EFFICIENT PARTICLE-BASED SIMULATION OF ION CHANNELS

BY

DAVID D. MARREIRO

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering in the Graduate College of the Illinois Institute of Technology

Approved

Advisor

Chicago, Illinois
December 2006
ACKNOWLEDGMENT

I am thankful for my advisor, Dr. Marco Saraniti, whose guidance and support have enabled me to complete this research. I would also like to thank Dr. Shela Aboud for her assistance, as well as my labmates and friends for their opinions and stimulating discussions. Finally, I would like to thank my family for giving me the opportunity to pursue my education, and the support to successfully complete it.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENT</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xiv</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION TO ION CHANNEL SIMULATION</td>
<td>1</td>
</tr>
<tr>
<td>1.1. Ion Channels</td>
<td>1</td>
</tr>
<tr>
<td>1.2. Simulation Approaches</td>
<td>2</td>
</tr>
<tr>
<td>1.3. Simulation Issues</td>
<td>3</td>
</tr>
<tr>
<td>2. SYSTEM COMPONENTS</td>
<td>5</td>
</tr>
<tr>
<td>2.1. Aqueous Solutions</td>
<td>5</td>
</tr>
<tr>
<td>2.2. Cell Membrane</td>
<td>8</td>
</tr>
<tr>
<td>2.3. Ion Channel Protein</td>
<td>11</td>
</tr>
<tr>
<td>3. SYSTEM SIMULATION</td>
<td>16</td>
</tr>
<tr>
<td>3.1. Particle-Based Computer Simulation</td>
<td>16</td>
</tr>
<tr>
<td>3.2. The Computer Model</td>
<td>34</td>
</tr>
<tr>
<td>4. VALIDATION OF THE SIMULATION METHODOLOGY</td>
<td>46</td>
</tr>
<tr>
<td>4.1. Benchmarks for Bulk Electrolyte Solutions</td>
<td>46</td>
</tr>
<tr>
<td>4.2. Benchmarks for Inhomogeneous Systems</td>
<td>58</td>
</tr>
<tr>
<td>4.3. Validation of the Force Field Scheme</td>
<td>63</td>
</tr>
<tr>
<td>4.4. Validation of the Integration Schemes</td>
<td>75</td>
</tr>
<tr>
<td>5. CHARGE TRANSPORT SIMULATION IN ION CHANNELS</td>
<td>79</td>
</tr>
<tr>
<td>5.1. Measurements In Real Systems</td>
<td>79</td>
</tr>
<tr>
<td>5.2. Charge Transport Simulation in Ion Channels</td>
<td>83</td>
</tr>
<tr>
<td>5.3. Results</td>
<td>93</td>
</tr>
<tr>
<td>5.4. Discussion</td>
<td>100</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Ion Physical Parameters</td>
<td>64</td>
</tr>
<tr>
<td>4.2</td>
<td>Ion–Ion Interaction Parameters</td>
<td>64</td>
</tr>
<tr>
<td>4.3</td>
<td>Conductivity Computed for Euler, Verlet-like and PC Integration Schemes.</td>
<td>77</td>
</tr>
<tr>
<td>5.1</td>
<td>Experimental Data On OmpF Channel Electrophysiology</td>
<td>82</td>
</tr>
<tr>
<td>5.2</td>
<td>Diffusion Coefficient Values Used in This Work. The Superscript '∞' Refers</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>To Infinite Dilution Or Bulk Values, While the Reduced Diffusion Value D'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Is Taken From [62].</td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>Side Chains and pKa of the Titratable Residues [13].</td>
<td>103</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Geometry and Polarity of a Water Molecule.</td>
<td>6</td>
</tr>
<tr>
<td>2.2</td>
<td>Hydrogen Bond Between Two Water Molecules.</td>
<td>7</td>
</tr>
<tr>
<td>2.3</td>
<td>Schematic Structure of a Phospholipid Molecule.</td>
<td>9</td>
</tr>
<tr>
<td>2.4</td>
<td>Chemical Structure of Palmitoyl-Oleyl-Phosphatidylcholine and Schematic Representation.</td>
<td>10</td>
</tr>
<tr>
<td>2.5</td>
<td>Self-Organized Lipid Nanostructures.</td>
<td>10</td>
</tr>
<tr>
<td>2.6</td>
<td>Molecular Structure of a Generic Amino-Acid.</td>
<td>12</td>
</tr>
<tr>
<td>2.7</td>
<td>The OmpF Porin Channel From <em>E. Coli</em>, Embedded In an Explicit POPC Membrane (A), and the Corresponding Top View of the OmpF Alone (B). The Picture Is Rendered With VMD [20] From the PDB Database Structure File 2omf.pdb [21, 22].</td>
<td>15</td>
</tr>
<tr>
<td>3.1</td>
<td>Flowchart of a Particle-Based Simulation.</td>
<td>19</td>
</tr>
<tr>
<td>3.2</td>
<td>Charge Assignment Schemes Used for the PM Force Computation.</td>
<td>20</td>
</tr>
<tr>
<td>3.3</td>
<td>Steady-State Ionic Energy of a 0.30 mol.L(^{-1}) KCl Solution Obtained With Euler, Verlet-Like and Predictor/Corrector Integration Schemes, for Different Timestep Values Between 1 and 200 fs, Compared to the Theoretical Value for the Kinetic Energy.</td>
<td>31</td>
</tr>
<tr>
<td>3.4</td>
<td>Portion of the Finite Difference Grid Used To Solve Poisson’s Equation. Points Labeled With (n, s, e, w, f, b) Are Neighbors of the Point Labeled (c), (i.e.) the Center of the Central Cell. The Points Shown With Primes (\left(n', s', e', w', f', b'\right)), Are Used for Interpolations. Characteristics Distances Are Also Depicted (Deltas).</td>
<td>44</td>
</tr>
<tr>
<td>3.5</td>
<td>Pointers To Nearest Neighbors for the Dirichlet and the Neumann’s Boundary Conditions.</td>
<td>45</td>
</tr>
<tr>
<td>4.1</td>
<td>Relation Between RDF and Ion Distribution in a Homogeneous Medium.</td>
<td>49</td>
</tr>
<tr>
<td>4.2</td>
<td>RDFs Between K(^+) and Cl(^-) for KCl Solutions With Concentrations Ranging From 0.10 To 1.00 mol.L(^{-1}). The HNC Algorithm Has Been Used With a Lennard-Jones Potential.</td>
<td>50</td>
</tr>
<tr>
<td>4.3</td>
<td>RDF Curve and 3(^{rd})-Order Polynomial Fit, From a 0.30 mol.L(^{-1}) KCl Solution Simulation</td>
<td>53</td>
</tr>
</tbody>
</table>
4.4 Experimental Setup for Measuring the Conductivity of an Electrolyte Solution ........................................... 57

4.5 Ionic Concentration Profiles Obtained from the Simulation of a 0.5 M KCl Solution in the Vicinity of a Membrane ......................................................... 61

4.6 Mean Electrostatic Potential Obtained from the Simulation of a 0.5 M KCl Solution in the Vicinity of a Membrane ................................................. 62

4.7 Simulation Results Obtained for the Grid Size $\Delta G$ Ranging Between 0.30 and 5.0 nm. The Dashed Lines Represent the Benchmark Values [54]. ............................................................... 65

4.8 Simulation Results Obtained for the Free-Flight Timestep $\Delta t_{ff}$ Ranging Between 1 and 75 fs. The Dashed Lines Represent the Benchmark Values [54]. ............................................................................. 67

4.9 Simulation Results Obtained for the Poisson Timestep $\Delta t_{Poiss}$ Ranging Between 0.1 and 30 ps. The Dashed Lines Represent the Benchmark Values [54]. ............................................................................. 69

4.10 RDF for Different Concentrations of KCl and NaCl. Comparison With Results From the HNC Analytical Model (Solid Lines) Shows Excellent Agreement [29]. ............................................................... 71

4.11 Osmotic Coefficient Versus Concentration for (a) KCl and (b) NaCl [29]. Results Are Compared With Experimental Values and With Results from the HNC. The Experimental Results Shown Here Use Pitzer Experimental Fit Equations [38] With Parameters Taken from [55]. ............................................................................. 72

4.12 Average Concentration of Anions and Cations in a 0.15 mol.L$^{-1}$ KCl Solution With a 2 nm Dielectric Membrane in the Center [29]. .... 73

4.13 Average Concentration of Anions and Cations in a KCl Solution With a 2 nm Dielectric Membrane in the Center. The Concentrations Are 0.30 mol.L$^{-1}$ On the Left and 0.15 mol.L$^{-1}$ On the Right Side [29]. ............................................................................. 74

4.14 Comparison of Different Integration Schemes Over a Range of Integration Timesteps. The RDF Peak Value for a Bulk KCl Solution of 0.30 M Is Used as a Benchmark and Compared to the Analytical HNC Result [54]. ............................................................................. 76

4.15 DL Potential Drop for Various KCl Concentrations, Obtained With All Three Integration Algorithms, Compared With the URGMC Analytical Predictions. An Integration Timestep of 20 fs Is Used in the BD Simulations. ............................................................................. 78
4.16 DL Total Surface Charge for Various KCl Concentrations, Obtained With All Three Integration Algorithms, Compared With the URMGC Analytical Results. An Integration Timestep of 20 fs Is Used in the BD Simulations. 78

5.1 Patch Clamp Measurement Setup With Op-Amp Voltage Clamp and Current Amplification [3]. 80

5.2 System Model Schematic Showing Geometry and Boundary Conditions. 85

5.3 Map of Potential and Force Experienced By a K⁺ Ion Along a Plane Perpendicular To OmpF Pore Axis. 87

5.4 Map of Potential and Force Experienced By a Cl⁻ Ion Along a Plane Perpendicular To OmpF Pore Axis. 87

5.5 Minimal Potential Path for K⁺ Ion Along Porin Channel Pore. 88

5.6 Potential Profile for a K⁺ Ion Along the Minimum Constraint Path in the Porin Channel Pore. 89

5.7 Pore Openings Model and Current Recording Procedure. 90

5.8 Porin Trimer Conductance As a Function of the Free-Flight Integration Timestep, Compared With an Experimental Measurement (Dashed Line) [59]. The Bath Concentration Is 1.00 M and the Applied Trans-Membrane Potential Is 0.5 V. 93

5.9 Dielectric Map of the Protein Area As Represented in the Simulation. Section Along $y - z$ Plane, A) Without Smoothing, B) With Smoothing. $\epsilon_{\text{protein}}$ Is Set To 6 in Both Cases. 95

5.10 I-V Curves Obtained With a Smoothed Dielectric Representation, With a 0.25 M KCl Bath and Applied Trans-Membrane Potentials Varying Between $-200$ and $+200$ mV. Diamonds: $\epsilon_{\text{protein}} = 5$, Triangles: $\epsilon_{\text{protein}} = 20$. 96

5.11 Channel Conductances Due To K⁺, Cl⁻, and Trimer Total Conductance As a Function of the Protein Relative Dielectric. The KCl Concentration Is Set to 0.25 M and the Trans-Membrane Voltage Is 500 mV. 97

5.12 Trimer Conductance As a Function of Bath KCl Concentration. Circles: BD Simulation With $\epsilon_{\text{protein}} = 6$, Squares: BD Simulation With $\epsilon_{\text{protein}} = 20$, Triangles: PNP Simulation [59], Diamonds: Experimental Measurements [62]. 98
5.13 Trimer Conductance As a Function of Bath KCl Concentration. Circles: BD Simulation With Bulk Diffusion $D^\infty$, Squares: BD Simulation With Reduced Diffusion $D''$, Triangles: PNP Simulation [59], Diamonds: Experimental Measurements [62].

5.14 OmpF Trimer View Showing Some of the Charged Residues Lining the Constriction Region (Generated With VMD [20]).

5.15 Computed Total Net Charge of OmpF As a Function of the pH.

B.1 GROMACS Simulation Results for a Bulk KCl 1.0 M Solution. The RDF Shows an Excellent Agreement With the HNC Benchmark.

B.2 Map of the Force On a K$^+$ Ion Along a Plane Perpendicular To the Pores of OmpF Obtained With the GROMACS P$^3$M Algorithm, Using a Custom Script.

C.1 RDFs for a 0.15 mol.L$^{-1}$ KCl Solution Obtained from the HNC Integral Computation. The Interaction Potential Uses the Empirical Lennard-Jones Formula and a Value of $5 \times 10^8$ Is Used for the Splitting Parameter $\alpha$.

D.1 Ionic Concentration Profiles Computed With the Standard DL Theory for a 1–1 Electrolyte Bulk Concentration of 0.5 M and a Surface Potential of 30 mV.

D.2 Mean Electrostatic Potential Computed With the Standard DL Theory for a 1–1 Electrolyte Bulk Concentration of 0.5 M and a Surface Potential of 30 mV.

D.3 Charge Accumulation In the DL As a Function of the Bulk Electrolyte Concentration for Values of the Surface Potential Ranging Between 5 and 150 mV.

D.4 Potential Drop Across the DL As a Function of the Bulk Electrolyte Concentration for Values of the Surface Potential Ranging Between 5 and 150 mV.

D.5 Illustration of the Three Regions Used for the Solution of the URMGC Equations.

D.6 Ionic Concentration Profiles Computed With the URMGC Model for a 1–1 Electrolyte Bulk Concentration of 0.5 M and No Surface Charge.

D.7 Mean Electrostatic Potential Computed With the URMGC Model for a 1–1 Electrolyte Bulk Concentration of 0.5 M and No Surface Charge.
D.8 Charge Accumulation In the DL As a Function of the Bulk Electrolyte Concentration With No Surface Charge Present. . . . . . 141

D.9 Potential Drop Across the DL As a Function of the Bulk Electrolyte Concentration With No Surface Charge Present. . . . . . . . . . . 141
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Alternating Current</td>
</tr>
<tr>
<td>AMU</td>
<td>Atomic Mass Unit</td>
</tr>
<tr>
<td>BC</td>
<td>Boundary Conditions</td>
</tr>
<tr>
<td>BD</td>
<td>Brownian Dynamics</td>
</tr>
<tr>
<td>CIC</td>
<td>Cloud In Cell</td>
</tr>
<tr>
<td>DL</td>
<td>Double Layer</td>
</tr>
<tr>
<td>DNA</td>
<td>DeoxyriboNucleic Acid</td>
</tr>
<tr>
<td><em>E. Coli</em></td>
<td><em>Escherichia Coli</em></td>
</tr>
<tr>
<td>FMM</td>
<td>Fast Multipole Method</td>
</tr>
<tr>
<td>GNU</td>
<td>Gnu is Not Unix</td>
</tr>
<tr>
<td>GROMACS</td>
<td>GROningen MAchine for Chemical Simulations</td>
</tr>
<tr>
<td>HNC</td>
<td>HyperNetted Chain</td>
</tr>
<tr>
<td>IP</td>
<td>Inverse Power</td>
</tr>
<tr>
<td>KCl</td>
<td>Potassium Chloride</td>
</tr>
<tr>
<td>LJ</td>
<td>Lennard-Jones</td>
</tr>
<tr>
<td>M</td>
<td>Molar [mol.L⁻¹]</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>MGC</td>
<td>Modified Gouy-Chapman</td>
</tr>
<tr>
<td>MSA</td>
<td>Mean Spherical Approximation</td>
</tr>
<tr>
<td>NaCl</td>
<td>Sodium Chloride</td>
</tr>
</tbody>
</table>

xii
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGP</td>
<td>Nearest Grid Point</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>OmpF</td>
<td>Outer Membrane Protein F</td>
</tr>
<tr>
<td>OZ</td>
<td>Ornstein-Zernike</td>
</tr>
<tr>
<td>P³M</td>
<td>Particle-Particle-Particle-Mesh</td>
</tr>
<tr>
<td>PB</td>
<td>Poisson-Boltzmann</td>
</tr>
<tr>
<td>PBE</td>
<td>Poisson-Boltzmann Equation</td>
</tr>
<tr>
<td>PC</td>
<td>Predictor/Corrector</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PM</td>
<td>Particle-Mesh</td>
</tr>
<tr>
<td>PNP</td>
<td>Poisson-Nernst-Planck</td>
</tr>
<tr>
<td>POPC</td>
<td>Palmitoyl-Oleyl-PhosphatidylCholine</td>
</tr>
<tr>
<td>PP</td>
<td>Particle-Particle</td>
</tr>
<tr>
<td>PY</td>
<td>Perkus-Yevick</td>
</tr>
<tr>
<td>RC</td>
<td>Resistor-Capacitor</td>
</tr>
<tr>
<td>RCSB</td>
<td>Research Collaboratory for Structural Bioinformatics</td>
</tr>
<tr>
<td>RDF</td>
<td>Radial Distribution Function</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Square</td>
</tr>
<tr>
<td>TSC</td>
<td>Triangular Shaped Cloud</td>
</tr>
<tr>
<td>URMGC</td>
<td>Unequal Radius Modified Gouy-Chapman</td>
</tr>
<tr>
<td>VMD</td>
<td>Visual Molecular Dynamics</td>
</tr>
</tbody>
</table>
ABSTRACT

The main objective of this work is to demonstrate the validity of a Poisson Particle-Particle–Particle-Mesh (P³M) coupled with a Brownian Dynamics (BD) engine simulation tool in modeling charge transport in biological ion channels. The challenges of ion channel modeling are presented with the underlying physical considerations. The details of the P³M force field scheme and its implementation are presented. The BD algorithm and the various integration schemes for the particle dynamics are presented and compared. The numerical model proposed for the electrolyte solution, the membrane and the ion channel are discussed. A set of benchmarks are defined and motivated to validate the numerical representation of the system. Analytical models are proposed for the electrolyte solution in bulk and interfacial conditions: the Hypernetted Chain (HNC) approximation is used in conjunction with the Ornstein-Zernike integral equation theory to describe the electrolyte solution, and provide a comparison with the simulation results. Similarly, the Gouy-Chapman Double Layer (DL) theory is used to compute analytical benchmarks for the membrane-solution interface simulation. The P³M BD simulation of bulk electrolyte solutions and membrane-solution interfaces is validated by comparison with the proposed analytical benchmarks. The range of valid numerical parameters for the system is determined by defining and applying an error analysis methodology. Subsequently, the well-studied OmpF porin channel from bacterium *E. Coli* is used as a test case to validate the proposed charge transport simulation approach. Potential mapping of the pore is performed for different ion types. Dynamic charge transport simulations are performed, and macroscopic channel conductance values are extracted and compared with published experimental measurements as well as other numerical models. The applicability of the P³M BD simulation is discussed, and compared with other numerical models from the computational cost standpoint. Finally, ways to improve the algorithmic efficiency and accuracy of the simulation are introduced.
CHAPTER 1
INTRODUCTION TO ION CHANNEL SIMULATION

1.1 Ion Channels

Ion channels are proteins embedded in the lipid membranes of biological cells. The interaction between the channels and their environment is quite complex and is responsible for the regulation of the ionic flux across the membrane. Examples of the crucial biological functions accomplished by ion channels are the generation and transmission of potentials in nerves and muscles, as well as the hormone release from endocrine cells [1].

The first demonstration in 1976 [2] of a reliable experimental method for the detection of the current flow through individual ion channels started a string of several refinements of the experimental setup. These new experiments have been successfully applied to a variety of membrane and cell configurations, both in vivo and in vitro [3]. The progress in experimental techniques triggered an increasing theoretical effort aimed at the understanding of the role of ion channels in the physiology of complex biological systems and their general role in the complex electrical equilibrium between the cell and its environment. Besides the theoretical aspects, important pharmacological advances have arisen from the knowledge on ion channels [4]. The interest for ion channels is also high from the engineering viewpoint. Indeed, ion channels are actually envisioned as a key component of a new generation of biosensors [5] that could integrate the extreme sensitivity of ion channels with the processing capabilities of modern microelectronics [6]. The solid-state electronics community is well aware of the fact that the constant reduction of feature sizes in transistors and microelectronic circuits will soon be inadequate to satisfy the requirements of emerging technologies. A good solution would be to increase the complexity of the basic components rather than their speed, and this is where much can be learned from ion channels. The
full understanding of ion channels would allow for the modification of their design for new applications [5] or the manufacturing of analogous structures reproducing their functionalities [7]. These prospects have motivated increasing research efforts especially in the field of computer modeling, due to the availability of reliable protein structural data, reproducible experimental data, and high performing software and hardware computational machinery.

1.2 Simulation Approaches

A hierarchy of simulative approaches have been applied to the study of ion channels during the last two decades. The implemented theoretical models range from the simplest to the most complex and computationally challenging. This section will give a brief classification of those models.

The simplest representation for a membrane–channel system is a simple RC circuit [1], this simplistic approach is able to model some measurements of the transient response of cell membranes to current steps, and helps in characterizing the type and density of ion channels, together with some membrane properties.

The next level of simulation complexity involves the continuum models, such as the Poisson-Boltzmann approach [8] and the Poisson-Nernst-Planck method [9], which have been used to define the electrostatic landscape of ion-channel systems. In these models, the ion electrodiffusion is modeled in terms of continuous fluxes.

On the other hand, particle-based approaches analyze ion fluxes as individual particle trajectories. The particle-based approaches can use either Brownian [10] or Newtonian [11] dynamics to model the trajectories and are the most favored for an accurate yet practical model of ion channels.

The highest order level models the quantum mechanical properties of the entire system through the Schrödinger equation. This approach would allow an exhaustive
description of the atomic interactions and would be very successful in describing the
details of the electrostatic and finite-size effects taking place in the constricted pore
region. However, this method is still impractical because of the many computational
and theoretical challenges which will be briefly introduced in the next section.

The question that may be asked is: why modeling ion channels? While it is
ture that experimental results have been obtained in an accurate and reproducible
way and have provided insight into the inner works of ion channels, while X-ray
crystallography has provided increasingly accurate protein structures for a variety
of channels, a computer model still permits the realization of experiments otherwise
impossible. In a computer model, all the parameters can be easily modified, making
a very large variety of experiments possible. Voltages, currents, ion concentrations,
temperatures, and every significant physical parameter can be changed to produce
the desired experimental setting. Different channel structures can be simulated, mak-
ing the prediction of the behavior of mutants [7] or man-made nanotubes [12] and
other channel-like structures possible. Finally, the biggest advantage of computer
experiments is that results are obtained at different modeling levels, allowing the
reproduction of macroscopic voltages and currents as well as the details of atomic
interactions and trajectories. No real physical experiment is currently capable of ac-
accessing that kind of information, hence the increasing interest for computer models.

1.3 Simulation Issues

The computer simulation of ion channels is a very challenging field simply
because of the nature of the system studied. The first issue is the sheer complexity of
the system. A simple system made of one ion channel surrounded by a few nanometers
of lipid membrane and surrounded by water is composed by a number of atoms in
the order of $10^5$. The number of atom-atom interactions that have to be resolved is
therefore quite large.
The second challenge is the nature of the phenomenon of ionic charge transport itself. Most of the physics defining the characteristics of the transport occur in the confined pore region, over distances measured in Ångstroms. The complete charge transport process is measured in microseconds, yet the ion transport in the pore itself relies upon atomic motion details occurring in fractions of a femtosecond. A model that is successful in describing ionic transport in an ion channel system must therefore be able to model the entire system for a long time, while at the same time following the system dynamics with space and time resolution of fractions of Å and femtoseconds, respectively.

Therefore, the simulation of ion channels demands for very efficient computer algorithms. This work is devoted to the development and improvement of particle-based computer simulation algorithms for the study of ionic channel transport properties.

This document is organized as follows. The next chapter will present a description of the physical and chemical characteristics of the components involved in ion channel study: aqueous solutions, membranes and proteins. Chapter 3 will then describe the computer simulation approach proposed for an efficient model of ion channels. The numerical models chosen for the simulated systems are then described in detail. Chapter 4 is dedicated to the validation of the proposed approach through extensive comparisons and test simulations. In chapter 5 the simulation of the complete ion channel itself and the results obtained are presented and discussed. Future developments of this project and concluding remarks are then offered in the final chapter.
Ion channels serve as an interface between two different media in biological cells. They naturally embed themselves in the cell membrane separating the inside and the outside of the cell, and allow ions in or out under specific conditions. A study of ion channels has therefore to include all three elements (aqueous solution, membrane, ion channel) in their natural configuration. This chapter will describe in detail the nature and the physics of each component of the system: the aqueous solutions, the membrane, and the ion channel protein itself.

2.1 Aqueous Solutions

A comprehensive study of the physics of electrolyte solutions is a strong prerequisite for any work related to biological systems. This is for a very simple reason: all living creatures and their constitutive tissues are composed mostly of water [13]. Hence the most common medium encountered in these studies will be aqueous solutions composed of water and ions.

Water has a very particular set of physical characteristics which play an important role in many biological functions. The components of living cells interact extensively with water, and this interaction determines their shapes and characteristics. Water is a source of mechanical stability for biomolecules and much of the metabolic reactions and processes occur in an aqueous environment. Ion channels do not derogate from this rule, since their function is to actually connect two separate electrolyte solutions.

2.1.1 Water Molecules Are Polar. Water molecules are composed of two hydrogen atoms covalently bonded to an oxygen atom [14]. In each O–H bond, one electron comes from the oxygen atom and the other one from the hydrogen. The bond is not
symmetric because of the difference in the electronegativity of oxygen and hydrogen. Electronegativity is the tendency for atoms to attract electrons. Oxygen is very electronegative in comparison to hydrogen, which tends to lose its electron, forming a simple proton (H\(^+\)). The resulting O–H bond is therefore polarized, since the electron pair is on average located closer to the oxygen atom than the hydrogen. This results in a partial positive charge \(\delta^+\) on each H and a negative charge \(2\delta^-\) on the oxygen.

By itself, this property of the bonds does not explain the polarity of the water molecules. The permanent electrostatic dipole of the water molecule comes from the misalignment of the two O–H bonds. The two covalent bonds make an angle of 104.5° and their polarities form an electrostatic dipole (Figure 2.1) [13]. Water is therefore a polar solvent and any other compound with a net electrostatic dipole will be easily soluble in water. Examples are ammonia and ethanol, both of which have dipoles from polarized N–H and O–H bonds. The substances that readily dissolve in water are called hydrophilic. On the contrary, molecules that exhibit little or no polarization like hydrocarbon chains are less soluble in water, and are then qualified as hydrophobic.

![Figure 2.1. Geometry and Polarity of a Water Molecule.](image)

2.1.2 **Hydrogen Bonding In Water.** A direct consequence of the polar nature of water molecules is the strong attraction between these molecules. The partial charges on both O and H atoms in different molecules attract each other forming a
weak bond between them (Figure 2.2). This bond is called the hydrogen bond \[13\], and each water molecule can form up to four bonds with four neighbors. Hydrogen bonds are longer than the covalent O–H bond, and much weaker. The energy required to break a hydrogen bond is about 20 kJ mol\(^{-1}\), while breaking the O–H bond takes 460 kJ mol\(^{-1}\). In ice, each molecule is hydrogen-bonded to its four neighbors resulting in a regular hexagonal lattice structure. These bonds give ice its high melting point and rigidity. In liquid water however, hydrogen bonds continuously break and reform, making it fluid.

![Hydrogen Bond Between Two Water Molecules](image)

**Figure 2.2.** Hydrogen Bond Between Two Water Molecules.

### 2.1.3 Aqueous Solutions of Ionic Compounds

In the study of ion channels, the solute of interest is made by ions that are atoms or molecules with a net electrostatic charge. The structurally simplest cases are the mono-atomic ions, such as potassium (K\(^+\)) or chloride (Cl\(^-\)) ions. When a crystal of KCl dissolves in water, each ion on the surface of the crystal interacts strongly with the polar water molecules. The electrostatic interaction is strong enough for an ion to break loose from the crystal, and move freely in the solution where it becomes surrounded by a solvation shell of water molecules. This shell is structurally stable because of the electrostatic interaction between the ion and the partial charges on the water molecules. Therefore, a positive ion such as K\(^+\) will attract the oxygen side of the water while the negative Cl\(^-\) will be surrounded by the H ends of the molecules. These hydration shells screen the
K$^+$–Cl$^-$ electrostatic interaction and allow the ions to separate and move freely in the solution. This dissolution process continues until an equilibrium is reached. At that point the solution is saturated, i.e. there are enough free ions in the solution to screen the effect of water on the remaining crystals, preventing their dissolution. Typical saturation concentrations do not occur normally in biological systems, and the next sections will focus only on moderately concentrated ionic solutions.

2.2 Cell Membrane

The boundaries of cells are formed by biological membranes preventing molecules inside to leak to the outside, and also unwanted substances from the outside to enter the cell. Cell membranes contain transport systems to control the flow of chemicals in and out of the cell. Some of the transport processes are produced by ion channels. To better understand how the channels are inserted in the membrane, and how the membrane structure is stabilized in water, it is important to understand the chemistry of the membrane molecules themselves.

Biological membranes are composed of lipid molecules. This class of substances contains three major types of molecules, of which the phospholipids are the most important. The other two groups, glycolipids and cholesterol realize some specific functions and are not as abundant in cell membranes. This section will elaborate on the molecular structure of phospholipids, and how they assemble to build cell membranes.

A phospholipid is an *amphipathic* molecule, that is a molecule that contains both a hydrophilic and a hydrophobic section (see Section 2.1.1). The hydrophobic part of the lipid molecule is composed by fatty acid chains. In phospholipids, two fatty acids are attached through ester bonds to the first two alcohol groups of a glycerol molecule [13]. The third alcohol group of the glycerol is attached to another
molecule through a phosphate group. Figure 2.3 shows the schematic structure of a phospholipid [15].

![Figure 2.3. Schematic Structure of a Phospholipid Molecule.](image)

The phospholipids therefore share the same common structure, with two apolar hydrocarbon chains forming the hydrophobic side of the molecule, and a strongly polarized and hydrophilic side with the charged phosphate and amine groups. Within the phospholipid group, the palmitoyl-oleyl-phosphatidylcholine (POPC) is one of the most commonly found in the membranes of bacteria like *E. Coli*. This work will use POPC molecules as building blocks of the cell membrane for the systems studied. Figure 2.4 shows the complete chemical structure of a POPC molecule and its schematic representation.

The left side of the structure is composed of the two hydrocarbon chains coming from the fatty acid molecules palmitate and oleate. They are bonded to a glycerol molecule (center) attached to the choline molecule through a phosphate group. The phosphate and tertiary amine groups bear net charges, making this end of the molecule stable in a polar solvent like water. The phospholipid molecules will therefore orient themselves naturally to present this end to the water and spontaneously build self-organized nanostructures like micelles and membranes [13] (figure 2.5).
A biological membrane is constructed by two parallel layers of lipids, with their hydrophilic heads pointing to the outside of the membrane, in contact with the aqueous solution. The hydrophobic tails are located in the interior of the membrane. The POPC bilayer is a very stable [16] structure that quickly reassembles if disturbed in normal biological conditions. The typical thickness of such a membrane is about 5 nm. The non-polar interior of the membrane is impermeable to water and ions. The membrane is therefore a strong dielectric barrier that can sustain an electric
potential. These trans-membrane potentials play a role in the ionic transport across the membrane through the channels.

2.3 Ion Channel Protein

Ion channels perform many highly specialized functions in cells. They are therefore classified in a number of families depending on which kind of ions they transport. Ion channels present three main properties: the permeation is the ability to allow ions to cross the strong dielectric barrier due to the lipid membrane, the selectivity is the capability of selecting one specific type of ions and not letting others go through the channel, while gating refers to the ability of the channel of modulating the ionic flow depending on external stimuli such as the trans-membrane voltage or the binding of an activating substance to the channel. This section will describe the chemical structure of ion channels, how they are organized and how they work. Finally, the ion channel used to evaluate the proposed simulation techniques is presented.

2.3.1 Protein Biochemistry. Ion channels are proteins, as such they are large macromolecules synthesized by the cells from small building blocks called amino-acids. All organisms use the same 20 amino-acids as building blocks to assemble their proteins. These 20 amino-acids are derivatives of carboxylic acids and share a common basic structure centered around a carbon atom called the α-carbon. This atom is bonded to an amine group $\text{H}_3\text{N}^+$, a carboxylic group $\text{COO}^-$, a hydrogen atom, and a side chain which is different for every amino-acid [13]. Figure 2.6 shows the generic structure of an amino-acid. The side chain is represented by the symbol R.

The amino-acids are bound together sequentially to form a peptide chain through a bonding process following a specific sequence coded in the DNA [17]. Therefore, the chemical composition of a protein is completely determined by the
sequence or the order in which the amino-acids are connected. This sequence is called the primary structure of the protein. The long peptide chain is a succession of molecule elements called residues, coming from the original amino-acids. Each residue presents a side-chain with specific chemical functional groups depending on the original amino-acid. These groups can be polar or apolar, and in some cases bear a net electric charge. Their organization in the chain form the basis for both the structure of the protein and its functionality when side-chains are accessible on the surface of the protein.

The 3D structural domains of the protein, called secondary structures, are obtained after the peptide chain folds into specific shapes. The protein folding is a process in which the disorganized peptide chain structures itself by reaching an energetically favored conformation, stabilized by several interactions between different segments of the chain. These interactions depend on the residues present, and can be hydrogen bonds, salt bonds (charge-charge interactions) or even covalent bonds (disulfide bonds). Although each interaction is weak when taken individually, the addition of many such interactions produces very stable structures. The resulting native conformation of the protein is much less sensitive to degradation than the unfolded chain. Two main secondary structural domains are encountered: \( \alpha \)-helices and \( \beta \)-sheets. As their names suggest, the former is a tubular structure formed by the folding of the chain into a helix whereas the latter is a planar sheet formed by parallel stretches of the chain bound together. The global assembly of these structural elements forms the 3D structure of the protein called tertiary structure.
Proteins can also exhibit a quaternary structural order. This is the case when several subunits, each made of a single peptide chain, bind together to form the protein. This is the case of the OmpF porin channel described later in this section.

2.3.2 Structure and Function. Ion channel proteins are composed of an arrangement of one or more cylinders forming pores that span the membrane. Ions crossing the channel will travel through these pores. While it is clear that the 3D structure of the protein is crucial in defining its function, the various properties of the ion channels such as selectivity or gating involve more complex structural properties than the static shape of the protein. Many ion channels feature a selectivity filter which is the basis of their ability of restricting the ionic flow to a particular ion type. The selectivity filter is usually made by an arrangement of side chains located inside the pore, providing the specific electrostatics and geometry needed to transport some ion species and block others. An accurate knowledge of the protein structure is therefore necessary for the understanding of the ion channel functionality.

The protein structures are obtained through X-ray scattering experiments on crystals of pure proteins. The data collected are correlated to the amino-acid sequence to determine the actual spatial arrangement of the residues and the 3D protein structure. The mapped proteins are available in an open library called the Protein Data Bank (PDB) [18]. Appendix A details the data formats used in the PDB and how to manipulate them to model protein structures in the simulation programs.

2.3.3 The *E. Coli* OmpF Channel. This paragraph is dedicated to the description of the ion channel chosen as a test system for the evaluation of the ion channel simulation techniques developed in this work. This channel is a porin channel from the bacterium *E. Coli*. It is a relatively large polypeptide made of three identical subunits containing 340 amino-acid residues each. OmpF is therefore a trimer, where each monomer is a hollow cylindrical barrel structure formed by 16 anti-parallel β-
strands. The structure also shows chain loops creating a constriction region inside the barrel. This constriction actually reduces the pore diameter to approximately 6 Å and forms the selectivity filter of the porin channel. Figure 2.7 shows side and top views of the trimer. The inner loops in the pore region are represented as shaded cylinders. It is important to note that the charged residue side chains in the constriction region give rise to strong electric fields that play a crucial role in the channel ionic transport properties. Finally, the ionization states of these residues change with the pH of the solution, indicating that the porin might operate as a pH-gated channel under certain conditions [19].
Figure 2.7. The OmpF Porin Channel From *E. Coli*, Embedded In an Explicit POPC Membrane (A), and the Corresponding Top View of the OmpF Alone (B). The Picture Is Rendered With VMD [20] From the PDB Database Structure File 2omf.pdb [21, 22].
CHAPTER 3
SYSTEM SIMULATION

The goal of a simulation code is to reproduce and predict the behavior and characteristics of a real system by implementing a physical model. The accurate reproduction of time-dependent properties of the system and its response to externally imposed conditions is the desired result of a successful simulation.

3.1 Particle-Based Computer Simulation

The study of microscopic systems such as ion channels involves modeling the behavior in time of the structure and its full environment. In the case of microscopic systems it makes sense to use a particle representation, as the system is naturally divided into molecules and atoms, i.e. “particles”, which are subunits suitable for a numerical representation. The whole system is then represented by a set of particles interacting with each other and with their environment. For instance, the E. Coli OmpF membrane channel in its biological environment contains 10,344 atoms [21], and a complete simulation must include the membrane molecules, the solvent, and the ions, raising the total number of simulated particles to hundreds of thousands.

The particle-based simulation tool described in this work is self-consistent. This means that the forces due to the interactions between the components of the system strictly depend on the spatial configuration of the components themselves. Therefore, the position of the particles and the force field they generate must be updated as the dynamics of the system evolves during the simulation.

The force field scheme consists in the algorithm and the parameters used for the computation of the force exerted on each particle due to the effect of all the other particles and the externally imposed boundary conditions. Various types of interactions exist between the components of the system, and they have to be modeled
in a computationally efficient manner. In particular, these interactions are expressed by electrostatic forces generated by charged particles or externally applied fields, by non-electrostatic forces such as van der Waals forces (see Section 3.2.1.1), or by chemical bonding.

The simplest way to compute the total electrostatic force acting on a particle is to loop over all the charged particles within the system, and determine for each one the influence of every other particle. This process is implemented in the so-called Particle–Particle (PP) algorithm, scales as $N^2$ [23], and its computational cost is prohibitive in the case of biological systems because of the long-range nature of the electrostatic interactions and the resulting large number of neighbors needed to be accounted for in the force calculation. Therefore, other approaches must be employed to produce a computationally efficient model.

An appealing alternative is to compute long-range forces on a grid, with the so-called Particle–Mesh (PM) approach. In other words, the simulated system is mapped on a discrete grid, and the forces are not computed at the position of each particle, but rather at each grid-point. The force calculated on the grid points is then interpolated at each particle position to obtain the force on the particle. This method is much faster than doing a direct computation of the force by summing the contributions of all particles in the system, and scales with the grid size rather than the particle number. The main shortcoming of this method is that variations of forces within a shorter range than the mesh size are not represented. In the case of systems where short-range interactions are important, such as the inter-molecular van der Waals forces, the PM method alone is not appropriate. In fact, since the resolution of the PM approach is dictated by the size of the grid cell, it would require an extremely fine grid to account accurately for short-range effects, thereby canceling the advantage of the method in terms of computational cost.
The option we have chosen is therefore to combine the PP and PM methods in a unified Particle-Particle-Particle-Mesh (P$^3$M) algorithm. Two main methods have been proposed for the computation of the force field schemes in biological systems. They are the Fast Multi-pole Method (FMM) [24, 25] and the P$^3$M method [23]. The latter option has been chosen because it has the ability to efficiently model the electrostatics of inhomogeneous systems with arbitrary boundary conditions. The Ewald summation method [26] solves the PM component in reciprocal space. One of the requirements for the Ewald method is the spatial periodicity of the charge distribution, which is at odds with the inhomogeneous nature of the ion channel systems. Studying ion channels requires the ability to handle non-periodic boundaries with externally applied potentials, which is incompatible with the Ewald method [27]. On the other hand, a real space Poisson solver easily fulfills these requirements. Even though the use of this technique has been limited in the past due to the assertion that the self-consistent solution of Poisson’s equation in particle-based simulation of ion channel systems is computationally prohibitive [28], it was proved to be applicable to this kind of simulations [29].

Within the P$^3$M method, the force on each particle is decomposed into two components [23]:

$$\vec{F}_i = \vec{F}_i^{PM} + \vec{F}_i^{PP},$$

(3.1)

where the long-range component $\vec{F}_i^{PM}$ of the total force $\vec{F}_i$ on a particle $i$ is obtained from the solution of Poisson’s equation on a grid in the direct (or position) space, and accounts for all the particles in the system far from particle $i$ as well as for the externally imposed boundary conditions and dielectric discontinuities (see Section 3.2.4).

Figure 3.1 shows the typical flowchart of a particle-based simulation. The iterative part of the simulation reproduces the evolution of the system with a sequence of simulation steps, and contains three parts: 1) the computation of the charge dis-
3.1.1 Charge Assignment. The charge assignment step consists in assigning a charge value to each point of the grid according to the distribution of the charged particles close to it. The relation between the particles charges and the charge assigned to each grid-point is described by a charge assignment scheme. Hockney [23] discusses a classification of such schemes with increasing order. Within these schemes, a spatial distribution of charge is set for each particle, and is used when assigning charge to the grid-point. The higher the order of the scheme, the larger and more complex the
extension of the particle charge distribution, and therefore the smoother the overall charge distribution of the system becomes. However, the computational cost of the charge assignment scheme also increases with the order of the scheme.

The simplest charge assignment scheme is the so-called Nearest Grid Point (NGP) scheme and is based on assigning the whole charge of a particle to the nearest mesh point. This approach is extremely efficient since only one mesh point is processed for each particle, but the charge distribution obtained on the grid is not a smooth function. Within the next higher order scheme the particle charge is distributed uniformly on 8 neighboring cells (in 3-D). This method is called Cloud In Cell (CIC) and represents a compromise between efficiency and smoothness of the charge density by assigning to the particle an homogeneous “cloud” of charge with a spatial extension.

Figure 3.2. Charge Assignment Schemes Used for the PM Force Computation.
of one grid cell. The last scheme that we implemented and tested is the so-called Triangular Shaped Cloud (TSC) method. It uses a linearly decreasing charge density cloud, with an extension of two grid cells in each dimension. The particle charge is therefore spread over 27 neighboring cells (in 3D) making the mesh charge density much smoother than the one obtained with the NGP and CIC schemes. Figure 3.2 summarizes the three charge assignment schemes in 2D.

The ionic solutions of interest for this work are very diluted, consequently the average number of particles in each cell for a typical mesh (see Section 3.2) is very low. This is why the TSC assignment scheme proved to give the most accurate results in accordance with the considerations exposed in [23]. Once a charge shape has been chosen for the particles, the corresponding weighting function is determined by,

\[
W(\vec{r} - \vec{r}_p) = \int_{V_p} S(\vec{r} - \vec{r})d\vec{r},
\]

where the function \( S(\vec{r}) \) represents the shape of the charge “cloud” associated with the particle and \( V_p \) is the volume of the grid cell. In one dimension, the weighting functions computed from equation 3.2 are given by the following relations for the three charge shapes mentioned above:

\[
W_{NGP}(x) = \begin{cases} 
1 & \frac{|x|}{\Delta \xi} \leq \frac{1}{2}, \\
0 & \text{else}
\end{cases}
\]

\[
W_{CIC}(x) = \begin{cases} 
1 - \frac{|x|}{\Delta \xi} & \frac{|x|}{\Delta \xi} \leq 1, \\
0 & \text{else}
\end{cases}
\]
\[
W_{TSC}(x) = \begin{cases} 
\frac{3}{4} - \left| \frac{x}{\Delta G} \right|^2 & \frac{1}{2} \leq \left| \frac{x}{\Delta G} \right| \\
\frac{1}{2} \left( \frac{3}{2} - \left| \frac{x}{\Delta G} \right| \right)^2 & \frac{1}{2} \leq \left| \frac{x}{\Delta G} \right| \leq \frac{3}{2} \\
0 & \text{else}
\end{cases}
\] (3.5)

where \( \Delta G \) is the mesh size. In three dimensions the weighting function is obtained as follows,

\[
W(\vec{r}) = W(x)W(y)W(z).
\] (3.6)

3.1.2 Force Calculation.

3.1.2.1 Poisson Solver. The next step in the computation of the PM component of the force is the determination of the forces on the mesh from the charge distribution. An efficient and robust Poisson solver developed in the framework of electron devices simulations [30] has been successfully adapted to ion channel modeling. During the simulation, Poisson’s equation has to be solved many times as moving particles change the overall charge distribution (see Section 3.1.1). Within this self-consistent framework, the availability of the previous solution to be used as the initial guess for the next one makes iterative solvers very attractive as opposed to direct ones. An iterative version of the multi-grid [31, 30] Poisson solver described in detail in [29] is therefore used in this work.

The computational cost of solving Poisson’s equation depends on both the grid size and the frequency at which the solution is updated. A finer grid contains more points for the same system size, therefore the solution computation will require more operations. The choice of the grid spacing and the Poisson timestep is a trade-off between accuracy and performance of the simulation.
Once Poisson’s equation has been solved for the charge distribution assigned to the grid points, the electrostatic potential is available at each mesh point. The electric field is then computed as the gradient of the potential:

\[ \vec{E}(\vec{r}_p) = -\vec{\nabla}V(\vec{r}_p), \]  

(3.7)

where \( V \) is the potential obtained from the Poisson solver at the mesh point located in \( \vec{r}_p \). As it is done in the Poisson solver, the differentiation of the potential takes into account the boundary conditions and the dielectric discontinuities in the system [29].

The long-range, PM force is then computed at the position of the charged particle. It is crucial to note that this computation has to be done using the same interpolation scheme that was chosen for the charge assignment to the mesh. Any inconsistency between the assignment and back-interpolation schemes leads to a simulation where the particle momentum is not conserved [23].

Therefore, the force on each particle is obtained by combining the electric field components on the neighboring nodes using the same weights determined when assigning the charges to the nodes themselves. For example, when using the TSC scheme, the electric field values present at the 27 closest nodes will therefore be taken into account in determining the long-range component of the force on each charged particle.

3.1.2.2 Short-Range Force Component. The short-range component of the force on a generic particle \( i \) is obtained from the direct summation of the pairwise forces between this particle and its neighbors. According to the nature of the particle–particle interactions present in the system and how they are modeled (see Section 3.2.1.1), the interaction between two particles \( i \) and \( j \) is composed of an electrostatic coulombic part and a non-electrostatic part:

\[ \vec{F}_{ij}^{PP} = \vec{F}_{ij}^C + \vec{F}_{ij}^W \]  

(3.8)
where $\vec{F}^W_{ij}$ is the empirical non-electrostatic potential defined as either the Lennard-Jones potential (equation 3.23) or the inverse power relation (equation 3.25), and represents the interaction due to the finite size of the ions. The total PP component of the force on particle $i$ as expressed in equation 3.1 is then obtained by summing the contributions of the neighboring particles $j$ as expressed by:

$$\vec{F}_{PP}^i = \sum_{j \neq i} 6 \left[ \vec{F}_{C_{ij}}^i + \vec{F}_{W_{ij}}^i \right]$$  \hspace{1cm} (3.9)

The short-range domain $\Omega_i$ is a small spherical region centered on the particle $i$. All the particles $j$ positioned within a given cutoff radius determining the size of this short-range domain will be included in the summation. Obviously, the size of this region should be chosen as small as possible for efficiency reasons. However, the minimum size for this region is given by the size of the charge cloud defined by the charge assignment scheme (see Section 3.1.1).

The P$^3$M algorithm therefore splits the electrostatic interaction between “close sources” and “far sources”. The “close sources” are all the particles with charge clouds overlapping the charge cloud of particle $i$. Therefore, the radius of the short-range domain $\Omega_i$ is determined by the charge assignment scheme, and should be at least equal to the diameter of the charge clouds. The “far sources” are all the charges in the system which are beyond the scope of the short-range domain. Their influence is taken into account by the PM computation.

One problem that arises from the splitting of the coulombic force into two different domains is that they overlap. More precisely, the short-range domain is fully included in the whole system when the long-range component of the force is computed. Therefore, the effect of the particles within the short-range domain is accounted for twice: in the solution of the mesh force and in the direct PP summation. Removing the effects of those charges from the PM solution would require a separate solution of Poisson’s equation for each particle at each iteration, resulting
in unreasonable simulation times. The solution is to introduce the so-called reference force that estimates analytically and cancels the effect of the close sources in the PM component.

The reference force is derived from the charge distribution associated to the particles $S(\vec{r})$, as follows [23]:

$$
\vec{R}_i = - \sum_{j \neq i} \frac{q_i q_j}{4\pi \epsilon_r \epsilon_0} \int \int S(\vec{r}_1)S(\vec{r}_2 - \vec{r}_1) \frac{(\vec{r}_1 - \vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|^3} d\vec{r}_1 d\vec{r}_2
$$

(3.10)

For efficiency purposes, the force is precomputed and tabulated before running the simulation, as are the other PP force components that are not required to be evaluated dynamically.

The complete equation of the force on one generic particle $i$ obtained with the P$^3$M force field scheme is then computed as follows:

$$
\vec{F}_i = \vec{F}_i^{PM} + \vec{F}_i^{PP} + \vec{R}_i
$$

(3.11)

where $\vec{F}_i^{PM}$ is the force interpolated from the mesh, $\vec{F}_i^{PP}$ is the PP component given by equation 3.9, and $\vec{R}_i$ is given by equation 3.10.

### 3.1.3 Particle Dynamics.

The last part of each simulation step consists in computing the new velocity and position for all the particles according to the force they experience. This part is implemented in a dynamics simulation engine by using one of several different approaches. This section will describe how the equation of motion translates in the framework of the implicit water model, in which the water is treated as a continuum dielectric medium and the ions as moving particles.

#### 3.1.3.1 Brownian Dynamics.

Within the implicit solvent approach, the water molecules are not modeled individually by the simulation code, but the solvent is considered as a continuum medium. The Brownian Dynamics (BD) formalism adopts this description of the aqueous solutions and models each ion explicitly as a Brownian
particle with its dynamics being tracked through the 6-dimensional phase space. In this framework, the dynamics of the ions is *explicitly* determined by the electrostatics of the system through Newtonian mechanics, whereas the solvent-ion interactions is *implicitly* modeled with the Langevin equation.

The implicit solvent approach has the inconvenient of not representing the polarization effects explicitly by accounting for the motion of the water molecules and their orientation around ions. The advantage of the method is that it reduces dramatically the computational burden by avoiding the representation of a very large number of solvent molecules.

Within this approach, it is possible to include other molecules such as ion channel proteins or lipid membranes, either macroscopically by creating dielectric discontinuities, or microscopically by inserting van der Waals particles subjected to appropriate dynamics. This work will make use of both approaches for modeling the ion channels.

### 3.1.3.2 The Langevin Equation.

Our implementation of the BD simulation kernel uses the strict Langevin equation [32, 33], which assumes Markovian random forces and neglects the spatial and temporal correlations of the ion motions [32],

\[
m_i \frac{d\vec{v}_i(t)}{dt} = -\alpha \vec{v}_i(t) + \vec{F}_i(\vec{r}_i(t)) + \vec{B}_i(t) \tag{3.12}
\]

where \(m_i\) is the mass of the \(i\)-th ion, \(\vec{v}_i(t)\) is its velocity at time \(t\), and \(\alpha\) is the friction coefficient. The notation used here is the same as the one used in the work of Gunsteren and Berendsen [34]. The electrostatic force exerted on the particle \(i\) due to all other particles and the externally applied fields is denoted \(\vec{F}_i\). It is obtained from the force field calculations previously presented in this chapter. The last term of the Langevin equation, \(\vec{B}_i\) is a force representing the molecular bombardment of the water molecules on the ions and is modeled with a Markovian random variable.
3.1.3.3 Discrete Representation. The simulation algorithm is based on a discrete representation of time. Each variable such as position, velocity and force is only known at a discrete set of points in time. Therefore, the dynamics is represented as a succession of frozen states. For each of these states the forces are computed and the updated velocities and positions are obtained according to the equation of motion. It is clear that the discrete representation of time is crucial to the efficiency and accuracy of the implemented models.

The integration scheme used for the Langevin equation 3.12 is based on the requirement of using the longest possible timesteps while maintaining energy stability. The reason for using long timesteps stems from the nature of the problem studied here. Ion transport through channels occurs over relatively long timescales whereas ion and atomic motions are very fast and have to be integrated accurately. By using longer integration timesteps, or free-flight timesteps, a given simulation will require less steps, therefore it will be less expensive computationally. However, the timestep must be chosen small enough to resolve the atomic collisions between particles. The short-range force increases very rapidly and small changes in position can induce very large changes in the force experienced by a particle. When the timesteps are too long, the inaccuracy in the particle trajectories results in a spurious heating of the particle population, that can become energetically unstable [23].

In this work, three different integration schemes are used: the standard Euler integration scheme, the Verlet-like method introduced by Gunsteren and Berendsen [34], which is a second order model that reduces to the Verlet algorithm [35] when the friction coefficient is null, and a novel Predictor/Corrector integration scheme proposed by [36]. These algorithms will be described in the next sections.
3.1.3.4 Euler Integration. The Euler method is a first order integration scheme that reduces the Langevin equation to the following relation:

\[
\vec{v}_i(t + \Delta t) = \vec{v}_i(t) - \Delta t \left[ \gamma \vec{v}_i(t) - \frac{\vec{F}_i}{m_i} - \sqrt{\frac{6 \gamma k_B T}{m_i \Delta t}} \vec{N}(0, 1) \right].
\] (3.13)

In this equation, \(\Delta t\) represents the integration timestep and \(\vec{N}(0, 1)\) is a 3-dimensional Gaussian random variable with zero mean and variance of 1. The inverse of the relaxation time is defined as \(\gamma = \alpha/m_i\), where \(\alpha\) is the friction coefficient in the Langevin equation. The trajectories are calculated following Newton’s laws of motion.

The main issue with this integration scheme is that a very small timestep is required to obtain a correct representation of the effects of short-range forces that fluctuate on femtosecond timescales. In fact, the timestep \(\Delta t\) has to be much smaller than the inverse of \(\gamma\) (i.e. smaller than the velocity relaxation time) in the Langevin equation 3.12, which is an estimate of the mean time between particle collisions in the system.

3.1.3.5 Verlet-Like Method. The Verlet-like method can provide more accurate results with longer timesteps than the Euler method. The key is that it accounts for the evolution of the force \(\vec{B}\) during the integration timestep. In this method, the force on the \(i\)th particle at time \(t_{n+1}\) is first expanded in a power series about the previous time \(t_n\),

\[
F_i(t_{n+1}) \sim F_i(t_n) + \dot{F}_i(t_n)(t_{n+1} - t_n),
\] (3.14)

where \(\dot{F}\) denotes the time derivative. The power series expansion is substituted into equation 3.12, and the resulting solution of the Langevin equation is,

\[
v_i(t_{n+1}) = v_i(t_n)e^{-\gamma \Delta t} + (m_i \gamma)^{-1}F_i(t_n)(1 - e^{-\gamma \Delta t}) \\
+ (m_i \gamma^2)^{-1}\dot{F}_i(t_n)(\gamma \Delta t - (1 - e^{-\gamma \Delta t})) \\
+ (m_i)^{-1}e^{-\gamma \Delta t}\int_{t_n}^{t_{n+1}} e^{-\gamma(t' - t_n)} B_i(t') dt',
\] (3.15)
where $\Delta t = t_{n+1} - t_n$ is the integration time-step. Note that the force $B_i(t)$ is left inside the integral. The ionic position is calculated with the expression,

$$x_i(t_{n+1}) = 2x_i(t_n) - x_i(t_{n-1})e^{-\gamma \Delta t} + \int_{t_n}^{t_{n+\Delta t}} v_i(t') dt' + e^{-\gamma \Delta t} \int_{t_{n-\Delta t}}^{t_n} v_i(t') dt'$$

(3.16)

and, finally, the updated particle position is written as,

$$x_i(t_{n+1}) = x_i(t_n)[1 + e^{-\gamma \Delta t}] - x_i(t_{n-1})e^{-\gamma \Delta t} + (m_i\gamma)^{-1}F_i(t_n)(\Delta t)[1 - e^{-\gamma \Delta t}]$$

$$+ (m_i\gamma^2)^{-1}\dot{F}_i(t_n)(\Delta t)[0.5\gamma \Delta t(1 + e^{-\gamma \Delta t}) - (1 - e^{-\gamma \Delta t})]$$

$$+ X^n_i(0, \Delta t) + e^{-\gamma \Delta t} X^n_i(0, -\Delta t).$$

(3.17)

where,

$$X^n_i(0, \Delta t) = (m_i\gamma)^{-1} \int_{t_n}^{t_{n+\Delta t}} [1 - e^{-\gamma(t_n + \Delta t - t')} ] B_i(t') dt'$$

(3.18)

is also a Markovian stochastic process with zero mean and variance $\Delta t$. $X^n_i(0, -\Delta t)$ is correlated with $X^n_i(0, \Delta t)$ through a bivariate Gaussian distribution. In the zero limit of the friction coefficient this set of equations corresponds to the trajectories obtained with the Verlet MD algorithm [35, 34].

3.1.3.6 Predictor/Corrector Method. The third integration scheme implemented for the Langevin equation is the novel Predictor/Corrector (PC) integration scheme proposed in [36]. In this algorithm, the Langevin equation is solved numerically at each timestep for the velocity and position of the Brownian particles as follows:

$$V_{n+1} = V_n e^{-\tau} + \frac{F_n + R_n}{m\gamma}(1 - e^{-\tau})$$

$$X_{n+1} = X_n + \frac{V_n}{\gamma}(1 - e^{-\tau})$$

$$+ \frac{F_n + R_n}{m\gamma^2}(\tau - 1 + e^{-\tau}).$$

(3.19)
where $V_n$, $X_n$, $V_{n+1}$ and $X_{n+1}$ are the particle velocities and positions at timesteps $n$ and $n + 1$, respectively, $F_n$ is the force at time $n\Delta t$, as computed with the P$^3$M force field, and $R_n$ is a random force with a Gaussian distribution of zero mean and variance $6\gamma k_B T/m\Delta t$ representing the bombardment of water molecules on the ions [34, 29]. This scheme operates in two steps [36]: from the positions and velocities at step $n$, positions and velocities at time $n + 1$ are predicted using equation 3.19. The forces are then recomputed at this position and a \textit{corrected} force is obtained by taking the average of the initial $F_n$ and this recomputed force. The final positions and velocities $X_{n+1}$ and $V_{n+1}$ for the particles are finally obtained by integrating the Langevin equation a second time (equation 3.19), using the \textit{corrected} force rather than the initial $F_n$.

It is important to note that the trajectories obtained with both Verlet-like and Predictor/Corrector schemes are not limited by the velocity relaxation time as it happens for the Euler integration, and therefore a much longer timestep can be used. Figure 3.3 shows how the steady-state ionic energy of a 0.30 mol.L$^{-1}$ KCl solution compares for the various integration schemes, over a range of timesteps. These simulations have been conducted at equilibrium, that is without any externally applied field. All algorithms show similar results for small timesteps up to approximately 10 fs. As the timestep is further increased, the Euler scheme presents a significant spurious heating of the ionic population, while the higher order Verlet-like and Predictor/Corrector hold closer to the theoretical thermodynamic average. These two schemes are reasonably stable for timesteps up to 100 fs and will be used the most in this work.

3.1.4 Time and Space Discretization. The numerical model for the system is defined by a limited set of parameters determining the time and space discretization. Indeed, in the computer code the physical system can only be represented by discrete
Figure 3.3. Steady-State Ionic Energy of a 0.30 mol.L$^{-1}$ KCl Solution Obtained With Euler, Verlet-Like and Predictor/Corrector Integration Schemes, for Different Timestep Values Between 1 and 200 fs, Compared to the Theoretical Value for the Kinetic Energy.

sets of numerical values, each representing a single state of the system in time. The intervals between each of these states in time, and each of these values in space are crucial parameters for the accuracy of the numerical representation.

3.1.4.1 Space Discretization. It was shown in Section 3.1.1 that for the purpose of computing the electrostatic potential in the entire system, the charge distribution is defined on a set of discrete points forming a mesh. The interval between each grid point, called grid spacing or cell size $\Delta G$ is the crucial parameter defining the space discretization scheme of the simulation. It is directly related to the charge assignment scheme and the accuracy of the Poisson solver, and it also determines the extension of the short-range domain of the P$^3$M force field scheme.

For these reasons, there are limitations to the range of acceptable values that can be used for the grid spacing of a particular system. A characteristic length of
the system is the Debye length $\lambda_D$, which is a measure of the spatial extension of electrical effects in the system. It can be defined as [37]:

$$\lambda_D = \sqrt{\frac{\epsilon k_B T}{nq^2}},$$

(3.20)

where $n$ is the number density in ions given by $n = N_A c \times 10^3$ with the concentration being $c$ and the Avogadro number being $N_A$. $\epsilon$ is the dielectric constant of water and $q$ the charge of an ion. The Debye length sets the upper limit of the grid spacing in the P$^3$M force field framework. The grid spacing should be smaller than this limit in order to resolve the spatial variations of the charge distribution [23].

3.1.4.2 Time Discretization. The evolution of the system in time is represented by a sequence of states. The time interval between each update of the system state is therefore the crucial parameter determining the accuracy of the representation of the system behavior in time.

The evolution of the charge density in the system has its specific time constants, and a fully self-consistent modeling of the system dynamics requires the computation of the mesh charge and potentials as often as needed to represent this evolution. As an estimator of the minimum frequency of field updates we suggest the plasma frequency of the ions:

$$\omega = \sqrt{\frac{c|q|}{\epsilon_r \epsilon_0 m}},$$

(3.21)

where $c$ is the ionic concentration, $m$ is the ion mass and $q$ is the magnitude of the ion charge. The plasma frequency represents the electrostatic response of the ions to a perturbation of the charge density. The Poisson solver will not be able to accommodate for charge distribution variations if the solution is not computed at least as often as the inverse of the plasma frequency. For an aqueous solution of 0.10 mol.L$^{-1}$ KCl this frequency is about 200 GHz, which corresponds to a minimum timestep of 5 ps for the Poisson solution update.
However, the integration of the Langevin equation of motion needs to be performed on the femtosecond timescale (see Section 3.1.3.3). Therefore, the Poisson solution and Langevin equation integration can be performed on different schedules. For computational efficiency purposes, the two processes are realized on two different intervals: the Poisson timestep denoted $\Delta t_{\text{Poiss}}$ is the time between two successive updates of the Poisson solution and the free-flight timestep $\Delta t_{\text{ff}}$ is the integration timestep of the Langevin equation defining the particle motion.

3.1.5 Averaging the Observables. The properties of interest of the system are extracted in different ways depending on their meaning.

One method for averaging quantities of interest like ionic energies, velocities, concentration, and fluxes is to extract them from the ensemble average at each simulation step. The accuracy of the ensemble-averaged measurements is improved by increasing the number of particles in the system. Methods to improve the accuracy of the measurements and estimate the amount of fluctuations attributable to the limited number of particles are presented in Section 4.1. Time-varying properties give a useful insight in transient phenomena like the establishment of a current or the ionic depletion of certain regions. The observation of the transient makes it possible to determine the time required for certain properties of the system to reach a steady-state, and start long time averages computed in order to reduce statistical fluctuations.
The structural characteristics of the ionic solutions for example require a very large amount of data to limit the statistical fluctuations. The Radial Distribution Function (RDF) is a good example of such a characteristic (see Section 4.1.2). Another example of a measure that requires to be averaged over a significant period of time is the particle currents. At each timestep only a very small number of particles actually cross the contact boundaries of the system which represent the effects of electrodes in the solution. Therefore, a reliable current reading can only be determined by accumulating the particle flux over a long simulation time.

3.2 The Computer Model

A section of the membrane and aqueous solution is represented by the simulation to study the behavior of the ion channel inserted in the center of the system. In order to apply a bias reproducing the effects of an external trans-membrane potential, Dirichlet boundary conditions are set on two opposite sides of the cubic simulation box. Neumann conditions are imposed on the other four boundaries [23]. The externally applied potential is accounted for directly in the solution of Poisson’s equation.

The ions are allowed to cross Dirichlet boundaries, and are specularly reflected by the Neumann ones. It is important to stress that the simulation does not use periodic boundary conditions, and therefore ions crossing to the outside of the box leave the system. An injection mechanism is therefore needed to have new ions enter the boundaries and compensate for those leaving. Since the system represents a small section of the volume of an electrolyte solution, the injection mechanism simply maintains constant the ionic concentrations at the Dirichlet boundaries. Each time the solution of Poisson’s equation is computed, the ionic population in the boundary cells is updated by removing the particles in excess or injecting the ones necessary to maintain the initial concentration at the boundaries. The velocity of the injected ions has a Maxwellian distribution in the directions parallel to the boundary and a
half-Maxwellian distribution in the normal direction. One artifact of this method is that the average velocity of the injected particles does not match the macroscopic flux and therefore the particles must relax their velocity and energy to the steady-state values. This process is fast, and when computing averages for the system, particles in the first rows of cells next to the injecting boundary are neglected in order to remove the small artifacts due to the injection mechanism.

3.2.1 Electrolyte Solution Modeling. The accurate description of ionic channels requires an accurate and efficient model of the electrolytic solution bathing them. Many theories have been developed to describe the physics underlying aqueous solutions, the solvent-solute interactions, and the chemical reactions occurring in this environment. This work makes use of the primitive or implicit model for water [38], which will be described in detail and briefly compared to other available approaches. An argument for using this relatively simple model will be made, its limitations and range of applicability will be discussed and related to the characteristics of the systems of interest for this work.

The aqueous solution considered in the next sections is a mixture of three components: water molecules, positive and negative ions. Autoprotolytic reactions\(^1\) will be neglected as well as the presence of other ionic species than the solutes.

\(^{1}\)An autoprotolytic reaction is the spontaneous breakage of water molecules into \(\text{H}^+\) and \(\text{HO}^-\) ions. Pure water naturally contains a concentration of \(10^{-7}\) M of each species.
3.2.1.1 Nature of the Interactions. Two basic types of interactions are present in this system: the long-range electrostatic coulombic interactions and the short-range interactions due to the finite size of the particles.

The Coulomb force between two charged particles $j$ and $i$ in vacuum is given by

$$F_{ij}^C = \frac{q_i q_j}{4 \pi \varepsilon_0 |\vec{r}_i - \vec{r}_j|^2 \hat{r}_{ij}},$$

(3.22)

where $\vec{r}_i$ and $\vec{r}_j$ are the positions of the particles, $\varepsilon_0$ is the permittivity of free space, $\hat{r}_{ij}$ is the unit vector in the direction of $j$ to $i$, and $q_i$ and $q_j$ represent the respective charges.

The non-electrostatic interaction has several components. Any two close atoms will interact even if they do not carry any net charge, because of the electron “cloud” surrounding their nuclei. When the distance between the two atoms is very small, this force is repulsive due to the overlapping of the electron orbitals of both atoms and shows a very steep increase when the separation distance is close to the atom diameter. For Lennard-Jones [39] particles this diameter is twice the radius $\sigma$ where the potential is minimal (see equation 3.23). This is the so called finite-size effect. Because of its quantum mechanical nature, it is very difficult to efficiently compute this component of the short-range interaction force, as a complete calculation should account for all the electrons in the cloud.

When the inter-atomic distance increases, another phenomenon gives rise to attractive components of the short-range force. One of these components is the electrostatic attraction between induced dipoles in each atom. Due to fluctuations of the electron cloud around its average position, an electrostatic dipole is formed by the electrons and the positively charged nucleus. The neutral atom then becomes a dipole and interacts with other atoms. This so-called dispersion force or London force originates from quantum mechanics but the classical representation of fluctuat-
ing electrostatic dipoles helps in visualizing how it produces an attractive contribution to the short-range force. For polar molecules like water, other interaction components such as dipole–dipole or dipole–induced dipole interactions are also present. However, in the following discussion only simple mono-atomic ions will be considered, and no permanent dipoles are involved. The sum of all the attractive components of the short-range interaction are regrouped under the term van der Waals interactions [39].

The addition of the strong repulsive interaction due to the finite-size effect, and the weak van der Waals attraction forms a characteristic inter-molecular or inter-atomic interaction. Exact formulas for the forces resulting from these interactions are not easily obtained, while empirical relations have been successfully used to reproduce experimental results. Two empirical inter-atomic potentials are used in this work, namely the Lennard-Jones (LJ) potential [39] and the Inverse-Power (IP) relation [40]. The empirical inter-atomic Lennard-Jones force exerted on atom $i$ by the atom $j$ is expressed by

$$
\tilde{F}_{ij}^{\text{LJ}} = \frac{24\epsilon_{ij}}{|\vec{r}_i - \vec{r}_j|} \left[ \frac{\sigma_{ij}}{|\vec{r}_i - \vec{r}_j|} \right]^{12} \left( \frac{\sigma_{ij}}{|\vec{r}_i - \vec{r}_j|} \right)^6 \vec{r}_{ij};
$$

(3.23)

where $\sigma_{ij}$ and $\epsilon_{ij}$ are two fitting parameters representing the maximum attraction distance and the strength of the interaction, respectively. For interactions between atoms of different species, these fitting parameters are given by the following equation [39]:

$$
\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j), \quad \epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j},
$$

(3.24)

where $\sigma_i$ and $\epsilon_i$ are tabulated values representing the distance at which the potential is minimal, and the depth of the potential well, respectively.

The Inverse Power relation is given by [40]:

$$
\tilde{F}_{ij}^{\text{IP}} = \frac{\beta_{ij} |q_i q_j|}{4\pi\epsilon_0 |\vec{r}_i - \vec{r}_j|^p (p + 1)} \left( \frac{s_i + s_j}{|\vec{r}_i - \vec{r}_j|} \right)^p \vec{r}_{ij}.
$$

(3.25)
where $\beta_{ij}$ is an adjustable parameter, $s_i$ is the diameter of atom $i$, and $p$ is a hardness parameter that represents the strength of the interaction.

### 3.2.1.2 The Electrostatic Model of the Electrolyte Solution.

Having defined the components of the system and the possible ways they can interact completes the requirements for building a computer model for the electrolytic solution. Many approaches have been developed and will be briefly cited before focusing more specifically on the primitive model, which is used in this work.

The models can be classified starting from the lowest level of detail for macroscopic models, up to the highest level of detail for microscopic models.

The simplest approach is fitting experimental measurements, such as the osmotic coefficient, with empirical equations. This method has been used mostly to describe high electrolyte concentrations for which there is no well-founded theoretical expression [38]. A more detailed, yet macroscopic description is based on classic statistical thermodynamics and yields relations which can be used to assess the quality of the more detailed microscopic models [38].

Three levels of modeling of the microscopic properties of the solution can be envisioned: the McMillan-Mayer level, which considers the solvent as a continuum medium and models only the behavior of the solute particles. This model is also called the primitive model [41] and will be used in this work. Monte Carlo and Brownian Dynamics simulations, as well as approximations applied to the integral equation theories of liquids, such as the hypernetted chain (HNC) approximation, are the tools of choice for this description of electrolyte solutions, and have been implemented in this study. At the next level, the Born-Oppenheimer level, the solvent molecules are accounted for explicitly and are described as particles. Molecular Dynamics simulations and integral equations are the theoretical tools to work at this descriptive
level. Finally, the most detailed level is the Schrödinger level, which involves the computation of the quantum mechanical behavior of all the particles in the system, is overwhelmingly complex, and does not provide much more details for the study of electrolyte solutions and ion channels than the previous levels [38].

3.2.1.3 The Implicit Model for Water. The primitive or implicit model for water is a microscopic model for an electrolyte solution. In this framework, the solute is treated as made of charged particles with a finite size, and water is represented by a continuous medium of uniform dielectric constant. This model attempts to represent the solution as a gaseous mixture of charged spheres in a given volume $V$ and at a specific temperature $T$ [41].

The simplest version of the primitive model assumes that the ions are hard spheres with a fixed diameter. The ionic charges are treated as point charges located at the center of the sphere, and the material in the sphere is assumed to have the same dielectric constant as the solvent. The interaction potential between ions is infinitely repulsive for inter-ionic separations smaller than the sphere diameter, which means that the ions cannot inter-penetrate. At larger separation the potential is purely coulombic.

The primitive model is a very efficient representation for computer simulations because the number of solute particles is much smaller than the number of water molecules. Indeed, a 1 M solution contains 55 water molecules for each solute particle. Furthermore, the ions simulated in this work are mono-atomic and therefore have a much simpler geometry than the water molecules. Treating water as a continuum simplifies the computation since the polarization effects can be accounted for in the electrostatics through a simple step-wise constant dielectric constant.
This model was first introduced for the study of electrolyte solutions by P. Debye and E. Hückel in 1923 [42] and the results they obtained are based on the assumption that the solution is very diluted [39]. However, certain improvements can be made to extend the validity range of the primitive model [39]. Attempting to describe or match experimental results obtained with a real electrolyte solution is intrinsically flawed as the primitive model is only exact in the limit of infinitely dilute concentrations, and the results from the Debye-Hückel theory become inaccurate for concentrations in excess of only $10^{-3}$ M. This limit is well below the values encountered in the systems of interest to this work and a more refined primitive model must be used.

The refinement consists in the introduction of semi-empirical corrections in the model. The limitation of the model to the representation of low concentrations comes from the description of the short-range interactions. At high concentrations, the ions get closer to each other and the influence of the short-range component of the interaction increases. Therefore, the range of applicability of the primitive model can be extended when using a more accurate representation of the particles behavior at close range. This improvement is obtained by replacing the hard sphere repulsion with an empirical force such as the one expressed by the Lennard-Jones or the inverse power relations (see section 3.2.1.1). This refinement extends the predictability of the structural and thermodynamical quantities up to molar concentrations, well beyond the typical concentrations found in biological systems. However, this model is still not applicable to real solutions, and any comparison with experimental data must take into consideration the fact that fitting parameters are used to define the interaction potentials.

Despite its limitations, the primitive model is a very useful tool to develop and compare theoretical methodologies such as particle-based computer models and
integral methods. It yields useful qualitative and in some cases quantitative results with a low computational burden. This makes possible the modeling of complex systems for simulation times that would be unachievable with methods based on explicit solvent.

3.2.2 Membrane Representation. The molecular structure of lipid bilayer membranes has been described in detail in Section 2.2. The main characteristics that should be included in a successful model are the impermeability to any particle (water or ions), the non-polar character of the interior of the membrane composed of fatty acid hydrocarbon chain and, finally, the strongly polar surfaces bearing charges in contact with the solution.

Lipid membranes are extremely stable two-dimensional structures. Lipid molecules can diffuse very fast along the surface. However, due to the very low dielectric constant of the apolar interior of the membrane, high electric potentials can be sustained by the membrane, making the crossing of charged particles almost impossible. Although the accurate description of all the physical properties of the bilayer can be arguably best achieved with an all-atom model, this study is more concerned with the behavior of the ion channel. Therefore, the proposed model will only reproduce the impermeability and dielectric properties of the membrane.

In an ion channel system, an explicit membrane model is only of interest to the study of the membrane–protein interaction, and generally affects more the stability of the protein insertion rather than its specific functionalities [43]. In particular, this is the case for the OmpF channel discussed in this document. The membrane is therefore represented in this work as an impermeable slab with a homogeneous low dielectric constant, or as an inhomogeneous slab with a higher effective dielectric in the region where the lipid polar heads reside [44].
3.2.3 Ion Channel Model. As it has been discussed in the introduction, the major challenge presented by ion channels simulation is the necessity of simulating charge transport behavior occurring over relatively long timescales (up to milliseconds), yet accurately resolving atomic motions that happen on a femtosecond timescale. An accurate model should include all the atoms of the system, together with the constrained bonded interactions and the unconstrained, long range, non-bonded interactions. The effects of the ions on the atoms of the residues inside the channel pore should be included as well. Indeed, the polarization of the residues lining the pore may play a significant role in the selectivity of the channel and, in order to account for this electrostatic effect, the motion of the residue atoms should be computed. The major problem arising from such approach is that a computer simulation would take an unrealistic time to model the characteristics of charge transport over a sufficiently long period of time. Developing an alternative and more efficient simulation is the main motivation for this work, and a simpler model for the ion channel will be tested and used for this reason.
The approach proposed here is to model the channel protein as a static structure. Each atom of the protein will be inserted in the numerical model as a rigidly bound particle with its specific charge and inter-particle empirical potential profiles. The coordinates used to position the protein atoms are obtained from an energy minimization procedure described below. The original atom coordinates from the crystallographic structure are first defined according to the data in the Protein Data Bank (PDB) [21] (see Appendix A). This preliminary structure is then input into the GROMACS [45] simulation tool (see Appendix B) to perform an all-atom energy minimization in order to relax the channel to a stable configuration. The outputs of the GROMACS simulation are the coordinates of all the atoms of the protein in their stable configuration, together with their partial charges computed from the electronegativity and bonding pattern information. In the proposed model, the set of atoms can be treated as individual charges fixed in space.

Of course, the problems posed by this approach arise from the lack of representation of the polarization and conformational changes occurring in the residues lining the pore when ions cross it. A more refined representation is envisioned and is part of the future work proposed for the continuation of this project (see Section 5.5). For the moment, the issue of the polarization of the pore residues is treated by assigning specific dielectric constants to different domains of the system (see Section 3.2.4).

3.2.4 Grid and Boundary Conditions. This section will describe the discrete grid used to compute the PM component of the force with the real-space Poisson solver. Figure 3.4 shows a portion of the discrete grid used to assign charges to the mesh, solve Poisson’s equation and interpolate the PM force on the particles using the charge assignment scheme described in section 3.1.1 [29].

Each cell has up to 26 neighbors, and the potential, dielectric constant, and 3D electric field components are considered constant within its volume.
Figure 3.4. Portion of the Finite Difference Grid Used To Solve Poisson’s Equation. Points Labeled With \( n, s, e, w, f, b \) Are Neighbors of the Point Labeled \( c \), \( i.e. \) the Center of the Central Cell. The Points Shown With Primes (\( n', s', e', w', f', b' \)), Are Used for Interpolations. Characteristics Distances Are Also Depicted (Deltas).

During the initialization phase of the simulation, the neighbors of each cell are determined and pointers to them are stored in fields associated to each grid-point. The linked list used to represent the structure of the grid is modified to account for the boundary conditions on the cells. These modifications make it possible to apply the same algorithm to each cell regardless of where it is located on the grid. Figure 3.5 shows how the linked list is set in the case of boundary cells in a 2D grid. On the left cell the Dirichlet condition is used, which sets a fixed potential on the cell. The pointer to the northern neighbor is therefore set to point to the cell itself. For the right cell, the Neumann condition is used corresponding to a zero perpendicular electric field. The northern neighbor is now considered to be the southern adjacent cell [29].
The code has been implemented in such a way that the grid can be inhomogeneous. Furthermore, the Poisson solver can handle dielectric discontinuities in an inhomogeneous stepwise constant dielectric medium. This is crucial for the implementation of the proposed implicit membrane model. The way the ion channel is represented is also dependent on the grid definition. The representation chosen is to use a much finer grid in the channel region of the simulation box than in the membrane or solution regions of the model. In this way it is possible to assign different dielectric constants to the cells containing protein atoms and the empty cells inside the pores. Several different sets of values have been tested and the influence of the protein and pore region dielectrics will be discussed in subsequent sections.
CHAPTER 4
VALIDATION OF THE SIMULATION METHODOLOGY

The previous chapter offered a description of the simulation program and the model chosen for the study of charge transport in ion channels. The P³M force field scheme and the BD simulation engine have been discussed in terms of their application to the systems of interest. This chapter will give a comprehensive validation of the proposed simulation algorithm. Simulations results were compared with other models, and with experimental results, in order to determine the validity and accuracy of the Poisson P³M force field scheme coupled with the BD engine.

4.1 Benchmarks for Bulk Electrolyte Solutions

A comprehensive validation of a simulation model and computer algorithm relies on a set of precisely defined benchmarks capable of reproducing specific properties of the system. These properties have been characterized both with the simulation tool and other numerical techniques, for the purpose of evaluating the quality of the model and establishing its range of applicability.

4.1.1 Kinetic Energy. The first benchmark property used for the assessment of the simulation accuracy is the ensemble average kinetic energy. The electrolyte solutions of interest to this work contain only mono-atomic ions. These particles exhibit three translational degrees of freedom, and their instantaneous kinetic energy is expressed by:

\[ E_K(t) = \frac{1}{2} \sum_{k=i,j}^N \sum_{l=1}^{N_k} \left( v_{x,l}(t)^2 + v_{y,l}(t)^2 + v_{z,l}(t)^2 \right), \]  

(4.1)

where \( N_k \) is the total number of ions of type \( k \), and \( v_{d,l} \) is the translation velocity of a generic ion \( l \) along the direction \( d \). The computed time average of the kinetic energy should be equal to the theoretical thermal energy of the system. For this model, the
average thermal energy of a particle has been expressed in Chapter 2 to be:

\[ E_{Th} = \frac{3}{2} k_B T, \quad (4.2) \]

where \( k_B \) is Boltzmann’s constant and \( T \) is the absolute temperature. A first measure of the simulation protocol stability is obtained by comparing the theoretical average thermal energy and the computed average kinetic energy of the ensemble of simulated particles.

The velocities and positions of all the particles in the simulation are recorded at each Poisson timestep. The time average of the kinetic energy is computed only after a certain time \( t_r \) to let the system relax to the steady state. The average kinetic energy can then be computed for one particle as follows:

\[ \overline{E_K} = \frac{1}{N} \frac{1}{N_{steps}} \sum_{n=1}^{N_{steps}} E_K(t_n), \quad (4.3) \]

where \( N \) is the number of particles in the system and \( N_{steps} \) is the total number of timesteps included in the time average computation after the steady-state condition has been reached. In order to make those comparisons possible for systems of different sizes, the energies are normalized to per-particle values. The numerical error for the simulation is now determined as follows:

\[ \epsilon = E_{Th} - \overline{E_K}, \quad (4.4) \]

The significance of this error can be evaluated by comparing the statistical variability of the average. The average kinetic energy can be considered as a statistical sample of \( N_{steps} \) values of the random variable \( E_K(t_n) \). The assumption will be made that this variable is normally distributed as the number of particles in the system is large. The sample standard deviation can therefore be obtained as follows:

\[ s = \sqrt{\frac{1}{N_{steps}} \sum_{n=1}^{N_{steps}} (E_K(t_n) - \overline{E_K})^2}. \quad (4.5) \]
Finally, the margin of error is computed for a given confidence level $\alpha$:

$$
\epsilon_\alpha = z_\frac{\alpha}{2} \frac{s}{\sqrt{N_{\text{steps}}}},
$$

(4.6)

where $z_\frac{\alpha}{2}$ is obtained from the standard normal table [46]. A value of 1.96 corresponds to a confidence level of 95%.

The results from the simulation are said to match the theoretical value with a probability greater or equal to $\alpha$ when $|\epsilon| \leq \epsilon_\alpha$. Following these guidelines for the comparison ensures that the numerical error is estimated correctly, and it reduces the influence of the system size (number of particles $N$) or the length of the simulation $N_{\text{steps}}$ on the measure of accuracy.

### 4.1.2 Radial Distribution Function.

The Radial Distribution Function (RDF) represents the reciprocal average position of all the particles. In other words, the RDF $g_{ss'}(r)$ represents the local density of ions of species $s'$ around any ion of species $s$, as a function of the distance $r$. The RDF is a very powerful concept for several reasons. First, it is possible to obtain it directly from X-ray scattering experiments on liquids [41]. Furthermore, all the thermodynamical quantities characteristic of the solution can be expressed using the RDF [41]. The distribution function can therefore be used as a link between scattering experiments and thermodynamical measurements. Finally, studying the RDF provides direct insight into the interactions underlying the structure of the solution.

In the case of mono-atomic ions, all the interactions are spherically symmetric and the RDF depends only upon the distance between ions. In a two-component mixture four different RDFs are defined, each depending on the ion separation: $g_{ss}(r)$, $g_{ss'}(r)$, $g_{s's}(r)$ and $g_{s's'}(r)$. For each RDF, the following property is valid: $g \to 0$ as $r \to 0$ since it is impossible for two ions to overlap as the repulsion diverges as $r \to 0$. On the other hand, since the influence of the central ion vanishes as $r$ increases,
it follows that \( g \to 1 \) as \( r \to \infty \). This means that the local density of ions \( s' \), \( \rho_{s'}(r) = \rho_{s'} g_{ss'}(r) \) becomes the bulk density \( \rho_{s'} \) as the inter-ionic distance increases.

Figure 4.1 demonstrates how the distribution of particles around any given ion is linked to the RDF.

The second benchmark property used in this work is the peak value of the cation–anion RDF. The RDF between ions of opposite signs presents a maximum (see for example figure C.1), that is highly dependent on the ionic concentration. Figure 4.2 shows the different RDFs obtained with an analytical model [38] (see Appendix C) for a range of concentrations.

The RDF peak value is therefore a good indicator of an accurate and stable ionic concentration, since any variation in ionic concentration will translate into a change of the peak value. The RDF is extracted from the simulation data by following the procedure used in [23]: the RDF is represented by a set of values for distances between 0 and 2 nm with increments \( dr = 5 \times 10^{-2} \) nm. The non-normalized RDF is thus computed at the time \( t_n \) and at the point \( r_k \) as follows:

\[
g^1_{s,s'}(r_k, t_n) = \sum_{l=1}^{N'_{s'}} \sum_{m>l} \delta \left( |\tilde{r}_l(t_n) - \tilde{r}_m(t_n)|, r_k \right),
\]

where \( N'_{s} \) is the number of particles of type \( s \) in the pair-correlation computation volume \( V_b \). \( N_{s'} \) is the total number of particles of type \( s' \) in the system, and \( t_n \) is the
Figure 4.2. RDFs Between K$^+$ and Cl$^-$ for KCl Solutions With Concentrations Ranging From 0.10 To 1.00 mol.L$^{-1}$. The HNC Algorithm Has Been Used With a Lennard-Jones Potential.

\[ g_{s,s'}(r_k; t_n) = \frac{2V_b}{N_s'(N_s' + N_s'')} V_k g_{s,s'}^1(r_k; t_n), \]

where \( V_k \) is the volume of a spherical region of radius \( r_k \) given by:

\[ V_k = \frac{4\pi}{3} ((r_k + dr)^3 - r_k^3) \]

The resulting instantaneous RDF is then time-averaged over \( N_{\text{steps}} \) steps in a similar fashion as what is done for the kinetic energy (see Section 4.1.1):

\[ g_{s,s'}(r_k) = \frac{1}{N_{\text{steps}}} \sum_{n=1}^{N_{\text{steps}}} g_{s,s'}(r_k; t_n) \]
This RDF can be directly compared to the RDF obtained with other approaches (see Appendix C). The peak value used as a benchmark is the maximum value of \( g_{s,s'}(r_k) \) when \( s \neq s' \).

One common issue with the calculation of the RDF is that the resulting curve is usually quite noisy. In this work, a statistical method has been implemented to yield an accurate estimate of both the peak value and the fluctuations resulting from the numerical technique used to compute the RDF.

### 4.1.2.1 Statistical Processing of the RDF Curves

The principle underlying the statistical enhancement method proposed is the determination of a least squares polynomial estimate of the RDF function in the peak region. The first step is to determine the maximum of the noisy RDF curve. This maximum occurs at distance \( r_m \). The samples \( x \) and \( y \) are then obtained by considering \( N \) points around the maximum from the raw RDF data:

\[
\begin{align*}
x[k] &= r_{k+m-N/2} \\
y[k] &= g_{s,s'}(r_{k+m-N/2})
\end{align*}
\]  

(4.12)

The RDF curve is then approximated by a 3rd-order polynomial function:

\[
p(x) = ax^3 + bx^2 + cx + d.
\]  

(4.13)

The coefficients \( a, b, c \) and \( d \) are determined by minimizing the square error between the sample and the fitting function defined by:

\[
E(a, b, c, d) \equiv \sum_{k=0}^{N-1} (y[k] - p(x[k]))^2
\]  

(4.14)
This minimization corresponds to the solution of the following linear system:

\[
\begin{align*}
\frac{\partial E}{\partial a} (a, b, c, d) &= 0 \\
\frac{\partial E}{\partial b} (a, b, c, d) &= 0 \\
\frac{\partial E}{\partial c} (a, b, c, d) &= 0 \\
\frac{\partial E}{\partial d} (a, b, c, d) &= 0
\end{align*}
\]

(4.15)

The solution \( \{a_0, b_0, c_0, d_0\} \) of this system is obtained by inverting the 4 \( \times \) 4 matrix obtained by computing the partial derivatives in equation 4.15:

\[
\begin{bmatrix}
\sum x[i]^6 & \sum x[i]^5 & \sum x[i]^4 & \sum x[i]^3 \\
\sum x[i]^5 & \sum x[i]^4 & \sum x[i]^3 & \sum x[i]^2 \\
\sum x[i]^4 & \sum x[i]^3 & \sum x[i]^2 & \sum x[i] \\
\sum x[i]^3 & \sum x[i]^2 & \sum x[i] & N
\end{bmatrix}
\begin{bmatrix}
a_0 \\
b_0 \\
c_0 \\
d_0
\end{bmatrix}
= 
\begin{bmatrix}
\sum x[i]^3 y[i] \\
\sum x[i]^2 y[i] \\
\sum x[i] y[i] \\
\sum y[i]
\end{bmatrix},
\]

(4.16)

where all the sums are over the \( N \) samples. The extrema of the polynomial fitting function are then obtained by solving \( p_0'(x) = 0 \):

\[
\{x_1, x_2\} = \left\{ \frac{-b_0 - \sqrt{b_0^2 - 3a_0c_0}}{3a_0}, \frac{-b_0 + \sqrt{b_0^2 - 3a_0c_0}}{3a_0} \right\}.
\]

(4.17)

The maximum value of \( p_0 \) will then be used as the estimate of the RDF peak value:

\[
RDF_{\text{peak}} = \max \left( p_0(x_1), p_0(x_2) \right).
\]

(4.18)

Besides the actual peak value used for benchmarking purposes, another quantity that can be extracted from the linear regression is the standard error between the sample and the fitting function, conveniently obtained from the error function (equation 4.14) as follows:

\[
s \equiv \sqrt{\frac{E(a_0, b_0, c_0, d_0)}{N - 1}}.
\]

(4.19)

Similarly to what was done for the thermal energy (see Section 4.1.1), the difference between the RDF peak values obtained from the simulation and an analytical
model [38] can be determined:

\[ \epsilon = RDF_{\text{peak}}^{HNC} - RDF_{\text{peak}}, \]  

(4.20)

where \( RDF_{\text{peak}}^{HNC} \) is the maximum value of the RDF between ions of opposite charge computed with the HNC analytical model [38] and \( RDF_{\text{peak}} \) is defined by equation 4.18. This difference can then be compared to the sample margin of error obtained with the \( t \)-distribution using \( N - 1 \) degrees of freedom:

\[ \epsilon_\alpha = t_\frac{\alpha}{2} \frac{s}{\sqrt{N}} \]  

(4.21)

For \( N = 16 \) and a confidence level of 95% the value for \( t_\frac{\alpha}{2} \) is 2.13. When \( |\epsilon| \leq \epsilon_\alpha \), the simulation results can be declared to match the analytical results obtained with the HNC approximation.

![RDF Curve and 3rd-Order Polynomial Fit, From a 0.30 mol.L\(^{-1}\) KCl Solution Simulation](image_url)

Figure 4.3. RDF Curve and 3rd-Order Polynomial Fit, From a 0.30 mol.L\(^{-1}\) KCl Solution Simulation

The least squares method has been tested for both 2nd-order (parabolic) and 3rd-order polynomial fits. The cubic fit was found to be a much better approximant
for the specific shape of the RDF curve in the region around the peak. Given the
grid used for the RDF function, a sample size of 16 points was found to be the best
option. Figure 4.3 shows a sample RDF curve obtained from a 0.30 mol.L\(^{-1}\) KCl
solution simulation, and the corresponding 3\(^{\text{rd}}\)-order polynomial fit obtained with the
procedure described in this section. The RDF peak value is shown with the associated
standard error.

4.1.2.2 Extraction of Thermodynamical Parameters. The derivation of
thermodynamic quantities from the RDF is based on the assumption that the total
potential energy of the system is pair-wise additive [41]. This means that for a system
of \(N\) particles, the total potential as a function of the particle positions is given by:

\[
U(r_1, ..., r_N) = \sum_{i<j} \sum_{j=1}^{N} u_{ij}(r_{ij}),
\]

(4.22)

where \(u_{ij}(r_{ij})\) is the total interaction potential between the particles \(i\) and \(j\). Since
the average configuration of the system is completely described in terms of the RDF
functions, it is possible to express the mean potential energy as a function of \(g\) [41]:

\[
\overline{U} = \frac{N^2}{2V} \int_{0}^{\infty} u(r)g(r)4\pi r^2dr,
\]

(4.23)

where the system is assumed to have a volume \(V\), and contains only one species. For
a two-component system, the average potential energy is:

\[
\overline{U} = \sum_{s,s'} \frac{N_s N_{s'}}{2V} \int_{0}^{\infty} u_{ss'}(r)g_{ss'}(r)4\pi r^2dr.
\]

(4.24)

The sum is over all pairs of species \((s, s')\), while \(N_s\) and \(N_{s'}\) represent the total
numbers of particles of types \(s\) and \(s'\), respectively.

The derivation of the mean potential energy makes it possible to express the
mean total energy of the system as the sum of the mean kinetic energy \(\overline{E}_K\) and \(\overline{U}\):

\[
E = \overline{E}_K + \overline{U}.
\]

(4.25)
For a two-component mixture of mono-atomic ions, the mean kinetic energy is expressed as [41]:

\[ E_K = \frac{3}{2} N_s k_B T + \frac{3}{2} N_{s'} k_B T. \] (4.26)

The mean total energy is then given by:

\[ E = \frac{3}{2} N_s k_B T + \frac{3}{2} N_{s'} k_B T + \sum_{s, s'} \frac{N_s N_{s'}}{2V} \int_0^\infty u_{ss'}(r) g_{ss'}(r) 4\pi r^2 dr. \] (4.27)

This expression can be rewritten using the densities \( \rho_s = N_s / V \), and expressing the energy per unit volume:

\[ \frac{E}{V} = \frac{3}{2} (\rho_s + \rho_{s'}) k_B T + \frac{1}{2} \sum_{s, s'} \frac{\rho_s \rho_{s'}}{2} \int_0^\infty u_{ss'}(r) g_{ss'}(r) 4\pi r^2 dr. \] (4.28)

### 4.1.3 The Osmotic Coefficient.

A useful relation that can be directly obtained from the RDF is the so-called pressure equation, which defines the osmotic pressure in terms of the RDFs and the interaction potentials. The osmotic pressure is the hydrostatic pressure that equilibrates the transfer of solvent through a semi-permeable membrane separating the electrolyte solution and a bath of pure solvent [39]. The osmotic pressure plays a crucial role in filtration and separation processes and can be measured experimentally [38]. It can be expressed by the following relation:

\[ p_{osm} = (\rho_s + \rho_{s'}) k_B T - \frac{1}{6} \sum_{s, s'} \rho_s \rho_{s'} \int_0^\infty r \frac{\partial u_{ss'}}{\partial r}(r) g_{ss'}(r) 4\pi r^2 dr, \] (4.29)

where \( \rho_s \) and \( \rho_{s'} \) are the densities in ions of species \( s \) and \( s' \) respectively. \( u_{ss'} \) is the total interaction potential between two ions of species \( s \) and \( s' \). The osmotic coefficient \( \phi \) can then be derived as follows [38]:

\[ \phi = \frac{p_{osm}}{\rho k_B T} = 1 - \frac{1}{6\rho k_B T} \sum_{s, s'} \rho_s \rho_{s'} \int_0^\infty r \frac{\partial u_{ss'}}{\partial r}(r) g_{ss'}(r) 4\pi r^2 dr, \] (4.30)

where \( \rho = \rho_s + \rho_{s'} \). It is interesting to note that the osmotic coefficient has a limit value of one as the concentration of solute goes to zero, in other words the osmotic coefficient of a pure solvent is one.
It should be noted that the osmotic coefficient is derived directly from the RDF and the inter-ionic potentials, and its determination does not add any information about the physics of the system that would not be already contained in the RDF. However, it is worthwhile to use the osmotic coefficient because it is obtained from an integral computation and therefore accounts for the behavior of the RDF at all distances, whereas the RDF peak value discussed in the previous section only accounts for the RDF at a specific (short) distance. Comparing the predicted osmotic coefficient with experimental measurements therefore allows the assessment of the correctness of the RDF as computed over the complete range of distances. The osmotic coefficient is obtained as follows:

\[
\phi = 1 - \frac{1}{6k_BT} \sum_{l=s,s'} \sum_{m=s,s'} \rho_l \rho_m \sum_{k=0}^{N_{points} - 1} \int_{r_k}^{r_k} \frac{u_{l,m}(r_k)g_{l,m}(r_k)4\pi r_k^3 dr_k}{}\]

where \( \rho = \rho_s + \rho_{s'} \) is the total ion density in the solution, \( N_{points} \) is the number of points used to represent the osmotic coefficient \( \phi \) and \( g_{l,m} \) are the RDFs.

4.1.4 Conductivity. Since the objective of the simulation is to represent the charge transport properties of ion channels, an accurate representation of the dynamic properties of electrolyte solutions is also a crucial requirement. In order to validate the correct representation of charge transport in electrolyte solutions, the solution conductivity is proposed as a benchmark.

An electrolyte solution contains mobile charges (ions) which are free to move when an electric field is present. The solution conductivity is measured experimentally by applying an electric potential between two reversible electrodes immersed in the solution. Figure 4.4 shows a simple setup used to measure the conductivity of a solution. Assuming the electrodes have a facing area \( A \), and are spaced at a distance \( l \) from each other, the conductivity \( \sigma \) of the solution is expressed as [39]:

\[
\sigma = \frac{l \times I}{A \times V},
\]
where $I$ is the current measured in the circuit and $V$ the potential difference applied between the electrodes. The unit of $\sigma$ is Siemens per meter ($\text{S.m}^{-1}$) or $\Omega^{-1}\text{m}^{-1}$. The ions in solution are accelerated by the field and move toward an electrode (anode or cathode depending on the ionic charge sign). Once the ions reach the electrode surfaces, they exchange charges through the electrode reactions [39]. In order to remove the effect of these reactions on the conductivity measurement, the potential applied is alternated and an AC conductance is measured. The frequency used is low (in the kHz range) so that the polarization effects in the solvent do not affect the measurement. In the system model proposed in this work, the electrodes are considered far from the system boundaries and their effect is accounted for by the injecting/ejecting contacts with Dirichlet BC.

![Experimental Setup for Measuring the Conductivity of an Electrolyte Solution.](image)

It is reasonable to use the infinite dilution as a first approximation in the description of electrolyte solutions with concentrations in the biological range. In
In this case, the motion of each ion and its contribution to the electric conduction is independent of the other ions. The molar conductance $\Lambda$ is then defined as $\Lambda = \sigma/c$ where $c$ is the molar concentration in the electrolyte. In the case of strong monovalent electrolytes such as KCl or NaCl, the equivalent conductance $\Lambda^*$ [38] is the same as the molar conductance $\Lambda$. The conductance extracted from the simulation results can then be compared directly to experimental measurements of $\Lambda^*$.

Within the Langevin framework, the only model parameters defining the ionic mobility are the limiting diffusion coefficients $D^\infty_k$. They represent the diffusion coefficients for ions of species $k$ in an infinitely dilute solution and are linked to the limiting equivalent conductance [38]:

$$\Lambda^\infty = \sum_{k=i,j} \frac{D^\infty_k e|z_k|F}{k_B T},$$  \hspace{1cm} (4.33)

where $e$ is the elementary charge, $z_k$ is the number of charges born by ions of species $k$, and $F$ is the Faraday number. The consistency of the model can therefore be tested by comparing the computed equivalent conductance $\Lambda^*$ to the limiting conductance $\Lambda^\infty$ corresponding to the simulation parameters $D^\infty_k$.

### 4.2 Benchmarks for Inhomogeneous Systems

The previous section defined the validation procedures used for the bulk electrolyte solution model, both in and out of equilibrium. As our simulation tools are designed to deal with highly inhomogeneous systems, a benchmark had to be selected in order to validate the correct representation of such systems as well. The proposed benchmark is based on an analytical model describing the properties of an electrolyte solution in the vicinity of a dielectric membrane, called the Double Layer (DL) theory.

#### 4.2.1 The Double Layer Theory for Membrane Systems

The structure of an electrolyte solution is modified in the region neighboring any surface in contact with the solution. The accurate description of this phenomenon is crucial for the
correct determination of local ionic densities and potential variations which affect the cell membrane transport and the solution-electrode charge transfer [39]. Although an approach considering all the particles present (ions and water molecules) will clearly yield the most accurate representation, it was shown that the primitive model (see chapter 3) can provide the key elements to the correct description of electrolyte solutions properties. The analogue of this model for the surface-electrolyte interface is the Gouy-Chapman electrical Double Layer (DL) theory [47, 48, 39]. The DL is created by the charge separation associated with the potential difference arising from the contact between two different polarizable media (solution and membrane) [49].

The classical theory of the DL developed by Gouy and Chapman [47, 48] describes the membrane-electrolyte system by means of the Poisson-Boltzmann equation (PBE) for the charge distribution near the surface. In this case, the ions are considered as point charges distributed following Boltzmann statistics [39]:

\[ n_i(x) = n_0 \exp \left[ -\frac{z_i e \psi(x)}{k_B T} \right], \quad (4.34) \]

where \( n_i(x) \) is the number density in ions of species \( i \), \( z_i \) the number of charges for such ions, \( e \) the elementary charge and \( \psi(x) \) the electrostatic potential as a function of \( x \). The bulk number density \( n_0 \) is given by:

\[ n_0 = 1000 \times c_0 N_A, \quad (4.35) \]

where \( c_0 \) is the electrolyte concentration in mol.L\(^{-1} \) and \( N_A \) the Avogadro number, taken to be \( 6.023 \times 10^{23} \) mol\(^{-1} \). Poisson’s equation can then be written as:

\[ \frac{d^2 \psi(x)}{dx^2} = -\frac{\rho(x)}{\epsilon_0 \epsilon_r}, \quad (4.36) \]

where \( \rho(x) \) is the total charge density at position \( x \) and \( \epsilon_r \) the relative dielectric of the solvent (water). Combining equations 4.34 and 4.36 yields:

\[ \frac{d^2 \psi(x)}{dx^2} = -\frac{1}{\epsilon_0 \epsilon_r} \sum_{k=i,j} z_i e n_i(x) \quad (4.37) \]
which gives the PBE [39] after rearranging the terms:

\[
\frac{d^2\psi(x)}{dx^2} = -\frac{n_0 e}{\varepsilon_0 \varepsilon_r} \sum_{k=i,j} z_i \exp \left[ -\frac{z_i e \psi(x)}{k_B T} \right]
\]

The PBE is then be solved analytically for the electrostatic potential and the ion density profiles.

As was done for the bulk electrolyte solution, this primitive model can be refined to yield more accurate results. Indeed, the simple point charge model for the ions was first improved by Stern [50], who proposed to use hard spheres with a given radius to represent the ions. The resulting model, the Modified Gouy-Chapman (MGC) model therefore assigns to both ions an unique closest approach distance to the membrane surface. This model was further refined by Valleau and Torrie [49, 51, 52] based on the observation that in most physical systems (in particular in the systems of interest to this work) the ions and counterions have different sizes. The proposed model called the Unequal Radius Modified Gouy-Chapman (URMGC) model assigns different radii \( r_i \) to each ionic species present.

The URMGC model was therefore chosen as an analytical benchmark to assess the performance of our simulation tools in the description of membrane inhomogeneous systems. The result of the analytical solution of the PBE yields both the electrostatic potential and ionic concentration profiles for a given system. A detailed discussion of this analytical resolution is offered in appendix D. In order to easily compare the simulation results to the analytical URMGC computations, two representative quantities are extracted from the results: the total charge accumulation in the DL and the total potential drop across the DL. The process used to extract these two quantities from the simulation results is described in the following paragraphs.
4.2.2 Double Layer Charge Accumulation. The first quantity extracted from the simulation results for the purpose of comparison with the analytical DL theory computations is the total surface charge accumulated in the DL. Indeed, the membrane surface results in a characteristic charge separation and a net charge density is present in the vicinity of the interface. In order to compute the amount of charge accumulated in the interfacial region, the ion concentrations are averaged in 0.05 nm-thick slices parallel to the membrane plane. Figure 4.5 shows the ionic concentration profiles as a function of the distance to the membrane obtained after the system settled into its steady-state configuration. The typical DL charge separation can be clearly seen, and the total surface charge accumulated is computed by integrating the total charge density across the DL region along the system axis. The resulting value in this case is $\sigma_{DL} = 9.58 \text{ mC.m}^{-2}$. This value will be compared to the results of the analytical URMGC computation (see appendix D).

![Figure 4.5. Ionic Concentration Profiles Obtained from the Simulation of a 0.5 M KCl Solution in the Vicinity of a Membrane.](image)
4.2.3 Double Layer Voltage Drop. The second quantity extracted from the ionic concentration profiles is the voltage drop across the DL. To obtain this quantity, the electrostatic potential has to be computed along the system axis. This is done by solving numerically the one dimensional Poisson’s equation for this system:

\[
\frac{d}{dx} \left( \epsilon(x) \frac{d}{dx} \psi(x) \right) = -\rho(x),
\]

where \(\epsilon(x)\) is the dielectric permittivity, \(\psi(x)\) is the electrostatic potential and \(\rho(x)\) is the charge density. The boundary conditions used here are fixed potentials at both ends of the system, corresponding to the potentials applied while simulating the electrolyte-membrane system. Furthermore, the charge density vanishes far from the interface (the bulk solution is electrically neutral) and the dielectric permittivity is constant in this region, therefore the potential reaches an asymptotic value as \(x\) increases. In these simulations, the dielectric is assumed to be 78 in water \((x > 0)\) and 2 in the membrane \((x < 0)\).

![Figure 4.6. Mean Electrostatic Potential Obtained from the Simulation of a 0.5 M KCl Solution in the Vicinity of a Membrane.](image)
Figure 4.6 shows the evolution of the electrostatic potential in the same 0.5 M KCl bath around the dielectric membrane. The result is consistent with the URMGC analytical model, showing a limited extension of the membrane effect into the bulk solution. This effect vanishes beyond about 1 nm from the interface. The DL voltage drop can be read directly from the potential curve, by taking the difference between the minimum (which is close to the effective surface potential) and the potential far from the surface (fixed to 0). The DL voltage drop is about 9 mV in this case, and it will be compared to the URMGC result to validate the simulation.

4.3 Validation of the Force Field Scheme

Having precisely defined the properties used to determine the accuracy of the simulation algorithm, the computer code can be tested. As discussed in Chapter 3, the computational error associated with the simulation is linked to the discrete scheme used to represent both time and space. Three main parameters define the discretization scheme used by the simulation program, and its influence on the accuracy and correctness of the computed results is evaluated using the bulk electrolyte solution benchmark properties described above. The first parameter is the grid spacing, which determines the spatial resolution of the Poisson solver and the extension of the short-range domain used in the force-field scheme (see Section 3.1). The next two parameters deal with the discrete subdivision of time and are the free-flight (see Section 3.1.3.3) and Poisson (see Section 3.1.2.1) timesteps, indicated by $\Delta t_{\text{ff}}$ and $\Delta t_{\text{Poiss}}$, respectively.

In this section, a rather traditional implementation of the Verlet-like algorithm is used to integrate the Langevin equation. In this first implementation, the numerical approximations initially proposed by van Gunsteren and Berendsen in 1982 [34] have been strictly implemented. These numerical approximations use truncated Taylor series to compute some mathematical functions that arise in the Verlet-like algorithm,
Table 4.1. Ion Physical Parameters

<table>
<thead>
<tr>
<th>Physical Parameters</th>
<th>Na(^{+})</th>
<th>K(^{+})</th>
<th>Cl(^{-})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass [amu]</td>
<td>22.9898</td>
<td>39.0983</td>
<td>35.4527</td>
</tr>
<tr>
<td>Diffusion Coefficient [nm(^2)/µs]</td>
<td>1.334×10(^{-9})</td>
<td>1.960×10(^{-9})</td>
<td>2.032×10(^{-9})</td>
</tr>
</tbody>
</table>

Table 4.2. Ion–Ion Interaction Parameters

<table>
<thead>
<tr>
<th>Lennard-Jones Parameters</th>
<th>Na(^{+})</th>
<th>K(^{+})</th>
<th>Cl(^{-})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well Depth (\epsilon) [eV]</td>
<td>2.034×10(^{-3})</td>
<td>3.773×10(^{-3})</td>
<td>6.505×10(^{-3})</td>
</tr>
<tr>
<td>Well Min. Position (\sigma) [nm]</td>
<td>0.24299</td>
<td>0.31426</td>
<td>0.40447</td>
</tr>
</tbody>
</table>

giving rise to some numerical errors for very long free-flight timesteps. Results presented later in this chapter (section 4.4) make use of an improved second implementation of this algorithm which reduces the numerical error and improve the computing efficiency by using analytical, tabulated values for these mathematical functions.

4.3.1 Grid Spacing. The test system used to determine the validity range of the force field scheme is a cubic simulation box containing a 0.30 mol.L\(^{-1}\) KCl solution. The size of the simulation box is 20 nm in each direction. The temperature is set to 298 K, and all the simulations are performed at equilibrium, that is without any external bias applied between the contacts. The dielectric constant of water in the primitive model is set to 78.0. Tables 4.1 and 4.2 show the physical properties used for the ions, as taken from literature [28, 53].

Simulations have been performed for a range of grid spacing (cell size \(\Delta G\)), from 0.30 to 5.0 nm. Figure 4.7 shows the results obtained for the three benchmarks [54]. The figure shows that the kinetic energy curve has a constant value of about 33 meV. This means that the computed kinetic energy is essentially independent of the grid spacing. The value obtained is off from the benchmark, but it will be shown subsequently that the value only depends on the free-flight timestep which
The RDF peak curve shows a breakdown from the benchmark when the grid spacing ΔG goes below a value of about 0.7 nm. This breakdown is corroborated by a corresponding increase in the osmotic coefficient above its benchmark value at the

**Figure 4.7. Simulation Results Obtained for the Grid Size ΔG Ranging Between 0.30 and 5.0 nm. The Dashed Lines Represent the Benchmark Values [54].**

regulates the accuracy of the Langevin equation integration scheme. The constant value obtained for the kinetic energy is consistent with what is expected from using a free-flight timestep Δt ff of 20 fs, which is used in this set of simulations (for the free-flight timestep sweep results, see Section 4.3.2).
same point. When looking at the equation of the osmotic coefficient (equation 4.30), it appears that when the RDF decreases, the integral value of the osmotic pressure also decreases, therefore the osmotic coefficient moves closer to unity.

This behavior can be explained by a combination of factors. First, the grid spacing should be set to a value lower than the Debye length (see Section 3.1.4). However, in the case of dilute electrolyte solutions, using a cell spacing smaller than the Debye length results in a number of particles per cell smaller than one. Under these conditions, the Poisson solver gives very inaccurate results because the charge distribution is not a smooth function. The use of the TSC charge assignment scheme helps in smoothing the distribution, but that is not enough, given the low number of particles per cell. Therefore, the use of larger grid sizes actually compensates for the low concentration and partially overcomes the Debye length limitation. It is also important to note that ionic solutions do not strictly obey the plasma physics framework in which the Debye length is defined, because the charged particle motion is restricted by the friction coefficient in the Langevin equation. The use of larger cells also improves the accuracy by increasing the short-range domain size used for the PP force computation. The short-range force and RDF are computed only for separations smaller than 2 nm, and a larger grid size improves greatly the accuracy by reducing the influence of high field gradients in the short-range domain. These gradients in the Poisson solution are not correctly compensated for by the reference force. Reducing the roughness of the mesh electrostatic potential therefore reduces the error associated with the field solution.
4.3.2 **Free-Flight Timestep.** The influence of the Langevin equation integration timestep $\Delta t_\text{ff}$ on the simulation accuracy is analyzed next. Simulations have been performed on the same test system for an extended range of free-flight timestep going from 1 fs to 75 fs. Figure 4.8 presents the results of the simulation compared to the three benchmarks.

![Graph](image-url)

Figure 4.8. Simulation Results Obtained for the Free-Flight Timestep $\Delta t_\text{ff}$ Ranging Between 1 and 75 fs. The Dashed Lines Represent the Benchmark Values [54].

The results basically show three main regions in all the curves. A stable region is visible for timesteps between 3 and 40 fs, in which the results for both RDF peak
value and osmotic coefficient are very accurate. These values are within 5% and 0.5% of the HNC results, respectively. The kinetic energy shows a slow decrease but remains within 15% of the analytical model. This decrease is due to the use of the first implementation of the Verlet-like algorithm (see Section 3.1.3.5) that produces kinetic energies lower than the thermodynamic average.

The region corresponding to values of $\Delta t_{\text{eff}}$ below 3 fs presents a breakdown from the benchmark. The average kinetic energy remains close to the theoretical value, but the increasingly large error bars show huge fluctuations of energy in the ionic population. The explanation for this error is likely an accumulation of numerical errors. When the equation of motion is integrated too frequently within the self-consistent approach, numerical errors tend to accumulate instead of canceling. The RDF peak value decreases consistently with an increase in the osmotic coefficient as it has been discussed in the previous section.

The last region occurs for free-flight timesteps higher than 40 fs. In this region the energy quickly increases to excessive values. The RDF peak value also increases with a subsequent decrease of the osmotic coefficient. The explanation for this behavior of the simulation is that the equation of motion is now inaccurately integrated. The ion trajectories are erroneous in the sense that the ions are allowed to get too close to each other. The inter-ionic potential is known to be very steep, and therefore the inter-ionic forces become very large. This in turn leads to excessive energy and a spurious numerical heating of the system. The ionic temperature obtained for $\Delta t_{\text{eff}} = 75$ fs is more than 500K. As ions get closer, the structure of the solution is changed and the RDF peak is higher since more particles are in close range. The osmotic pressure is also increased leading to a larger deviation from unity in the osmotic coefficient.
4.3.3 Poisson Timestep. The last parameter that defines the discrete scheme used in the simulation is the so-called Poisson timestep $\Delta t_{\text{Poiss}}$. This timestep regulates the frequency at which the charge distribution of the system is mapped on the grid and the PM force computed (see Section 3.1.4). Simulations have been performed for the same test system, with a value of 20 fs for the free-flight timestep and a grid spacing of 1 nm. The results obtained for values of $\Delta t_{\text{Poiss}}$ ranging from 0.1 to 30 ps are shown on figure 4.9.

Figure 4.9. Simulation Results Obtained for the Poisson Timestep $\Delta t_{\text{Poiss}}$ Ranging Between 0.1 and 30 ps. The Dashed Lines Represent the Benchmark Values [54].
The simulation results obtained for the different values of the Poisson timestep show very little influence of this parameter. The energy is perfectly stable and presents the value expected with a free-flight timestep value of 20 fs. The RDF peak value shows little error over most of the range (less than 5% error) and the osmotic coefficient is also within 0.5% of the benchmark over the entire range. Some instability can however be noticed for values of $\Delta t_{\text{Poiss}}$ above 5 ps.

The inverse of the plasma frequency (see section 3.1.4.2) for this system is around 3 ps, some instability is therefore expected when going above this value as the solution of Poisson’s equation is not computed often enough in order to follow accurately the evolution of the charge distribution. The results are correct because the system simulated is a homogeneous ionic solution, therefore charge fluctuations are dampened by the influence of the solvent through the high dielectric constant value and the friction in the Langevin equation. One can expect more constraints on the Poisson timestep when dealing with highly inhomogeneous systems such as a membrane–protein complex. Therefore, the value suggested for the Poisson timestep in this system is 2 ps, which is below the plasma oscillation period but still high enough to save computer time. The Poisson solver is then called only every 100 free-flight timesteps, making the simulation very efficient.

### 4.3.4 Bulk Electrolyte Solutions

This section will present various results obtained with the proposed simulation tool. They extend the first set of validation simulations exhibited in the previous sections which was based on a single test system.

A bulk electrolyte solution has been successfully simulated for a variety of concentrations and ionic species. The simulated system is a cubic box of $20 \times 20 \times 20$ nm$^3$. Two facing sides of the box bear injecting “electrodes” however no potential bias is applied in order to simulate a solution at equilibrium. The temperature is 298 K and the dielectric constant for water is set to 78. As a benchmark, the RDF and
the osmotic coefficient are computed and compared to results obtained with the same solutions from both the HNC model and experimental values.

The RDF for KCl and NaCl electrolyte solutions are shown in Figure 4.10 where they are also compared to the numerical HNC solution. As can be seen, the RDFs computed with the simulations show an excellent agreement with the HNC RDF curves, for both electrolytes and various concentrations.

The osmotic coefficient $\phi$ has been obtained using both the HNC (see Appendix C) and simulated RDF curves for KCl and NaCl solutions at different concentrations, and is compared to both experimental and HNC results in Figure 4.11.

The agreement between the HNC and the BD simulation results is excellent. There is a significant discrepancy between the two theoretical models and the experimental values that can be explained by several factors. First, both theories use the primitive model for water, which departs significantly from real solutions. This is especially apparent here because the equilibrium pressure, which is a crucial component
Figure 4.11. Osmotic Coefficient Versus Concentration for (a) KCl and (b) NaCl [29]. Results are compared with experimental values and with results from the HNC. The experimental results shown here use Pitzer experimental fit equations [38] with parameters taken from [55].

of the osmotic coefficient, differs between real solutions and the primitive model [38]. The second point is that the inter-ionic potential used here is the Lennard-Jones potential for which the parameters vary greatly in literature especially for NaCl [53] and do not necessarily match the behavior of real solutions.
4.3.5 Dielectric Slab Simulation. The next sample system that was tested is composed by two electrolyte solutions separated by an impermeable dielectric slab. This is a simplified numerical model of a biological membrane separating two solutions. The simulation allows to determine the charge distribution at the water–lipid interface. The relative dielectric constant of the membrane is set to 2 while the one of the surrounding electrolyte is 80. An external bias is applied at the electrodes, producing a large potential drop across the slab. Due to the dielectric discontinuity, ionic charges accumulate or deplete the interfacial region of the membrane. A plot of the average concentration of anions and cations along the axis normal to the membrane surface is shown in figure 4.12.

![Figure 4.12](image.png)

Figure 4.12. Average Concentration of Anions and Cations in a 0.15 mol.L\(^{-1}\) KCl Solution With a 2 nm Dielectric Membrane in the Center [29].

The membrane thickness is set to 2 nm and the electrolyte bath is a 0.15 mol.L\(^{-1}\) KCl solution on each side. The results show that, for a low voltage bias (0.1 V, figure 4.12a), the ion distribution is approximately homogeneous on both sides of the membrane. When the voltage is increased, K\(^{+}\) ions accumulate on one side of
the membrane and move away from the other, while Cl\textsuperscript{–} ions assume the opposite configuration as a consequence of Gauss’ law.

An advantage of the proposed simulation tool is that it can handle non-periodic, inhomogeneous systems. An example of such a system is shown in figure 4.13, which presents a plot of the average ionic concentration for the same system with an asymmetric concentration of KCl solution.

![Figure 4.13. Average Concentration of Anions and Cations in a KCl Solution With a 2 nm Dielectric Membrane in the Center. The Concentrations Are 0.30 mol.L\textsuperscript{–1} On the Left and 0.15 mol.L\textsuperscript{–1} On the Right Side [29].](image)

This set of dielectric slab simulations has been performed using the same grid as the bulk electrolyte simulations. The Euler integration scheme (see Section 3.1.3.4) was used with free-flight timestep value of 5 fs. The Poisson timestep was set to 2 ps. The total simulation times were 5 ns in each case with averages taken over the last 4 ns of each simulation.
4.4 Validation of the Integration Schemes

4.4.1 Comparing Euler, Verlet-Like and Predictor/Corrector algorithms for Homogeneous Systems. In order to select the integration scheme providing the best efficiency/accuracy trade-off, simulations were performed for a test system using a range of timesteps with each integration scheme. The test system proposed is the same as the one used in the previous section, a homogeneous 0.30 M KCl solution. In this set of simulations, the grid spacing is kept constant at 1 nm and the temperature is 298 K. The standard benchmarks for equilibrium bulk solutions are used to compare the integration schemes.

Figure 4.14 shows the peak value of the radial distribution function (RDF) between K$^+$ and Cl$^-$ ions, compared to the HNC analytical model [38] for a range of integration timestep values. It can be seen that both the Verlet-like (second implementation) and the PC algorithms give very accurate results with timesteps up to about 100 fs. Beyond that point, inaccuracies in the determination of the particle trajectories begin to impair the structure of the electrolyte solution. Figure 3.3 presented earlier shows the evolution of the average kinetic energy of the ionic population as the timesteps are increased, for each integration scheme. This evolution also shows the outstanding stability and accuracy of both second-order Verlet-like and Predictor/Corrector schemes for timesteps in the 100 fs range.

4.4.2 Conductivity Measurements. The integration schemes have also been compared using systems out of equilibrium to validate the dynamic representation of electrolyte solutions, which is crucial for charge transport simulations in realistic conditions. The test system used for this purpose is a homogeneous KCl solution subjected to an applied field. In this case, a steady-state ionic current is established and the conductivity of the solution is computed.
Figure 4.14. Comparison of Different Integration Schemes Over a Range of Integration Timesteps. The RDF Peak Value for a Bulk KCl Solution of 0.30 M Is Used as a Benchmark and Compared to the Analytical HNC Result [54].

Table 4.3 gives the equivalent conductance calculated from the simulation of a bulk KCl solution subjected to an applied constant voltage, and each value is the average of the equivalent conductance of 0.15, 0.30 and 0.60 M solutions, each simulated with 0.25 and 0.50 V applied voltage. The error estimate is the standard deviation of this set of 6 simulations. The free-flight timestep used in these simulations is 100 fs for the Verlet-like and PC algorithms, and 50 fs for the Euler algorithm. The results are consistent with the limiting conductivity computed from the diffusion coefficients (see section 4.1.4).

4.4.3 Electrolyte Structure in the Membrane Vicinity. This last set of results is concerned with the simulation of the inhomogeneous dielectric membrane with an applied bias, and the DL benchmarks outlined in section 4.2.1 are used to validate the
Table 4.3. Conductivity Computed for Euler, Verlet-like and PC Integration Schemes.

<table>
<thead>
<tr>
<th>Integration Scheme</th>
<th>$\Lambda^*$ [S.cm$^{-2}$.mol$^{-1}$]</th>
<th>Estimated Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euler</td>
<td>139.4</td>
<td>±7.80</td>
</tr>
<tr>
<td>Verlet-like</td>
<td>145.3</td>
<td>±3.00</td>
</tr>
<tr>
<td>PC</td>
<td>146.2</td>
<td>±5.74</td>
</tr>
<tr>
<td>Obtained from $D^\infty$</td>
<td>149.9</td>
<td>N/A</td>
</tr>
</tbody>
</table>

results. Figure 4.15 shows the voltage across the double layer formed in the vicinity of a dielectric membrane in KCl solutions, as a function of the concentration. Figure 4.16 shows the total surface charge accumulated in the double layer as a function of the ionic concentration.

The evolution of the DL voltage drop with respect to the concentration is in good agreement with that can be expected from the URMGC calculations (solid lines). All three integration schemes give similar results, therefore validating their use for the correct description of the electrolyte structure close to the dielectric membrane. The discrepancy observed in Figure 4.16 between the values obtained with the Gouy-Chapman theory and the BD simulations comes from the fact that in the former, ions are represented as point charges with an infinite potential, preventing them to approach the membrane surface closer than a given radius for each ion type [52], while in the BD simulations the ions are represented with a soft core Lennard-Jones potential [29]. Therefore, much denser charge accumulations are made possible by the URMGC model, especially at higher concentrations, while in the BD simulation the soft core repulsion prevents the particles from aggregating into a very dense layer next to the membrane, leading to lower DL surface charges at high concentrations.
Figure 4.15. DL Potential Drop for Various KCl Concentrations, Obtained With All Three Integration Algorithms, Compared With the URMGC Analytical Predictions. An Integration Timestep of 20 fs Is Used in the BD Simulations.

Figure 4.16. DL Total Surface Charge for Various KCl Concentrations, Obtained With All Three Integration Algorithms, Compared With the URMGC Analytical Results. An Integration Timestep of 20 fs Is Used in the BD Simulations.
CHAPTER 5

CHARGE TRANSPORT SIMULATION IN ION CHANNELS

The previous chapter offered a comprehensive validation of the proposed simulation approach. It was shown that a correct description of electrolyte solutions can be achieved, while validity ranges were offered for the main simulation parameters. This chapter builds on these results to present the application of this approach to the simulation of charge transport in ion channel systems. The mechanisms of charge transport are introduced with a discussion of experimental methods and results. The system numerical model is reviewed briefly and the chosen charge transport simulation methodology is proposed and validated. The computation results are then presented, with a discussion on their meaning and the information that can be extracted from them regarding the inner works of ion channel systems.

5.1 Measurements In Real Systems

Although the existence of ion channels was inferred from studies of the membrane potential and ionic permeability as early as the 1950s [56], the actual study of individual ion channel transport properties has not been possible until the development of the patch clamp technique by Neher and Sakmann in 1976. This technique allowed to study the electrophysiology of individual ion channels in vivo, and confirmed previous conclusions relying on statistical techniques used to infer the characteristics of individual ion channels based on experiments performed on several hundred or thousands of them. The patch clamp technique not only allowed the measurement of the voltage response of a single channel, but it also allowed the study of its response to external stimuli such as ligands, specific mechanical or osmotic stresses. These studies helped discover, categorize and characterize an increasingly large number of channels [3].
5.1.1 Patch Clamp Measurement Setup. The single channel recording setup relies on the use of a glass micropipette to isolate a small patch of cell membrane containing a single ion channel. The setup can then be used to record channel activity as a part of the cell, or to remove the entire patch containing the channel by withdrawing the pipette from the cell. The technique relies on a tight seal between the pipette and the membrane, which is the result of an extremely fine pipette tip (diameter about 1 μm), thermally treated and coated in order to reduce electrical noise [3]. In order to record current and voltages, a pair of Ag–AgCl electrodes is used. One electrode is inserted in the micropipette, while the other (the reference electrode) is put in contact with the extracellular bath either directly or through a salt bridge to prevent contamination of the setup due to Ag⁺ ions. The current flowing through the channel can then be measured by ways of an op-amp circuit which maintains the desired trans-membrane voltage $V_{\text{ref}}$ while amplifying the current. The output voltage of the circuit is then proportional to the current flowing inside the channel. Figure 5.1 shows a sketch of the patch clamp setup and the schematic of the amplifier circuit used to measure the ion channel current.

![Figure 5.1. Patch Clamp Measurement Setup With Op-Amp Voltage Clamp and Current Amplification [3].](image)
While this technique uses living cells, and therefore could be qualified as an *in vivo* technique, other techniques based on an artificially constructed lipid bilayer have been developed [57]. Ion channels can then be inserted in this artificial bilayer through a process called *reconstitution* [56]. These techniques have the advantage that they do not require micropipettes, a microscope and an extremely accurate manipulator. In these cases, the bilayer is formed across an aperture (50 to 250 μm in diameter) etched into a thin Teflon™ (tetrafluoroethylene) film separating two chambers containing the bath solutions. Each solution is connected to the corresponding electrode by salt bridges. The bilayer can be either painted [58, 59], or created by lowering the aperture in the bath, whereby monolayers re-attach from each side to form the bilayer, and seal the aperture [57, 60, 61]. The ion channel is then reconstituted into the membrane by stirring free channels into one bath until an insertion is detected by means of a real-time voltage clamp measurement. The solution with channels is removed from the bath and replaced with an electrolyte so that the system is ready to perform single channel electrophysiology measurements.

5.1.2 Experimental Results. The current flowing through single ion channels are extremely small: usually less than 1 nA. This implies that extreme care must be taken during the measurements to avoid interference from electromagnetic parasitic signals, and any other electromagnetic or chemical contamination, as the system is essentially sensitive to a single ion or molecule. In spite of its sensitivity, the patch clamp technique and its derivatives are now the tools of choice for the study of ion channel electrophysiology [3]. Results are now reliably and reproducibly obtained for a variety of channels under a number of conditions. The growing wealth of data collected forms the basis on which channel models have been developed. This work is part of the effort in developing numerical models for ion channel charge transport, which seek to deepen our understanding of the physics and chemistry underlying transport processes as well as inquire into potential engineering applications for ion channels.
Table 5.1. Experimental Data On OmpF Channel Electrophysiology.

<table>
<thead>
<tr>
<th>Reference</th>
<th>KCl Concentration [mol.L⁻¹]</th>
<th>$V_{\text{ref}}$ [mV]</th>
<th>$G_{\text{trimer}}$ [nS]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[60]¹</td>
<td>0.10</td>
<td>-150</td>
<td>0.90</td>
</tr>
<tr>
<td>[60]¹</td>
<td>1.00</td>
<td>-150</td>
<td>3.33</td>
</tr>
<tr>
<td>[61]¹</td>
<td>1.00</td>
<td>+120</td>
<td>3.30</td>
</tr>
<tr>
<td>[59]²</td>
<td>1.00</td>
<td>-25 to +25</td>
<td>4.15</td>
</tr>
<tr>
<td>[62]²</td>
<td>0.10</td>
<td>-200 to +200</td>
<td>0.63</td>
</tr>
<tr>
<td>[62]²</td>
<td>0.25</td>
<td>-200 to +200</td>
<td>1.25</td>
</tr>
<tr>
<td>[62]²</td>
<td>0.50</td>
<td>-200 to +200</td>
<td>1.89</td>
</tr>
<tr>
<td>[62]²</td>
<td>1.00</td>
<td>-100 to +100</td>
<td>3.73</td>
</tr>
</tbody>
</table>

This work focuses on the *E. Coli* OmpF channel, which has been described in detail in section 2.3.3. Table 5.1 presents data taken from literature, pertaining to the electrophysiology of the wild-type OmpF (no mutations). The applied transmembrane voltage used for the measurements is denoted by $V_{\text{ref}}$, and $G_{\text{trimer}}$ is the measured trimer conductance. This data will be used in the following sections as experimental comparisons to the simulation results. The following should be noted regarding the details of the experiments and the data extraction:

1. In this group of experiments, the applied voltage was kept constant; the trimer conductance is estimated from the current traces at times when all three pores of an individual protein are open.

2. In this group of experiments, the applied voltage is ramped between the values indicated; the trimer conductance is estimated by a linear fit of the current–voltage curve slope, also when all three pores are open.
5.2 Charge Transport Simulation in Ion Channels

The physical considerations underlying the structure and properties of the ion channel, membrane and electrolyte solution have been described in detail in chapter 2 while the numerical representation chosen for each component of the system was presented and motivated in section 3.2. The following section will therefore focus on the specific choices made for the simulation parameters.

5.2.1 Ion Channel System Numerical Model. Within the BD algorithm, the friction coefficient is the only input parameter that governs the dynamics of the simulated ions. It is obtained from the experimental diffusion coefficients through the Einstein relation, which is applicable to BD models for molecular-sized particles [39]:

$$\gamma_i^\infty = \frac{k_B T}{D_i^\infty},$$  \hspace{1cm} (5.1)

where $\gamma_i^\infty$ is the friction coefficient for the ion species $i$, $D_i^\infty$ is its experimental diffusion coefficient, $k_B$ is Boltzmann’s constant, and $T$ is the system temperature. The superscript $^\infty$ in equation 5.1 refers to the fact that the diffusion (and hence friction) coefficients used in the model are valid for bulk, infinitely dilute ionic solutions [38]. Whether using this diffusion coefficient is a valid choice within the P3M framework is tested in this work and will be discussed in section 5.4.

We have chosen to represent the membrane as an impermeable dielectric barrier (see section 3.2.2). This simple representation is likely to be a suitable choice in the case of relatively large pore channels such as porins, where the transport may not significantly be affected by the external surface of the channel in contact with the membrane. On the other hand, the model chosen for the channel protein is an all-atom representation where the protein atoms are represented as fixed soft spheres with a Lennard-Jones [39] interaction potential, and the protein volume is given a specific relative dielectric value. Furthermore, the protein atoms bear electrostatic
charges which are computed with the GROMACS [45] force field, using the standard protonation states assumed for a neutral pH. The porin molecule bears a total of $-36$ elementary charges in this case.

The following assignments were made for the relative dielectric constants: the water dielectric constant was set to $\epsilon_{\text{water}} = 78$ throughout the system, while the dielectric membrane was chosen as having a relative dielectric of $\epsilon_{\text{membrane}} = 2$. The value of the dielectric constant of the protein is treated as a parameter of the simulation, as its effect is studied in this work. The values for the protein dielectric $\epsilon_{\text{protein}}$ were taken between 5 and 30. The dielectric constant is taken to be stepwise constant throughout the system grid (see chapter 3). The grid uses a cell spacing of 1 nm in the bulk solution regions, and gradually reduces to a finer 0.4 nm grid spacing in the channel region. In some cases, the dielectric map was smoothed by the application of a simple moving average over the three-dimensional grid. In those cases, the final dielectric constant in a cell was obtained as the arithmetic unweighted average of the preset cell dielectric value and that of its 26 neighboring cells in space. For the cells adjoining to the system boundaries, the average is taken on the dielectric values of the cells that are part of the system only. The simulations making use of a system grid obtained through this smoothing process will be referred to as smoothed dielectric simulations.

The system boundaries are non-periodic, which allows for the direct application of a trans-membrane electrostatic potential and avoids artifacts due to the application of periodic boundary conditions to extremely inhomogeneous systems. The system is enclosed in a box, the channel being positioned at the center, which is partitioned by the 4 nm-thick dielectric membrane in the $x$-$y$ plane (see figure 5.2). The ionic charge transport therefore occurs along the vertical $z$ axis. The two opposite boundaries normal to the $z$ axis are assigned Dirichlet Boundary Conditions (BC)
Figure 5.2. System Model Schematic Showing Geometry and Boundary Conditions.

(see section 3.2.4), where the potential value is fixed in the first plane of cells adjacent to the system boundary. Furthermore, to account for the correct representation of the electrolyte and the effect of far electrodes, the concentration in ions and counter-ions is kept constant in these “contact” cells through an appropriate injection/ejection algorithm (see section 3.2). In particular, in order to minimize the disruptive effect of the newly injected ions on the thermodynamic properties of the bath, they are injected with velocities following a Maxwellian distribution in the directions parallel to the contact cells, and a half-Maxwellian in the direction normal to the contact cells. This injection process also ensures that the steady-state current flux established after a transient period does not deplete the system. The remaining four system boundaries are assigned Neumann BC, with the component of the electric field normal to the boundary set to zero and a specular reflection of the particles hitting the boundary. In order to reduce the artifacts due to the boundaries, the box is made relatively large in order to secure a buffer zone between the injecting contacts and the channel area. This buffer is more than 15 nm on each side of the membrane resulting in a
total size of 41.6 nm for the box. Similarly, a margin of 7 nm was reserved around the channel giving dimensions of 21.6 nm in $x$ and $y$ directions.

The time discretization of the system is determined by the free-flight timestep, and by the time interval between two consecutive solutions of Poisson’s equation, the Poisson timestep, which is chosen to be 2 ps, much longer than the free-flight timestep. This is done to save computation time, as the overall changes in the system charge distribution are much slower than the individual particle motions. Indeed, one estimator of the minimum frequency required for electric field updates is the plasma frequency of the ions, which is in the order of a few picoseconds in the case of biologically relevant aqueous solutions (see section 3.1.4.2).

5.2.2 Preliminary Results. In order to analyze the electrostatics of the pore interior, the approach chosen is to use an ion ($K^+$ or $Cl^-$) as a test particle to measure the force it experiences within the pore. The process is easy to implement in the simulator since the test particle has simply to be successively placed at the positions in which the data is desired. The complete P$^3$M force field scheme is used to determine the electrostatic potential in the whole simulation box. After each required coordinate has been processed, the data is collected in a 3D map. It is then possible to visualize the potential and the force at each point, either with a 3D representation or by producing 2D slices showing the force and potential over a specific plane. This allows the close examination of the electrostatics of the pore and the influence of both the type of the permeating ion and of the distribution of the dielectric constant.

Figure 5.3 shows a 2D slice of the OmpF porin channel, on a plane perpendicular to the pore axis. Both potential and force experienced by a $K^+$ ion are shown in a color coded map. The picture shows clearly the three openings of the trimer. Figure 5.4 shows the same results obtained for a $Cl^-$ test particle. The values for both force and potential are higher in this case due to the larger finite-size of the
chloride ion, resulting in more constricted pores. The value of the dielectric constants were set to 2 for the inside of the protein and the membrane, and 78 in the electrolyte bath and the pore regions. The force field scheme uses the Lennard-Jones empirical particle-particle potentials, and no external bias is applied to the simulation domain.

Having obtained the potential and force values everywhere in the simulation domain, it is now possible to extract the profile of the possible ionic permeation pathways along the pores. A minimization algorithm has been implemented following a modified Dijkstra critical path algorithm in order to extract the most favorable path in the force or potential distribution. This path can be extracted by considering
either the sum of the force experienced at each point along the path, or the sum of
the potential. None of these two criteria yields the actual thermodynamically favored
path for the ion in the real, dynamic system, but the paths allow the observation
of the potential and force profiles along a path that crosses the pore assuming that
the test ion does not change the electrostatic structure of the system, which is also
assumed motionless. These results can therefore be considered as possible pathways
for the ion permeation, even though they are not necessarily the most likely ones. The
analysis of these paths allows to study of the potential profile experienced by each
ion along a path that actually crosses the channel, and also estimates the influence
of the dielectric constants used in the model.
Figure 5.5 shows the minimal potential path for K\(^+\) computed in one of the OmpF porin pores in 3D. The plot also shows a force iso-surface representing the surface of the channel molecule in the pore region. The path has been computed by obtaining the lowest potential while moving the test-ion down in the \(z\) direction, starting from a position located at the top opening of the pore. The successive positions obtained are shown superimposed on the 3D force iso-surface. It can be seen that the K\(^+\) ion follows a spiraling path in the pore, as previously discussed in [28].

![Potential Profile](image)

Figure 5.6. Potential Profile for a K\(^+\) Ion Along the Minimum Constraint Path in the Porin Channel Pore.

Once the path coordinates have been obtained, it is possible to look at the potential profile along this path. Figure 5.6 shows the distribution of the electrostatic potential experienced by a K\(^+\) ion along the minimum potential path shown in figure 5.5. This simulation uses a dielectric constant of 2 for the inside of the protein, and a value of 78 for the bulk solution and the inside of the pore. The picture clearly shows three separate potential wells in an area that corresponds to the constriction region of the pore.
5.2.3 Ionic Current Computation Methodology. In order to compute the ionic currents in the channel and extract the related charge transport properties, the current flux of each ionic species was recorded at the inner and outer openings of each pore in the channel. At the end of each free-flight timestep, particles crossing either boundary were recorded, and the total current flux was updated depending on the charge of the particle and the crossing direction. The resulting current flux curves for each particle are linearly increasing over time, and their slope, extracted with a least squares linear regression, yields the average current going through the pore opening. Finally, the total channel current is given by the average of the inner and outer currents in each of the three pores. Figure 5.7 shows how the current recording boundaries are set.

![Figure 5.7. Pore Openings Model and Current Recording Procedure.](image)

5.2.4 Validation. Preliminary work has been presented in chapter 4 to validate the use of the extended primitive model for the electrolyte solutions of interest at concentrations in the biological range (up to 1 M). The bulk electrolyte solution was simulated at equilibrium and the RDF of the ions in solution was computed. The RDF was compared with results obtained with analytical computations for primitive model ionic solutions using the integral equations [63, 38] and the Hypernetted Chain (HNC) approximation [38, 41] as their closure relation (see appendix C). It was
Table 5.2. Diffusion Coefficient Values Used in This Work. The Superscript $'∞'$ Refers To Infinite Dilution Or Bulk Values, While the Reduced Diffusion Value $D'$ Is Taken From [62].

<table>
<thead>
<tr>
<th></th>
<th>K⁺</th>
<th>Cl⁻</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D^∞$</td>
<td>[m².s⁻¹]</td>
<td>$1.960 \times 10^{-9}$</td>
</tr>
<tr>
<td>$D'$</td>
<td>[m².s⁻¹]</td>
<td>$0.630 \times 10^{-9}$</td>
</tr>
</tbody>
</table>

also shown that the use of the Verlet-like [34] integration scheme for the Langevin equation provided the best trade-off between accuracy and computational efficiency. The RDF and kinetic energy benchmarks for bulk electrolyte solutions at equilibrium have shown a correct and accurate representation of the thermodynamics and the structure of the solutions.

Further work was conducted in order to validate an improved implementation of the Verlet-like integration scheme. The bulk solution was simulated under an applied voltage to compute conductivity and the results were successfully compared to the experiment, demonstrating the possibility of using free-flight timesteps in the 100 fs range while maintaining excellent accuracy in the description of the dynamic transport properties of the ions. Furthermore, to validate the accurate representation of the electrolyte solution structure in the vicinity of a cell membrane, simulations were performed with a dielectric membrane under an externally applied trans-membrane voltage. The resulting ionic distributions in the vicinity of the membrane and the associated electrostatic potential were successfully compared to the analytical results of the Gouy-Chapman Double Layer (DL) theory [39], therefore validating the use of free-flight timesteps values up to 100 fs in the simulation of inhomogeneous systems as well.

In order to ensure that the integration scheme of the Langevin equation is used within its range of validity in the particular case of the ion channel system, a set of test
simulations was also run to choose an integration timestep that satisfies the accuracy requirements while providing the best computational efficiency. Figure 5.8 shows the evolution of the porin trimer conductance as a function of the integration free-flight timestep. The other simulation parameters are kept constant: KCl bath concentration of 1.00 M, protein region dielectric of $\epsilon_{\text{protein}} = 6$, applied trans-membrane potential of 0.5 V and reduced diffusion coefficients for the ions ($D_i = D'_i$, see table 5.2 for the numerical values). The trimer conductance is compared to the experimental value taken from [59], denoted by the dashed line in figure 5.8. It can be seen that for timestep values 60 fs and below, the conductance remains stable and within the numerical error of the experimental value (given the set of used parameters). If the timestep is greater than 60 fs, the conductance value gradually decreases. In this case, the less frequent integration of the Langevin equation and the consequent inaccuracies in the determination of the individual ion trajectories reduce the simulation accuracy. The duration of the free-flight timestep value retained for all subsequent simulations was therefore fixed to 50 fs, which is well within the acceptable range of integration timesteps for the Verlet-like algorithm (see chapter 4). This value ensures that the structure of the electrolyte both in bulk and in the vicinity of membranes is correctly represented. Choosing one of the largest acceptable integration timesteps obviously reduces the computational cost of the charge transport simulations.
5.3 Results

The system model is defined by several physical and numerical parameters. The difficulty presented by the time-scale of charge transport phenomena in ion channels can be artificially sidestepped by using extremely high concentrations and unrealistic voltages. Indeed, using larger voltages (of the order of 1 V) and concentrations in the molar range could provide statistically meaningful results with shorter simulation times. However, as a trans-membrane potential in excess of only 0.5 V effectively disrupts the bilayer structure [61], only voltages no greater than 0.5 V were used in this work, and between $-200$ and $200$ mV in most cases. Similarly, the very high efficiency of the simulation and the choice of optimal parameters for the BD algorithm allowed to restrain the concentrations used to a maximum of 1 M, resulting in the accumulation of statistically relevant results in runs of 10–100 nanoseconds. This work was therefore restricted to the representation of realistic conditions for the
ion channel system, and only the influence of the model parameters (dielectric and diffusion) which are not determined by the experimental setup (i.e., concentration and applied voltage) were tested.

5.3.1 The Influence of the Dielectric Representation. Since the polarizability of the residues inside the channel pore is not explicitly represented in the model, the polarization effects are only accounted for by means of a dielectric constant. The validity of this choice is arguable, however the relatively wide opening of the OmpF pore allows the ions to remain hydrated as they move through the channel, in which case the implicit solvent and polarization representation is a reasonable approximation. The value of the protein dielectric constant affects how charged groups on the outside as well as inside the protein will influence the ionic dynamics inside and around the pore, and will change the charge transport properties of the channel.

The system model uses the same dielectric value as water (78) inside the lumen of the pore, however the first layer of cells that include the protein surface residues will have a different value. In the case of a smoothed dielectric map (see section 5.2.1), these cells will have a value between 40 and 50, which is the average of the water and protein dielectric constant. When smoothing is not used, those cells will be assigned the same dielectric value as the protein body, $\epsilon_{\text{protein}}$, which is chosen in the range of 5 to 20 since proteins are not a medium as polar as water. Figure 5.9 shows a cross section of the protein and membrane regions, in both smoothed and non-smoothed grid cases. It can be seen that the smoothing operation results in a much higher dielectric for the protein surface regions, as well as less sharp interfaces, particularly in the pore region.

5.3.1.1 The Protein/Water Interface in the Pore Lumen. In order to study the relative importance of the internal and surface dielectric representations of the protein, the following test was conducted: a first simulation was performed with a
smoothed dielectric model, assigning $\varepsilon_{\text{protein}} = 5$. The porin trimer total current was measured with the trans-membrane potential varying between $-200$ and $+200$ mV, and with a KCl concentration of 0.25 M. A second simulation was performed with the same configuration, but a value of $\varepsilon_{\text{protein}} = 20$ was assigned instead. Because of the smoothing, the protein surface has very similar dielectric properties in both configurations, while the inner values are very different.

Figure 5.10 shows the two I-V curves obtained in these two cases. It can be seen that they are almost the same, which leads to the conclusion that the charge transport in the pore is mostly influenced by the dielectric representation of the surface of the protein (inside the pores), while the innermost part of the protein (and the associated dielectric constant) does not significantly affect the charge transport properties of the channel in this model. This is due to the fact that the inner charges inside the protein are effectively screened by the polarizable side chains on the loops inside the pores, preventing them from interacting strongly with the mobile ions in the pore, and therefore do not influence the charge transport in a measurable way.
Figure 5.10. I-V Curves Obtained With a Smoothed Dielectric Representation, With a 0.25 M KCl Bath and Applied Trans-Membrane Potentials Varying Between -200 and +200 mV. Diamonds: $\epsilon_{\text{protein}} = 5$, Triangles: $\epsilon_{\text{protein}} = 20$.

5.3.1.2 The Protein Region Dielectric Constant. To assess the effect of the dielectric value chosen for the protein region, and especially for the surface within the pore constriction region, tests were conducted where the external conditions are maintained constant (KCl concentration of 0.25 M and trans-membrane voltage of 500 mV), while the dielectric constant of the protein region was varied between 5 and 30. The diffusion coefficient used in this case is the bulk diffusion $D^\infty$. In this case, no smoothing was used in the dielectric map so that the dielectric constant was the same across the entire protein body and surface regions. Figure 5.11 shows the evolution of the trimer conductance due to $K^+$, $Cl^-$ and the trimer total conductance as a function of $\epsilon_{\text{protein}}$. It can be seen that for $\epsilon_{\text{protein}}$ values greater than 10, the porin conductance is almost unchanged, while a significant drop in conductance is observed when $\epsilon_{\text{protein}}$ is taken below 10.
To further study the importance of the protein dielectric properties, two sets of tests were performed under different operating conditions. Two series of computations, one using $\varepsilon_{\text{protein}} = 6$ and the other using $\varepsilon_{\text{protein}} = 20$ (no smoothing) were compared to experimental and PNP simulation results taken from literature [59, 62]. The KCl bath concentration was varied between 0.10 and 1.00 M and the ionic diffusion coefficient was set to the reduced diffusion $D'$ (see table 5.2). Each point was obtained by averaging the pore channel conductance computed under the effect of a trans-membrane potential ranging between $-500$ and $+500$ mV. Figure 5.12 shows the results of these computations inclusive of the standard deviation.

It can be seen that the results obtained with $\varepsilon_{\text{protein}} = 6$ is relatively close to experimental measurements and PNP results, matching within the standard deviation in most cases. On the other hand, the results produced with $\varepsilon_{\text{protein}} = 20$ show a consistently higher channel conductance (particularly at concentrations higher than
0.5 M). This result is consistent with what was shown in figure 5.11, and can be explained as follows. It is known that the porin constriction region bears charged side-chains from aspartate, glutamate and arginine [64] which give rise to a strong electric field inside the pore [62]. The dielectric constant assumed for this region modulates the strength of the field and its effect on the ion dynamics inside the pore. A higher dielectric value reduces the field and facilitates the ionic motion, while a lower dielectric value increases the field. A stronger field increase the forces on the ions and reduces their mobility in the pore. Comparison with experimental values shows that under the operating conditions used here (reduced diffusion $D'$) the lower value of 6 for the protein dielectric provides a better representation.

5.3.2 The Influence of the Diffusion Coefficients. The diffusion coefficient is a parameter of the simulation having a direct influence on the charge transport properties of the channel. Unlike the dielectric constant, which influences the ion
motion by modifying the electrostatic interaction between the charged groups on the protein surface and the ions, the diffusion coefficient modifies the dynamic behavior of the ions in the Langevin equation framework. Indeed, a smaller diffusion coefficient means a higher friction coefficient (according to equation 5.1), and thus a lower ionic mobility. Several mappings of the diffusion coefficient aimed at describing accurately the ion transport in the channel pores have been described in literature [58, 62]. The diffusion coefficient is set to specific “local” values depending on the area of the pore that is considered (pore entrance on the intracellular side, pore constriction region, and pore exit on the extracellular side) such as in [58], or a single, system-wide diffusion coefficient value can be used as a fitting parameter to match experimental measurements [62].

In order to evaluate what diffusion coefficient provides the most appropriate description of the ion motion in the channel, two sets of values for the diffusion coefficients ($D^1$ and $D^0$) are used as inputs, and the channel conductance is computed and compared to experimental results. In both sets, the KCl concentration is varied between 0.10 and 1.00 M, while $\epsilon_{\text{protein}} = 6$. Again, each point is obtained by averaging the pore channel conductance computed by varying the trans-membrane potential values between $-500$ and $+500$ mV. Figure 5.13 shows the results of these tests, the error bars represent the standard deviation associated with this set of computations.

It can be seen that using the bulk diffusion results in very high trimer conductances values, three to four times larger than experimental values under the same conditions. The use of the reduced diffusion value ($D^r$, see table 5.2) produces results that are much closer to experiment. The reduced diffusion was proposed as a fit to experimental data for the PNP model [62]. It can be seen that the actual numerical value of the diffusion coefficients is about three times smaller than the bulk diffusion $D^\infty$, which explains the lower conductance observed in the OmpF channel.
5.4 Discussion

The main purpose of this work is to show the suitability of a non-periodic P$^3$M force field scheme coupled with a Brownian Dynamics simulation engine for the correct description of charge transport in ion channels. The OmpF channel is supposedly large enough for the ions to be considered hydrated inside the channel. This implies the possibility of using bulk diffusion and dielectrics, and describing the ions within the primitive model. The results obtained here with the P$^3$M BD simulations, as well as continuum models [62, 58], and their comparison with experimental data [62] (Figures 5.12 and 5.13) showed that this rough approximation does not provide a model close enough to the real system. However, it was also shown that the parameters used by this algorithm can be adjusted to provide very accurate results by carefully tuning dielectric constants and diffusion coefficients in order to account correctly for the polarization in the pore region.
The proposed simulative approach is therefore able to provide insight in the processes governing the charge transport in OmpF. Much knowledge was gained in the process of developing and validating the algorithm. The details of the dielectric representation of the protein within the BD framework have been thoroughly studied and are now better understood. The meaning of the diffusion coefficient in bulk solution, and the possibility of extending it to the pore transport has been studied as well. This knowledge however raises a new range of questions, which are already being fathomed by experiments as well as other simulation approaches. Some of these questions will be discussed in the next sections, together with their bearing on the proposed P3M BD simulative approach.

5.4.1 Protonation States of the Protein Residues. The OmpF channel is represented as a static, all-atom structure in the proposed approach. Each atom is represented by a Lennard-Jones soft-core particle with a fixed partial charge obtained from the tabulated GROMACS [45] force-field. The values of these charges on the side-chains of the residues lining the pore is an absolutely critical component of the charge transport model. Figure 5.14 shows a top view of OmpF where some of these residues have been highlighted. It can be seen that the positively charged arginine residues ARG42, ARG82 and ARG132 are effectively facing the negatively charged glutamate GLU117 and aspartate ASP113. This gives rise to a strong transverse electric field across the constriction region [62]. It is therefore crucial to know the correct protonation states of the residues lining the pore as they undoubtedly have a strong influence on the charge transport process [65].

Computing the actual protonation states of each protein residue is a complex task as it involves expressing the interactions between neighboring residues, as well as accounting for specific molecular conformations, sometimes including water molecules [13, 19]. An estimate of the protonation states of the residues can be computed
Figure 5.14. OmpF Trimer View Showing Some of the Charged Residues Lining the Constriction Region (Generated With VMD [20]).

by using the standard pKa of the seven residues bearing titratable side chains [13]. The relative fraction of the residues that are charged can be computed depending on the pH, and the total charge of the porin is then obtained. Figure 5.15 shows the total charge computed for OmpF as a function of the pH using the standard pKa for the titratable residues (see table 5.3). It can be seen that the porin bears a net charge of $-36$ elementary charges at neutral pH $= 7$, consistently with the total charge assigned by the GROMACS force field. While this overall negative net charge in a neutral environment could explain some of the OmpF selectivity towards cations [66], only the accurate modeling of charge and shape of the pore can account for differences in transport for cations and anions. Furthermore, the protonation state of the key residues is a dynamic property that can effectively change during transport, adding another degree of complexity to the process.
Table 5.3. Side Chains and pKa of the Titratable Residues [13].

<table>
<thead>
<tr>
<th>Residue</th>
<th>pKa</th>
<th>Acid Form</th>
<th>Basic Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASP</td>
<td>3.9</td>
<td>R—COOH</td>
<td>R—COO⁻</td>
</tr>
<tr>
<td>GLU</td>
<td>4.1</td>
<td>R—COOH</td>
<td>R—COO⁻</td>
</tr>
<tr>
<td>HIS</td>
<td>6.0</td>
<td>R—N⁺R'</td>
<td>R—N'R'</td>
</tr>
<tr>
<td>CYS</td>
<td>8.4</td>
<td>R—SH</td>
<td>R—S⁻</td>
</tr>
<tr>
<td>LYS</td>
<td>10.5</td>
<td>R—NH₃⁺</td>
<td>R—NH₂</td>
</tr>
<tr>
<td>TYR</td>
<td>10.5</td>
<td>R—OH</td>
<td>R—O⁻</td>
</tr>
<tr>
<td>ARG</td>
<td>12.5</td>
<td>R—NH₃⁺</td>
<td>R—NH</td>
</tr>
</tbody>
</table>

Figure 5.15. Computed Total Net Charge of OmpF As a Function of the pH.
5.4.2 Conformation Changes and Polarization. The protein structure is represented as a set of fixed charged particles, while the actual structure is very mobile. Charged side-chain groups change their conformation in the presence of an ion, altering its hydration state as it travels in the pore narrow region. It can be reasonably assumed that the strong fields present in this region affect the geometry of charged and polar groups by stretching and rotating bonds. This molecular polarization is not accurately represented by the fixed dielectric constant, if at all.

Furthermore, another aspect of ion channel transport that is linked to the conformational changes of key groups in the pore is gating. In gated channels (there are arguments for the presence of voltage gating in OmpF [66]), the displacement of charged or polar groups induced by the electric field effectively closes the channel. This effect is obviously not represented by a static structure. However, the validity of using the crystal structure to represent the open state of the channel is corroborated by the matching of experimental conductance results shown here, as well as the small deviation from the crystal structure observed in MD simulations [64].

5.4.3 Ion Diffusion and Ion-Solvent Interaction. Finally, the validity of using an implicit solvent model for the ions is argued here. Indeed, the results have shown that the bulk diffusion parameters, while correctly modeling ionic transport in bulk solutions under a variety of conditions, are however erroneous in the case of transport inside a narrow channel. In fact, the bulk diffusion assumes a complete hydration shell around the ions as it represents the motion of ions in an infinitely diluted solution. While the OmpF pore is (relatively) wide, the complete hydration shell is certainly not retained during transport. The modality of the changes in the hydration shell and the possible substitution of water by pore-lining residues, and the associated changes in the dynamics of charge transport are clearly not represented by the proposed approach.
5.5 A Step Closer Toward Global Simulation

It is important to note that the model proposed here is extremely less costly in terms of computation than an all-atom MD would be. The use of empirical values for the dielectric constants and diffusion coefficients was proved successful in capturing the physics of charge transport inside the channel pores. The model can reasonably be applied to the study of the charge transport properties of channels similar to OmpF, both wild-type and engineered, as well as analogous nanoscale structures such as decorated nanotubes, without the staggering cost of MD simulations.

However, the simulation approach proposed in this work should not be interpreted as a “close” model able to predict the complete electrophysiology of an arbitrary, non-characterized ion channel. The proposed approach is very useful in providing a prediction tool for the description of a known channel. These predictions can be comfortably extended to channels with minor structural alterations (mutations) or to a variety of system-level configurations such as ionic solutions, membrane geometry, and applied voltages. This flexibility will prove extremely useful in the Computer Aided Design of ion channel-based devices, which show a promising future in engineering and biomedical applications.

This work has enhanced our knowledge of ion channel modeling issues. The P$^3$M BD simulation approach was extensively tested in its application to charge transport in ion channels. This work makes even more sense as a part of a hierarchical approach to the modeling of ion channels and nanostructures in general, as outlined in [67]. Every level of the hierarchy, from the most detailed atomic level including all the chemical, electrical and even quantum physical phenomena, up to the system level, make sense as a part of a global effort to enhance our understanding and utilization of ion channels. This work is meaningful as a part of this effort to discover, characterize and eventually exploit the properties of ion channels.
CHAPTER 6
CONCLUSION

The main contribution of this work is an assessment of the validity of the particle-based \( \text{P}^3\text{M BD} \) simulation of charge transport in ion channels. While a number of simulative approaches have already been applied to the problem of ionic transport in channels, the novel approach presented in this work proposes appealing features which other approaches lack.

The continuum flux-based methods such as PNP do not account directly for the effects due to ion size, and make use of fitting parameters to produce quantitative transport properties. While being computationally less expensive than the proposed approach, they are subject to severe limitations relative to the dielectric representations of the protein, which have been discussed in this work. On the other hand, the computationally intensive all-atom MD simulations present other issues. While the physics of molecular polarization effects and hydration are fully accounted for, this model suffers from the difficulty of including externally applied boundary conditions in a satisfactory way. Furthermore, the nature of the ion channel transport problem makes MD simulations excessively expensive in terms of computing resources.

The proposed approach relies on an extremely efficient real-space Poisson solver originally developed for semiconductor device simulations. This solver correctly accounts for arbitrary externally applied fields on non-periodic systems, which are crucial requirements for ion channel systems. The dynamics of the ions is modeled by the full Langevin equation, and several numerical integration schemes have been implemented and validated in this work. Therefore, the resulting simulative approach presents an attractive middle ground, being more detailed in its description than the continuum methods while remaining less computationally expensive than the full-atom MD simulation approach.
The proposed approach was first applied to bulk electrolyte solutions. In order to validate the model, precise thermodynamic and structural benchmarks were defined. The HNC approximation was used to implement an analytical model for ionic solutions and then successfully validate our BD description of bulk solutions. The dynamics of the system was validated by simulating ionic currents and comparing them with conductance benchmarks. A further step was taken in the validation process by modeling inhomogeneous systems. To that end, a dielectric membrane subjected to applied potentials was simulated. The analytical URMGC model was used to provide a benchmark. Three different numerical integration schemes for the Langevin equation of motion were implemented: the Euler scheme, the Verlet-like scheme and the Predictor/Corrector scheme. Their respective merits in terms of efficiency and accuracy were assessed using the proposed validation methodology, and an optimal algorithm and set of parameters were defined for the simulation of charge transport in ion channels.

The OmpF channel from *E. Coli* was chosen as a test case for the proposed simulation approach. The numerical model was first validated by computing potential profiles for Potassium and Chlorine ions inside the porin channels. A simulation methodology for the study of charge transport was then defined and employed to compute the channel conductance under a variety of solution concentrations and applied trans-membrane voltages. The parameters of the model were then studied. In particular, the influence of the dielectric mapping of the channel protein was studied in depth. The diffusion coefficient governing the ion dynamics was also studied. The computed charge transport properties of OmpF were compared to experimental measurements to determine the validity of the approach. It was shown that the numerical model can be adjusted to provide an accurate model for ion channel charge transport, while keeping a large computational advantage on MD simulations. Much knowledge was gained regarding the issues involved in implicit solvent models and
polarization. Finally, a reflection on the meaning of this work in the broader field of ion channel study and applications was offered.
APPENDIX A
PROTEIN DATA BANK
A.1 General Description

The Protein Data Bank (PDB) is a worldwide repository of macromolecular structures, providing free access to 3D macromolecular structural data. The PDB is managed by the Research Collaboratory for Structural Bioinformatics (RCSB) and was established in 1971 at the Brookhaven National Laboratories as an archive for biological macromolecular crystal structures [18]. Since the 1980s, the number of available structures increased at a very fast pace. The development of techniques for protein crystallography such as the Nuclear Magnetic Resonance (NMR) helped produce more accurate structures, and more easily. The improvements of the available computational machinery also improved the capability to process and distribute the data collected in the experiments. The database contains more than 31,000 structures as of June 2005 (http://www.rcsb.org/pdb/).

The dramatic development of the World Wide Web has allowed increased connectivity and facilitated the communication between researchers in structural biochemistry. The process of making a new structure available has been simplified and accelerated, making the expansion of the PDB even faster. The success of the PDB relies on a unified file format for the description of 3D molecular structures.
A.2 File Formats and Standards

The PDB file format provides a self-contained, exhaustive description of a protein structure. The single PDB file contains all the information about the structure, protein name and classification, contributor name and publication reference, experimental technique used to determine the structure, DNA amino-acid residue sequence, protein secondary structural domains (see Chapter 2), tertiary and quaternary structure definition and, finally, a complete list of all the atoms and their coordinates. Each data entry is a line, with a 6-character header defining the information contained in the line.

The PDB file starts with a header giving the identification of the protein and its PDB identifier. The *E. Coli* OmpF porin bears the PDB identifier 2OMF. After a set of optional remarks describing the experiment setup used for the determination of the structure, the primary protein sequence is given in terms of the amino-acid residue sequence. Each residue is described by its standard 3-letter code. Next comes the description of the various secondary structural domains of the protein, signaled by headers like HELIX or SHEET. Finally, all the atoms of the protein are listed in lines starting with the explicit header ATOM, followed by the atom sequence number, the atom name, the residue code of which it is part, and the atomic 3-dimensional coordinates.

The next section will provide some example sections of the PDB file 2omf.pdb [68] describing the porin channel.
A.3 Example: OmpF Porin PDB File

Here are the first lines of the files giving the name and information about the protein described in the file:

```
HEADER INTEGRAL MEMBRANE PROTEIN PORIN 28-FEB-95 2OMF 2OMF 2
TITLE OMPF PORIN
```

The next sample portion is the section of the PDB file describing the primary residue sequence. Here is the complete description of the 340 residues composing one of the three monomers forming the OmpF trimer:

```
SEQRES   1 340  ALA  GLU  ILE  TYR  ASN  LYS  ASP  GLY  ASN  LYS  VAL  ASP  LEU  2OMF  115
SEQRES   2 340  TYR  GLY  LYS  ALA  VAL  GLY  LEU  HIS  TYR  PHE  SER  LYS  GLY  2OMF  116
SEQRES   3 340  ASN  GLY  GLU  ASN  SER  TYR  GLY  GLY  GLY  ASN  ASP  GLY  ASP  2OMF  117
SEQRES   4 340  TYR  ALA  ARG  LEU  GLY  PHE  LYS  GLY  GLU  THR  GLN  ILE  ASN  2OMF  118
SEQRES   5 340  SER  ASP  LEU  THR  GLY  TYR  GLY  GLN  TRP  GLU  TYR  ASP  PHE  2OMF  119
SEQRES   6 340  GLN  GLY  ASN  SER  GLU  GLY  ALA  ASP  ALA  GLN  THR  GLY  2OMF  120
SEQRES   7 340  ASN  LYS  THR  ARG  LEU  ALA  PHE  ALA  GLY  LEU  LYS  TYR  ALA  2OMF  121
SEQRES   8 340  ASP  VAL  GLY  SER  ASP  TYR  GLY  ARG  ASN  TYR  GLY  GLY  VAL  2OMF  122
SEQRES   9 340  VAL  TYR  ASP  ALA  LEU  GLY  TYR  THR  ASP  MET  LEU  PRO  GLU  2OMF  123
SEQRES  10 340  PHE  GLY  GLY  ASP  THR  ALA  TYR  ARG  ASP  ASN  THR  PHE  VAL  2OMF  124
SEQRES  11 340  GLY  ARG  VAL  GLY  GLY  VAL  ALA  THR  TYR  ARG  ASN  SER  ASN  2OMF  125
SEQRES  12 340  PHE  PHE  GLY  LEU  VAL  ASP  GLY  LEU  ASN  PHE  ALA  VAL  GLN  2OMF  126
SEQRES  13 340  TYR  LEU  GLY  LYS  ASN  GLU  ARG  ASP  THR  ALA  ARG  ARG  SER  2OMF  127
SEQRES  14 340  ASN  GLY  ASP  GLY  VAL  GLY  SER  ILE  SER  TYR  GLY  TYR  2OMF  128
SEQRES  15 340  GLU  GLY  PHE  GLY  ILE  VAL  GLY  ALA  THR  GLY  ALA  ALA  ASP  2OMF  129
SEQRES  16 340  ARG  THR  ASN  LEU  GLN  GLU  ALA  GLN  PRO  LEU  GLY  ASN  GLY  2OMF  130
SEQRES  17 340  LYS  LYS  ALA  GLU  GLN  THR  ALA  THR  GLY  LEU  LYS  TYR  ASP  2OMF  131
SEQRES  18 340  ALA  ASN  ASN  ILE  TYR  LEU  ALA  ALA  ASN  TYR  GLY  GLU  THR  2OMF  132
SEQRES  19 340  ARG  ASN  ALA  THR  PRO  ILE  THR  ASN  LYS  PHE  THR  ASN  THR  2OMF  133
SEQRES  20 340  SER  GLY  PHE  ALA  ASN  LYS  THR  LYS  LEU  VAL  ALA  GLU  VAL  2OMF  134
SEQRES  21 340  ALA  GLN  TYR  GLN  PHE  ASP  PHE  GLY  LEU  ARG  PRO  SER  ILE  2OMF  135
SEQRES  22 340  ALA  TYR  THR  LYS  SER  LYS  ALA  LYS  ASP  VAL  GLU  GLY  ILE  2OMF  136
SEQRES  23 340  GLY  ASP  VAL  ASP  LEU  VAL  ASN  TYR  PHE  GLU  VAL  GLY  ALA  2OMF  137
SEQRES  24 340  THR  TYR  TYR  PHE  ASP  ALN  LYS  ASP  GLY  MET  MET  THR  TYR  VAL  2OMF  138
SEQRES  25 340  TYR  ILE  ILE  ASN  GLN  ILE  ASP  SER  ASP  ASN  ASP  LEU  GLY  2OMF  139
SEQRES  26 340  VAL  GLY  SER  ASP  ASP  THR  VAL  ALA  VAL  GLY  ILE  VAL  TYR  2OMF  140
SEQRES  27 340  GLN  PHE  2OMF  141
```
Finally, here is a sample section showing the atomic coordinates of the atoms belonging to the first three residues of the monomer:

<table>
<thead>
<tr>
<th>ATOM</th>
<th>1</th>
<th>N</th>
<th>ALA</th>
<th>1</th>
<th>-1.906</th>
<th>59.835</th>
<th>18.950</th>
<th>1.00</th>
<th>20.54</th>
<th>2OMF 187</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATOM</td>
<td>2</td>
<td>CA</td>
<td>ALA</td>
<td>1</td>
<td>-2.032</td>
<td>60.264</td>
<td>17.533</td>
<td>1.00</td>
<td>21.15</td>
<td>2OMF 188</td>
</tr>
<tr>
<td>ATOM</td>
<td>3</td>
<td>C</td>
<td>ALA</td>
<td>1</td>
<td>-0.689</td>
<td>60.841</td>
<td>17.144</td>
<td>1.00</td>
<td>23.34</td>
<td>2OMF 189</td>
</tr>
<tr>
<td>ATOM</td>
<td>4</td>
<td>O</td>
<td>ALA</td>
<td>1</td>
<td>-0.071</td>
<td>61.551</td>
<td>17.925</td>
<td>1.00</td>
<td>25.09</td>
<td>2OMF 190</td>
</tr>
<tr>
<td>ATOM</td>
<td>5</td>
<td>CB</td>
<td>ALA</td>
<td>1</td>
<td>-3.111</td>
<td>61.320</td>
<td>17.397</td>
<td>1.00</td>
<td>12.49</td>
<td>2OMF 191</td>
</tr>
<tr>
<td>ATOM</td>
<td>6</td>
<td>N</td>
<td>GLU</td>
<td>2</td>
<td>-0.215</td>
<td>60.500</td>
<td>15.956</td>
<td>1.00</td>
<td>24.63</td>
<td>2OMF 192</td>
</tr>
<tr>
<td>ATOM</td>
<td>7</td>
<td>CA</td>
<td>GLU</td>
<td>2</td>
<td>1.061</td>
<td>61.005</td>
<td>15.495</td>
<td>1.00</td>
<td>26.39</td>
<td>2OMF 193</td>
</tr>
<tr>
<td>ATOM</td>
<td>8</td>
<td>C</td>
<td>GLU</td>
<td>2</td>
<td>0.862</td>
<td>62.403</td>
<td>15.012</td>
<td>1.00</td>
<td>24.19</td>
<td>2OMF 194</td>
</tr>
<tr>
<td>ATOM</td>
<td>9</td>
<td>O</td>
<td>GLU</td>
<td>2</td>
<td>0.450</td>
<td>62.604</td>
<td>13.892</td>
<td>1.00</td>
<td>28.92</td>
<td>2OMF 195</td>
</tr>
<tr>
<td>ATOM</td>
<td>10</td>
<td>CB</td>
<td>GLU</td>
<td>2</td>
<td>1.596</td>
<td>60.165</td>
<td>14.349</td>
<td>1.00</td>
<td>27.76</td>
<td>2OMF 196</td>
</tr>
<tr>
<td>ATOM</td>
<td>11</td>
<td>CG</td>
<td>GLU</td>
<td>2</td>
<td>2.852</td>
<td>60.727</td>
<td>13.705</td>
<td>1.00</td>
<td>35.18</td>
<td>2OMF 197</td>
</tr>
<tr>
<td>ATOM</td>
<td>12</td>
<td>CD</td>
<td>GLU</td>
<td>2</td>
<td>3.549</td>
<td>59.695</td>
<td>12.838</td>
<td>1.00</td>
<td>41.92</td>
<td>2OMF 198</td>
</tr>
<tr>
<td>ATOM</td>
<td>13</td>
<td>OE1</td>
<td>GLU</td>
<td>2</td>
<td>4.169</td>
<td>58.758</td>
<td>13.419</td>
<td>1.00</td>
<td>44.58</td>
<td>2OMF 199</td>
</tr>
<tr>
<td>ATOM</td>
<td>14</td>
<td>OE2</td>
<td>GLU</td>
<td>2</td>
<td>3.465</td>
<td>59.818</td>
<td>11.589</td>
<td>1.00</td>
<td>39.28</td>
<td>2OMF 200</td>
</tr>
<tr>
<td>ATOM</td>
<td>15</td>
<td>N</td>
<td>ILE</td>
<td>3</td>
<td>1.178</td>
<td>63.375</td>
<td>15.850</td>
<td>1.00</td>
<td>28.14</td>
<td>2OMF 201</td>
</tr>
<tr>
<td>ATOM</td>
<td>16</td>
<td>CA</td>
<td>ILE</td>
<td>3</td>
<td>0.998</td>
<td>64.764</td>
<td>15.478</td>
<td>1.00</td>
<td>28.31</td>
<td>2OMF 202</td>
</tr>
<tr>
<td>ATOM</td>
<td>17</td>
<td>C</td>
<td>ILE</td>
<td>3</td>
<td>2.218</td>
<td>65.361</td>
<td>14.799</td>
<td>1.00</td>
<td>29.77</td>
<td>2OMF 203</td>
</tr>
<tr>
<td>ATOM</td>
<td>18</td>
<td>O</td>
<td>ILE</td>
<td>3</td>
<td>2.144</td>
<td>66.475</td>
<td>14.285</td>
<td>1.00</td>
<td>36.25</td>
<td>2OMF 204</td>
</tr>
<tr>
<td>ATOM</td>
<td>19</td>
<td>CB</td>
<td>ILE</td>
<td>3</td>
<td>0.677</td>
<td>65.628</td>
<td>16.706</td>
<td>1.00</td>
<td>27.77</td>
<td>2OMF 205</td>
</tr>
<tr>
<td>ATOM</td>
<td>20</td>
<td>CG1</td>
<td>ILE</td>
<td>3</td>
<td>1.817</td>
<td>65.531</td>
<td>17.709</td>
<td>1.00</td>
<td>26.83</td>
<td>2OMF 206</td>
</tr>
<tr>
<td>ATOM</td>
<td>21</td>
<td>CG2</td>
<td>ILE</td>
<td>3</td>
<td>-0.616</td>
<td>65.177</td>
<td>17.348</td>
<td>1.00</td>
<td>28.47</td>
<td>2OMF 207</td>
</tr>
<tr>
<td>ATOM</td>
<td>22</td>
<td>CD1</td>
<td>ILE</td>
<td>3</td>
<td>1.605</td>
<td>66.388</td>
<td>18.922</td>
<td>1.00</td>
<td>33.36</td>
<td>2OMF 208</td>
</tr>
</tbody>
</table>
APPENDIX B
THE GROMACS SIMULATION TOOL
B.1 General Description

GROMACS is a software package used to perform molecular dynamics simulations and energy minimization [69]. GROMACS stands for Gröningen Machine for Chemical Simulation, from the name of the Dutch university where it is being developed. The software is freely distributed under the GNU General Public License (http://www.gnu.org).

While it was primarily designed for the simulation of biomolecules like proteins and lipids, GROMACS is also very efficient in modeling non-bonded interactions like the ones between ions in a solution. It includes specifically optimized algorithms to perform various types of simulations such as energy minimization, MD or even BD simulations. Force field schemes like P3M are also implemented. In GROMACS however, the implementation of the P3M solves for the mesh potential in the reciprocal space. The boundary conditions have to be periodic in this case, which is a major restriction for the study of ionic channels. Moreover, the GROMACS algorithm does not support inhomogeneous meshes and varying dielectric constant. However, it is possible to use GROMACS with homogeneous systems such as bulk solutions to compare its results to what has been obtained with our simulation code.
B.2 Usage

GROMACS includes a set of file formats and executables in order to prepare and run a simulation. Every definition file is written in text format making it easy to setup a simulation.

The first step is to obtain a starting structure, and tools are provided to build the structure data from PDB files. The executable `pdb2gmx` takes a PDB file as an input and generates a GROMACS `.gro` configuration file containing the atom coordinates, together with a `.top` topology file containing the information about the molecules and chemical bonds present in the system.

The next step is to setup a simulation by writing the simulation parameters in a `.mdp` definition file. this file is then checked and processed together with the configuration and topology files by the executable `grompp`, which is a preprocessing code. It checks for the consistency of the information provided in the files previously described, and produces a binary simulation definition file `.tpr`.

Finally, the simulation code `mdrun` can be started with the `.tpr` file as an input. It produces an output configuration file, a trajectory file `.trr`, and a log file. This last step is the longest one since it performs the actual simulation. The GROMACS package also contains post-processing modules for example the extraction of the RDF is done by the executable `g_rdf`. Other tools permit to build some simple systems such as a water box, and then insert ions to produce an electrolyte solution. The details on how to use each program provided in the GROMACS package are given in the GROMACS user manual [70]. The next section will give some GROMACS usage examples.
B.3 Some Examples

B.3.1 Bulk Electrolyte Solution. The first simulation performed with GROMACS was a simple bulk electrolyte solution simulation. The system consisted of a $10 \times 10 \times 10$ nm$^3$ box containing a 1.0 mol.L$^{-1}$ KCl solution. The simulation uses the GROMACS P$^3$M algorithm with periodic boundary conditions. The simulation temperature is set to 298 K and the ion parameters used were set to the Lennard-Jones values used in this work (see table 4.2). The sequence of operations to perform this simulation with GROMACS is described next.

1. Generate a simulation box containing water only. The executable genbox produces the configuration and topology files. The inputs are the size of the box and the water model to use, in this case spc216.gro, which is provided with the GROMACS package.

2. Produce a binary topology file .tpr for use as an input to the program genion which will add the ions to the simulation box. The program needs as an input the total numbers of ions of each type to be added to the system, expressed as:

\[ N_{K^+} \equiv N_{Cl^-} \equiv N_A c V_b, \]  

(B.1)

where $N_A$ is the Avogadro number, $c$ is the solution concentration and $V_b$ is the volume of the simulation box.

3. Now that the ions have been placed in the system, the water molecules can be removed by hand from the configuration and topology files, since the implicit water model will be used in the simulation. The topology file has also a #include statement to add the non-bonded interaction parameters from the provided ffgmx.itp. The procedure used to include specific Lennard-Jones parameters is to make a copy of this file, my_ffgmx.itp, in which the parameters for the interaction potential are modified for both $K^+$ and $Cl^-$. The units
for these values must be modified since GROMACS uses values expressed in kJ.mol\(^{-1}\).nm\(^6\) for the parameter \(c_6\) and kJ.mol\(^{-1}\).nm\(^{12}\) for the parameter \(c_{12}\), representing the coefficients of the \(r^{-6}\) and \(r^{-12}\) terms in the Lennard-Jones formula, respectively. The values \(\epsilon\) and \(\sigma\) provided in table 4.2 must therefore be converted and combined to produce the GROMACS \(c_6\) and \(c_{12}\) parameters.

4. The simulation definition file \textit{bd.mdp} can now be edited. The options which values must be set include the force field scheme used (P\(^3\)M here), the temperature, the friction coefficient, the timestep and the total simulation time. The GROMACS user manual provides examples of how to write the definition file and the list of all parameters and their descriptions.

5. The \textit{grompp} preprocessor is run in order to check the topology, the configuration, and the parameters, and to produce the binary simulation run file \textit{bd.tpr}.

6. The simulation itself is run with the program \textit{mdrun}.

7. The resulting trajectories \textit{traj.trr} are finally processed with \textit{g.rdf} to obtain the final RDF curves, which is the desired result here. Figure B.1 shows the RDF obtained with GROMACS.

\textbf{B.3.2 Force Map of the OmpF Channel.} The second type of simulations that have been performed with GROMACS are 3D force maps. GROMACS does not have such a function implemented in the distribution tools. A specific C code has been developed to take care of this process. For each data point, the test ion is positioned at the right coordinates and added to the structure and topology files by the implemented code. The required informations are written by the C code into the \textit{.gro} and \textit{.top} files, by appending a line for the test ion. The simulation definition file is similar to the one used in the simulation described in the previous section, except that it runs only one simulation step. The C code then runs each GROMACS
Figure B.1. GROMACS Simulation Results for a Bulk KCl 1.0 M Solution. The RDF Shows an Excellent Agreement With the HNC Benchmark.

This mapping script was run for a system containing the OmpF channel protein only, in order to make a comparison with the results obtained with the simulator proposed in this work. Figure B.2 shows a force map for a K\(^+\) test ion on the OmpF porin channel. In this simulation, the dielectric constant is set to 78.0 everywhere and the boundary conditions are periodic. Even though the simulation proposed in this
work uses inhomogeneous dielectric constants and non-periodic boundary conditions, the GROMACS results show a similar force profile in 3D for the OmpF channel.

Figure B.2. Map of the Force On a K\(^{+}\) Ion Along a Plane Perpendicular To the Pores of OmpF Obtained With the GROMACS P\(^3\)M Algorithm, Using a Custom Script.
APPENDIX C
HNC ALGORITHM IMPLEMENTATION
The Ornstein-Zernike (OZ) integral equation theory [63] provides an analytical technique to determine the RDF of a solution given an inter-particle potential. It was previously shown (see section 4.1.2) that the knowledge of the RDF makes it possible to obtain all the thermodynamical parameters of interest. It is therefore possible to obtain the thermodynamical and structural characteristics of an electrolyte solution analytically by solving the OZ integral equations.

C.1 The Ornstein-Zernike Equation

The Ornstein-Zernike (OZ) equation is a non-linear integral equation that expresses the RDF in terms of the pair-wise intermolecular potential [38]. The issue with electrolyte solutions is that they cannot be treated as a simple two-body problem with only pair-wise interactions. When considering a pair of two particles $i$ and $j$ of respective species $s$ and $s'$, their interaction not only comprises the direct, pair-wise interaction, but also an infinite series of indirect interactions through the neighboring particles. This infinite series can be expressed by the so-called cluster expansion of the interaction between pairs of particles. A detailed analysis [38] separates this expansion into a sum of convolution terms making up the total particle correlation function $h_{ss'}$ whereas all the other terms make up the direct correlation function $c_{ss'}$. The relationship between the total and direct correlation functions is the OZ equation [63]:

$$h_{ss'}(r_{ij}) = c_{ss'}(r_{ij}) + \sum_l \rho_l \int h_{sl}(r_{ik})c_{ls'}(r_{jk})dr_3$$  \hspace{1cm} (C.1)

It can be seen that this equation relates the two functions $c_{ss'}$ and $h_{ss'}$ through a convolution integral. The total correlation $h_{ss'}$ and the RDF $g_{ss'}$ are directly related by definition [38]:

$$h_{ss'}(r_{ij}) \equiv g_{ss'}(r_{ij}) - 1$$  \hspace{1cm} (C.2)

Since both correlation functions $h_{ss'}$ and $c_{ss'}$ have to be determined simultaneously, a closure relation is needed to solve equation C.1. This closure relation will relate $c_{ss'}$
to $h_{ss'}$ (or $g_{ss'}$), and the closed system of equations formed by the OZ equation and the closure relation is then solved for all the correlation functions.

Three types of closure relations have been proposed [38]: the Hypernetted Chain approximation (HNC) [71], the Percus-Yevick (PY) [72, 73] and the Mean Spherical Approximation (MSA) [74, 75]. In this work the HNC is adapted to the case of an electrolyte solution. For a detailed discussion of the other two closure relations the reader is referred to reference [38].

C.2 The HNC Approximation

The HNC approximation is a particular simplification of a general closure relation obtained from the graph theory that can be written as:

$$g_{ss'} = \exp[-\beta u_{ss'} + \gamma_{ss'} + B_{ss'}] \quad (C.3)$$

where $\beta = 1/k_B T$, $u_{ss'}$ is the pair potential, $\gamma_{ss'}$ is the sum of all the series graphs and $B_{ss'}$ represents the sum of all the bridge graphs [38, 71].

The HNC approximation consists of neglecting the bridge graphs and expressing the sum of the series graphs as:

$$\gamma_{ss'} = h_{ss'} - c_{ss'} \quad (C.4)$$

yielding:

$$g_{ss'} = \exp[-\beta u_{ss'} + h_{ss'} - c_{ss'}] \quad (C.5)$$

The combination of the OZ equation C.1 with equation C.2 and the closure relation equation C.5 forms the HNC integral equation to be solved. The reader is referred to work of A. Allnatt [71] for a detailed discussion of the graph theory applied to the derivation of the HNC equations.
C.3 Numerical Solution

The original algorithm proposed by Allnatt [71] has been developed by Ng [76] for the solution of HNC integral equations for a single-component plasma. This algorithm forms the basis of the HNC solution for multi-component coulombic systems in water. The non-linear, convolution-type equation is solved in the Fourier space with an iterative process that successively solves the OZ equation and then applies the HNC closure relation until convergence is reached for all the correlation functions. The algorithm, adapted to two-components electrolyte solutions by Belloni [77], is described as follows:

1. An initial guess for the direct correlation function is computed as follows:

\[ c_{ss}^0 = \exp[-\beta u_{ss}] - 1. \] \hspace{1cm} (C.6)

2. The direct correlation function \( c_{ss} \) is then Fourier transformed using the 3D Fourier transform proposed by F. Lado [78] for liquid state computations.

3. The OZ equation expressed in terms of the Fourier transforms of the sum of the series graphs \( \hat{\gamma}_{ss} \) and \( \hat{c}_{ss} \) is solved:

\[ \hat{\gamma}_{ss'} = \sum_{k=s,s'} \rho_k \hat{c}_{sk}(\hat{\gamma}_{ks'} + \hat{c}_{ks'}) \] \hspace{1cm} (C.7)

Note that for a two component system each function is actually a 2×2 matrix of 4 functions:

\[ \hat{C} = \begin{bmatrix} \hat{c}_{ss} & \hat{c}_{ss'} \\ \hat{c}_{s's} & \hat{c}_{s's'} \end{bmatrix} \] \hspace{1cm} (C.8)

The resulting OZ equation C.7 is therefore a matrix equation.

4. The sum of the series graph \( \gamma_{ss'} \) can be now obtained in real space through the inverse 3D Fourier transform of \( \hat{\gamma}_{ss'} \).
5. The HNC closure relation is applied by combining equation C.4 and C.5, which allows the computation of the pair-correlation function $g_{ss'}$. The new direct correlation function $c_{ss'}^{n+1}$ is then determined using equation C.2 and C.4:

$$c_{ss'}^{n+1} = g_{ss'} - 1 - \gamma_{ss'}.$$  \hfill (C.9)

6. The convergence criteria is now verified following Ng [76]:

$$\epsilon_{th} \geq \left[ \int_0^\infty |c_{ss'}^n - c_{ss'}^{n+1}|^2 dr \right]^{1/2}$$  \hfill (C.10)

The process is iterated until each of the four functions ($ss, ss', s's, s's'$) converges to an RMS threshold $\epsilon_{th}$, set to $10^{-14}$ as suggested in [76].

This algorithm as it is presented here does not converge for coulombic interaction potentials that decay as $1/r$, because the Fourier transform integral diverges. A way around this issue has been originally proposed by Ng and Belloni [76, 77] and is described in detail in [79]. The idea is to split the coulombic interaction potential into a short-range term $u_{ss'}^{SR}$ and a long-range term $u_{ss'}^{LR}$ using the error function:

$$u_{ss'}^{SR}(r) = \frac{q_sq_{s'}}{4\pi\epsilon_r\epsilon_0r} \left[ 1 - \text{erf}(\alpha r) \right]$$  \hfill (C.11)

$$u_{ss'}^{LR}(r) = \frac{q_sq_{s'}}{4\pi\epsilon_r\epsilon_0r} \left[ \text{erf}(\alpha r) \right]$$  \hfill (C.12)

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-u^2) du$$  \hfill (C.13)

where $\alpha$ is a splitting parameter determining the spatial extension of the short-range part of the coulombic interaction. The total interaction potential is thus written as a sum of three components:

$$u_{ss'} = u_{ss'}^* + u_{ss'}^{SR} + u_{ss'}^{LR}$$  \hfill (C.14)

where $u_{ss'}^*$ represents non-coulombic interactions such as the short-range Lennard-Jones potential. These three parts are inserted into the equations following the
method described in [79] to make the mathematical computation of the Fourier transforms possible.

C.4 HNC Computation Example

In this work, a HNC analytical solver has been implemented to provide comparison and validation of particle-based simulations of electrolyte solutions. The algorithm described in the previous section is implemented by discretizing each function on a uniform grid in real space with a spacing of 0.1 Å. The functions are computed for values of the ionic distance of up to 2 nm and the Fourier transforms are represented on 1024 points. Computations have been successfully obtained for KCl and NaCl solutions with concentrations ranging between 0 and 2.25 mol.L\(^{-1}\), which is well above typical values of biological systems, and beyond the limit of validity for the primitive model. Figure C.1 shows the RDFs obtained for a 0.15 mol.L\(^{-1}\) KCl solution using Lennard-Jones empirical interaction potentials. The HNC results obtained in this work show a perfect agreement with published results [28].

![Figure C.1](image)

Figure C.1. RDFs for a 0.15 mol.L\(^{-1}\) KCl Solution Obtained from the HNC Integral Computation. The Interaction Potential Uses the Empirical Lennard-Jones Formula and a Value of \(5 \times 10^8\) Is Used for the Splitting Parameter \(\alpha\).
APPENDIX D

ELECTRICAL DOUBLE LAYER CALCULATIONS
D.1 The Classical Theory of the Electrical Double Layer

Within the classical theory of the Double Layer (DL) developed by Gouy and Chapman [47, 48], the ions are treated as point charges set in an homogeneous one-dimensional dielectric medium where only Coulombic interactions are modeled. They are distributed according to the Boltzmann distribution, and the system formed by the electrolyte solution in contact with the membrane is described by the Poisson-Boltzmann equation (PBE), which is derived (see section 4.2.1):

\[
\frac{d^2\psi(z)}{dz^2} = -\frac{n_0 e}{\varepsilon_0 \varepsilon_r} \sum_{k=i,j} z_i \exp \left[ -\frac{z_i e \psi(z)}{k_B T} \right],
\]

where \(\psi(z)\) is the electrostatic potential. This non-linear differential equation can be integrated analytically [39] for the potential and the ion distribution along \(z\).

D.1.1 Analytical Derivations for the Classical DL. In this section, the analytical solution of the classical DL PBE is shown, with some examples of DL charge distributions and potential profiles. The case of a simple electrolyte (where the ions bear unitary charges such as in KCl or NaCl) is considered here. In this case, \(z_i = -z_j = |Z|\). Equation D.1 then becomes:

\[
\frac{d^2\psi(z)}{dz^2} = -\frac{n_0 |Z| e}{\varepsilon_0 \varepsilon_r} \left[ e^{-|Z| e \psi(z)/k_B T} - e^{-|Z| e \psi(z)/k_B T} \right],
\]

which can be rewritten using the definition of the hyperbolic sine:

\[
\frac{d^2\psi(z)}{dz^2} = -\frac{2n_0 |Z| e}{\varepsilon_0 \varepsilon_r} \sinh \left( \frac{|Z| e \psi(z)}{k_B T} \right).
\]

Now, since the derivative of the hyperbolic cosine can be expressed as:

\[
\frac{d}{dz} \left[ \cosh \left( \frac{|Z| e \psi(z)}{k_B T} \right) + C_1 \right] = \sinh \left( \frac{|Z| e \psi(z)}{k_B T} \right) \frac{|Z| e}{k_B T} \frac{d \psi(z)}{dz},
\]

and noting that:

\[
2 \frac{d \psi(z)}{dz} \frac{d^2 \psi(z)}{dz^2} = \frac{d}{dz} \left( \frac{d \psi(z)}{dz} \right)^2,
\]
equation D.3 can be multiplied on both sides by $2(d\psi(z)/dz)$ to give:

$$2 \frac{d\psi(z)}{dz} \frac{d^2\psi(z)}{dz^2} = \frac{4n_0 k_B T}{\epsilon_0 \epsilon_r} \sinh \left( \frac{|Z| e \psi(z)}{k_B T} \right) \left| \frac{Z| e}{k_B T} \frac{d\psi(z)}{dz} \right|.$$  \hspace{1cm} (D.6)

By substituting equation D.4 and D.5 and integrating once, the following relation is obtained:

$$\left( \frac{d\psi(z)}{dz} \right)^2 = \frac{4n_0 k_B T}{\epsilon_0 \epsilon_r} \left[ \cosh \left( \frac{|Z| e \psi(z)}{k_B T} \right) + C_1 \right].$$  \hspace{1cm} (D.7)

The integration constant $C_1$ can be replaced by $-1$ under the assumption that $\phi(z)$ vanishes monotonically as $z \to \infty$, that is $\psi(z) \to 0$ and $d\psi(z)/dz \to 0$ as $z \to \infty$. This assumption is based on the fact that the electrolyte solution becomes homogeneous, and therefore electrically neutral, far from the interface. Noting that:

$$\sinh^2 \frac{x}{2} = \frac{\cosh x - 1}{2},$$  \hspace{1cm} (D.8)

equation D.7 then becomes:

$$\frac{d\psi(z)}{dz} = \pm \left( \frac{8n_0 k_B T}{\epsilon_0 \epsilon_r} \right)^{1/2} \sinh \left( \frac{|Z| e \psi(z)}{2k_B T} \right).$$  \hspace{1cm} (D.9)

Since [80]:

$$\int \frac{dx}{\sinh x} = \ln \left[ \tanh \frac{x}{2} \right] + C_2,$$  \hspace{1cm} (D.10)

equation D.9 is rewritten to show the required integrand:

$$\frac{|Z| e \psi(z)}{2k_B T} = \pm \left( \frac{2|Z|^2 e^2}{\epsilon_0 \epsilon_r} \right)^{1/2},$$  \hspace{1cm} (D.11)

and finally integrated to yield:

$$\ln \left[ \tanh \left( \frac{|Z| e \psi(z)}{4k_B T} \right) \right] = \pm \left( \frac{2|Z|^2 e^2}{\epsilon_0 \epsilon_r} \right)^{1/2} z + C_2,$$  \hspace{1cm} (D.12)

and

$$\tanh \left( \frac{|Z| e \psi(z)}{4k_B T} \right) = \exp \left[ \pm \left( \frac{2|Z|^2 e^2}{\epsilon_0 \epsilon_r} \right)^{1/2} z + C_2 \right].$$  \hspace{1cm} (D.13)
The only possible choice for the sign in the exponential on the right hand side of the equation is negative, since $\psi(z) \to 0$ as $z \to \infty$. If the potential at the surface is defined as $\Delta \psi \equiv \psi(z = 0)$, the integration constant can be determined as follows:

$$\exp[C_2] = \tanh \left( \frac{|Z|e\Delta \psi}{4k_BT} \right). \quad (D.14)$$

Defining the Debye-Hückel screening factor $[39]$ for a symmetric monovalent electrolyte as:

$$\kappa \equiv \sqrt{\frac{2|Z|^2e^2n_0}{\varepsilon_0\varepsilon_rk_BT}}. \quad (D.15)$$

and substituting the integration constant, equation D.13 becomes:

$$\tanh \left( \frac{|Z|e\psi(z)}{4k_BT} \right) = \exp[-\kappa z] \tanh \left( \frac{|Z|e\Delta \psi}{4k_BT} \right). \quad (D.16)$$

Equation D.16 relates the electrostatic potential to the potential at the electrolyte-membrane interface and the ionic density $n_0$. By applying the definition of the inverse hyperbolic tangent:

$$\tanh^{-1}(x) = \frac{1}{2} \ln \frac{1+x}{1-x}, \quad (D.17)$$

equation D.16 yields the explicit mean electrostatic potential:

$$\psi(z) = \frac{2k_BT}{|Z|e} \ln \frac{1 + \tanh(|Z|e\Delta \psi/4k_BT) \exp(-\kappa z)}{1 - \tanh(|Z|e\Delta \psi/4k_BT) \exp(-\kappa z)}. \quad (D.18)$$

Finally, the ionic density profiles can be expressed by replacing the expression of the mean electrostatic potential into the Boltzmann distribution expression (equation 4.34):

$$n_i(z) = n_0 \left( 1 + Z_i \tanh(|Z|e\Delta \phi/4k_BT) \exp(-\kappa z) \right)^2. \quad (D.19)$$
D.1.2 Results for the Classical DL. In this section, some results obtained with the standard DL theory are presented and a discussion is offered about their applicability to the real system and their use as a benchmark for BD simulations.

Figure D.1. Ionic Concentration Profiles Computed With the Standard DL Theory for a 1—1 Electrolyte Bulk Concentration of 0.5 M and a Surface Potential of 30 mV.

Figure D.1 shows a typical set of ionic concentration profiles created by the surface in contact with a simple 1—1 electrolyte (such as KCl). The surface potential $\Delta \psi$ is maintained to a fixed value of 30 mV in this example, while the bulk concentration is 0.5 M. It can be seen that the effects of the surface vanish at a distance not greater than 1 nm due to the strong screening of the surface potential by the ion accumulations near the surface. The charge separation (electrical DL) is clearly visible, and the associated potential drop (shown in figure D.2) proves the limited spatial extent of the surface effects. Although spatially limited, these effects give rise to local charge concentrations that are much higher than the bulk concentration, leading to high fields in the interfacial region. This charge accumulation has a crucial effect in the determination of charge transport properties of membranes [49] which is the main objective of this work.
In order to obtain quantities easier to compare than the ionic densities and potential curves, the following definitions are proposed. The DL surface charge is defined as the total charge accumulated in the double layer, within the DL thickness $r_{DL}$ of the surface:

$$\sigma_{DL} \equiv \int_{z=0}^{z=r_{DL}} \sum_{k=i,j} z_k e n_k(z) dz.$$  \hfill (D.20)

Similarly, the DL potential drop is defined as the potential difference between the surface and the DL outer limit, also defined by $r_{DL}$:

$$V_{DL} \equiv \Delta \psi - \psi(r_{DL}).$$  \hfill (D.21)

Figure D.3 shows the evolution of the DL surface charge $\sigma_{DL}$ as a function of the bulk electrolyte concentration for various values of the surface potential $\Delta \psi$. Figure D.4 shows the DL voltage drop $V_{DL}$ as a function of the concentration for the same values of the surface potential. The DL thickness $r_{DL}$ was chosen as 0.5 nm here, which corresponds to the approximate ionic diameters. It can be seen that the evolution of both quantities is almost linear for concentration values above 0.5 M. This really
means that the shape of the DL is not significantly different for concentrations beyond 0.5 M, and that most of the potential drop has occurred within $r_{DL}$ of the surface. It appears that the DL is increasingly diffuse when the concentration is reduced to 0.5 M and below. In those cases the potential drop is less significant within the DL, meaning that, as expected, the effect of the surface extends farther out in a low concentration solution. It can also be noted that the charge accumulations are negative, since the surface potential is positive and therefore anions are concentrated toward the surface, as can be seen in figure D.1.
Figure D.3. Charge Accumulation In the DL As a Function of the Bulk Electrolyte Concentration for Values of the Surface Potential Ranging Between 5 and 150 mV.

Figure D.4. Potential Drop Across the DL As a Function of the Bulk Electrolyte Concentration for Values of the Surface Potential Ranging Between 5 and 150 mV.
D.2 The URMGC Model for the Electrical Double Layer

Although the standard Gouy-Chapman theory of the electrical DL provides some keys to the understanding of the physical properties of an electrolyte solution in contact with a membrane, some of its limitations are also shown clearly by the results. Within this model, the ion–ion interactions are purely coulombic, as is their interaction with the surface. There is no finite-size effect included in this model, allowing for unphysical ionic concentrations and potentials very close to the surface. In order to provide a better model for the electrolyte-surface physics, the standard Gouy-Chapman DL theory was refined by Stern [50] who proposed to include some finite-size effects by attributing to the ions a single closest approach radius. This means that the ions interact with the surface as uniformly charged hard spheres. A depleted zone is therefore formed between the DL and the surface in this case. This model was further refined by Valleau and Torrie [49, 51, 52] by noting that in most physical systems the ions and counterions have different sizes. The resulting model is called the Unequal Radius Modified Gouy-Chapman (URMGC) model, and assigns different radii \( r_k \) to each ionic species present in solution. The analytical calculation of the DL properties with the URMGC approach is presented in the following section.

D.2.1 Analytical Calculations for the URMGC Model. The only ion–ion interaction accounted for by the URMGC model is the Coulombic interaction. The ions are still considered as point charges within the electrolyte while being assigned closest approach radii \( r_k \) in their interaction with the surface. This approach generates three different regions within the system as sketched in figure D.5. Let us assume that the negative charges will have a larger radius \( r_j \) than the positive ones, as is the case in KCl. In this case, the regions are defined as follows: region A for \( z < r_i \), region B for \( r_i < z < r_j \) and region C for \( z > r_j \). In a fashion similar to the standard DL theory, it is assumed that the ions are distributed according to Boltzmann statistics
Figure D.5. Illustration of the Three Regions Used for the Solution of the URMGC Equations.

(equation 4.34). Furthermore, The surface at \( z = 0 \) is assumed to bear a surface charge of density \( \sigma \). The solution of the URMGC model \([52]\) therefore consists in solving the PBE (equation D.1) within three integration domains connected by the usual electrostatic continuity relations at the boundaries \( z = r_i \) and \( z = r_j \).

The notations used in this section are the same as in the classical DL solution (section D.1): \( n_0 \) is the bulk ion number density and the electrolyte is symmetric with the relation between the charge numbers being \( z_i = -z_j = |Z| \). Following \([52]\), the following definitions are proposed to simplify the notation:

\[
\begin{align*}
  a &= \frac{8\pi n_0|Z|e}{\varepsilon_0\varepsilon_r}, \\
  b &= \frac{|Z|e}{k_B T}.
\end{align*}
\] (D.22)

The PBE can then be expressed as follows, in each of the regions (see figure D.5):

region A: \( \frac{d^2\psi(z)}{dz^2} = 0 \) (D.23)
region B: \( \frac{d^2\psi(z)}{dz^2} = \frac{a}{2} \exp\left(b\psi(z)\right) \) (D.24)
region C: \( \frac{d^2\psi(z)}{dz^2} = a \sinh\left(b\psi(z)\right). \) (D.25)
Let us consider region C first, and recognize that the differential equation is of the same form as in the standard DL theory. Accordingly, it can be integrated using the condition that the potential vanishes monotonically as \( z \to \infty \) to give the general solution:

\[
\exp \left(-\frac{b|\psi(z)|}{2}\right) = \tanh \left(\frac{\sqrt{ab}(z - \alpha)}{2}\right)
\]  

(D.26)

where \( \alpha \) is an integration constant to be determined later using the joining conditions at \( z = r_j \). Now, equation D.24 for region B can be integrated once to yield [52]:

\[
\frac{d\psi(z)}{dz} = \pm \sqrt{\frac{a}{b}} \exp \left(b\psi(z)\right) - \beta_1^2,
\]  

(D.27)

where \( \beta_1 \) is an integration constant. This constant can be imaginary, in which case a subsequent integration of the differential equation for region B will yield a different expression from the one obtained here [52]. In the case of \( \beta_1 \) real (as is always the case in the configurations considered here), the second integration yields:

\[
\exp \left(-\frac{b\psi(z)}{2}\right) = \frac{1}{\beta_1} \cos \left(\frac{\beta_1}{2} \sqrt{ab}(z - \beta_2)\right),
\]  

(D.28)

where \( \beta_2 \) is the second integration constant. Finally, equation D.23 is immediately integrated twice to yield:

\[
\psi(z) = \gamma_1 + \gamma_2 z,
\]  

(D.29)

where \( \gamma_1 \) and \( \gamma_2 \) are the integration constants. After some algebraic manipulations, the expression of the mean electrostatic potential can be given in each region, as a function of the integration constants \( \alpha, \beta_1, \beta_2, \gamma_1 \) and \( \gamma_2 \):

region A: \( \psi(z) = \gamma_1 + \gamma_2 z \)  

(D.30)

region B: \( \psi(z) = -\frac{2}{b} \log \left[ \frac{1}{\beta_1} \cos \left(\frac{\beta_1}{2} \sqrt{ab}(z - \beta_2)\right)\right] \)  

(D.31)

region C: \( \psi(z) = \mp \frac{2}{b} \log \left[ \tanh \left(\frac{\sqrt{ab}}{2}(z - \alpha)\right)\right]. \)  

(D.32)

The sign in equation D.32 is determined by the sign of the surface charge \( \sigma \) at \( z = 0 \) (or by the surface potential). The set of integration constants can be found using
the continuity of the potential and its derivative at \( z = r_i \) and \( z = r_j \), and the forcing condition at \( z = 0 \) (surface charge or potential). For practical purposes, the integration constants and the surface charge have been expressed as functions of the electrostatic potential \( \psi_{r_i} \equiv \psi(z = r_i) \):

\[
\alpha = r_i - \frac{2}{\sqrt{ab}} \tanh^{-1} \left[ \exp \left( -\frac{b|\psi_{r_i}|}{2} \right) \right] \quad (D.33)
\]

\[
\beta_1 = \sqrt{2 - \exp(-b\psi_{r_i})} \quad (D.34)
\]

\[
\beta_2 = r_i + \frac{2}{\beta_1 \sqrt{ab}} \tanh \left[ \frac{2}{\beta_1} \sinh \left( \frac{b\psi_r}{2} \right) \right] \quad (D.35)
\]

\[
\gamma_2 = \beta_1 \sqrt{\frac{a}{b} \tan \left( \frac{\beta_1}{2} (r_j - \beta_2) \sqrt{ab} \right)} \quad (D.36)
\]

\[
\gamma_1 = -\frac{2}{b} \log \left[ \frac{1}{\beta_1} \cos \left( \frac{\beta_1}{2} (r_j - \beta_2) \sqrt{ab} \right) \right] - r_j \beta_1 \sqrt{\frac{a}{b} \tan \left( \frac{\beta_1}{2} (r_j - \beta_2) \right)}. \quad (D.37)
\]

To simplify the resolution of this set of equations, the surface charge is itself expressed as a function of \( \gamma_2 \), itself dependent on \( \psi_{r_i} \):

\[
\sigma(\psi_{r_i}) = -\frac{\epsilon_0 e\gamma_2}{a\pi}. \quad (D.38)
\]

A numerical value for \( \psi_{r_i} \) is found by solving \( \sigma(\psi_{r_i}) = \sigma \), and used in equations D.33 to D.37 to obtain the values of all the integration constants. The mean electrostatic potential is thus completely determined from the knowledge of the surface charge at \( z = 0 \). Finally, the ionic density curves can be obtained by applying the Boltzmann distribution to the mean electrostatic potential:

\[
n_k(z) = \begin{cases} 
0 & \text{if } z < r_k, \\
n_0 \exp \left( -\frac{z_k e \psi(z)}{k_B T} \right) & \text{otherwise}
\end{cases} \quad (D.39)
\]
D.2.2 Results for the URMGC DL model. This section presents some results of the URMGC computations and discusses their applicability as a benchmark for the BD simulation.

Figure D.6. Ionic Concentration Profiles Computed With the URMGC Model for a 1–1 Electrolyte Bulk Concentration of 0.5 M and No Surface Charge.

Figure D.6 shows a typical set of concentration profiles computed with the URMGC model. The ion radii used in the model are the same as the Lennard-Jones parameters for KCl (see table 4.2). The bulk concentration used in these computations is 0.5 M and the surface charge at the interface is set to zero. As in the standard DL theory, the effects of the surface rapidly vanish as the distance increases, and are negligible after only 1 nm. Furthermore, the difference between the radii of the \( \text{K}^+ \) and \( \text{Cl}^- \) ions (\( \text{K}^+ \) being smaller) induces a positive charge accumulation in the DL which results in a negative potential at the surface as can be seen in the mean electrostatic potential profile (figure D.7). The electrostatic potential shows that the asymptotic value of zero is essentially attained in about 1 nm, while the potential increases sharply in the DL, especially in the region B (see figure D.5) where only potassium ions are present.
Figure D.7. Mean Electrostatic Potential Computed With the URMGC Model for a 1–1 Electrolyte Bulk Concentration of 0.5 M and No Surface Charge.

As was done in the case of the standard DL theory, integral quantities are computed to help compare the results with the BD simulations. The definition of the total accumulated surface charge is the same as in the previous section (see equation D.20) while the DL potential drop is this time defined as the electrostatic potential change from the surface to a point infinitely far from the interface: \( V_{DL} \equiv \psi(z = \infty) - \psi(0) \), which reduces to \( V_{DL} = -\psi(0) \). It is important to note that in the URMGC model, the potential at the surface is not a parameter, but it is rather derived by the application of a surface charge constraint. In order to compare the results to the model used for the cell membrane in the BD simulations, the surface charge is set to zero in all the URMGC computations. Figures D.8 and D.9 show the evolutions of the above-defined total accumulated surface charge and potential drop, respectively. The results show moderate increases in both quantities, which values are significantly lower than the results of the standard DL theory computations. The main reason is that the ions are not allowed to pack as densely against the surface due to the excluded volume attributed to the ions. The URMGC model was used as
a benchmark to validate the BD simulation results which were found (see chapter 4) to be in good agreement, both quantitative and qualitative.

Figure D.8. Charge Accumulation In the DL As a Function of the Bulk Electrolyte Concentration With No Surface Charge Present.

Figure D.9. Potential Drop Across the DL As a Function of the Bulk Electrolyte Concentration With No Surface Charge Present.
BIBLIOGRAPHY


[47] Gouy, G. “Sur la Constitution de la Charge Electrique à la Surface d’un Elec-


[68] Cowan, S. “The Refined Structure of OmpF Porin from E. Coli at 2.4 Angstroms Resolution.” *To be Published*.


