

Master Thesis

Computational Modelling of Ion Conduction through ORF3a Channel

Author: Atena Salek

1st supervisor: Prof. M. Noelia Faginas-Lago
2nd supervisor: Prof. Andrea Lombardi
3rd supervisor : Dr. Juan José Nogueira

July 9, 2023



UNIVERSITÀ DEGLI STUDI DI PERUGIA

DIPARTIMENTO DI CHIMICA, BIOLOGIA E BIOTECNOLOGIE

CORSO DI LAUREA MAGISTRALE IN SCIENZE CHIMICHE

TESI DI LAUREA

**Modellazione computazionale della conduzione ionica
attraverso il canale ORF3a**

Laureando: Atena Salek

Relatori:

Prof. M. Noelia Faginas-Lago

Correlatori:

Dr. Juan José Nogueira

Prof. Andrea Lombardi

ANNO ACCADEMICO 2021/2023

Abstract

In the midst of the ongoing Covid-19 pandemic, the emergence of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a global health crisis. One of the key viral components, known as ORF3a, has garnered significant interest due to its intriguing characteristics as an ion channel and transmembrane protein. Viroporins, a class of proteins that alter ion channel functions, including membrane permeability and remodeling, have been associated with various stages of viral life cycles, such as virion morphogenesis, viral particle replication, and virus release.

In this study, we investigate the ion channel activity of the SARS-CoV-2 ORF3a protein through the application of molecular dynamics simulations. Our experimental study unveiled the ability to observe the migration of ions during their passage through the channel. Therefore, classical molecular dynamics simulations of ORF3a are employed to elucidate the mechanism by which ions traverse the channel. We uncover that ion passage is facilitated by a solvation shell, which aids in the migration of ions into the channel. Notably, the transport of ions is mediated by a water channel spanning across the protein.

Taking our investigation further, we aim to identify the specific residues within ORF3a that are involved in interactions with ions, thus facilitating their flow into the channel. By elucidating these critical molecular interactions, we can gain deeper insights into the underlying mechanisms of ion channel activity in the context of SARS-CoV-2. Our findings contribute to the understanding of the vital role played by ORF3a in the viral life cycle, potentially opening doors for the development of targeted therapeutic strategies aimed at disrupting this important viral protein.

SUMMARY

L'emergere della grave sindrome respiratoria acuta Coronavirus 2 (SARS-CoV-2), e la sua evoluzione in pandemia di Covid-19, hanno causato una crisi sanitaria globale. Uno dei componenti virali chiave, noto come ORF3a, è stato oggetto di notevole interesse per le sue interessanti caratteristiche di potenziale canale ionico e proteina trans-membrana. Le viroporine, quali ORF3a, una classe di proteine che alterano le funzioni dei canali ionici, quindi la permeabilità e il rimodellamento della membrana cellulare, sono state associate a varie fasi dei cicli di vita virali, come la morfogenesi del virione, la replicazione virale delle particelle e rilascio di virus. In questo lavoro di tesi, è stata studiata l'attività del canale ionico del SARS-CoV-2, la proteina ORF3a, attraverso l'applicazione di simulazioni di dinamica molecolare. Le nostre osservazioni hanno rivelato la presenza di canali cationici non selettivi all'interno della membrana della cellula, veri e propri pori macromolecolari che facilitano il rapido movimento degli ioni. Per convalidare le nostre scoperte, sono state effettuate simulazioni di dinamica molecolare classica del sistema ORF3a-membrana, al fine di chiarire il meccanismo mediante il quale gli ioni attraversano il canale. Abbiamo dimostrato che il passaggio degli ioni è facilitato da un guscio di solvatazione, che aiuta nella migrazione degli ioni nel canale. In particolare, il trasporto di ioni è mediato da un canale d'acqua che attraversa la tutta membrana cellulare. Sviluppi ulteriori dello studio mirano a identificare i residui specifici all'interno di ORF3a che sono direttamente coinvolti nelle interazioni con gli ioni, facilitandone così il flusso nel canale. Chiarire queste interazioni molecolari critiche, permetterà di ottenere informazioni più approfondite sui meccanismi alla base dell'attività dei canali ionici nel contesto dell'interazione del SARS-CoV-2 con le cellule. I risultati ottenuti contribuiscono alla comprensione di il ruolo cruciale svolto da ORF3a nel ciclo di vita virale, facilitando lo sviluppo di strategie terapeutiche mirate, volte a bloccare questa importante proteina virale.

Contents

List of Figures	iii
List of Tables	v
1 Introduction	1
1.1 Covid-19	1
2 Theoretical Background	5
2.1 Statistical Averages and the Ergodic Hypothesis	5
2.2 Exploring the Equations of Motion	6
2.3 Relationship Between Temperature, Pressure, and the Virial Equation of State	7
2.3.1 Kinetic Energy-Temperature Correlation:	8
2.3.2 Virial-Pressure Relationship:	9
2.4 Stages of the MD Simulation Process	10
2.4.1 Velocity-Verlet Algorithm	10
2.5 Periodic Boundary Conditions	12
2.6 Temperature Regulation in the Isothermal Ensemble	13
2.7 Control of Pressure in the Isothermal-Isobaric Ensemble	14
2.8 Force Field	15
2.8.1 Bonded Intractions	15
2.8.2 Non-Bonded Interactions	16
3 Computational Details	19
3.1 Design and construction of model	19
3.2 Steps of Molecular Dynamics Simulations	21
3.2.1 Equilibration Runs	22
3.2.2 Production Runs	23

CONTENTS

4	Discussion and Results	25
4.1	System Model	25
4.2	Equilibration Analysis	26
4.3	Classical Molecular Dynamics Production - RMSD	28
4.4	Analysis of Ion Conduction Patterns: Insights from Graphical Representation	30
4.5	Exploring Interactions of Positive and Negative Residues with Ions	34
4.6	Water Solvation along Conduction	38
5	Conclusion	43
	References	45

List of Figures

1.1	SARS-CoV-2 virus structure	2
1.2	The dimeric ORF3a protein (PDB ID: 6XDC). The six transmembrane helices form ion channels in the host membrane.	3
2.1	Van der Waals, repulsive, and Lennard-Jones potential energy functions depict with ε and σ . The term ε denotes the potential depth and σ refers to the distance at which the energy is zero.	18
3.1	The components of the simulation box. protein is represented in blue and red (cartoon), lipid bilayer in cyan (line), chloride ion in cyan, sodium in yellow, potassium in brown, and calcium in orange (sphere). The positive z-axis is shown in the blue arrow.	21
3.2	The segment of the NAMD input file specifies the application of the external electric field.	23
4.1	Simulation system depicts the bilayer membrane and the directional external electric field from intracellular to extracellular Domains.	26
4.2	(a): Exploring various forms of energies (including total Energy, kinetic energy, and potential energy) during the processes of minimization, heating, and equilibration. (b): Analyzing temperature variations during heating. (c): Investigating the density increase during the equilibration phase. . . .	28
4.3	(a): RMSD of the aligned channel with respect to its initial position. (b): RMSD of aligned resid 61 to 66 with respect to its initial position.	29
4.4	Representation of the C_α in a protein structure: with a focus on residues histidine (HSD) and serine (SER)	30
4.5	Permeability profiles: migration of cations and anions across the intracellular ($z > 40$) to extracellular ($z < 40$) regions.	32

LIST OF FIGURES

4.6	Spatial arrangement of key amino acid residues (Arg, His, Lys, His, Glu, Asp) in the ion channel.	35
4.7	Exploring ion interactions with key amino acid residues: Arg, His, Lys, Glu, Asp residues within the ion channel.	37
4.8	Temporal variation of water molecule counts.	38
4.9	Spatial distribution of water molecules in the X-Z plane.	40
5.1	Scheme of NAMD program	55
5.2	Input parameter: part (1)	56
5.3	Input parameter: part (2)	57
5.4	Input parameter: part (2)	59

List of Tables

3.1 System info	20
---------------------------	----

Abbreviations

MD : Molecular Dynamics

CMD : Classical Molecular Dynamics

RMSD : Root Mean Square Deviation

CHARMM : Chemistry at Harvard Macromolecular Mechanics

AMBER : Assisted Model Binding with Energy Refinement

NAMD : Nanoscale Molecular Dynamic

OPM : Operations of Protein in Membranes

LIST OF TABLES

1

Introduction

1.1 Covid-19

The Coronavirus Disease 2019 (COVID-19) is a respiratory illness that spreads easily and is caused by a recently discovered coronavirus known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). (1, 2) There have been three outbreaks of coronavirus since 2002, including those caused by severe acute respiratory syndrome coronavirus (SARS-CoV), detected in 2002, and Middle East respiratory syndrome coronavirus (MERS-CoV), detected in 2012. During the year 2002, 8,098 people became infected with the SARS-CoV virus, and 774 of them died due to this illness after the virus had spread via air travel routes from the Guangdong province in China.(3)The MERS-CoV virus spread throughout the Arabian Peninsula, leading to a large number of public health threats in the region, and was exported to 27 countries, with over 2,494 people infected and 858 deaths registered. It was discovered in December 2019 that a previously unknown coronavirus, called SARS-CoV-2, had been isolated and sequenced from Wuhan, Hubei province of China. According to the World Health Organization (WHO), the disease has been confirmed in more than 700 million cases throughout the world, leading to more than 6 million deaths.(3, 4, 5) In a recent study, an emerging pathogen was identified as a variant of the betacoronavirus genus which is related closely to several bat coronaviruses. Since Sars-Cov-2 appeared to be spreading rapidly from human to human, the WHO declared a Public Health Emergency of International Concern (PHEIC) in 2020 as a result of its rapid spreading across multiple continents.(6, 7)

Virion surfaces are protruded by a transmembrane spike glycoprotein (S protein), which is used as a vehicle for Coronavirus to enter into host cells. In order for the virus to infect the host cells, this protein facilitates the attachment of the virus to a type of receptor

on the cell surface of the host.(6, 7, 8) By doing so, the virus can fuse with the host's membrane. There are also 29 proteins encoded by the viral genome of Cov2 which include 16 structural proteins (NSP1-NSP16), 4 nonstructural proteins (S, E, M, and N), and the presence of 9 accessory Open Reading Frames (ORFs), namely 3a, 3b, 6, 7a, 7b, 8, 9b, and 9c, is identified. These ORFs represent specific DNA or RNA sequences that have the potential to be translated into proteins.(9, 10) The protein S exhibits a binding affinity for the angiotensin-converting enzyme ACE2, which is abundantly expressed in China horse-shoe bats, human skin, and civets.(Figure 1.1) This specific interaction plays a pivotal role in facilitating viral cellular entry, enabling the virus to gain access to host cells. (11, 12) Many of the vaccines and therapeutic drugs are developed based on virus-encoded proteins, such as spike, RNA-dependent RNA polymerase, and main protease.(13, 14) Several of these targets have been analyzed at high resolution, and some of them have been complexed with potential drugs or neutralizing antibodies, to provide mechanistic insight into their function, which has been used to develop highly effective vaccines and drugs using structure-guided design. There is still a need to develop additional knowledge about the target proteins of SARS-CoV-2 in order to better understand its virology and to design alternative treatments leading to minimizing the risk of developing resistance or introducing new viruses.(15, 16, 17) ORF3a is identified as a viroporin, a viral transmembrane

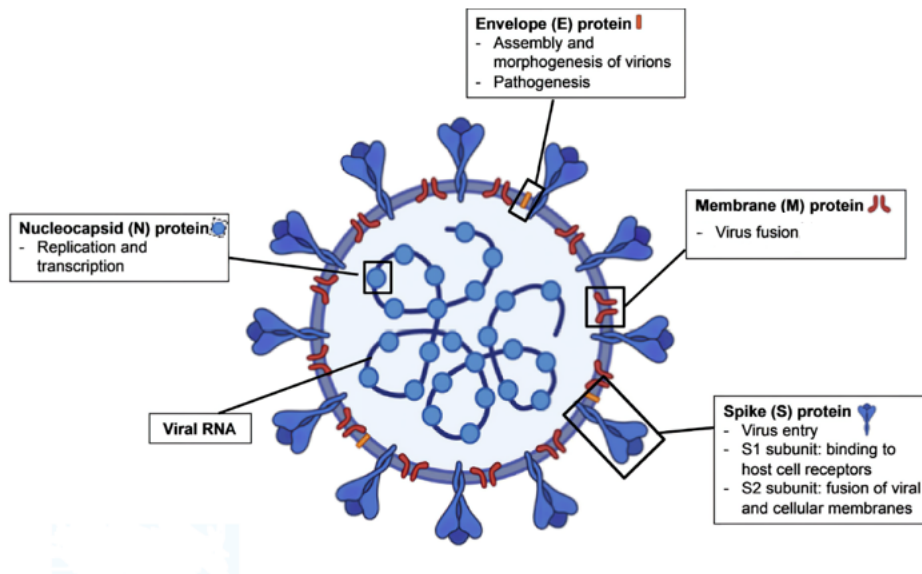


Figure 1.1: SARS-CoV-2 virus structure

protein that possesses ion channel properties within cell membranes. As a hydrophobic protein, ORF3a oligomerizes in the cell membrane, forming hydrophilic pores that enable the passage of ions and small molecules. This unique characteristic of ORF3a viroporin holds significance in viral replication and pathogenesis, as it influences the permeability of cell membranes. (18) Understanding the role of ORF3a viroporin is essential because it helps to comprehend how the virus and our bodies interact in intricate ways. This knowledge is crucial because it enables scientists to develop targeted treatments and preventive measures that specifically address the function of ORF3a viroporin. By studying it, it is possible to discover ways to combat the virus more effectively and safeguard the health.(19, 20) In a recent cryo-EM study(21), the structure of SARS-CoV-2 ORF3a was determined for the first time (PDB ID: 6XDC), see Figure 1.2. The structure comprises three transmembrane helices (TM1-TM3, residues 41-132) followed by a cytosolic domain that consists of two cell walls arranged in an antiparallel configuration. Upon dimerization, ORF3a forms six transmembrane helices which act as an ion channel. These helices contain polar/charged residues within the channel, which conduct cations. Furthermore, ion channels are important therapeutic targets, and a great deal of work has already been done in developing drugs that target ion channels. Due to its role in pathogenicity, ORF3a can be another potential drug target for disease management.(14, 22, 23)

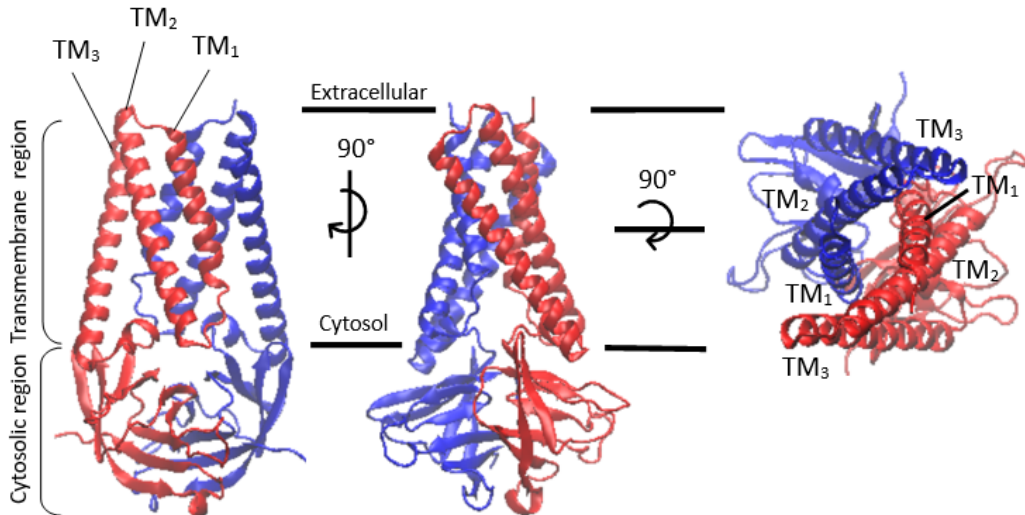


Figure 1.2: The dimeric ORF3a protein (PDB ID: 6XDC). The six transmembrane helices form ion channels in the host membrane.

The ORF3a protein present in SARS-CoV-2 plays a crucial role in the virus's interaction

with host cells. It generates a specialized channel, known as a non-selective cationic channel, that permits the movement of positively charged ions across the protective membrane of the host cell. Remarkably, this channel is not limited to a specific type of ion and allows the passage of various cations, including sodium (Na^+), potassium (K^+), and calcium (Ca^{2+}).⁽²⁴⁾ The ORF3a protein in SARS-CoV-2 has therefore two critical purposes. In the first place, it alters the ion balance in the cellular environment, potentially activating signaling pathways that support virus propagation. This could assist the virus in evading the host's immune response and establishing infection. Additionally, enveloped viral particles may be released from infected cells through the ORF3a channel. During viral replication within infected cells, the ORF3a channel aids in the release of newly formed viral particles by enabling their exit through the host cell membrane.⁽²⁵⁾⁽²¹⁾ This channel plays a critical role in the development of potential treatments for COVID-19 since it is believed that targeting the channel can potentially disrupt the viral life cycle and prevent the virus from causing any disease.^(14, 26)

It is our aim in this work to investigate the ionic permeability through the ORF3a channels by means of computational methods. In fact, the results presented in Reference ⁽²¹⁾ suggest that ORF3a of SARS-CoV-2 might function as a non-selective cationic channel with a large pore and high single-channel conductance which refers to a high capacity to conduct ions, potentially influencing cellular processes and the virus's interaction with the host. In addition, the study of the ion solvation sphere could provide crucial insights into the dynamic movement of the solvation shell as it traverses the channel during each simulation. Finally, we can possibly obtain valuable insights into the selection of key protein residues, by systematically investigating the interactions between ions and residues. Such an analysis will shed light on the precise influence these interactions have on the intricate process of ion movement through the channel.

2

Theoretical Background

2.1 Statistical Averages and the Ergodic Hypothesis

Molecular dynamics (MD) simulations offer a means to comprehend the temporal evolution occurring at the atomic level. By employing statistical averaging, we establish a crucial connection between microscopic observations and the macroscopic properties exhibited by systems. Within a macroscopic system in equilibrium, comprising multiple states W and maintained at a constant temperature T , the average value $\langle A \rangle$ of a particular variable of interest can be determined. This average value is obtained by summing the possible values A_n of the variable, each weighed by their corresponding probabilities ρ_n .(27, 28)

$$\langle A \rangle = \sum_{n=1}^W \rho_n A_n \quad (2.1)$$

As the quantity $n \rightarrow \infty$, the distribution becomes increasingly continuous and can be mathematically represented as an integral. This integral incorporates a partition function ρ_n , which is defined by a Boltzmann distribution containing the Boltzmann constant k_B and the potential energy of state U_n .

$$\langle A \rangle = \int_0^\infty A(n) \rho(n) dn \quad \rho(n) = \frac{e^{\frac{-U(n)}{k_B T}}}{\int e^{\frac{-U(n)}{k_B T}} dn} \quad (2.2)$$

The quantity $\langle A \rangle$ described here is referred to as an ensemble average. Within MD simulation, the time average of the observable A can be approximated by computing the average of its values at equally spaced time intervals, considering a total of M time steps.(27, 28, 29)

$$\bar{A} = \frac{1}{M} \sum_{i=1}^M A(t_n) \quad (2.3)$$

When $M \rightarrow \infty$, this expression can be viewed as performing an integration across the entire duration of the simulation, denoted as t_{tot} .(30, 31)

$$\lim_{M \rightarrow \infty} \bar{A} = \frac{1}{t_{tot}} \int_0^{t_{tot}} A(t') dt' \quad (2.4)$$

According to the ergodic hypothesis, when a significant number of states W are sampled during an adequate number of timesteps M , ensemble averaging and time averaging become interchangeable methods for computing the average value of an observable quantity A . This observable is reliant on the coordinates and momenta of a system composed of multiple particles and is influenced by the system's phase space. In the realm of MD simulations, the iterative propagation of atomic positions and velocities generates a trajectory that characterizes the system's evolution over time. By considering the ensemble of all points i within the phase space, one can accurately predict various system properties, where each point represents a potential configuration of the system.

2.2 Exploring the Equations of Motion

In classical molecular dynamics simulations, the displacement of particles is examined in response to the cumulative forces \mathbf{F}_i exerted upon them. By integrating Newton's equations of motion, a trajectory is generated, offering a comprehensive understanding of the temporal evolution of positions, velocities, and accelerations. The profound connection between acceleration and the forces acting on individual atoms is elucidated through Newton's second law.(32)(33)

$$\mathbf{F}_i = m_i \mathbf{a}_i \quad (2.5)$$

These equations can be reformulated using the Hamiltonian formulation, which employs generalized coordinates \mathbf{q}_i and momenta \mathbf{p}_i instead of Cartesian coordinates \mathbf{r}_i and velocities \mathbf{v}_i . This more general approach expands the understanding of dynamic systems by incorporating a broader range of coordinate systems and their associated momenta.

$$\mathbf{p}_i = m_i \mathbf{v}_i \quad (2.6)$$

The expressions defining the kinetic energy of a particle K_i and the force acting on it \mathbf{F}_i can be reformulated using momentum \mathbf{p}_i as an alternative representation.

$$K_i = \frac{1}{2} m_i \mathbf{v}_i^2 = \frac{1}{2} \frac{\mathbf{p}_i^2}{m_i} \quad (2.7)$$

$$\mathbf{F}_i = \frac{d\mathbf{p}_i}{dt} \quad (2.8)$$

The initial equation of motion is expressed as the derivative of K_i with respect to \mathbf{p}_i , yielding the displacement \mathbf{q}_i in dt .

$$\frac{d\mathbf{q}_i}{dt} = \frac{dK_i}{d\mathbf{p}_i} \quad (2.9)$$

Within an isolated system, the force \mathbf{F}_i can be defined as the negative gradient of the potential energy surface.

$$\mathbf{F}_i = \nabla U \equiv -\frac{\partial U}{\partial \mathbf{q}_i} \quad (2.10)$$

By combining Equation 2.8 and Equation 2.10, the second Hamilton equation for an isolated system can be derived.

$$\frac{d\mathbf{p}_i}{dt} = -\frac{\partial U}{\partial \mathbf{q}_i} \quad (2.11)$$

To numerically integrate Newton's equations of motion for an atomic ensemble in Cartesian coordinates space, each atom i is assigned a position \mathbf{r}_i , a velocity \mathbf{v}_i , and an acceleration vector \mathbf{a}_i . These three quantities are interconnected as follows:

$$\mathbf{v}_i = \frac{d\mathbf{r}_i}{dt} \quad \text{and} \quad \mathbf{a}_i = \frac{d\mathbf{v}_i}{dt} = \frac{d^2\mathbf{r}_i}{dt^2} \quad (2.12)$$

By performing numerical integration from the initial time $t = 0$ to the final time t , we can determine the time evolution of both \mathbf{r}_i and \mathbf{v}_i .

$$\mathbf{r}_i(t) = \int_0^t \mathbf{v}_i(t') dt' + \mathbf{r}_i(0) \quad \text{and} \quad \mathbf{v}_i(t) = \int_0^t \mathbf{a}_i(t') dt' + \mathbf{v}_i(0) \quad (2.13)$$

2.3 Relationship Between Temperature, Pressure, and the Virial Equation of State

The equation of state provides a comprehensive description of the interplay between various thermodynamic properties in a system. The force acting on an atom i is the cumulative sum of all forces \mathbf{F}_{ij} , encompassing interactions between atoms within the system, pressure-related forces \mathbf{F}_p , and additional external contributions (\mathbf{F}_{ext} , influences from applied electric fields). (34, 35)

$$\mathbf{F}_i = \sum_{i \neq j}^N \mathbf{F}_{ij} + \mathbf{F}_p + \mathbf{F}_{ext} \quad (2.14)$$

The derivation of the Virial equation of state begins with the time derivative of $\mathbf{r}_i \cdot \mathbf{p}_i$. By utilizing the relationships presented in Equation 2.6 and Equation 2.12, we can derive the following expression:

$$\frac{d}{dt}(\mathbf{r}_i \cdot \mathbf{p}_i) = m_i \frac{d}{dt}(\mathbf{r}_i \frac{d\mathbf{r}_i}{dt}) = \underbrace{\mathbf{r}_i m_i \frac{d^2\mathbf{r}_i}{dt^2}}_{m_i \mathbf{a}_i = \mathbf{F}_i} + \underbrace{m_i \left(\frac{d\mathbf{r}_i}{dt}\right)^2}_{m_i \mathbf{v}_i^2} \quad (2.15)$$

Within the expression, the second term encompasses a force \mathbf{F}_i , while the final term corresponds to a contribution from kinetic energy.

$$2\left(\frac{1}{2}m_i \mathbf{v}_i^2\right) = 2K \quad (2.16)$$

The equation represented by Equation 2.15 can be expanded to encompass a summation across all particles within a system.

$$\sum_{i=1}^N m_i \frac{d}{dt}(\mathbf{r}_i \frac{d\mathbf{r}_i}{dt}) = \sum_{i=1}^N \mathbf{r}_i \mathbf{F}_i + 2 \sum_{i=1}^N K_i \quad (2.17)$$

In an ergodic system, the ensemble average can be computed as the sum of individual averages for each term. For an isolated system devoid of external forces, only forces originating from atomic interactions and pressure are relevant. Fluctuations in particle velocities and positions occur randomly and exhibit a symmetric distribution around their respective mean values. Furthermore, the time derivative of their product follows a random distribution centered around an average value of zero. (36, 37)

$$\underbrace{\left\langle \sum_{i=1}^N m_i \frac{d}{dt}(\mathbf{r}_i \frac{d\mathbf{r}_i}{dt}) \right\rangle}_{=0} = \underbrace{\left\langle \sum_{i=1}^N \mathbf{r}_i \mathbf{F}_i \right\rangle}_W + 2\langle K \rangle \quad (2.18)$$

The term denoted on the right side of the equation represents the system's virial W .

$$0 = \langle W \rangle + 2\langle K \rangle \quad (2.19)$$

In a perfectly isolated system, the total energy remains conserved, while variables such as pressure and temperature exhibit fluctuations.

2.3.1 Kinetic Energy-Temperature Correlation:

The temperature T of a particle system is determined by the average kinetic energy $\langle K \rangle$, which, in turn, depends on the average velocities v_i of the particles.

$$\langle K \rangle = \left\langle \sum_{i=1}^N \frac{1}{2} m_i \mathbf{v}_i^2 \right\rangle = \frac{3}{2} N k_B T \quad (2.20)$$

In an N -atomic system at equilibrium, the given equation expresses that each atom with three degrees of freedom contributes $\frac{1}{2}k_B T$ (k_B corresponding to the Boltzmann constant) to the overall kinetic energy. This equation mathematically represents the equipartition principle in statistical mechanics, which allocates an equal share of energy to each degree of freedom in the system. (38, 39)

2.3.2 Virial-Pressure Relationship:

In Equation 2.14, it is stated that within an isolated system, the total force acting on its particles comprises the combined effects of a pressure force \mathbf{F}_P and the summation of all atom-atom interaction contributions \mathbf{F}_{int} .(40)

$$\langle W \rangle = \left\langle \sum_{i=1}^N \mathbf{r}_i \mathbf{F}_i \right\rangle = \left\langle \sum_{i=1}^N \mathbf{r}_i \mathbf{F}_p \right\rangle + \left\langle \sum_{i=1}^N \mathbf{r}_i \mathbf{F}_{int} \right\rangle \quad (2.21)$$

The application of pressure on a system primarily affects its surface, resulting in a force contribution of zero for bulk atoms. In the scenario of a cubic simulation box, the force exerted on each face of the cube demonstrates a direct proportionality to the applied pressure p and the surface area of the cube L^2 . Additionally, the total pressure force manifests a proportional relationship with the cubic volume V of the system.(41)

$$F = pL^2 \quad \text{and} \quad \left\langle \sum_{i=1}^N \mathbf{r}_i \mathbf{F}_p \right\rangle = -3pL^3 = -3pV \quad (2.22)$$

Having established these relationships, it becomes possible to reformulate the equation of state (Equation 2.20) to express the interdependence among a system's particle count N , temperature T , pressure p , volume V , and internal forces.

$$3Nk_B T + \left\langle \sum_{i=1}^N \mathbf{r}_i \mathbf{F}_{int} \right\rangle + 3pV = 0 \quad (2.23)$$

By utilizing the correlation between force \mathbf{F} and potential energy U , as depicted in Equation 2.10, it is possible to express the virial equation of state in terms of the potential energy associated with particle interactions. When dealing with a noninteracting system, this contribution becomes negligible, resulting in Equation 2.24 reducing to the ideal gas law.

$$\frac{pV}{k_B N T} = -\frac{1}{3k_B N T} \left[1 + \left\langle \sum_{i=1}^N \mathbf{r}_i \left(\frac{-\partial U_{int}}{\partial \mathbf{r}_i} \right) \right\rangle \right] \quad (2.24)$$

In a system consisting of N particles, the interaction potential U_{int} , which is influenced by the coordinates of all particles \mathbf{r}_i , can be expanded into a series of terms representing x-body interactions.

$$U_{int}(r_1, \dots, r_N) = \sum_{i < j}^N U_{ij}(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i < j < k}^N U_{ijk}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots \quad (2.25)$$

By substituting Equation 2.24 into Equation 2.25 and categorizing the terms based on interaction order, the virial equation is derived. This equation introduces the concept of volume virial coefficients (B, C, \dots) that relate the macroscopic properties of pressure, volume, and temperature to the interaction potentials originating from x-body interactions.

$$\frac{pV}{k_B NT} = 1 + \frac{NB}{V} + \frac{N^2 C}{V^2} + \dots \quad (2.26)$$

As the order of interactions x increases, the impact of the interaction terms on the average virial becomes progressively smaller. It is customary to truncate the expansion, typically at $x = 2$, as the higher-order contributions become negligible. This simplification allows for more practical estimation of the total force arising from particle interactions.[\(41, 42\)](#)

2.4 Stages of the MD Simulation Process

The objective of MD simulations is to investigate the dynamic behavior of molecular systems at the atomic level. This involves iteratively propagating the positions of atoms (denoted as $\mathbf{r}(t)$) over a time interval from an initial time point t to a final time point $t + \Delta t$. The time interval Δt is chosen to be significantly smaller than the characteristic timescale of motion within the system. This propagation is achieved by numerically integrating Hamilton's equations of motion using specific algorithms known as integrators. Examples of integrators include the Verlet algorithms, such as Simple, Leap-Frog, and Velocity Verlet, as well as other techniques like Runge-Kutta and Gear-Predictor correctors. The Velocity Verlet algorithm, renowned for its stability and time-reversal properties, has been implemented in the Nanoscale Molecular Dynamics (NAMD) program.[\(39, 43, 44\)](#)

2.4.1 Velocity-Verlet Algorithm

NAMD employs the Velocity-Verlet integration method to advance atomic velocities $\mathbf{v}_i(t)$ and positions $\mathbf{r}_i(t)$. The implementation involves computing the intermediate positions at a given time point $t + \frac{\Delta t}{2}$ around the initial time t . The accelerations at each timestep

t are then calculated based on the corresponding forces $F(\mathbf{r}_i(t))$, as described by [Equation 2.14](#).[\(32, 45, 46\)](#)

$$\mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}_i(t) + \frac{\Delta t \mathbf{a}_i(t)}{2} \quad (2.27)$$

The obtained velocities are subsequently employed to determine the positions \mathbf{r}_i at a given time $t + \Delta t$.

$$\begin{aligned} \mathbf{r}_i(t + \Delta t) &= \mathbf{r}_i(t) + \Delta t \mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) \\ &= \mathbf{r}_i(t) + \Delta t \mathbf{v}_i + \frac{\Delta t^2 \mathbf{a}_i(t)}{2} \end{aligned} \quad (2.28)$$

The updated positions are sequentially utilized to modify the accelerations $\mathbf{a}_i(t + \Delta t)$ and propagate the corresponding quantity \mathbf{v}_i .

$$\begin{aligned} \mathbf{v}_i(t + \Delta t) &= \mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) + \frac{\Delta t \mathbf{a}_i(t + \Delta t)}{2} \\ &= \mathbf{v}_i(t) + \frac{\Delta t [\mathbf{a}_i(t) + \mathbf{a}_i(t + \Delta t)]}{2} \end{aligned} \quad (2.29)$$

In the following the different steps to run dynamics are explained:

1. To compute $\mathbf{r}(t + \Delta t)$, it is necessary to have a collection of initial positions, referred to as $\mathbf{r}_i(t)$, as well as velocities denoted by $\mathbf{v}_i(t)$. Based on these initial conditions, internal coordinates referred to as r_j , which encompass bond lengths, bond angles, and dihedral angles, can be determined. At the start of a simulation with $t = 0$, it is necessary to provide input coordinates obtained from experimental databases or quantum mechanical calculations. In cases where initial velocities are not explicitly specified, they are assigned probabilistically using values $\rho \mathbf{v}_i$ derived from a Boltzmann distribution of velocities denoted as \mathbf{v}_i , and they correspond to the desired simulation temperature T . The assigned values are adjusted to ensure that the total momentum denoted as \mathbf{p} is maintained at zero.

$$\rho \mathbf{v}_i = \left(\frac{m_i}{2\pi k_B T} \right) e^{-\frac{m_i v_i^2}{2k_B T}} \quad (2.30)$$

$$\mathbf{p} = \sum_i^N m_i \mathbf{v}_i^2 \quad (2.31)$$

2. In classical molecular dynamics simulations, force field parameters are employed to calculate the system's potential energy by summing the interaction potential terms. This potential energy surface enables the determination of forces acting on particles based on its functional form.

3. The force magnitude is determined by the gradient of the potential energy surface at a specific point r . If U only depends on one given internal coordinate, r , the force can be obtained by taking the derivative with respect to that coordinate.

$$F = -\frac{dU(r)}{dr} \quad (2.32)$$

4. Given the knowledge of the forces, F_i , acting on individual particles, the corresponding accelerations \mathbf{a}_i can be determined (as described in [Equation 2.5](#)).
5. After the evaluation of \mathbf{a}_i , an integrator is employed to update the positions and velocities of atoms.

2.5 Periodic Boundary Conditions

The maximum number of atoms that can be simulated in MD falls within the range of 10^5 to 10^6 . For a cubic simulation box, the ratio of surface atoms to the total number of atoms is determined by the length of the box. ([47](#), [48](#))

$$N \propto L^3 \quad N_s \propto 6L^2 \implies N_s \propto 6N^{\frac{2}{3}} \quad (2.33)$$

In a system with 10^6 atoms, approximately 6% of these atoms are located at the surface. However, real physical systems often consist of a much larger number of particles. For example, one mole of atoms in a cubic box contains only about $7 \times 10^{-6}\%$ surface atoms. In the case of the smaller system, the contribution of surface interactions to bulk properties cannot be ignored. To simulate bulk behavior, periodic boundary conditions are applied, creating an infinite system by replicating the original simulation box in a three-dimensional array. Each image atom behaves as if it were in the original box, identified by the cell vector $\mathbf{n}(0,0,0)$. Interactions are considered not only within the primary cell but also between periodic images separated by a multiple by \mathbf{n} of the box length (L). The prime notation in [Equation 2.34](#) indicates the exclusion of interactions between an atom and itself within the original cell.

$$U = \sum_{\mathbf{n}}' \sum_{i=1}^N \sum_{j=1}^N U(|\mathbf{r}_{ij} + \mathbf{n}L|) \quad (2.34)$$

The computation of infinite sums is avoided by considering only short-range interactions within a defined cutoff radius. Beyond this radius, the contributions of these interactions are assumed to be negligible. However, long-range interactions cannot be treated in the same manner. Their calculation requires methods that approximate infinite sums. The

Ewald summation approach is commonly used to evaluate long-range electrostatic potentials.

2.6 Temperature Regulation in the Isothermal Ensemble

In order to simulate realistic experimental conditions, it is often preferable to conduct simulations at a constant temperature T rather than maintaining constant total energy. In a closed system with a constant volume V and temperature, the canonical distribution function indicates that lower energy states are more probable than higher energy states. (49, 50)

$$\rho(i) = \frac{e^{-\frac{E_i}{k_B T}}}{\sum_{i=1}^N e^{-\frac{E_i}{k_B T}}} \quad (2.35)$$

In an NVT simulation, temperature control is implemented through periodic adjustments at specified time intervals Δt . The aim is to maintain a constant temperature throughout the simulation. This can be achieved by two methods: rescaling the velocities of particles within the system based on Equation 2.20, or allowing energy exchange between the system and an external thermostat or heat bath, resulting in fluctuations in the total energy. The Andersen method is an example of the latter approach, which involves simulating random collisions with particle walls and reassigning velocities of a fraction of particles to values chosen from a Boltzmann distribution of velocities for the target temperature. The number of particles to which rescaling is applied is governed by the collision frequency parameter f . Temperature regulation can be achieved through the adjustment of particle velocities using a scaling factor λ , which is the basis for methods such as velocity rescaling and the Berendsen barostat. In the velocity rescaling method, the scaling factor λ_{VR} is determined by the ratio between the actual temperature $T(t)$ and the target temperature T_{set} . The Berendsen algorithm includes an additional coupling parameter, $\Delta t/\tau$, as part of its formulation represented by the parameter λ_B . τ is known as the "rise time," which controls the strength of coupling between the heat bath and the system. As $\tau \rightarrow \infty$ the coupling is gradually removed, and the system becomes decoupled from the heat bath. (51)

$$\mathbf{v}_i^{new} = \lambda \mathbf{v}_i^{old} \quad (2.36)$$

$$\lambda_{VR} = \left(\frac{T_{set}}{T(t)} \right)^{\frac{1}{2}} \quad \lambda_B = \left[1 + \frac{\Delta t}{\tau} \left(\frac{T_{set}}{T(t)} - 1 \right) \right]^{\frac{1}{2}} \quad (2.37)$$

In the present simulations, the temperature was maintained using the Langevin thermostat which expands upon the equation of motion for momenta (Equation 2.11) by incorporating

two additional forces: a friction coefficient γ and a random force $R(t)$. These forces account for the system's viscosity and random collisions among particles. The random force is generated from a Gaussian distribution with a mean value of zero and an average value of zero. So, It is important to highlight that the width of the Gaussian distribution for the random force and the value of the friction coefficient rely on the temperature.(52)

$$\frac{d\mathbf{p}_i}{dt} = \frac{\partial U}{\partial \mathbf{q}_i} - \gamma \mathbf{p}_i + R(t) \quad (2.38)$$

2.7 Control of Pressure in the Isothermal-Isobaric Ensemble

In order to capture realistic experimental conditions, it is often preferable to simulate systems in the isothermal-isobaric ensemble, where both temperature and pressure are kept constant. This ensemble provides a more accurate representation by considering the probabilities of states with specific values of temperature and pressure, denoted as ρ and T respectively. (53, 54)

$$\rho(i) = \frac{e^{-\frac{E_i + p v_i}{k_B T}}}{\sum_{i=1}^N e^{-\frac{E_i + p v_i}{k_B T}}} \quad (2.39)$$

The probability of finding the system in a specific state at pressure p is influenced by the volume of the simulation cell. The relationship between pressure p and volume V is described by the virial equation of state (Equation 2.26). To control both pressure and temperature, a combination of a barostat for adjusting the cell volume and a thermostat for modulating temperature can be used. In our simulations, the pressure was regulated using the Langevin Nosé-Hoover algorithm. This algorithm extends the concept of the Andersen barostat, where a closed system is connected to a piston with mass M and subjected to an external pressure α , resulting in uniform expansion or contraction of the simulation box. The α piston corresponds to the desired pressure of the system. The equations of motion for this method are formulated as follows:(55, 56)

$$\frac{d\mathbf{q}_i}{dt} = \frac{\partial K}{\partial \mathbf{p}_i} + \frac{1}{3} \mathbf{q}_i \frac{d \ln V}{dt} \quad (2.40)$$

$$\frac{d\mathbf{p}_i}{dt} = \frac{\partial U}{\partial \mathbf{q}_i} - \frac{1}{3} \mathbf{p}_i \frac{d \ln V}{dt} \quad (2.41)$$

$$\frac{d^2 V}{dt^2} = \frac{1}{M} \left[\underbrace{\left(\frac{2}{3} \sum_{i=1}^N K_i - \frac{1}{3} \sum_{i=1}^N \mathbf{F}_i \cdot \mathbf{r}_i \right)}_{=p(t)} - \underbrace{\alpha}_{=p_{set}} \right] \quad (2.42)$$

$$\frac{d^2V}{dt^2} = \frac{1}{M} (p(t) - p_{set}) \quad (2.43)$$

The Langevin Barostat method incorporates a collision frequency γ and a random force $R(t)$ into Equation 2.42, which governs the fluctuations in the volume of the simulation box. Similar to the Langevin thermostat, the random force is sampled from a Gaussian distribution with a mean value and average of zero.(56)

$$\frac{d^2V}{dt^2} = \frac{1}{M} \left(p(t) - p_{set} + \gamma \frac{dV}{dt_R(t)} \right) \quad (2.44)$$

2.8 Force Field

The Born-Oppenheimer approximation enables the separation of nuclear and electronic components in a system. Classical potential models in molecular mechanics force fields focus on the nuclear terms and approximate the electronic contributions. These force fields incorporate parameters for calculating interatomic interaction energies, which are derived from experimental data or quantum mechanical calculations. Atom types are assigned to each atom based not only on their atomic number but also on their chemical environment. For instance, a carbon atom in an aromatic ring is categorized differently from a carbonyl carbon. In simulations of biological systems, popular force fields include CHARMM and the Assisted Model Binding with Energy Refinement (AMBER). (57) The total potential energy, U , is typically divided into bonded U_b and non-bonded U_{nb} contributions. (58, 59)

$$U = \sum_{i=1}^N U_{i,b} + \sum_{i=1}^N U_{i,nb} \quad (2.45)$$

2.8.1 Bonded Intractions

The interactions between covalently bonded atoms can be divided into bond U_{bond} , angle U_{angle} , and dihedral angle $U_{dihedral}$ terms, representing 2-, 3-, and 4-body interactions, respectively. These terms contribute to the overall potential energy of the system and play a crucial role in describing the behavior of molecules.(44)

$$\sum_{i=1}^N U_{i,b} = \sum_{i>j}^N U_{ij,bond} + \sum_{i>j>k}^N U_{ijk,angle} + \sum_{i>j>k>l}^N U_{ijkl,dihedral} \quad (2.46)$$

Bond terms in molecular simulations are described by a harmonic potential, where the spring constant represents the bond stiffness. The interatomic distance between atoms i and j is denoted as R_{ij} , while the equilibrium distance is represented by R_0 . This

harmonic oscillator approximation accurately captures the bond stretching contribution and quantifies the bond's strength and deviations from its equilibrium position.

$$U_{ij,bond} = k_b (R_{ij} - R_\theta)^2 \quad (2.47)$$

Angle terms in molecular simulations are modeled using a harmonic potential, similar to the bond terms. The angular bending term, represented by U_{angle} , involves three atoms i , j , and k . The angle θ is measured between the vectors \mathbf{r}_{ij} and \mathbf{r}_{jk} , while θ_0 denotes the equilibrium angle. Additionally, the Urey-Bradley term accounts for a spring-like interaction between two atoms i and k that are not directly bonded. The separation distance between them is denoted by R_{ik} , while the equilibrium distance and spring constant are represented by R_{ub} and k_{ub} , respectively.[\(45\)](#)

$$U_{ijk,angle} = k_\theta(\theta - \theta_0) + k_{ub}(R_{ij} - R_{ub})^2 \quad (2.48)$$

Dihedral terms in molecular simulations account for the interactions between atom pairs that are part of three consecutive bonds forming a dihedral angle. These interactions are computed using an angular potential between two planes formed by atoms i , j , k , and j , k , l , where ψ represents the angle enclosed by the planes.

$$\text{If } a \geq 0 : \quad (2.49)$$

$$U_{ijkl,dihedral} = k_d (1 - \cos(a\psi + \varphi)) \quad (2.50)$$

For periodic dihedral potentials, the integer $a > 0$ specifies the periodicity order. In such cases, φ denotes the phase shift angle and k_d serves as a multiplicative constant.

2.8.2 Non-Bonded Interactions

The total nonbonded potential can be divided into three components: long-range electrostatic interactions denoted as $U_{coulomb}$, short-range van der Waals interactions labeled as U_{vdW} , and repulsive interactions referred to as U_{rep} .[\(60\)](#)

$$\sum_{i=1}^N U_{i,nb} = \sum_{i>j}^N U_{ij,Coulomb} + \sum_{i>j}^N U_{ij,vdW} + \sum_{j>j}^N U_{ij,rep} \quad (2.51)$$

Electrostatic interactions, known as Coulomb interactions, arise from the electrostatic forces between point charges q of two atoms i and j . These interactions are influenced by the dielectric constant ε , which represents the solvent's screening effect.[\(61, 62\)](#)

$$U_{ij,Coulomb} = \frac{1}{4\pi\varepsilon} \frac{q_i q_j}{R_{ij}} \quad (2.52)$$

The energy term related to electrostatic interactions exhibits attraction between opposite charges and repulsion between equal charges. Although the strength of this interaction decreases with increasing distance between atoms (R_{ij}), it remains significant even for atoms that are far apart.

The Ewald summation method is employed to accurately compute electrostatic interactions. This approach resolves the slow convergence and dependency on summation order. By introducing Gaussian charge distributions and considering their interactions with point charges, the long-range electrostatic interactions are effectively transformed into short-range interactions.(63, 64) This simplifies the calculation process and allows for the accurate determination of electrostatic energies. Additionally, a compensation scheme involving additional Gaussian functions with opposite signs is employed to account for the introduced charge distributions. Overall, the Ewald summation method enables the decomposition of the electrostatic energy into three distinct terms, facilitating the analysis of electrostatic interactions. (65, 66)

$$U_{Coulomb} = U_{direct} + U_{reciprocal} + U_{correction} \quad (2.53)$$

$$U_{direct} = \frac{1}{2} \sum_n \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{R_{ij} + |\mathbf{n}|} \text{erfc}(\alpha R_{ij} + |\mathbf{n}|) \quad (2.54)$$

$$\text{erfc}(x) = 1 - \text{erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (2.55)$$

$$U_{reciprocal} = \frac{1}{2\pi U} \sum_{i=1}^N \sum_{j=1}^N q_i q_j \sum_{\mathbf{m} \neq 0} \frac{e^{-\left(\frac{\pi \mathbf{m}}{\alpha}\right)^2 + 2\pi \mathbf{m} R_{ij}}}{\mathbf{m}^2} \quad (2.56)$$

$$U_{correction} = -\frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^N q_i^2 \quad (2.57)$$

The inclusion of the Gaussian error function term (Equation 2.54) in Equation 2.53 ensures a rapid decay as $\mathbf{n} \rightarrow \infty$ increases. The reciprocal space vector \mathbf{m} is determined through the Fourier transform of the real space cell vector \mathbf{n} . The correction term $U_{correction}$ takes into consideration the width of the Gaussian function distributions used in the computation.(67, 68, 69)

Van der Waals' and repulsion interactions, The van der Waals' potentials U_{vdW} which are responsible for attractive interactions, arise from the interaction between fluctuating instantaneous and permanent dipole moments.(67, 70)

$$U_{vdW} \propto -R_{ij}^{-6} \quad (2.58)$$

The repulsion energy term U_{rep} arises from the repulsion between overlapping electronic orbitals at very short separations, and it increases as the distance decreases. This contribution cannot be accurately described without employing quantum mechanics.

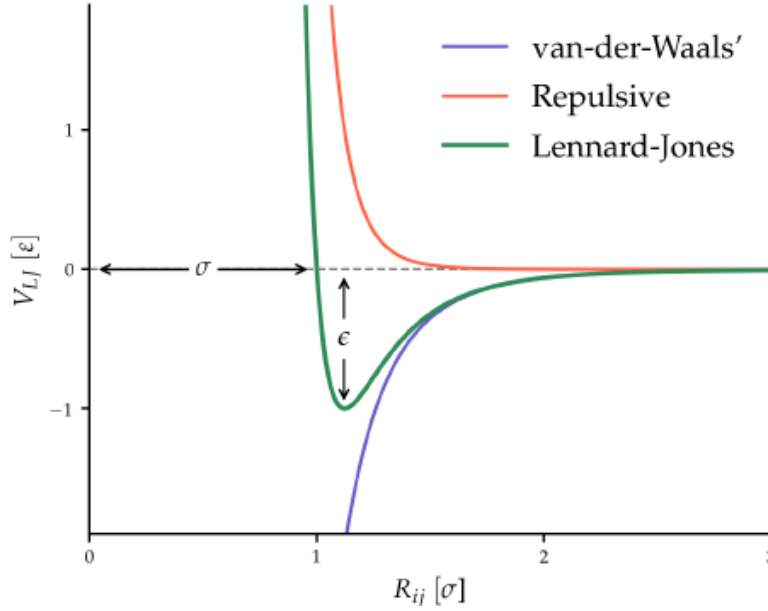


Figure 2.1: Van der Waals, repulsive, and Lennard-Jones potential energy functions depict with ε and σ . The term ε denotes the potential depth and σ refers to the distance at which the energy is zero.

The short-range interactions are typically represented using model potentials that incorporate both the repulsive and attractive energy terms mentioned above. One commonly used model is the Lennard-Jones potential, U_{LJ} , as shown in Figure 2.1. At short distances, the repulsive term dominates the interaction energies, while at larger distances, the attractive van der Waals' term becomes more significant. This leads to a minimum in the potential energy curve at intermediate separations. The parameter ε and σ represent the well depth which measures how strongly the two particles attract each other and the distance at which the intermolecular potential between the two particles is zero. The values of ε and σ can be calculated based on the respective parameters of the two atoms.(67)

$$U_{LJ} = \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}} \right)^6 \right] \quad (2.59)$$

3

Computational Details

The computational goal of this thesis work is to massively run molecular dynamics simulations of the protein embedded in a lipid bilayer and solvated by molecules of water. In this way, we can focus on utilizing the ORF3a channel as a model system to investigate the conduction of ions. Here, we describe the methods used to study the Sars-Cov-2 ORF3a and provide comprehensive details of the classical molecular dynamics (MD) simulations conducted, including details on the trajectory processing methods employed. Furthermore, the computational approaches employed in constructing the SARS-CoV-2 ORF3a model will be outlined.

3.1 Design and construction of model

It involves preparing the initial state of the system for input into the simulation package, which includes constructing a starting structure. There are a variety of ways to prepare a system depending on not only its composition but also the information available about the initial structure of the system. Due to this, each system requires a unique set of tools and considerations during this step. The preparation of a system consists of two logical components, which are to construct the system's configuration in the desired chemical state and to apply the necessary force field parameters for the system to function as desired. During the system preparation step in molecular dynamics simulations, it is necessary to have a good starting structure. At the thermodynamic state point of interest, the starting structure should be similar to the equilibrium structure of the system and avoid energetically unfavorable configurations.

In order to prepare the input files for the simulation system, we utilized the CHARMM GUI input which was done by using the CHARMM-GUI Bilayer Builder generator to

prepare the input files from the Sars-CoV-2 ORF3a structure obtained from PDB code (<https://www.rcsb.org/>).[\(71, 72\)](#) This was accomplished by embedding the protein into a DOPE (dioleoylphosphatidylethanolamine) lipid bilayer. The protein was first correctly aligned along the z-axis using the Operations of Protein in Membranes (OPM) database. Then, in order to create a simulation box, we need to determine the type and size of the box. There are several factors that can affect the length of the Z-axis, such as the thickness of the water. As a result, the default value for the thickness of water has been established at 22.5 Å. To establish the length of XY, it was set to 60 Å. This table illustrates some of the many features we applied to build up the structure [Table 3.1](#):

Calculated Number of Lipids (n)		
Lipid Type	UpperLeaflet Number	LowerLeaflet Number
DOPE	36	42
Calculated XY System Size		
Protein Area (Å ²)	1364.62079	961.78849
Lipid Area (Å ²)	2282.4	2662.8
# of Lipids (<i>n</i>)	36	42
Total Area (Å ²)	3647.02079	3624.58849

Table 3.1: System info

This simulation [Figure 3.1](#) (PBD ID: 6XDC) was carried out using the following ionic condition [\(21\)](#); NaCl 0.15M, KCl 0.15M, and CaCl₂ 0.15M.

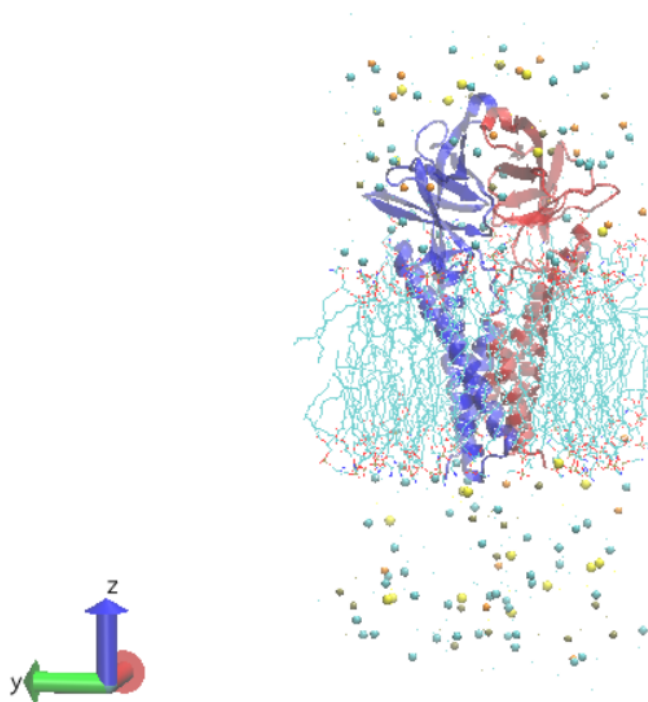


Figure 3.1: The components of the simulation box. protein is represented in blue and red (cartoon), lipid bilayer in cyan (line), chloride ion in cyan, sodium in yellow, potassium in brown, and calcium in orange (sphere). The positive z-axis is shown in the blue arrow.

3.2 Steps of Molecular Dynamics Simulations

MD were performed using the NAMD software package. It is feasible to study biomolecules through the use of the many force fields available, such as OPLS, AMBER, CHARMM, GROMOS, etc. (57, 73). As a first step, we performed an energy minimization for 5000 steps in order to drive the system to the local minimum. It is crucial in molecular simulations that the minimization or relaxation process be performed in order to obtain a local minimum in the energy of the starting structure, which prevents the simulation from "blowing up" due to the large forces on atoms. Standard minimization algorithms such as steepest descent are employed during this process. To begin the simulation, velocities must be assigned randomly because minimization only provides the final positions of atoms. It is usually done in a way that ensures the correct Maxwell-Boltzmann distribution at the desired temperature. It is not crucial what the exact assignment process is since the question of the correct distribution will quickly arise naturally from the equations of motion. Nevertheless, Newtonian dynamics conserve the momentum of the center-of-mass of the

simulation box, so the center-of-mass velocity is usually removed from all particles after the random assignment has been conducted to protect the simulation box from drifting. (48) We carried out the minimization and heating step by using the canonical ensemble (NVT ensemble) at 303.15 K and 1.01 bar with a timestep of 2 fs, in which we used the equation of state. By using the SHAKE algorithm (74), it was possible to keep the lengths of hydrogen bonds fixed. We set the van-der-Waals’ cutoff radius, switching distance, and the pairlist distance to be 12.0 Å, 10 Å, and 16 Å, respectively. In both the equilibration and production steps, it is acknowledged that electrostatic interactions are calculated using the particle-mesh Ewald (PME) method (75), with a grid spacing of 1.0 units. Force calculations for protein and lipids were based on the CHARMM36 force field and those of water on the TIP3P model. (76, 77) The Lennard-Jones parameters for the ions were taken from the parameterization made by Joung and Cheatham (78). Another important factor is that the system was powered with an external electric field of -23.065 kcal/mole in order to accelerate the crossing of the ions by controlling the electric field.

3.2.1 Equilibration Runs

A molecular simulation involves running simulations in a particular thermodynamic ensemble and gathering unbiased data for analysis at a specific state point. There are four steps involved in this process in order to bring the system to the appropriate state point and relax away from any artificially induced metastable starting state. Our aim is to sample the most relevant configurations of the equilibrium ensemble. It should be noted, however, starting in a less stable configuration may require a longer equilibration time to reach the most relevant configuration space, which can be extremely difficult for biomolecules or systems at phase equilibrium. It is necessary to bring the system to the target state point during the equilibration process.(39, 45)The first two steps of equilibration were carried out using the NVT (constant number, volume, and temperature) ensemble, and the last two steps were carried out using the isothermal-isobaric ensemble (NPT), constant number, temperature, and pressure. During the simulation, the temperature and pressure were adjusted with the help of a Langevin thermostat (damping coefficient 1 ps^{-1}), and barostat (in the last two equilibration steps, the piston period was 50 fs, and the piston decay was 25 fs in up to 100000 running steps).

3.2.2 Production Runs

Once equilibration has been completed, the production phase begins, in which data is collected for analysis. Prior to production, it is important to perform an equilibration that is appropriate to the target production ensemble before moving forward with production, and data should never be collected and recorded immediately after a change in conditions. Analyzing the production requires the computation of observables and the assessment of the uncertainty that is associated with those properties.[\(39, 48\)](#)

The production runs were carried out using the NPT (constant number, temperature, and pressure) ensemble. Temperature and pressure were controlled with a Langevin thermostat (damping coefficient 1 ps^{-1}) and barostat (piston period 50 fs, piston decay 100 fs). The production runs within our system till 461 ns. It is important to keep in mind that the simulation of biomolecules requires a very long simulation, therefore it is intended to speed up the permeation of ions through the simulation by using an external electric field. The electric field is a physical quantity that describes the influence of electric charges on other charged particles or objects. By applying an external electric field to biomolecules, the simulation can be accelerated, allowing for faster results. This approach is particularly useful when capturing rare events or phenomena that may occur over long timescales.[\(79\)](#) The specific implementation of the external electric field can be customized based on the study objectives, including factors such as field strength and direction. [Figure 3.2](#) illustrates the key commands utilized to apply an external electric field. The electric field in this simulation was set up to drive the ions from the intracellular region to the extracellular region. The `eFieldNormalized` command plays a crucial role in adjusting the electric field vector by considering the cell basis vectors. When set to "yes," this command ensures that the constant electric field forces applied are not taken into account during the pressure calculation. This approach allows for a more accurate assessment of the pressure solely based on the inherent molecular interactions, excluding any influence from the externally applied electric field forces. In addition, take into account that NAMD uses the unit of kcal/(mole) as a unit of measurement.

```
1 #external field parameter
2 eFieldOn yes
3 eField 0.0 0.0 -23.06549
4 eFieldNormalized yes
```

Figure 3.2: The segment of the NAMD input file specifies the application of the external electric field.

4

Discussion and Results

This chapter is dedicated to presenting the results of the data analyses performed to evaluate the influence of external electric field strength on the ion conduction and to quantify ion flux. Through a meticulous examination simulation results, we have gained valuable insights into multiple aspects, including water solvation, and the intricate interplay between ions and positively and negatively charged residues.

4.1 System Model

In this investigation, we applied the external electric field to the SARS-CoV-2 ORF3 ion channel with the aim of exploring its impact on ion conduction dynamics (Figure 4.1). The ion channel plays a critical role in facilitating various viral functions and is, therefore, a key component in viral infection. Our computational findings exhibited a significant augmentation in ion conduction when the electric field was applied. By subjecting the ion channel to an external electric field, featuring a positive charge in the intracellular space and a negative charge in the extracellular space, our objective was to expedite the process of ion conduction. The empirical observations underscore the critical role of the external electric field in enhancing the conduction of ions within the SARS-CoV-2 ORF3 ion channel. This phenomenon holds potential implications for viral pathogenesis, particularly in relation to vital processes including viral replication, assembly, and release. The observation that anions predominantly pass through a channel intended for cations indicates a potential flaw in the model. However, due to the identified failure in our simulations, we cannot claim the participation of anion conduction in the viral life cycle. Through thorough and

comprehensive analysis, it is possible to shed light on various aspects of ion conduction mechanisms and identify where the theoretical model is failing

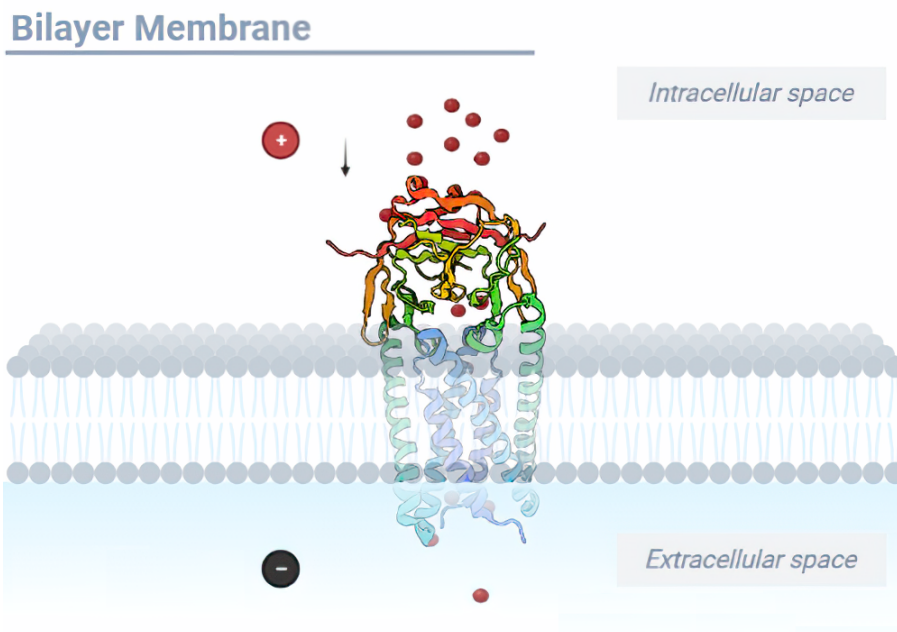


Figure 4.1: Simulation system depicts the bilayer membrane and the directional external electric field from intracellular to extracellular Domains.

4.2 Equilibration Analysis

Figure 4.2(a) illustrates the time-dependent variation of the total energy throughout three distinct stages: minimization, heating, and equilibration. During the initial minimization step, the energy exhibits a decreasing trend. This reduction in energy is expected as the system strives to reach a state of lower potential energy by adjusting its configuration. Following the minimization phase, the heating stage is implemented, resulting in an increase in energy. The heating process involves the gradual introduction of thermal energy into the system, which leads to an overall rise in total energy.

Upon entering the equilibration phase, the energy stabilizes around the value of approximately -75000 kcal/mole, demonstrating a relatively constant level. Equilibration is a vital process enabling the system to reach a stable state characterized by minimal energy fluctuations. By observing the temporal energy changes throughout these different stages (minimization, heating, and equilibration), valuable insights can be gained regarding the system's behavior. The potential energy, represented by the green line, remains constant

around -110000 kcal/mole between 0.6 ns and 4 ns, indicating the establishment of a thermodynamically stable configuration with a minimized potential energy surface. On the other hand, the kinetic energy initially experiences an increase within the first 0.6 ns of the simulation, surpassing 25,000 kcal/mole. This notable rise in kinetic energy signifies heightened molecular motion within the system. However, after 0.6 ns, the kinetic energy stabilizes, maintaining a constant plateau. This indicates that the average molecular motion and associated energy levels remain relatively constant during this period.

The plot in Figure 4.2(b) provides a detailed representation of the time-dependent evolution of temperature during the heating step. The plotted data reveals a continuous and consistent increase in temperature throughout the duration of the heating phase. Throughout the equilibration step, the system's density exhibited a gradual rise, as demonstrated by the temporal behavior illustrated in Figure 4.2(c). The density plot gradually approached a stable value after 2.7 ns, indicating the establishment of equilibrium. In this state, the system's density exhibited fluctuations around the anticipated density of water, demonstrating a balanced and steady condition.

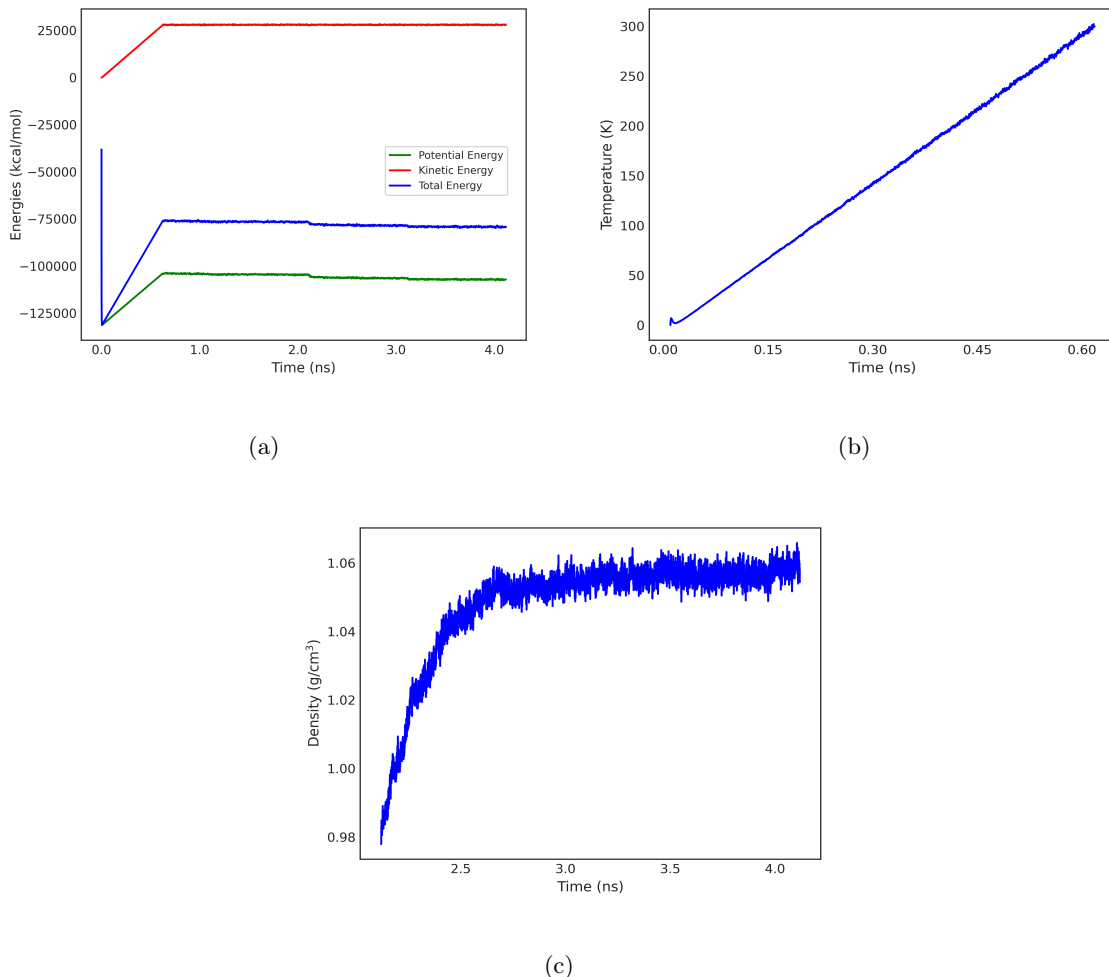


Figure 4.2: (a): Exploring various forms of energies (including total Energy, kinetic energy, and potential energy) during the processes of minimization, heating, and equilibration. (b): Analyzing temperature variations during heating. (c): Investigating the density increase during the equilibration phase.

4.3 Classical Molecular Dynamics Production - RMSD

During the initial phase of the simulation, spanning a duration of 200 ns, the system was allowed to evolve without the presence of an external electric field. This period served as a baseline for observing the behavior and dynamics of the ion channel structure in its natural state. Subsequently, from 200 ns to 461 ns, an external electric field was introduced with a strength of -23.06 kcal/mole, exerting a directional force from the intracellular domain towards the extracellular domain. To assess the structural stability and deviations of

the ion channel during the simulation, the RMSD analysis was employed. The RMSD value quantifies the average displacement of the ion channel's atoms from their initial positions over time. By aligning the ion channel structures to their starting configuration, it becomes possible to evaluate the extent of structural changes and fluctuations throughout the simulation. Snapshots of the trajectory were captured during the 461 ns production run. In Figure 4.3(a), the RMSD plot depicts the deviation of the aligned ion channel structure relative to its initial position. Notably, the observed continuous increase in

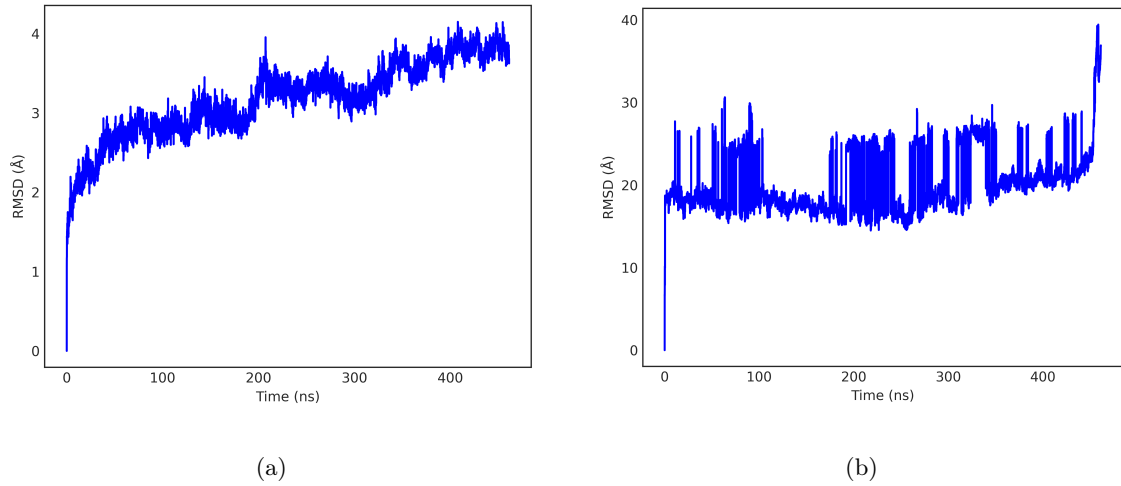


Figure 4.3: (a): RMSD of the aligned channel with respect to its initial position. (b): RMSD of aligned resid 61 to 66 with respect to its initial position.

RMSD of the protein during the entire simulation signifies that the protein structure failed to attain a stable and converged state. The RMSD value demonstrated a persistent and gradual increase, reaching up to 4 Å, which indicates that the system's geometry had not achieved complete equilibration. This behavior can likely be attributed to the application of an electric field intended to accelerate the movement of ions. This analysis provides valuable insights into the behavior of the ion channel under the influence of the external electric field, shedding light on its structural dynamics, stability, and the effects of the applied electric field on its conformational changes. Residues 61 to 66 are located in a specific region of the helix within the extracellular domain. Upon visual inspection, it was observed that these residues undergo significant motion. As a result, a more in-depth analysis of these particular residues will be conducted. The graph in Figure 4.3(b) provides insights into the dynamic behavior of these residues throughout the simulation period. During the initial timeframe of 0 to 450 ns, the graph exhibits a pattern of continuous

increase, specifically within the range of 20 Å to 30 Å. This trend indicates a significant motion of the analyzed residues during that period. However, beyond the 450 ns mark, a notable and sudden increase occurs, leading to a distinct change in the graph. The increase in RMSD values is indicative of deviations from the expected structural behavior of the protein. These deviations lead to the deformation of the protein and subsequently result in the crashing of the simulation.

4.4 Analysis of Ion Conduction Patterns: Insights from Graphical Representation

The present investigation focuses on comprehensively exploring the conduction behavior of cations (Na^+ , K^+ , Ca^{2+}) and anions (Cl^-) within a biological system through computational analysis. The primary objective of this analysis is to gain insights into the dynamic behavior and movement of cations and anions within the biological system. Specifically, the investigation seeks to elucidate the conduction mechanisms and associated processes by closely examining the behavior of these ions.

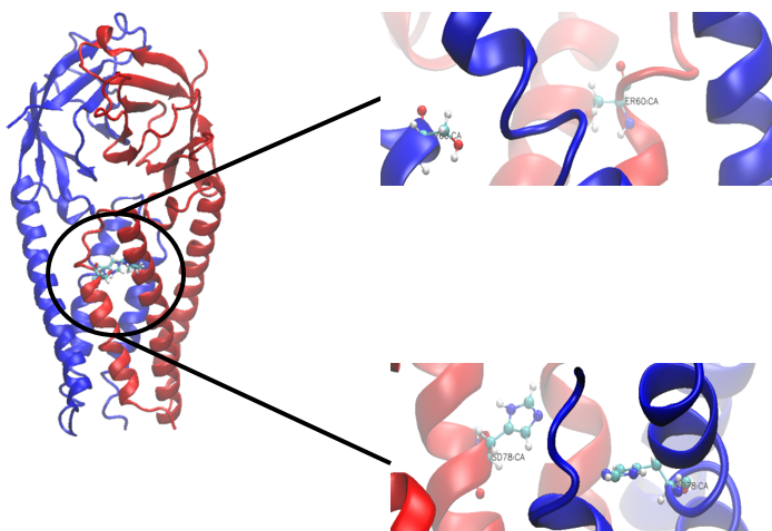


Figure 4.4: Representation of the C_α in a protein structure: with a focus on residues histidine (HSD) and serine (SER)

To achieve this objective, the analysis follows a systematic approach encompassing the following key steps:

1. Selection of cylindrical region: In order to analyze ion conduction in a specific region, a cylindrical area was defined within the pore by selecting the C_α atoms of residues HSD and

SER. This selection of residues was done in an arbitrary manner to establish the boundaries of the cylinder and allow for a concentrated investigation of ion conduction phenomena in this defined area.

2. Recording of ion positions: The analysis records the z-coordinate positions of the selected ions within the defined selection cylinder. Each ion is associated with its corresponding trajectory frame and z-coordinate value, enabling precise tracking of ion movement and comprehensive examination of their spatial distribution throughout the simulation.

3. Quantitative assessment of ion behavior: The analysis tracks the unique indices of the plotted ions and quantifies their total count. This count provides a measure of the number of ions involved in the conduction process. Additionally, the analysis determines the ionic flux, which represents the rate of ion movement within the system, calculated by dividing the ion count by the simulation time. These quantitative measures yield valuable insights into the efficiency and kinetics of ion conduction processes.

The analysis of the scatter plot generated from the simulation data provides valuable insights into the behavior of ions within the ion channel under the influence of the external electric field. The scatter plot represents the z-coordinate positions of ions as they traverse the system over a 460 ns simulation period.

In Figure 4.5, each color-dotted line represents the trajectory of an ion capable of traversing from the intracellular domain ($z > 40$) to the extracellular domain ($z < 40$). Upon careful examination of the scatter plot in Figure 4.5(a), a notable observation arises, the presence of an orange-dotted area, prominently centered around the 60 Å on the z-coordinate axis. This area indicates the occurrence of ion trapping or hindered movement within the ion channel, including the attempted entry of chloride ions around 100 ns. It suggests that the applied external electric field, along with other factors in the biological system, affects the smooth passage of Cl^- ions through the channel. Electrostatic interactions play a significant role in the observed trapping or hindered movement. Within the ion channel, charged residues or regions exert attractive forces on the Cl^- ions, leading to their temporary entrapment or hindered motion. These electrostatic interactions are influenced by the specific arrangement of charged amino acids or ions within the ion channel's environment. Furthermore, the channel's structural features, such as irregularities or constrictions, create energetic barriers or steric hindrances that impede the movement of ions. These structural features can result from amino acid residues, water molecules, or other biomolecular components comprising the ion channel.

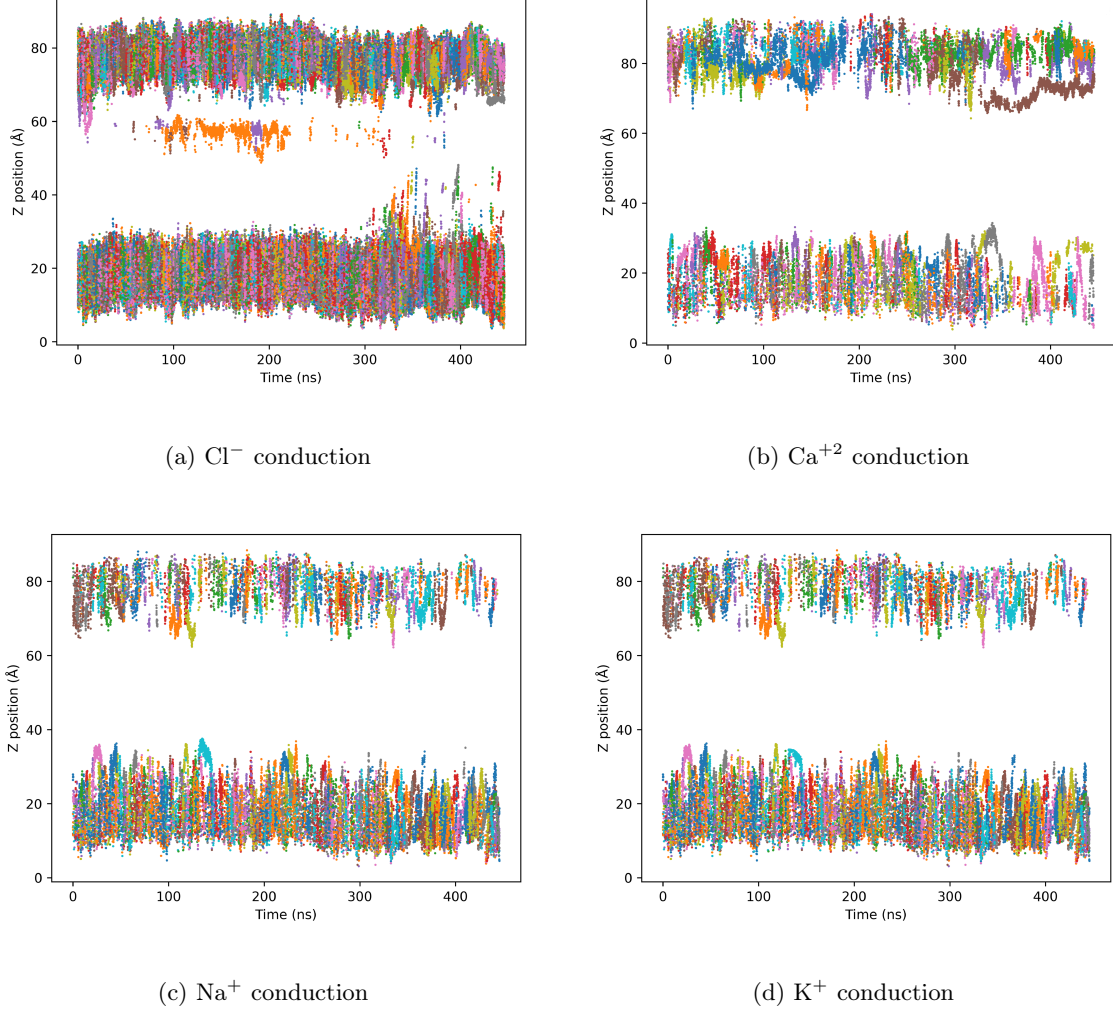


Figure 4.5: Permeability profiles: migration of cations and anions across the intracellular ($z > 40$) to extracellular ($z < 40$) regions.

Further analysis of the scatter plot (Figure 4.5(b)) reveals distinct patterns related to Ca^{+2} ions positioned around 80 Å along the z-coordinate axis. Two significant observations stand out: a continuous blue dotted line spanning from 100 ns to 200 ns and a separate brown dotted line occurring between 350 ns and 461 ns. These lines indicate intriguing phenomena, as they imply that the ions fail to enter the ion channel despite their proximity to the channel's gate during these time intervals. The persistent blue-dotted line signifies that, between 100 ns and 200 ns, the calcium ions consistently remain close to the ion channel's gate without progressing further inside. This observation suggests the existence of

a barrier or an energetically unfavorable threshold that prevents the ions from translocating across the gate and entering the channel. Similarly, the occurrence of the brown-dotted line between 350 ns and 461 ns suggests a similar phenomenon, where the Ca^{+2} ions fail to enter the ion channel despite their proximity to the gate.

Lastly, upon close examination of the scatter plot (Figure 4.5(c) and 4.5(d)), a significant observation arises of the absence of Na^+ and K^+ ions within both the proximity of the cavity and the ion channel throughout the simulation. This absence indicates distinct behavior compared to calcium ions and suggests the presence of specific barriers or restrictions preventing the presence or movement of ions near the cavity or within the ion channel.

The behavior of ions within an ion channel is influenced by multiple factors. One critical factor is the interplay of electrostatic interactions. Charged residues or regions within the ion channel exert attractive or repulsive forces on ions due to their charges. These electrostatic interactions are determined by the specific arrangement of charged amino acids or ions present in the ion channel environment. Furthermore, the structural features of the ion channel can be modulated by the applied electric field. The external electric field induces conformational changes in the channel's structure, leading to alterations in its geometry. These changes can create energetic barriers or steric hindrances, impeding the movement of ions. These barriers physically block the pathway of ions or impose conformational restrictions that hinder their entry into the channel. we hypothesize that the presence of an external electric field can be a factor inducing electrostatic repulsion between positively charged ions and the charged residues near the ion channel's gate. This repulsion, influenced by the external electric field, may contribute to the observed absence of cations within the channel but further study is required to investigate. However, it is crucial to acknowledge that the simulation encountered a critical issue resulting in the loss of protein integrity. This simulation failure likely affected the accurate representation of cations within the ion channel, ultimately leading to their absence in the depicted results presented in Figure 4.5(a).

As a consequence, the cations face difficulties in traversing the gate and entering the channel. The intensified repulsion acts as a hindrance, preventing the ions from passing through the gate and entering the channel. It is worth noting that the specific behavior of each ion may vary due to their unique properties and interactions with the charged residues near the gate. However, all cations experience increased repulsive forces that impede their entry into the ion channel.

4.5 Exploring Interactions of Positive and Negative Residues with Ions

Figure 4.6 visually depicts the spatial arrangement of Arginine (Arg), Lysine (Lys), Histidine (His), Aspartic acid (Asp), and Glutamine (Glu) residues within the ion channel. We specifically chose these residues in order to investigate the reasons behind the observed cation blockage and the absence of conduction within the channel. Within this selected group of residues, Arg, His, and Lys are negatively charged protein residues, while Asp and Glu are positively charged protein residues. In our simulation, we observed the sites of residue then the chosen negative residues have the ability to capture and restrict the movement of cations. By focusing on these residues, we can explore the interactions between the negative residues and cations. Once we observed the positions of positive residues on the protein, it helps to provide valuable insights into the specific residues that interact with anions.

The positively charged amino acids, namely Arg, His, and Lys, exhibit stronger coordination with Cl^- ions (Figure 4.7). Their positive charges enable favorable electrostatic interactions with the negatively charged ions. The coordination of Cl^- by Arg and Lys residues exhibit a distinct and pronounced peak, indicating on average, each Cl^- ion is coordinated by approximately 1.3 and 1 residues, respectively in Figure 4.7(a) and 4.7(c). This pronounced spike indicates a robust and specific interaction between chloride ions and these residues within that specific region of the ion channel in Figure 4.6. Moreover, The coordination of Cl^- by His residue (Figure 4.5) shows the interactions occurring in the extracellular domain, specifically around -40 \AA . In contrast, the negative residues Glu and Asp demonstrate stronger coordination with Ca^{+2} ions. The negative charges on these residues facilitate favorable electrostatic interactions with the positively charged Ca^{+2} ions, resulting in a higher affinity for coordination. On the other hand, Na^+ and K^+ ions display weaker coordination within the ion channel. This can be attributed to the smaller charges carried by Na^+ and K^+ ions compared to Ca^{+2} ions. Additionally, as discussed in the previous analysis, they are not able to approach the channel closely. Regarding the specific residues, Figure 4.7(b) illustrates the interactions of ions with the ASP residue predominantly in the intracellular region, approximately around 40 \AA . In Figure 4.7(d), the interactions of ions with the GLU residue are observed in both the intracellular region (around 40 \AA) and the extracellular region (-40 \AA). This suggests that the GLU residue contributes to ion interactions in both regions of the ion channel.

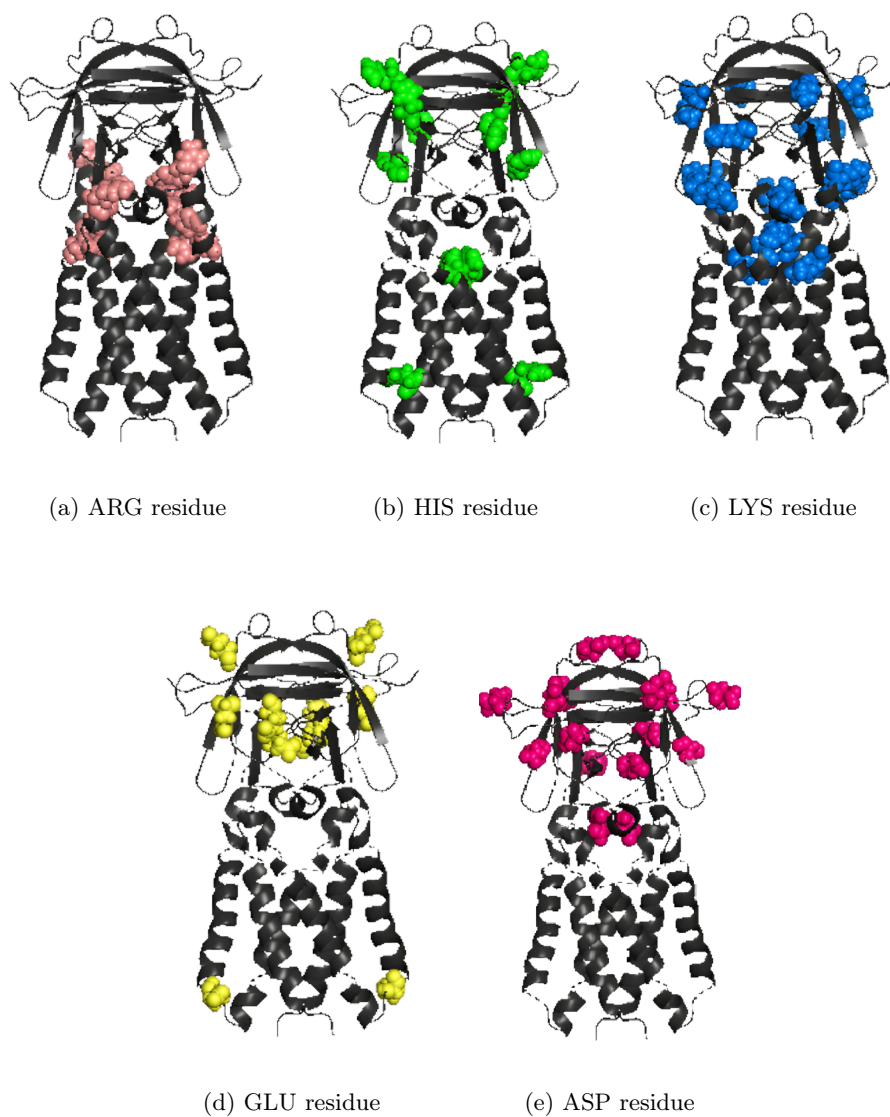
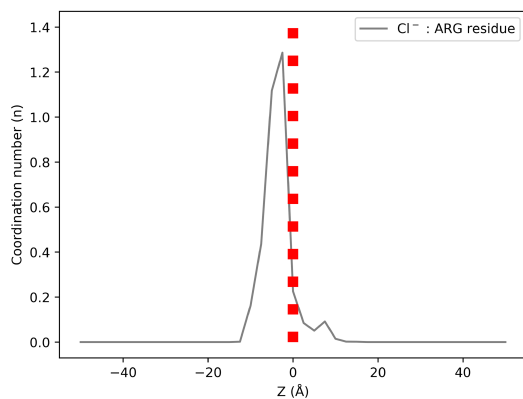
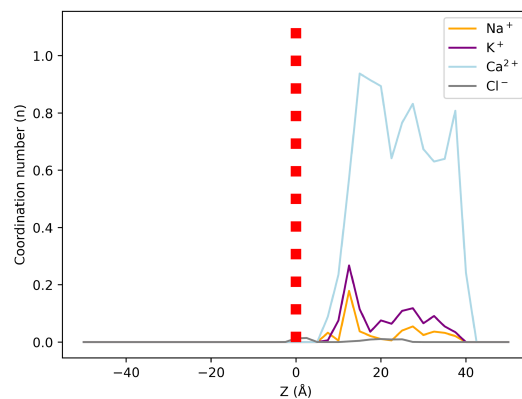


Figure 4.6: Spatial arrangement of key amino acid residues (Arg, His, Lys, His, Glu, Asp) in the ion channel.

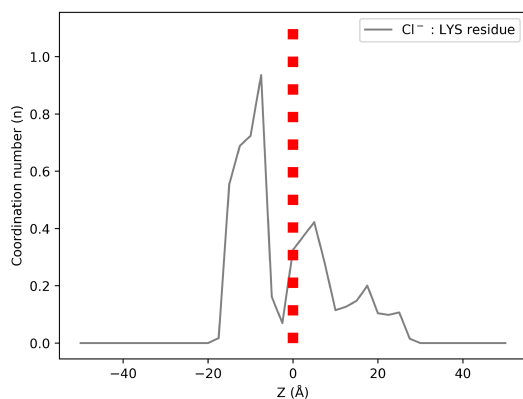
The coordination number, denoting the count of water molecules surrounding an ion within a defined range, serves as a quantitative measure for assessing the ion's solvation characteristics (Figure 4.7(f)). Upon analyzing the positions along the z-axis in the extracellular domain, specifically around -40 Å, a clear and discernible pattern becomes evident in the coordination numbers of ions. This pattern indicates the occurrence of interactions between ions and water molecules in that region. Between 0 Å and -20 Å, only Cl^- ions are able to traverse the channel, resulting in the observation of solvation exclusively for Cl^- . On the other hand, cations are unable to permeate, leading to a sudden drop in the coordination of water molecules to 0. Beyond the 0 Å along the z-axis, there is a sharp increase in coordination numbers indicating a greater propensity for these ions to interact with water molecules. Specifically, for anions, the coordination numbers exhibit a gradual increase from approximately 3 to nearly 6 which indicates the interaction of water molecules with Cl^- in the intracellular region around 40 Å.



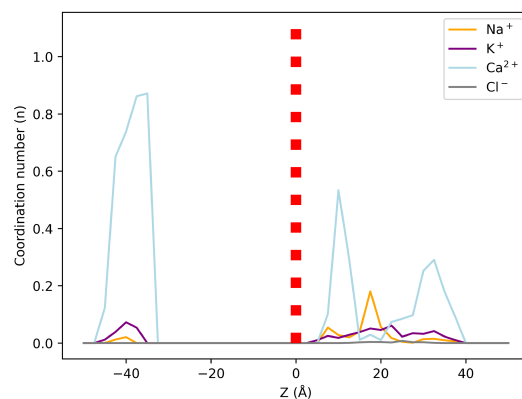
(a) Coordination number analysis of Cl^- ions with respect to ARG residue.



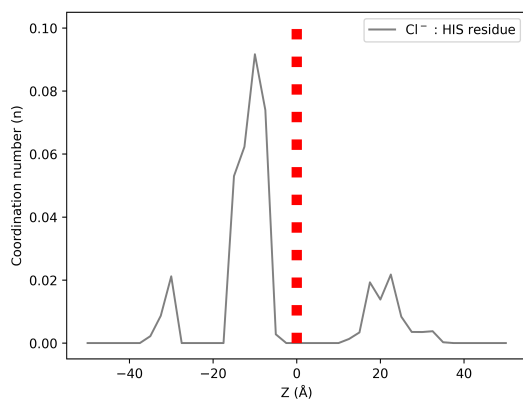
(b) Coordination number analysis of ions with respect to ASP residue.



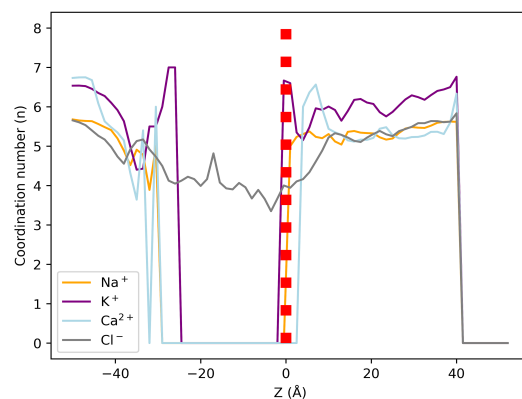
(c) Coordination number analysis of Cl^- ions with respect to LYS residue.



(d) Coordination number analysis of ions with respect to GLU residue.



(e) Coordination number analysis of Cl^- ions with respect to HIS residue.

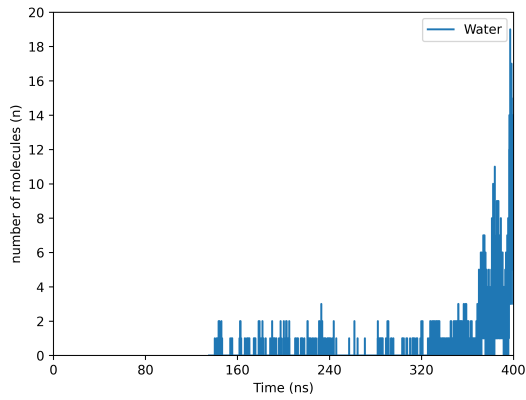


(f) Coordination number analysis of ions with respect to water molecules.

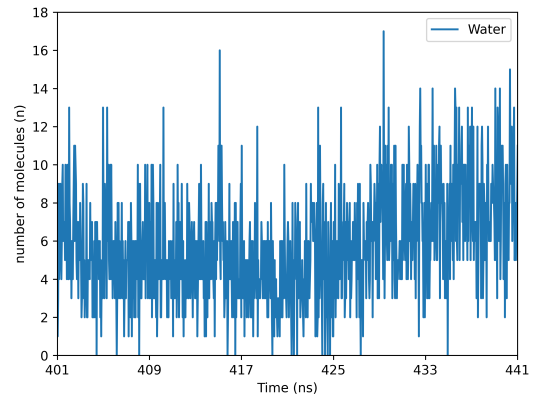
Figure 4.7: Exploring ion interactions with key amino acid residues: Arg, His, Lys, Glu, Asp residues within the ion channel.

4.6 Water Solvation along Conduction

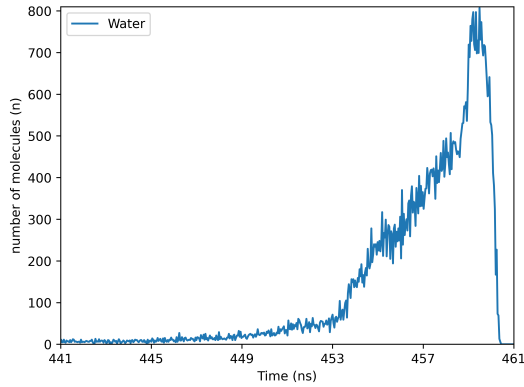
The primary aim of this study is to gain a comprehensive understanding of the behavior and motion of water molecules within a specific Z range. This is achieved by extracting the positional information of water molecules at each frame of the trajectory, enabling the tracking of their movements as the simulation progresses.



(a) Water molecule counts over time (0 ns to 400 ns).



(b) Water molecule counts over time (401 ns to 441 ns).



(c) Water molecule counts over time (441 ns to 461 ns).

Figure 4.8: Temporal variation of water molecule counts.

The plots depicted in Figure 4.8 offer a comprehensive analysis of the temporal dynamics of water molecule crossings through the channel. Initially, there is a gradual increase in

the number of water molecules, with approximately 20 molecules counted by the end of the 400 ns period. This indicates the occurrence of water molecule crossings within the specified Z range. Beyond the 441 ns mark, there is a continued rise in the number of water molecules, reaching a peak value of around 800. This peak signifies a substantial influx or accumulation of water molecules within the region of interest, reflecting a high rate of crossings. However, following this peak, there is a sudden and significant decrease in the number of water molecules, ultimately dropping to zero at approximately 461 ns. This sharp decline indicates the removal or depletion of water molecules within the region of interest. This phenomenon is visually represented in Figure 4.8(c), where the decline in the number of water molecules is clearly observed. The cause of this trend can be attributed to the deformation of the protein, which eventually leads to the simulation crashing. The structural changes and instability of the protein have an impact on the accurate representation of the ion channel, resulting in a reduced number of water molecule crossings.

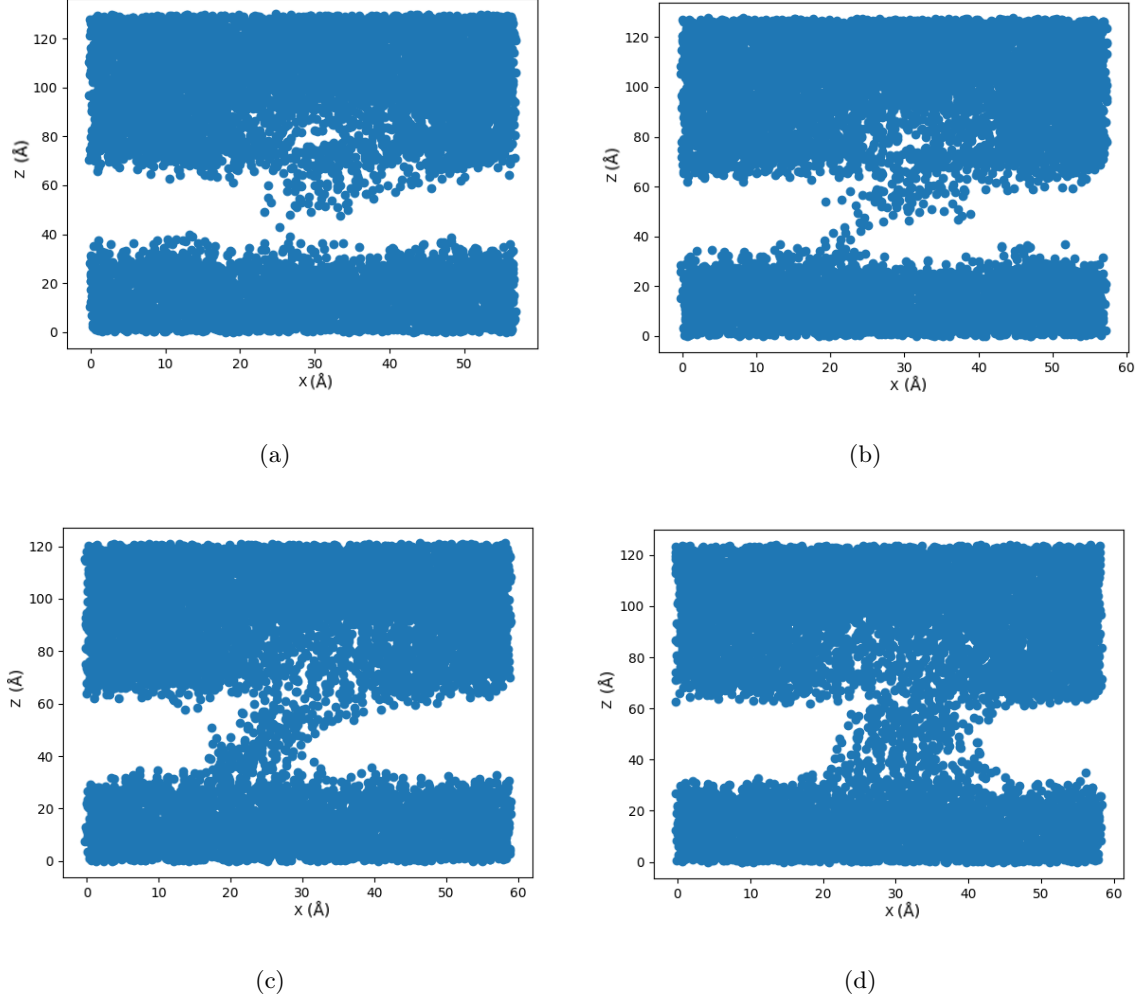


Figure 4.9: Spatial distribution of water molecules in the X-Z plane.

Figure 4.9 presents an analysis of water molecule behavior within a defined Z range throughout the simulation period. The individual frame plots offer a visual representation of the distribution and dynamics of water molecules in the X-Z plane, allowing for the observation of their spatial arrangement and movement. Additionally, the X-Z coordinate plot, depicted by blue dotted lines, highlights the spatial displacement of water molecules within a specific range of 20 \AA to 40 \AA along the X-axis. This plot effectively demonstrates the progressive nature of water molecule motion within this region as the simulation progresses. Notably, the plot reveals an increase in both the magnitude and frequency of water molecule displacements over time, providing valuable insights into their dynamic behavior

within the system.

Conclusion

In this thesis, our focus was on investigating ion conduction within the ion channel of SARS-CoV-2 ORF3a. Using classical molecular dynamics simulations spanning 461 ns, we aimed to analyze the presence and transport of cationic and anionic species within the ORF3a channel. A significant aspect of our study was understanding the interactions between ions and specific residues within the channel. By identifying positive and negative residues, we gained insights into the ion-residue interactions and their influence on ion conduction processes. Moreover, we observed that positive amino acids within the channel cavity effectively trapped anions, impeding their conduction. Conversely, negative residues showed a strong affinity for cations, hindering their movement through the channel. This unexpected behavior raises questions about the force field employed in our simulations, necessitating further exploration. Another intriguing finding was the observation of water molecule entry into the ion channel. The presence of water molecules in the channel suggested the possibility of ions traveling within the solvation shell along the channel. However, the expected conduction of cations was not observed until the protein structure underwent distortion. This distortion, likely caused by the application of a large electric field, ultimately resulted in the simulation crashing. Notably, during the simulation, we observed the transport of Cl⁻ ions and a substantial influx of water molecules.

In conclusion, this thesis contributes valuable insights into the process of ion conduction within the channel cavity. The unexpected behavior observed in ion transport, influenced by the presence of positive and negative residues, highlights the importance of further investigating the role of the force field.

References

- [1] YI ZHANG AND TATIANA G KUTATELADZE. **Molecular structure analyses suggest strategies to therapeutically target SARS-CoV-2.** *Nature communications*, **11**(1):2920, 2020. 1
- [2] CHERYL YI-PIN LEE, RAYMOND TP LIN, LAURENT RENIA, AND LISA FP NG. **Serological approaches for COVID-19: epidemiologic perspective on surveillance and control.** *Frontiers in immunology*, **11**:879, 2020. 1
- [3] DAVID E GORDON, GWENDOLYN M JANG, MEHDI BOUHADDOU, JIEWEI XU, KIRSTEN OBERNIER, KRIS M WHITE, MATTHEW J O’MEARA, VERONICA V REZELJ, JEFFREY Z GUO, DANIELLE L SWANEY, ET AL. **A SARS-CoV-2 protein interaction map reveals targets for drug repurposing.** *Nature*, **583**(7816):459–468, 2020. 1
- [4] NA ZHU, DINGYU ZHANG, WENLING WANG, XINGWANG LI, BO YANG, JINGDONG SONG, XIANG ZHAO, BAORYING HUANG, WEIFENG SHI, ROUJIAN LU, ET AL. **A novel coronavirus from patients with pneumonia in China, 2019.** *New England journal of medicine*, 2020. 1
- [5] WORLD HEALTH ORGANIZATION. (N.D.). **Who coronavirus (COVID-19) dashboard.** World Health Organization. <https://covid19.who.int/>. 1
- [6] ALEXANDRA C WALLS, YOUNG-JUN PARK, M ALEJANDRA TORTORICI, ABIGAIL WALL, ANDREW T MCGUIRE, AND DAVID VEESLER. **Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein.** *Cell*, **181**(2):281–292, 2020. 1, 2
- [7] BEN HU, HUA GUO, PENG ZHOU, AND ZHENG-LI SHI. **Characteristics of SARS-CoV-2 and COVID-19.** *Nature Reviews Microbiology*, **19**(3):141–154, 2021. 1, 2

-
- [8] LANYING DU, YUXIAN HE, YUSEN ZHOU, SHUWEN LIU, BO-JIAN ZHENG, AND SHIBO JIANG. **The spike protein of SARS-CoV—a target for vaccine and therapeutic development.** *Nature Reviews Microbiology*, **7**(3):226–236, 2009. 2
 - [9] FRANCIS K YOSHIMOTO. **The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19.** *The protein journal*, **39**(3):198–216, 2020. 2
 - [10] DAVID A JAMISON JR, S ANAND NARAYANAN, NÍDIA S TROVÃO, JOSEPH W GUARNIERI, MICHAEL J TOPPER, PEDRO M MORAES-VIEIRA, VIKTORIJA ZAKSAS, KESHAV K SINGH, EVE SYRKIN WURTELE, AND AFSHIN BEHESHTI. **A comprehensive SARS-CoV-2 and COVID-19 review, Part 1: Intracellular overdrive for SARS-CoV-2 infection.** *European Journal of Human Genetics*, **30**(8):889–898, 2022. 2
 - [11] WENTAO NI, XIUWEN YANG, DEQING YANG, JING BAO, RAN LI, YONGJIU XIAO, CHANG HOU, HAIBIN WANG, JIE LIU, DONGHONG YANG, ET AL. **Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19.** *Critical Care*, **24**(1):1–10, 2020. 2
 - [12] RAMESWARI CHILAMAKURI AND SAURABH AGARWAL. **COVID-19: characteristics and therapeutics.** *Cells*, **10**(2):206, 2021. 2
 - [13] FILIPPO SCIALO, AURORA DANIELE, FELICE AMATO, LUCIO PASTORE, MARIA GABRIELLA MATERA, MARIO CAZZOLA, GIUSEPPE CASTALDO, AND ANDREA BIANCO. **ACE2: the major cell entry receptor for SARS-CoV-2.** *Lung*, **198**:867–877, 2020. 2
 - [14] DAVID M KERN, BEN SORUM, SONALI S MALI, CHRISTOPHER M HOEL, SAVITHA SRIDHARAN, JONATHAN P REMIS, DANIEL B TOSO, ABHAY KOTECHA, DIANA M BAUTISTA, AND STEPHEN G BROHAWN. **Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs.** *Nature structural & molecular biology*, **28**(7):573–582, 2021. 2, 3, 4
 - [15] MART M LAMERS AND BART L HAAGMANS. **SARS-CoV-2 pathogenesis.** *Nature reviews microbiology*, **20**(5):270–284, 2022. 2

-
- [16] MELISSA R GITMAN, MARYIA V SHABAN, ALBERTO E PANIZ-MONDOLFI, AND EMILIA M SORDILLO. **Laboratory diagnosis of SARS-CoV-2 pneumonia.** *Diagnostics*, **11**(7):1270, 2021. 2
- [17] IGOR DE ANDRADE SANTOS, VICTORIA RIQUENA GROSCHE, FERNANDO RODRIGUES GOULART BERGAMINI, ROBINSON SABINO-SILVA, AND ANA CAROLINA GOMES JARDIM. **Antivirals against coronaviruses: candidate drugs for SARS-CoV-2 treatment?** *Frontiers in microbiology*, **11**:1818, 2020. 2
- [18] CARLOS CASTAÑO-RODRIGUEZ, JOSE M HONRUBIA, JAVIER GUTIÉRREZ-ÁLVAREZ, MARTA L DEDIEGO, JOSE L NIETO-TORRES, JOSE M JIMENEZ-GUARDEÑO, JOSE A REGLA-NAVA, RAUL FERNANDEZ-DELGADO, CARMINA VERDIA-BÁGUENA, MARIA QUERALT-MARTÍN, ET AL. **Role of severe acute respiratory syndrome coronavirus viroporins E, 3a, and 8a in replication and pathogenesis.** *MBio*, **9**(3):e02325–17, 2018. 3
- [19] MARTINA BIANCHI, ALESSANDRA BORSETTI, MASSIMO CICCOTZI, AND STEFANO PASCARELLA. **SARS-Cov-2 ORF3a: mutability and function.** *International journal of biological macromolecules*, **170**:820–826, 2021. 3
- [20] RITESH GORKHALI, PRASHANNA KOIRALA, SADIKSHYA RIJAL, ASHMITA MAINALI, ADESH BARAL, AND HITESH KUMAR BHATTARAI. **Structure and function of major SARS-CoV-2 and SARS-CoV proteins.** *Bioinformatics and biology insights*, **15**:11779322211025876, 2021. 3
- [21] DAVID M KERN, BEN SORUM, SONALI S MALI, CHRISTOPHER M HOEL, SAVITHA SRIDHARAN, JONATHAN P REMIS, DANIEL B TOSO, ABHAY KOTECHA, DIANA M BAUTISTA, AND STEPHEN G BROHAWN. **Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs.** *BioRxiv*, pages 2020–06, 2020. 3, 4, 20
- [22] RIMANSHEE ARYA, SHWETA KUMARI, BHARATI PANDEY, HIRAL MISTRY, SUBHASH C BIHANI, AMIT DAS, VISHAL PRASHAR, GAGAN D GUPTA, LATA PANICKER, AND MUKESH KUMAR. **Structural insights into SARS-CoV-2 proteins.** *Journal of molecular biology*, **433**(2):166725, 2021. 3
- [23] FRANCISCO J BARRANTES. **Structural biology of coronavirus ion channels.** *Acta Crystallographica Section D: Structural Biology*, **77**(4):391–402, 2021. 3

-
- [24] WEI LU, BO-JIAN ZHENG, KE XU, WOLFGANG SCHWARZ, LANYING DU, CHARLOTTE KL WONG, JIADONG CHEN, SHUMING DUAN, VINCENT DEUBEL, AND BING SUN. **Severe acute respiratory syndrome-associated coronavirus 3a protein forms an ion channel and modulates virus release.** *Proceedings of the National Academy of Sciences*, **103**(33):12540–12545, 2006. 4
 - [25] KIMBERLY A DOLAN, MANDIRA DUTTA, DAVID M KERN, ABHAY KOTECHA, GREGORY A VOTH, AND STEPHEN G BROHAWN. **Structure of SARS-CoV-2 M protein in lipid nanodiscs.** *Elife*, **11**:e81702, 2022. 4
 - [26] CONOR MCCLENAGHAN, ALEX HANSON, SUN-JOO LEE, AND COLIN G NICHOLS. **Coronavirus proteins as ion channels: current and potential research.** *Frontiers in Immunology*, **11**:573339, 2020. 4
 - [27] YONG DUAN, LU WANG, AND PETER A KOLLMAN. **The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation.** *Proceedings of the National Academy of Sciences*, **95**(17):9897–9902, 1998. 5
 - [28] RON O DROR, ROBERT M DIRKS, JP GROSSMAN, HUAFENG XU, AND DAVID E SHAW. **Biomolecular simulation: a computational microscope for molecular biology.** *Annual review of biophysics*, **41**:429–452, 2012. 5
 - [29] QINGFENG ZHONG, PRESTON B MOORE, DENNIS M NEWNS, AND MICHAEL L KLEIN. **Molecular dynamics study of the LS3 voltage-gated ion channel.** *FEBS letters*, **427**(2):267–270, 1998. 5
 - [30] KI RAMACHANDRAN, GOPAKUMAR DEEPA, AND KRISHNAN NAMBOORI. *Computational chemistry and molecular modeling: principles and applications*. Springer Science & Business Media, 2008. 6
 - [31] QINGFENG ZHONG, QING JIANG, PRESTON B MOORE, DENNIS M NEWNS, AND MICHAEL L KLEIN. **Molecular dynamics simulation of a synthetic ion channel.** *Biophysical journal*, **74**(1):3–10, 1998. 6
 - [32] RALF SCHNEIDER, AMIT RAJ SHARMA, AND ABHA RAI. **Introduction to molecular dynamics.** *Computational Many-Particle Physics*, pages 3–40, 2008. 6, 11
 - [33] HOLGER FEHSKE, RALF SCHNEIDER, AND ALEXANDER WEISSE. *Computational many-particle physics*, **739**. Springer, 2007. 6

-
- [34] WILLIAM C SWOPE, HANS C ANDERSEN, PETER H BERENS, AND KENT R WILSON. **A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters.** *The Journal of chemical physics*, **76**(1):637–649, 1982. 7
 - [35] LOUP VERLET. **Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules.** *Physical review*, **159**(1):98, 1967. 7
 - [36] WILLIAM W WOOD, JEROME J ERPENBECK, GEORGE A BAKER JR, AND JD JOHNSON. **Molecular dynamics ensemble, equation of state, and ergodicity.** *Physical Review E*, **63**(1):011106, 2000. 8
 - [37] K CHO AND JD JOANNOPOULOS. **Ergodicity and dynamical properties of constant-temperature molecular dynamics.** *Physical Review A*, **45**(10):7089, 1992. 8
 - [38] DAAN FRENKEL AND BEREND SMIT. **Molecular simulation: from algorithms to applications**, 2000. 9
 - [39] MICHAEL P ALLEN AND DOMINIC J TILDESLEY. *Computer simulation of liquids*. Oxford university press, 2017. 9, 10, 22, 23
 - [40] DAAN FRENKEL, BEREND SMIT, AND MARK A RATNER. *Understanding molecular simulation: from algorithms to applications*, **2**. Academic press San Diego, 1996. 9
 - [41] TETSUO TOMINAGA AND SIDNEY YIP. **Molecular dynamics study of pressure in molecular systems.** *The Journal of chemical physics*, **100**(5):3747–3756, 1994. 9, 10
 - [42] H BEKKER AND P AHLSTRÖM. **The virial of angle dependent potentials in molecular dynamics simulations.** *Molecular Simulation*, **13**(6):367–374, 1994. 10
 - [43] JAMES C PHILLIPS, ROSEMARY BRAUN, WEI WANG, JAMES GUMBART, EMAD TAJKHORSHID, ELIZABETH VILLA, CHRISTOPHE CHIPOT, ROBERT D SKEEL, LAXMIKANT KALE, AND KLAUS SCHULTEN. **Scalable molecular dynamics with NAMD.** *Journal of computational chemistry*, **26**(16):1781–1802, 2005. 10
 - [44] STEWART A ADCOCK AND J ANDREW MCCAMMON. **Molecular dynamics: survey of methods for simulating the activity of proteins.** *Chemical reviews*, **106**(5):1589–1615, 2006. 10, 15

-
- [45] ANDREW R LEACH. *Molecular modeling: principles and applications*. Pearson education, 2001. 11, 16, 22
 - [46] CHRISTIAN HOLM. **Simulation methods in physics 1**. *Institute for Computational Physics. University of Stuttgart*, **2013**, 2012. 11
 - [47] FARHANG RADJAI AND CHARLES VOIVRET. **Periodic boundary conditions**. *Discrete Numerical Modeling of Granular Materials*, pages 181–198, 2011. 12
 - [48] M. SCOTT SHELL. **Principles of modern molecular simulation methods — sites.engineering.ucsb.edu**. <https://sites.engineering.ucsb.edu/~shell/che210d/assignments.html>. 12, 22, 23
 - [49] M SRI HARISH AND PUNEET KUMAR PATRA. **Temperature and its control in molecular dynamics simulations**. *Molecular Simulation*, **47**(9):701–729, 2021. 13
 - [50] HANS C ANDERSEN. **Molecular dynamics simulations at constant pressure and/or temperature**. *The Journal of chemical physics*, **72**(4):2384–2393, 1980. 13
 - [51] HERMAN JC BERENDSEN, JPM VAN POSTMA, WILFRED F VAN GUNSTEREN, ARHJ DINOLA, AND JAN R HAAK. **Molecular dynamics with coupling to an external bath**. *The Journal of chemical physics*, **81**(8):3684–3690, 1984. 13
 - [52] JOS THIJSSSEN. *Computational physics*. Cambridge university press, 2007. 14
 - [53] MASATAKA YAMAUCHI, YOSHIHARU MORI, AND HISASHI OKUMURA. **Molecular simulations by generalized-ensemble algorithms in isothermal–isobaric ensemble**. *Biophysical reviews*, **11**:457–469, 2019. 14
 - [54] CUNKUI HUANG, CHUNLI LI, PHILLIP YK CHOI, K NANDAKUMAR, AND LARRY W KOSTIUK. **A novel method for molecular dynamics simulation in the isothermal–isobaric ensemble**. *Molecular Physics*, **109**(2):191–202, 2011. 14
 - [55] DAVID BROWN AND SYLVIE NEYERTZ. **A general pressure tensor calculation for molecular dynamics simulations**. *Molecular Physics*, **84**(3):577–595, 1995. 14
 - [56] SCOTT E FELLER, YUHONG ZHANG, RICHARD W PASTOR, AND BERNARD R BROOKS. **Constant pressure molecular dynamics simulation: The Langevin piston method**. *The Journal of chemical physics*, **103**(11):4613–4621, 1995. 14, 15

-
- [57] J LEE, X CHENG, JM SWAILS, MS YEOM, PK EASTMAN, JA LEMKUL, S WEI, J BUCKNER, JC JEONG, Y QI, ET AL. **BROOKS, CL, 3RD, MACKERELL, AD, JR., KLAUDA, JB & IM, W. 2016. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field.** *J Chem Theory Comput*, **12**:405–13, 2016. 15, 21
- [58] OLGUN GUVENCH AND ALEXANDER D MACKERELL. **Comparison of protein force fields for molecular dynamics simulations.** *Molecular modeling of proteins*, pages 63–88, 2008. 15
- [59] ANTHONY K RAPPÉ, CARLA J CASEWIT, KS COLWELL, WILLIAM A GODDARD III, AND W MASON SKIFF. **UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations.** *Journal of the American chemical society*, **114**(25):10024–10035, 1992. 15
- [60] PEDRO GONNET. **A simple algorithm to accelerate the computation of non-bonded interactions in cell-based molecular dynamics simulations.** *Journal of Computational Chemistry*, **28**(2):570–573, 2007. 16
- [61] JUN PEI, LIN FRANK SONG, AND KENNETH M MERZ JR. **Pair Potentials as Machine Learning Features.** *Journal of Chemical Theory and Computation*, **16**(8):5385–5400, 2020. 16
- [62] FRANZ SCHWABL. **The Coulomb Potential.** *Advanced Quantum Mechanics*, pages 161–180, 2005. 16
- [63] ABDULNOUR Y TOUKMAJI AND JOHN A BOARD JR. **Ewald summation techniques in perspective: a survey.** *Computer physics communications*, **95**(2-3):73–92, 1996. 17
- [64] BRAD A WELLS AND ALAN L CHAFFEE. **Ewald summation for molecular simulations.** *Journal of chemical theory and computation*, **11**(8):3684–3695, 2015. 17
- [65] A GRZYBOWSKI, E GWÓZDŹ, AND A BRÓDKA. **Ewald summation of electrostatic interactions in molecular dynamics of a three-dimensional system with periodicity in two directions.** *Physical Review B*, **61**(10):6706, 2000. 17

-
- [66] MARTIN FITZNER, LAURENT JOLY, MING MA, GABRIELE C SOSSO, ANDREA ZEN, AND ANGELOS MICHAELIDES. **Communication: Truncated non-bonded potentials can yield unphysical behavior in molecular dynamics simulations of interfaces.** *The Journal of Chemical Physics*, **147**(12):121102, 2017. 17
- [67] DAAN FRENKEL AND BEREND SMIT. *Understanding molecular simulation: from algorithms to applications*, **1**. Elsevier, 2001. 17, 18
- [68] MAHFOUD BELHADJ, HOWARD E ALPER, AND RONALD M LEVY. **Molecular dynamics simulations of water with Ewald summation for the long range electrostatic interactions.** *Chemical Physics Letters*, **179**(1-2):13–20, 1991. 17
- [69] KWANGHO NAM, JIALI GAO, AND DARRIN M YORK. **An efficient linear-scaling Ewald method for long-range electrostatic interactions in combined QM/MM calculations.** *Journal of Chemical Theory and Computation*, **1**(1):2–13, 2005. 17
- [70] DOMINIC J TILDESLEY AND MICHAEL P ALLEN. *Computer simulation of liquids*. Clarendon Oxford, 1987. 17
- [71] JAEWOON JUNG, CHIGUSA KOBAYASHI, KENTO KASAHARA, CHENG TAN, AKIYOSHI KURODA, KAZUO MINAMI, SHIGERU ISHIDUKI, TATSUO NISHIKI, HIKARU INOUE, YUTAKA ISHIKAWA, ET AL. **New parallel computing algorithm of molecular dynamics for extremely huge scale biological systems.** *Journal of computational chemistry*, **42**(4):231–241, 2021. 20
- [72] DJURRE H DE JONG, GURPREET SINGH, WF DREW BENNETT, CLEMENT ARNAREZ, TSJERK A WASSENAAR, LARS V SCHAFER, XAVIER PERIOLE, D PETER TIELEMAN, AND SIEWERT J MARRINK. **Improved parameters for the martini coarse-grained protein force field.** *Journal of chemical theory and computation*, **9**(1):687–697, 2013. 20
- [73] KENNO VANOMMESLAEGHE, ELIZABETH HATCHER, CHAYAN ACHARYA, SIBSANKAR KUNDU, SHIJUN ZHONG, JIHYUN SHIM, EVA DARIAN, OLGUN GUVENCH, P LOPES, IGOR VOROBYOV, ET AL. **CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.** *Journal of computational chemistry*, **31**(4):671–690, 2010. 21

-
- [74] JEAN-PAUL RYCKAERT, GIOVANNI CICCOTTI, AND HERMAN JC BERENDSEN. **Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.** *Journal of computational physics*, **23**(3):327–341, 1977. 22
- [75] ULRICH ESSMANN, LALITH PERERA, MAX L BERKOWITZ, TOM DARDEN, HSING LEE, AND LEE G PEDERSEN. **A smooth particle mesh Ewald method.** *The Journal of chemical physics*, **103**(19):8577–8593, 1995. 22
- [76] WILLIAM L JORGENSEN, JAYARAMAN CHANDRASEKHAR, JEFFRY D MADURA, ROGER W IMPEY, AND MICHAEL L KLEIN. **Comparison of simple potential functions for simulating liquid water.** *The Journal of chemical physics*, **79**(2):926–935, 1983. 22
- [77] ROBERT B BEST, XIAO ZHU, JIHYUN SHIM, PEDRO EM LOPES, JEETAIN MITTAL, MICHAEL FEIG, AND ALEXANDER D MACKERELL JR. **Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles.** *Journal of chemical theory and computation*, **8**(9):3257–3273, 2012. 22
- [78] IN SUK JOUNG AND THOMAS E CHEATHAM III. **Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations.** *The journal of physical chemistry B*, **112**(30):9020–9041, 2008. 22
- [79] ZHOUTING JIANG, LE YOU, WENHUI DOU, TINGTING SUN, AND PENG XU. **Effects of an electric field on the conformational transition of the protein: A molecular dynamics simulation study.** *Polymers*, **11**(2):282, 2019. 23

Appendix

NAMD input file

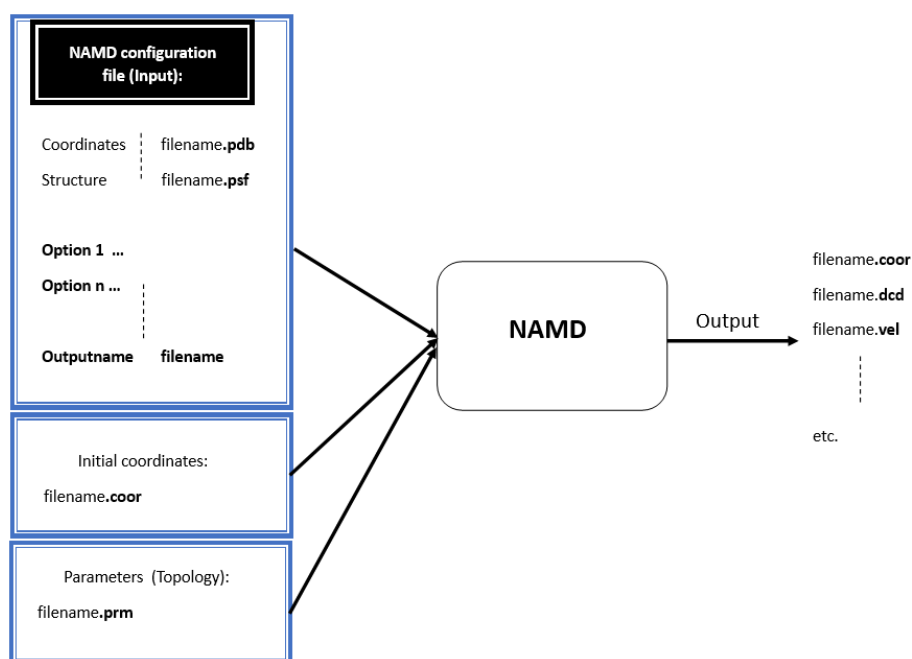


Figure 5.1: Scheme of NAMD program

NAMD is a popular molecular dynamics simulation program that requires various parameters to be defined for running simulations accurately. Here's a brief description of the important parameters you mentioned:

Structure and Coordinates:

This parameter specifies the molecular structure and initial conditions of the system under study. It typically involves providing a file that contains information about the atoms, bonds, angles, and other molecular properties.

binCoordinates, binVelocities, and extendedSystem:

```

1 structure          step5_input.psf
2 coordinates        step5_input.pdb
3
4 set temp            303.15;
5 outputName          MD_efield; # base name for output from this run
6                                # NAMD writes two files at the end, final
7                                # coord and vel
8                                # in the format of first-dyn.coor and
9                                # first-dyn.vel
10
11 set inputname        MD_efield;
12 binCoordinates       $inputname.restart.coor; # coordinates from last run (binary)
13 binVelocities        $inputname.restart.vel;  # velocities from last run (binary)
14 extendedSystem       $inputname.restart.xsc;  # cell dimensions from last run (binary)
15
16
17 dcdfreq              5000;
18 dcdUnitCell          yes;          # the file will contain unit cell info in the
19                                # style of
20                                # charmm dcd files. if yes, the dcd files
21                                # will contain
22                                # unit cell information in the style of
23                                # charmm DCD files.
24 xstFreq              5000;          # XSTFreq: control how often the extended
25                                # system configuration
26                                # will be appended to the XST file
27 outputEnergies        5000;          # 5000 steps = every 10ps
28                                # The number of timesteps between each energy
29                                # output of NAMD
30 outputTiming          5000;          # The number of timesteps between each timing
31                                # output shows
32                                # time per step and time to completion
33 restartfreq          5000;          # 5000 steps = every 10ps
34
35 # Force-Field Parameters
36 paraTypeCharmm        on;          # We're using charmm type parameter file(s)
37                                # multiple definitions may be used
38                                # but only one file per definition

```

Figure 5.2: Input parameter: part (1)

These parameters are binary output files generated from a previous run. `binCoordinates` contains the atomic coordinates at a particular time step, `binVelocities` contains the corresponding velocities, and `extendedSystem` contains additional information such as periodic boundary conditions.

dcdfreq, restartfreq, dcdUnitCell, xstreq:

These parameters determine the frequency at which certain output files are generated during the simulation. `dcdfreq` specifies the frequency of writing trajectory snapshots in the DCD (CHARMM) format, `restartfreq` specifies the frequency of writing restart files, `dcdUnitCell` controls the output of unit cell information in the DCD files, and `xstreq` specifies the frequency of writing coordinates and velocities in the XST format.

paraTypeCharmm:

This parameter is used to specify that the Charmm force field is being used in the simulation. The Charmm force field is a widely used set of parameters for describing molecular interactions in biomolecular systems.

```

1 # Nonbonded Parameters
2 exclude          scaled1-4      # non-bonded exclusion policy to use
3                               # "none,1-2,1-3,1-4,or scaled1-4"
4                               # 1-2: all atoms pairs that are bonded are going to
5                               # be ignored
6                               # 1-3: 3 consecutively bonded are excluded
7                               # scaled1-4: include all the 1-3, and modified
8                               # 1-4 interactions
9                               # electrostatic scaled by 1-4scaling factor 1.0
10                              # vdW special 1-4 parameters in charmm parameter file.
11 1-4scaling        1.0
12 switching         on
13 vdwForceSwitching on;          # New option for force-based switching of vdW
14                               # if both switching and vdwForceSwitching are on
15                               # CHARMM force
16                               # switching is used for vdW forces.
17
18 # You have some freedom choosing the cutoff
19 cutoff            12.0;        # may use smaller, maybe 10., with PME
20 switchdist        10.0;        # cutoff - 2.
21                               # switchdist - where you start to switch
22                               # cutoff - where you stop accounting for
23                               # nonbond interactions.
24                               # correspondence in charmm:
25                               # (cutnb,ctofnb,ctonnb = pairlistdist,cutoff,switchdist)
26 pairlistdist       16.0;        # stores the all the pairs with in the distance it should
27                               # be larger
28                               # than cutoff( + 2.)
29 stepspercycle      20;          # 20 redo pairlists every ten steps
30 pairlistsPerCycle  2;           # 2 is the default
31                               # cycle represents the number of steps between
32                               # atom reassignments
33                               # this means every 20/2=10 steps the pairlist
34                               # will be updated
35
36 # Integrator Parameters
37 firsttimestep      200000000;
38 timestep           2.0;         # fs/step
39 rigidBonds         all;         # Bound constraint all bonds involving H are fixed in length
40 nonbondedFreq      1;           # nonbonded forces every step
41 fullElectFrequency 1;           # PME every step
42
43 wrapWater          on;          # wrap water to central cell
44 wrapAll            on;          # wrap other molecules too
45 if { $boxtype == "hexa" } {
46   wrapNearest      on;          # use for non-rectangular cells (wrap to the nearest image)
47 } else {
48   wrapNearest      off;         # use for non-rectangular cells (wrap to the nearest image)
49 }
50
51 # PME (for full-system periodic electrostatics)
52 PME                yes;
53 PMEInterpOrder     6;           # interpolation order (spline order 6 in charmm)
54 PMEGridSpacing      1.0;        # maximum PME grid space / used to calculate grid size
55
56 # Constant Pressure Control (variable volume)
57 useGroupPressure    yes;         # use a hydrogen-group based pseudo-molecular
58                               # viral to calcualte pressure and
59                               # has less fluctuation, is needed for
60                               # rigid bonds (rigidBonds/SHAKE)
61 useFlexibleCell     yes;         # yes for anisotropic system like membrane
62 useConstantRatio    yes;         # keeps the ratio of the unit cell in the x-y plane
63                               # constant A=B

```

Figure 5.3: Input parameter: part (2)

exclude:

This parameter specifies the non-bonded exclusion policy.

The value "scaled1-4" indicates that non-bonded interactions between atoms involved in 1-4 interactions are scaled by a factor of 1.0.

It includes 1-3 interactions and modified 1-4 interactions in addition to excluding 1-2 interactions.

In the CHARMM force field, special 1-4 parameters are defined in the parameter file.

1-4scaling:

This parameter sets the scaling factor for 1-4 interactions.

A value of 1.0 means that the electrostatic interactions involved in 1-4 interactions are scaled by a factor of 1.0.

switching:

This parameter enables the use of force-based switching for van der Waals (vdW) interactions.

When enabled, CHARMM force switching is used for vdW forces.

vdwForceSwitching:

This parameter is a new option for force-based switching of vdW interactions.

If both "switching" and "vdwForceSwitching" are enabled, CHARMM force switching is used for vdW forces.

cutoff:

This parameter sets the cutoff distance for nonbonded interactions.

Interactions beyond this distance are not accounted for in the simulation.

A value of 12.0 Å is specified, but it can be adjusted based on the system and simulation requirements.

switchdist:

This parameter sets the distance where the switching of nonbonded interactions begins.

pairlistdist:

This parameter specifies the distance within which all pairs should be stored.

stepspercycle and pairlistsPerCycle:

These parameters control the frequency of updating the pairlists, which store the interacting pairs.

In this case, pairlists are updated every 20 steps, and the pairlistsPerCycle value is set to 2.

```

1 langevinPiston      on;          # Nose-Hoover Langevin piston pressure control
2 langevinPistonTarget 1.01325;    # target pressure in bar 1atm = 1.01325bar
3 langevinPistonPeriod 50.0;      # oscillation period in fs. correspond to
4                                # pgamma T=50fs=0.05ps
5                                # f=1/T=20.0(pgamma)
6 langevinPistonDecay  100.0;     # oscillation decay time. smaller value corresponds
7                                # to larger random
8                                # forces and increased coupling to the Langevin temp
9                                ↪ bath.
10 langevinPistonTemp  $temp;      # Equall or smaller than piston period
11                                # coupled to heat bath
12 # Constant Temperature Control
13 langevin            on;          # langevin dynamics
14 langevinDamping     1.0;        # damping coefficient of 1/ps (keep low)
15 langevinTemp        $temp;     # random noise at this level
16 langevinHydrogen    off;       # don't couple bath to hydrogens
17
18 constraints on
19 consRef restrain_ca.pdb
20 consKFile restrain_ca.pdb
21 consKCol B
22
23 #external field parameter
24 eFieldOn yes
25 eField 0.0 0.0 -30.06549
26 eFieldNormalized yes
27
28 #margin 10.0
29 # run
30 run                50000000;    # 100ns

```

Figure 5.4: Input parameter: part (2)

"Lagevin" Parameter:

to set NPT ensemble (both thermostat and barostat are set to on)