equations of Motion. Available at: https://www.compchems.com/molecular-dynamics-equations-of-motion/#md-algorithm-summary. Accessed 27 Oct. 2024.

25. Spitznagel, B., Pritchett, P. R., Messina, T. C., Goadrich, M., & Rodriguez, J. "An Undergraduate Laboratory Activity on Molecular Dynamics Simulations." Biochemistry and Molecular Biology Education, vol. 44, no. 2, 2016, pp. 130–139. https://doi.org/10.1002/bmb.20939.

26. Phillips, J. C., et al. NAMD: Scalable Molecular Dynamics. University of Illinois at Urbana-Champaign, Theoretical and Computational Biophysics Group, 2024. Available at: https://www.ks.uiuc.edu/Research/namd/. Accessed 27 Oct. 2024.

27. Plimpton, S. J., et al. LAMMPS: Large-scale Atomic/Molecular Massively Parallel Simulator. Sandia National Laboratories, 2024. Available at: https://lammps.sandia.gov/. Accessed 27 Oct. 2024.

28. Case, D. A., et al. AMBER: Assisted Model Building with Energy Refinement. University of California, San Francisco, 2024. Available at: https://ambermd.org/. Accessed 27 Oct. 2024.

29. Abraham, M. J., et al. GROMACS: GROningen MAchine for Chemical Simulations. University of Groningen, 2024. Available at: https://www.gromacs.org/. Accessed 27 Oct. 2024.

30. Humphrey, W., Dalke, A., and Schulten, K. VMD: Visual Molecular Dynamics. University of Illinois at Urbana-Champaign, Theoretical and Computational Biophysics Group, 2024. Available at: https://www.ks.uiuc.edu/Research/vmd/allversions/what_is_vmd.html. Accessed 27 Oct. 2024.

31. Pettersen, E. F., et al. UCSF Chimera – A Visualization System for Exploratory Research and Analysis. University of California, San Francisco, Resource for Biocomputing, Visualization, and Informatics, 2024. Available at: https://www.cgl.ucsf.edu/chimera/. Accessed 27 Oct. 2024.

32. Schrödinger, LLC. PyMOL Molecular Graphics System. 2024. Available at: https://www.pymol.org. Accessed 27 Oct. 2024.

33. Rosignoli, Serena, and Alessandro Paiardini. “Boosting the Full Potential of PyMOL with Structural Biology Plugins.” Biomolecules vol. 12,12 1764. 27 Nov. 2022, doi:10.3390/biom12121764

34. Jo, S., et al. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. CHARMM-GUI Development Team, 2024. Available at: https://www.charmm-gui.org/. Accessed 27 Oct. 2024.

35. Chemical Computing Group ULC. Chemical Computing Group Products. 2024. Available at: https://www.chemcomp.com/en/Products.htm. Accessed 27 Oct. 2024.

36. Hanwell, M. D., et al. Avogadro: An Open-Source Molecular Builder and Visualization Tool. 2024. Available at: https://avogadro.cc/. Accessed 27 Oct. 2024.

37. BIOVIA, Dassault Systèmes. Discovery Studio: Visualization Software. 2024. Available at: https://www.3ds.com/products/biovia/discovery-studio/visualization. Accessed 27 Oct. 2024.

38. Martínez, L et al. “PACKMOL: a package for building initial configurations for molecular dynamics simulations.” Journal of computational chemistry vol. 30,13 (2009): 2157-64. doi:10.1002/jcc.21224

39. Gaussian, Inc. Gaussian 16 Release Notes. 2024. Available at: https://gaussian.com/g16new/. Accessed 27 Oct. 2024.

40. Knapp, B, and W Schreiner. “Graphical user interfaces for molecular dynamics-quo vadis?.” Bioinformatics and biology insights vol. 3 103-7. 23 Sep. 2009

41. Samsonov, Sergey et al. “A molecular dynamics approach to study the importance of solvent in protein interactions.” Proteins vol. 73,2 (2008): 515-25. doi:10.1002/prot.22076

42. Martín-García, Fernando et al. “Comparing molecular dynamics force fields in the essential subspace.” PloS one vol. 10,3 e0121114. 26 Mar. 2015, doi:10.1371/journal.pone.0121114

43. Hénin, J., Lelièvre, T., Shirts, M. R., Valsson, O., & Delemotte, L. "Enhanced Sampling Methods for Molecular Dynamics Simulations." LiveCoMS Perpetual Review, v1.0, 29 Aug. 2022. Available at: https://github.com/jhenin/Methods-for-enhanced-sampling-and-free-energy-calculations.

44. Doltsinis, Nikos L. "Molecular dynamics beyond the Born-Oppenheimer approximation: mixed quantum-classical approaches." NIC Series 31 (2006): 389-409.

45. Bruce, N. J., Ganotra, G. K., Kokh, D. B., Sadiq, S. K., & Wade, R. C. "New Approaches for Computing Ligand–Receptor Binding Kinetics." Current Opinion in Structural Biology, vol. 49, 2018, pp. 1–10. https://doi.org/10.1016/j.sbi.2017.10.001.

46. Wong, C. F. "Flexible Ligand–Flexible Protein Docking in Protein Kinase Systems." Biochimica et Biophysica Acta, vol. 1784, no. 2, 2008, pp. 244–251. https://doi.org/10.1016/j.bbapap.2007.10.005.

47. Qi, Ruxi et al. “Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example.” Methods in molecular biology (Clifton, N.J.) vol. 1777 (2018): 101-119. doi:10.1007/978-1-4939-7811-3_5

48. Center for Computational Biophysics, University of Illinois at Urbana-Champaign. Coarse-Grained (CG) Modeling Resources. 2024. Available at: https://www.ks.uiuc.edu/Research/CG/. Accessed 27 Oct. 2024.

49. Teo, I., Perilla, J., Shahoei, R., McGreevy, R., & Harrison, C. GPU Accelerated Molecular Dynamics Simulation, Visualization, and Analysis. University of Illinois at Urbana-Champaign, Beckman Institute for Advanced Science and Technology, Theoretical and Computational Biophysics Group, 19 May 2014. Available at: www.ks.uiuc.edu/Training/Tutorials/.

50. Navarro CA, Hitschfeld-Kahler N, Mateu L. A Survey on Parallel Computing and its Applications in Data-Parallel Problems Using GPU Architectures. Communications in Computational Physics. 2014;15(2):285-329. doi:10.4208/cicp.110113.010813a

51. Introduction to Machine Learning for Molecular Simulation Nicholas E. Jackson, Brett M. Savoie, Antonia Statt, and Michael A. Webb Journal of Chemical Theory and Computation 2023 19 (14), 4335-4337 DOI: 10.1021/acs.jctc.3c00735

52. Brault, Pascal. “Multiscale molecular dynamics simulation of plasma processing: Application to plasma sputtering.” Frontiers in Physics, vol. 6, 13 June 2018, https://doi.org/10.3389/fphy.2018.00059.

53. Stocker, U., Juchli, D., & van Gunsteren, W. F. (2003). Increasing the Time Step and Efficiency of Molecular Dynamics Simulations: Optimal Solutions for Equilibrium Simulations or Structure

Refinement of Large Biomolecules. Molecular Simulation, 29(2), 123–138. https://doi.org/10.1080/0892702031000065791

54. Aksimentiev, A., & Schulten, K. (2005). Imaging α-Hemolysin with molecular dynamics: Ionic conductance, osmotic permeability, and the electrostatic potential map. Biophysical Journal, 88(6), 3745-3761. https://doi.org/10.1529/biophysj.104.058727

55. Di Marino, D., Bonome, E. L., Tramontano, A., & Chinappi, M. (2015). All-atom molecular dynamics simulation of protein translocation through an α-hemolysin nanopore. The Journal of Physical Chemistry Letters, 6(15), 2963-2968. https://doi.org/10.1021/acs.jpclett.5b01077

56. Heerema, S., Dekker, C. Graphene nanodevices for DNA sequencing. Nature Nanotech 11, 127–136 (2016). https://doi.org/10.1038/nnano.2015.307

57. Spitznagel, B., Pritchett, P. R., Messina, T. C., Goadrich, M., & Rodriguez, J. (2016). An undergraduate laboratory activity on molecular dynamics simulations. Biochemistry and Molecular Biology Education, 44(2), 130–139. https://doi.org/10.1002/bmb.20939

58. Axel Kohlmeyer & Josh Vermaas, (2022). TopoTools: Release 1.9 https://doi.org/10.5281/zenodo.598373

59. Yadav, Prakarsh, Zhonglin Cao, and Amir Barati Farimani. "DNA Detection with Single-Layer Ti_3C_2 MXene Nanopore." ACS Nano, vol. 15, no. 3, 2021, pp. 4861-4869. https://doi.org/10.1021/acsnano.0c09595.

60. Comer, Jeffrey R et al. “Modeling nanopores for sequencing DNA.” Methods in molecular biology (Clifton, N.J.) vol. 749 (2011): 317-58. doi:10.1007/978-1-61779-142-0_22

61. Valle, João V. L., et al. "Accuracy of TIP4P/2005 and SPC/Fw Water Models." The Journal of Physical Chemistry B, vol. 128, no. 4, American Chemical Society, 2024, pp. 1091-1097. J. Phys. Chem. B, https://doi.org/10.1021/acs.jpcb.3c07044.

62. Phillips, James C et al. “Scalable molecular dynamics with NAMD.” Journal of computational chemistry vol. 26,16 (2005): 1781-802. doi:10.1002/jcc.20289

63. Kaszab, Eszter, et al. "Metagenomics Revealing New Virus Species in Farm and Pet Animals and Aquaculture." Genomics and Biotechnological Advances in Veterinary, Poultry, and Fisheries, edited by Yashpal Singh Malik, Debmalya Barh, Vasco Azevedo, and S. M. Paul Khurana, Academic Press, 2020, pp. 29-73. https://doi.org/10.1016/B978-0-12-816352-8.00002-3.

64. Murphy, M. C., I. Rasnik, ., T. Ha. 2004. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. Biophys. J. 86:2530–2537.

65. Shi, Liuyong et al. “The Influence of Electric Field Intensity and Particle Length on the Electrokinetic Transport of Cylindrical Particles Passing through Nanopore.” Micromachines vol. 11,8 722. 25 Jul. 2020, doi:10.3390/mi11080722

66. Chuang, Po-Yen, and Jyh-Ping Hsu. "Influence of Shape and Charged Conditions of Nanopores on Their Ionic Current Rectification, Electroosmotic Flow, and Selectivity." Colloids and Surfaces A:

Physicochemical and Engineering Aspects, vol. 658, 2023, article 130696. https://doi.org/10.1016/j.colsurfa.2022.130696.

67. Nathan Schmid, Andreas P Eichenberger, Alexandra Choutko, Sereina Riniker, Moritz Winger, Alan E Mark, and Wilfred F van Gunsteren. Definition and testing of the gromos force-field versions 54a7 and 54b7. European biophysics journal, 40:843–856, 2011.

68. Wells, D. B., Abramkina, V., and Aksimentiev, A. (2007) Exploring transmembrane transport through α-hemolysin with grid-steered molecular dynamics J Chem Phys 127, 125101.

69. Microscopic Kinetics of DNA Translocation through Synthetic Nanopores Aksimentiev, Aleksij et al. Biophysical Journal, Volume 87, Issue 3, 2086 – 2097

70. Uplinger, James et al. “K(+) , Na(+) , and Mg(2+) on DNA translocation in silicon nitride nanopores.” Electrophoresis vol. 33,23 (2012): 3448-57. doi:10.1002/elps.201200165

APPENDICES

APPENDIX A

```
#!/bin/bash
#SBATCH --job-name=lammps
#SBATCH --output=lammps-%j.out
#SBATCH --error=lammps-%j.err
#SBATCH --partition=general
#SBATCH --nodes=2
#SBATCH --ntasks=2
#SBATCH --ntasks-per-node=1
#SBATCH --cpus-per-task=20

#Load LAMMP module
module load LAMMPS/23Jun2022-foss-2021b-kokkos

# 2 nodes, 1 MPI task/node, 20 threads/task
mpirun -np 2 -ppn 1 lmp_kokkos_omp -k on t 20 -sf kk -in in.lj

```

Figure 37: SAMPLE BATCH SCRIPT FOR LAMMPS ON BUDDY

```
#!/bin/bash
#SBATCH --partition=sooner_test
#SBATCH --time=01:00:00 # up to 48 hours
#SBATCH --container=el9-devel
#SBATCH --nodes=1
#SBATCH --ntasks=20
#SBATCH --ntasks-per-node=20
#SBATCH --mem=10G
#SBATCH --chdir=/home/mwatson2/soonertest9_19
#SBATCH --output=/home/mwatson2/soonertest9_19/jobname_%J_stdout.txt
#SBATCH --error=/home/mwatson2/soonertest9_19/jobname_%J_stderr.txt
#SBATCH --job-name=container_test
#SBATCH --mail-user=mwatson2@uco.edu
#SBATCH --mail-type=ALL
#
################################################
module load NAMD/3.0-20240614-foss-2022a-mpi
mpirun namd3 min.namd
```

Figure 38: SAMPLE BATCH SCRIPT FOR NAMD ON SCHOONER SUPERCOMPUTER. Sbatch sample file running min.namd

APPENDIX B

Simulation of water box in LAMMPS

```
units real #masses are in grams per mole, distances in Ã…ngstroms, time in femtoseconds, energies in Kcal/mole.
atom_style full #each atom is a dot with a mass and a charge that can be linked by bonds, angles, dihedrals and/or impropers
#bond_style, angle_style, and dihedral_style commands define the potentials for the bonds, angles, and dihedrals used in the simulation, here harmonic.
bond_style harmonic
angle_style harmonic
dihedral_style harmonic
pair_style lj/cut/coul/long 12 #atoms interact through both a Lennard-Jones (LJ) potential and Coulombic interactions. The value of 12Ã… is the cutoff.
kspace_style pppm 1e-5 #defines the long-range solver for the (long) Coulombic interactions. The pppm style refers to particle-particle particle-mesh
special_bonds lj 0.0 0.0 0.5 coul 0.0 0.0 1.0 angle yes  #cancels the Lennard-Jones interactions between the closest atoms of the same molecule.

#create a 3D simulation box of dimensions 9x3x3nm^3, and make space for 9 atom types (2 for the water molecule, and 7 for the polymer molecule)
#7 bond types, 8 angle types, and 4 dihedral types.
region box block -45 45 -15 15 -15 15
create_box 9 box &
bond/types 7 &
angle/types 8 &
dihedral/types 4 &
extra/bond/per/atom 3 &
extra/angle/per/atom 6 &
extra/dihedral/per/atom 10 &
extra/special/per/atom 14
#extra/x/per/atom commands are here for memory allocation.
#keep that in mind if one day you see the following error message ERROR: Molecule topology/atom exceeds system topology/atom.

#file contains all the parameters (masses, interaction energies, bond equilibrium distances, etc). types 8 and 9 are for water and the atoms of types 1 to 7 are for the polymer molecule
include PARM.lammps

#create water molecules
molecule h2omol H2O-SPCFw.mol #molecule template
```

Figure 39: First part of water box in LAMMPS code.

```
molecule h2omol H2O-SPCFw.mol #molecule template
create_atoms 0 random 1050 87910 NULL mol h2omol 454756 overlap 1.0 maxtry 50 #maxtry 50 asks LAMMPS to try at most 50 times to insert the molecules


group H2O type 8 9 #organize the atoms of types 8 and 9 of the water molecules in a group named H2O
minimize 1.0e-4 1.0e-6 100 1000 #energy minimization is mandatory here given the small overlap value of 1 Ã…ngstrom chosen in the create_atoms command.
reset_timestep 0    #reset_timestep command is optional. It is used here because the minimize command is usually performed over an arbitrary number of steps

#Allows us to impose both a temperature of 300K(with a damping constant of 100fs), and a pressure of 1 atmosphere (with a damping constant of 1000fs).
#With the iso keyword, the three dimensions of the box will be re-scaled simultaneously.

fix mynpt all npt temp 300 300 100 iso 1 1 1000 #control the temperature of the molecules with a NosÃ©-Hoover thermostat and the pressure of the system with a NosÃ©-Hoover barostat [29, 30, 31]

#print the atom positions in a .lammpstrj file every 1000 steps (i.e. 1 ps), print the temperature volume,
#and density every 100 steps in 3 separate data files, and print the information in the terminal every 1000 steps
dump mydmp all atom 1000 dump.lammpstrj #print the atom positions in a .lammpstrj file every 1000 steps
variable mytemp equal temp
variable myvol equal vol
fix myat1 all ave/time 10 10 100 v_mytemp file temperature.dat #print the temperature
fix myat2 all ave/time 10 10 100 v_myvol file volume.dat #print the volume
variable myoxy equal count(H2O)/3 #corresponds to the number of atoms divided by 3, i.e. the number of molecules.
variable mydensity equal ${myoxy}/v_myvol
fix myat3 all ave/time 10 10 100 v_mydensity file density.dat #print the density
thermo 1000 #print the information in the terminal every 1000 steps:

#set the timestep to 1.0 fs, and run the simulation for 20 ps
timestep 1.0
run 30000

write_data H2O.data
```

Figure 40: Second part of water box in LAMMPS code.

SIMULATION OF PEG MOLECULE SOLVATED IN WATER IN LAMMPS

```
units real #masses are in grams per mole, distances in Ã…ngstroms, time in femtoseconds, energies in Kcal/mole.
atom_style full #each atom is a dot with a mass and a charge that can be linked by bonds, angles, dihedrals and/or impropers
#bond_style, angle_style, and dihedral_style commands define the potentials for the bonds, angles, and dihedrals used in the simulation, here harmonic.
bond_style harmonic
angle_style harmonic
dihedral_style harmonic
pair_style lj/cut/coul/long 12 #atoms interact through both a Lennard-Jones (LJ) potential and Coulombic interactions. The value of 12Ã… is the cutoff.
kspace_style pppm 1e-5 #defines the long-range solver for the (long) Coulombic interactions. The pppm style refers to particle-particle particle-mesh
special_bonds lj 0.0 0.0 0.5 coul 0.0 0.0 1.0 angle yes  #cancels the Lennard-Jones interactions between the closest atoms of the same molecule.

#create a 3D simulation box of dimensions 9x3x3nm^3, and make space for 9 atom types (2 for the water molecule, and 7 for the polymer molecule)
#7 bond types, 8 angle types, and 4 dihedral types.
region box block -45 45 -15 15 -15 15
create_box 9 box &
bond/types 7 &
angle/types 8 &
dihedral/types 4 &
extra/bond/per/atom 3 &
extra/angle/per/atom 6 &
extra/dihedral/per/atom 10 &
extra/special/per/atom 14
#extra/x/per/atom commands are here for memory allocation.
#keep that in mind if one day you see the following error message ERROR: Molecule topology/atom exceeds system topology/atom.

#file contains all the parameters (masses, interaction energies, bond equilibrium distances, etc). types 8 and 9 are for water and the atoms of types 1 to 7 are for the polymer molecule
include PARM.lammps

#create water molecules
molecule h2omol H2O-SPCFw.mol #molecule template
```

Figure 41: First part of solvating PEG molecule code.

```
molecule h2omol H2O-SPCFw.mol #molecule template
create_atoms 0 random 1050 87910 NULL mol h2omol 454756 overlap 1.0 maxtry 50 #maxtry 50 asks LAMMPS to try at most 50 times to insert the molecules


group H2O type 8 9 #organize the atoms of types 8 and 9 of the water molecules in a group named H2O
minimize 1.0e-4 1.0e-6 100 1000 #energy minimization is mandatory here given the small overlap value of 1 Ã…ngstrom chosen in the create_atoms command.
reset_timestep 0    #reset_timestep command is optional. It is used here because the minimize command is usually performed over an arbitrary number of steps

#Allows us to impose both a temperature of 300K(with a damping constant of 100fs), and a pressure of 1 atmosphere (with a damping constant of 1000fs).
#With the iso keyword, the three dimensions of the box will be re-scaled simultaneously.

fix mynpt all npt temp 300 300 100 iso 1 1 1000 #control the temperature of the molecules with a NosÃ©-Hoover thermostat and the pressure of the system with a NosÃ©-Hoover barostat [29, 30, 31]

#print the atom positions in a .lammpstrj file every 1000 steps (i.e. 1 ps), print the temperature volume,
#and density every 100 steps in 3 separate data files, and print the information in the terminal every 1000 steps
dump mydmp all atom 1000 dump.lammpstrj #print the atom positions in a .lammpstrj file every 1000 steps
variable mytemp equal temp
variable myvol equal vol
fix myat1 all ave/time 10 10 100 v_mytemp file temperature.dat #print the temperature
fix myat2 all ave/time 10 10 100 v_myvol file volume.dat #print the volume
variable myoxy equal count(H2O)/3 #corresponds to the number of atoms divided by 3, i.e. the number of molecules.
variable mydensity equal ${myoxy}/v_myvol
fix myat3 all ave/time 10 10 100 v_mydensity file density.dat #print the density
thermo 1000 #print the information in the terminal every 1000 steps:

#set the timestep to 1.0 fs, and run the simulation for 20 ps
timestep 1.0
run 30000

write_data H2O.data
```

Figure 42: Second part of solvating PEG molecule code.

APPENDIX C

MINIMIZATION, HEATING, AND EQUILIBRATION OF DNA SYSTEM IN NAMD

```
numsteps                    1000
structure                   ssdna+ions.psf
coordinates                 ssdna+ions.pdb
outputName                  min

cellBasisVector1        60. 0.  0.
cellBasisVector2        0.  60. 0.
cellBasisVector3        0.  0.  180.
cellOrigin          0.  0.  20.

minimization            on

set temp                    288
temperature         $temp

langevin                    off
langevinDamping             1.0
langevinTemp                $temp
langevinHydrogen            off

outputEnergies              100
outputTiming                100
xstFreq                     100
dcdFreq                     100
restartFreq                 100

### COMMON ###
switching                   on
switchDist                  10
cutoff                      12
pairlistdist                14

binaryOutput                yes
binaryRestart               yes
wrapAll             yes
wrapNearest         yes
comMotion                   yes

paraTypeCharmm              on

parameters                  /c32b1/toppar/par_all27_na.prm
parameters                  /c32b1/toppar/par_all27_prot_lipid.prm

timestep                    1
nonBondedFreq               2
fullElectFrequency          4
stepsPerCycle               20

exclude                     scaled1-4
1-4scaling                  1
```

Figure 43: Code used to minimize ssDNA.

```
numsteps                    10000
structure                   ssdna+ions.psf
coordinates                 ssdna+ions.pdb
outputName                  heat

set last            min
bincoordinates              ${last}.coor
binvelocities               ${last}.vel
set xsc                     ${last}.xsc
extendedSystem              $xsc

set temp                    295

langevin                    on
langevinDamping             5.0
langevinTemp                $temp
langevinHydrogen            off

outputEnergies              1000
outputTiming                1000
xstFreq                     1000
dcdFreq                     1000
restartFreq                 1000

Pme                         on
PmeGridSizeX                48
PmeGridSizeY                48
PmeGridSizeZ                200

### COMMON ###
switching                   on
switchDist                  10
cutoff                      12
pairlistdist                14

binaryOutput                yes
binaryRestart               yes
wrapAll             yes
wrapNearest         yes
comMotion                   yes

paraTypeCharmm              on

parameters                  ../c32b1/toppar/par_all27_na.prm
parameters                  ../c32b1/toppar/par_all27_prot_lipid.prm

timestep                    1
nonBondedFreq               2
fullElectFrequency          4
stepsPerCycle               20

exclude                     scaled1-4
1-4scaling                  1

```

Figure 44: Code used to heat system.

```
numsteps                        100000
structure                       ssdna+ions.psf
coordinates                     ssdna+ions.pdb
outputName                      eq
set last            heat
bincoordinates                  ${last}.coor
binvelocities                   ${last}.vel
set xsc                         ${last}.xsc
extendedSystem                  $xsc
set temp                        295
langevin                        on
langevinDamping                 1.0
langevinTemp                    $temp
langevinHydrogen                off
langevinPiston                  on
langevinPistonTarget            1.01325
langevinPistonTemp              $temp
langevinPistonPeriod            200
langevinPistonDecay             100
useGroupPressure                yes
margin              3
outputEnergies                  1000
outputTiming                    1000
xstFreq                         1000
dcdFreq                         1000
restartFreq                     1000
Pme                             on
PmeGridSizeX                    48
PmeGridSizeY                    48
PmeGridSizeZ                    200
### COMMON ###
switching                       on
switchDist                      10
cutoff                          12
pairlistdist                    14
binaryOutput                    yes
binaryRestart                   yes
wrapAll             yes
wrapNearest         yes
comMotion                       yes
paraTypeCharmm                  on
parameters                      ../c32b1/toppar/par_all27_na.prm
parameters                      ../c32b1/toppar/par_all27_prot_lipid.prm
timestep                        1
nonBondedFreq                   2
fullElectFrequency              4
stepsPerCycle                   20
exclude                         scaled1-4
1-4scaling                      1

```

Figure 45: Code used to equilibrate system.

APPENDIX D

SCHOLARLY PRESENTATIONS

Micah Watson, Fernando Salazar, "Using the TOPOTOOLS in VMD to Generate a Water Box for Molecular Dynamics Simulation". *2024 Oklahoma Research Day,* University of Central Oklahoma, Oklahoma, March 2024. Poster Presentation.

Micah Watson, Fernando Salazar, Kevin Knopp, Jeffery Moore, "Electromechanical Linear Translation of DNA through Nanopores Membranes Vs. ssDNA Translocation through Different Nanopore Sheets". *2023 Oklahoma Academy of Science Fall Technical Meeting*, University of Science and Arts, Chickasha, Oklahoma, November 2023. Poster Presentation.

APPENDIX E

NUCLEOTIDES PERMEATED VS TIME(NANOSECOND) FOR SILICON NITRATE

Time(ns)	Nucleotides permeated
0	0
0.005	0
0.01	0
0.015	0
0.02	0
0.025	0
0.03	0.061
0.035	0.275
0.04	0.486
0.045	0.636
0.05	0.790
0.055	0.819
0.06	1.629
0.065	2.157
0.07	3.106
0.075	3.806
0.08	4.340
0.085	5.190
0.09	5.433
0.095	6.829
0.1	7.777
0.105	8.438
0.11	9.500
0.115	10.313
0.12	11.375

IONIC CURRENT WITH DNA IN SILICON NITRATE SYSTEM

Time(ns)	current due to K+(nA)	current due to Cl−(nA)

0.055	38.192	3.953
0.155	27.882	2.159
0.255	23.647	1.654
0.355	34.781	1.970
0.455	40.731	3.562
0.555	72.896	4.045
0.655	52.248	1.932
0.755	49.988	1.776
0.855	40.468	1.669
0.955	35.424	1.769
1.055	41.525	2.145
1.155	24.959	1.803
1.255	33.035	1.275
1.355	26.324	1.934
1.455	28.880	1.492
1.555	29.645	2.068
1.655	33.079	1.660
1.755	26.402	1.715
1.855	26.954	1.653
1.955	30.616	1.594
2.055	29.431	1.121
2.155	24.325	1.371
2.255	32.351	1.120
2.355	23.945	1.188
2.455	21.350	1.136
2.555	23.970	1.800
2.655	25.836	1.395
2.755	25.990	1.874
2.855	33.108	2.410
2.955	26.830	1.352
3.055	29.020	1.283
3.155	25.009	1.206
3.255	27.664	1.197
3.355	27.832	1.497
3.455	30.227	1.226

IONIC CURRENT WITH OPEN SILICON NITRATE SYSTEM

Time(ns)	current due to K+(nA)	current due to Cl−(nA)
0.055	84.233	6.495
0.155	44.596	1.607

0.255	35.661	1.515
0.355	31.722	2.575
0.455	37.972	1.339
0.555	31.263	1.117
0.655	27.843	4.035
0.755	33.127	1.479
0.855	38.903	1.821
0.955	36.140	2.376
1.055	26.060	1.197
1.155	25.140	1.937
1.255	29.814	1.652
1.35	35.257	1.528