



National Technical University of Athens
School of Mechanical Engineering
Fluids Department
Parallel CFD & Optimization Unit

**Modelling using Molecular Dynamics and Optimization of
Proton Conductivity in (Polymerized) Ionic Liquids for
Proton Exchange Membrane Fuel Cells.**

Diploma Thesis

Matthaios N. Chatzopoulos

Academic Supervisor :
Kyriakos C. Giannakoglou, Professor NTUA

Industrial Supervisor :
Dr. Konstantinos Gkagkas, Technology Manager Toyota Motor Europe

Athens, February 2022

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my professor Kyriakos C. Giannakoglou for giving me the chance to conduct my diploma thesis on this industrial subject in collaboration with Toyota Motor Europe. His advice, guidance and corrections were really valuable lessons for me. Throughout my studies in NTUA, he was an inspiring figure as an engineer and teacher. I am really grateful for the opportunity that I was given to develop my engineering insight under his academic supervision on this very interesting and novel topic.

Secondly, I would like to wholeheartedly thank my industrial supervisor Dr. K. Gkagkas for placing his trust in me by assigning me this challenging topic. His insightful guidance and his unique problem solving style have been indisputably of great importance. He and Dr. Marco Di Genaro created a very warm working environment full of enthusiasm about science. I am really thankful for that and for assisting me by any means possible. I want to thank all the TME engineers for all the meaningful dialogues they provided during my internship.

Thirdly, I would like to deeply thank all of the members of the PCOpt/NTUA research team and especially Dr. Varvara Asouti for her support before and during my internship at TME. Also, I am thankful to Petros Lamprinidis for assisting me and teaching me a lot of useful skills before starting my internship.

I would also like to express how grateful I am for having the chance to study and learn inside the NTUA community. The interaction with the teaching staff and fellow students combined with hard work really shaped me as an engineer. I want to deeply thank all of my friends for supporting me throughout the academic years, but more specifically my fellow students Michalis Tsagkaris and Dimitris Tolis for accompanying me along this difficult journey full of unforgettable experiences.

Last but not least, I would like to thank my family for always believing in me and supporting me with their love.

In loving memory of my godmother and my grandfather, my inspiring figures for studying and working hard.



National Technical University of Athens
School of Mechanical Engineering
Fluids Department
Parallel CFD & Optimization Unit

Modelling using Molecular Dynamics and Optimization of Proton Conductivity in (Polymerized) Ionic Liquids for Proton Exchange Membrane Fuel Cells.

Diploma Thesis

Matthaios N. Chatzopoulos

Academic Supervisor :
Kyriakos C. Giannakoglou, Professor NTUA

Industrial Supervisor :
Dr. Konstantinos Gkagkas, Technology Manager Toyota Motor Europe

Athens, February 2022

Abstract

Working towards carbon neutrality, the automotive industry have been focused on the development of highly efficient proton exchange membrane fuel cells (PEMFCs). One of the most important components of a PEMFC is the proton exchange membrane (PEM) itself, which is typically made from nafionTM. Regardless of the membrane material used, it is desirable for the fuel cell to operate at temperatures even above 120 °C, because the system becomes more efficient at higher temperatures. However, this cannot be achieved when nafionTM is used as the PEM material, because it requires high relative humidity and, thus, temperatures below 100 °C to operate properly.

The replacement of nafionTM membranes by a new class of PEMs produced through polymerization of the anion or the cation of an ionic liquid (IL) could be the remedy to the previous problem. ILs are organic salts of which the melting point is lower than 100 °C. However, though current membranes of this type are able to operate efficiently at very high temperatures, they unfortunately have about two orders of magnitude lower proton conductivity than nafionTM membranes. For this

reason, membranes that have been produced by polymerizing ILs are yet unable to replace nafionTM PEMs. Therefore, they firstly need to be optimized by using the conductivity as objective function.

In this Diploma Thesis, alternative pairs of cations and anions (ILs) with higher proton conductivity are searched by using an evolutionary algorithm, for solving the previous problem. Throughout this process, it is assumed that the conductivity of an IL is proportional to the conductivity of the respective polymerized IL membrane. This realistic assumption is made to avoid the prohibitive computational cost of simulating membranes. The calculation of conductivity is done through a Molecular Dynamics (MD) software. All the simulations were performed by the open source software GROMACS. The previous software is further supported by codes and linking scripts which are responsible for evaluating each IL and for the creation of the topology and geometry for each ion based on the optimization parameters. The optimization parameters for an ion consist of the type of different chemical elements which are used (e.g. C, N, S, O), the type and position of single, double, triple bonds, branches and the position from which a proton is removed or added for the final chemical substance to be an ion and not a neutral molecule.

The previous process is applied by using both computationally expensive all atom models (AA) and approximate, but inexpensive, coarse-grained (CG) models. These models are developed and their results are validated by comparing with available experimental data, before they are used in the optimization process. The optimization was performed by using the evolutionary algorithm based software EASY of the PCOpt/NTUA.

Major part of this diploma thesis was carried out at the research premises of Toyota Motor Europe in Brussels, Belgium, during a six month long internship there.



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Μηχανολόγων Μηχανικών
Τομέας Ρευστών
Μονάδα Παράλληλης Υπολογιστικής Ρευστοδυναμικής
& Βελτιστοποίησης

Μοντελοποίηση με χρήση Μοριακής Δυναμικής και
Βελτιστοποίηση της Πρωτονιακής Αγωγιμότητας σε
(Πολυμερισμένα) Ιονικά Υγρά για Κυψέλες Καυσίμου
Μεμβράνης Ανταλλαγής Πρωτονίων.

Διπλωματική Εργασία

Ματθαίος Ν. Χατζόπουλος

Επιβλέπων :

Κυριάκος Χ. Γιαννάκογλου, Καθηγητής ΕΜΠ

Βιομηχανικός Επιβλέπων :

Δρ. Κωνσταντίνος Γκαγκάς, Toyota Motor Europe

Αθήνα, Φεβρουάριος 2022

Περίληψη

Στην προσπάθειά της να μειώσει τις εκπομπές διοξειδίου του άνθρακα, η αυτοκινητοβιομηχανία αναπτύσσει κυψέλες καυσίμου με την τεχνολογία μεμβράνης ανταλλαγής πρωτονίων. Η μεμβράνη είναι ένα από τα βασικότερα εξαρτήματα σε μία κυψέλη καυσίμου και το υλικό το οποίο έχει επικρατήσει να χρησιμοποιείται στις μεμβράνες είναι το nafionTM. Ανεξαρτήτως του υλικού της μεμβράνης, μία κυψέλη καυσίμου είναι επιθυμητό να λειτουργεί σε υψηλές θερμοκρασίες, άνω των 120 °C, εξαιτίας του αυξημένου βαθμού απόδοσης της μηχανής σε αυτές. Ωστόσο, αυτό είναι αδύνατον να πραγματοποιηθεί με χρήση του nafionTM, καθώς απαιτεί συνθήκες υψηλής υγρασίας για να λειτουργήσει και, επομένως, θερμοκρασίες μικρότερες από 100 °C.

Για να λυθεί το προηγούμενο πρόβλημα, εξετάζεται η αντικατάσταση της μεμβράνης από nafionTM με μία άλλη μεμβράνη που θα προκύψει πολυμερίζοντας το ανιόν ή το κατιόν σε ένα ιονικό υγρό. Τα ιονικά υγρά είναι οργανικά άλατα με σημείο τήξης μικρότερο των 100 °C. Αν και οι τρέχουσες μεμβράνες αυτού του τύπου μπορούν

να λειτουργούν αποτελεσματικά σε πολύ υψηλές θερμοκρασίες, έχουν περίπου δύο τάξεις μεγέθους μικρότερη πρωτονιακή αγωγιμότητα σε σχέση με τις μεμβράνες από nafionTM. Για τον λόγο αυτό, οι μεμβράνες που προκύπτουν από πολυμερισμό ιονικών υγρών δεν είναι ακόμη σε θέση να αντικαταστήσουν τις μεμβράνες από nafionTM και απαιτείται βελτίωση/βελτιστοποίησή τους με την αγωγιμότητα ως συνάρτηση-στόχο.

Στη διπλωματική αυτή εργασία αναζητούνται εναλλακτικά ζεύγη κατιόντων και ανιόντων (ιονικά υγρά) με υψηλότερη πρωτονιακή αγωγιμότητα, εφαρμόζοντας στοχαστικές μεθόδους βελτιστοποίησης, ώστε να λυθεί το προηγούμενο πρόβλημα. Κατά τη διαδικασία αυτή θεωρείται πως η αγωγιμότητα ενός ιονικού υγρού και της αντίστοιχης μεμβράνης που παράγεται από αυτό είναι μεγέθη ανάλογα. Η ρεαλιστική αυτή παραδοχή γίνεται λόγω του απαγορευτικά υψηλού υπολογιστικού κόστους που έχει η προσομοίωση μεμβρανών. Ο υπολογισμός της αγωγιμότητας πραγματοποιείται μέσω προσομοίωσης μοριακής δυναμικής (Molecular Dynamics ή MD) και με χρήση του λογισμικού ανοικτού κώδικα GROMACS. Το λογισμικό αυτό συνοδεύει ένα πλήθος λογισμικών, καθώς και συνδετικών κωδίκων, για να είναι εφικτή η αξιολόγηση του κάθε ιονικού υγρού, αλλά και η παραγωγή γεωμετρίας και τοπολογίας ιόντων από τις παραμέτρους της βελτιστοποίησης. Οι παράμετροι της βελτιστοποίησης για ένα ιόν είναι το είδος των διαφορετικών χημικών στοιχείων που θα χρησιμοποιηθούν (π.χ. C, N, S, O), το είδος και η θέση των μονών, διπλών, τριπλών δεσμών, των διακλαδώσεων καθώς και η θέση αφαίρεσης ή προσθήκης ενός πρωτονίου, ώστε αυτό που θα προκύψει να είναι ιόν και όχι ουδέτερο μόριο.

Η προηγούμενη διαδικασία εφαρμόζεται τόσο με χρήση ακριβέστερων μοντέλων που συμπεριλαμβάνουν όλα τα άτομα (All atom ή AA models), όσο και με χρήση πιο προσεγγιστικών, αλλά υπολογιστικά φθηνότερων, αδρομερών μοντέλων (Coarse-grained ή CG models). Τα μοντέλα αυτά αναπτύσσονται και τα αποτελέσματά τους επαληθεύονται με χρήση διαθέσιμων πειραματικών δεδομένων προτού χρησιμοποιηθούν στη διαδικασία βελτιστοποίησης. Η βελτιστοποίηση πραγματοποιήθηκε με χρήση του λογισμικού εξελικτικών αλγορίθμων EASY της ΜΠΥΡΒ του ΕΜΠ.

Το μεγαλύτερο μέρος της διπλωματικής εργασίας πραγματοποιήθηκε στις ερευνητικές εγκαταστάσεις της Toyota Motor Europe στις Βρυξέλλες του Βελγίου κατά τη διάρκεια εξαμήνης πρακτικής άσκησης.

Acronyms

AA	All Atoms
AMBER	Assisted Model Building with Energy Refinement
CG	Coarse Grained
CL	Catalyst Layer
CL& P	Canongia Lopez & Padua force field
DFT	electronic structure Density Functional Theory
EA	Evolutionary Algorithm
EASY	Evolutionary Algorithms SYstem
FC	Fuel Cell
FF	Force Field
GDL	Gas Diffusion Layer
GFCs	Gas Flow Channels
GROMOS	GRoningen MOlecular Simulation
ILs	Ionic Liquids
ILM	Ionic Liquid Monomer
MARTINI	MARrink's Toolkit INItiative (force field)
MD	Molecular Dynamics
MOO	Multi-Objective Optimization

MSD	Mean Square Displacement
NPAAFF	Non-Polarizable All Atom Force Field
NTUA	National Technical University of Athens
OPLS	Optimized Potentials for Liquid Simulations
PAAFF	Polarizable All Atom Force Field
PCOpt	Parallel CFD & Optimization unit
PDB	Protein Databank File
PEM	Proton Exchange Membrane
PEMFC	Proton Exchange Membrane Fuel Cell
PILs	Polymerized ionic liquids
Poly-ILs	Polymerized Ionic Liquids
PILM	Polymerized Ionic Liquid Membrane
PIM	Polarizable Ion Model
SAPT	Symmetry Adapted Perturbation Theory
SMILES	Simplified Molecular-Input Line-Entry System
SOO	Single-Objective Optimization
TME	Toyota Motor Europe
VACF	Velocity Auto-Correlation Function

Contents

Contents	i
1 Introduction	1
1.1 Introduction to Molecular Dynamics	1
1.2 Introduction to PEM fuel cells	4
1.3 The PEM of a fuel cell	4
1.4 Polymerized ionic liquid membranes	5
1.5 Literature survey on the force fields in MD	7
1.6 Force fields in classical MD simulations	9
1.7 Non-polarizable all atom force fields (NPAAFF)	10
1.7.1 The typical structure of a NPAAFF	10
1.7.2 AMBER force field	14
1.7.3 OPLS All Atom (OPLS-AA) force field	14
1.7.4 CL&P force field	15
1.8 Polarizable all atom force field (PAAFF)	15
1.8.1 Advantages of PAAFF over NPAAFF	15
1.8.2 The Polarized Ion Model (PIM)	16
1.8.3 Other Polarizable all atom force fields	17
1.8.4 Conclusions on Polarizable all atom force fields	18
1.9 Charge scaling methods	18
1.10 Coarse-grained models	19
1.10.1 Basic principles of coarse grained models	19
1.10.2 Brief introduction to the MARTINI 2.0 force field	20
1.11 Scope and structure of the thesis	22
2 AA simulations of ionic liquids	25
2.1 The structure of a typical AA MD simulation	25
2.1.1 Input files and prerequisites of the simulation	25
2.1.2 The main section of the simulation	27
2.1.3 Post-processing and Convergence	28

2.2	Convergence studies for the simulation parameters	30
2.2.1	Information about the simulated ionic liquid pairs	30
2.2.2	Determining the proper size for the simulation box	32
2.2.3	Determining the total number of molecules in the system	32
2.2.4	Determining the cut-off distances for the non-bonded interactions	34
2.2.5	Determining the proper frequency and time duration for the MSD method	36
2.3	Results and discussion about the AA MD simulations	36
3	Coarse grained simulations of ionic liquids	41
3.1	Coarse grained models based on Martini 2.0	42
3.1.1	A simplified coarse grained model	42
3.1.2	Parametric studies regarding the bonded terms	43
3.1.3	Problems and restrictions of the model	44
3.2	An improved coarse grained model	44
3.2.1	Modifying the simulation parameter files	44
3.2.2	The Martini 3.0 force field	45
3.2.3	Mapping process for the $CAT_1^+ AN_i^-$ pairs	45
3.2.4	Results for the $CAT_1^+ AN_i^-$ ionic liquid pairs	47
3.3	Simulation of the $Rmim^+ AN_1^-$ ionic liquid pairs	50
3.3.1	Mapping process for the $Rmim^+ AN_1^-$ pairs	50
3.3.2	Results for the $Rmim^+ AN_1^-$ ionic liquid pairs	51
4	Optimization by using CG models	53
4.1	Evolutionary Algorithms	53
4.2	EASY	55
4.3	Parameterization of CG Ions	55
4.3.1	Computational Cost and Configuration Selection	55
4.3.2	Design Variables	56
4.4	Results of the CG Optimization	58
4.5	Validation by using AA models	64
5	Optimization using AA Models	67
5.1	The challenges of AA optimization	67
5.2	Design Variables	68
5.3	AA Topology Generation from SMILES Strings	70
5.4	The Optimization Setup	71
5.5	Results of the AA Optimization	73
5.6	Validation using more accurate AA input data	75

6 Overview and Conclusions	79
6.1 Overview	79
6.2 Conclusions	80
6.3 Future Work Proposals	81
Bibliography	i
A Appendix-All atom simulations of ionic liquids	vii
B Appendix-Coarse grained simulations of ionic liquids	xi

Chapter 1

Introduction

1.1 Introduction to Molecular Dynamics

Molecular Dynamics (MD) is the study of how molecules move, deform and interact over time. Predictions derived by the movements of molecules are very important in chemistry, physics, biology and engineering. MD predictions can either be microscopic or macroscopic. For example, predicting the magnitude of a force that acts upon an important group of atoms in a macromolecule could give to a scientist useful information, regarding the mechanism of a microscopic phenomenon, like the folding of a protein. On the contrary, a macroscopic prediction could be, for example, the calculation of the density of a liquid material.

MD, as a simulation technique, involves the direct simulation of systems of interacting spheres and the extraction of thermodynamic or physical properties [9]. These spheres interact only with forces from distance as real atoms do. When they approach close to each other, strong repulsive forces act upon them prohibiting any contact, as it happens in reality. Therefore, only the centers of these spheres are important for the simulation. Usually, the radius of each sphere should be equal to the corresponding real atom radius, only for visualization purposes. A typical simulation starts with an initial configuration of molecules in a simulation space, which can often be a cube (see figure 1.1), usually called simulation box in MD. To calculate the dynamic state of the configuration in the next timestep the Newton's 2nd law

$$\mathbf{M} \frac{d^2 \mathbf{q}}{dt^2} = F(\mathbf{q}) = -\nabla U(\mathbf{q}) \quad (1.1)$$

is used. \mathbf{M} is a diagonal mass matrix, \mathbf{q} and $\dot{\mathbf{q}}$ represent position and velocity

vectors of the atoms in the system and U is a potential energy function. A very important aspect in MD simulations is to properly model the energy function U . Many different techniques and models exist for the previous reason and they affect significantly the accuracy level of the simulation. In MD, the models regarding the formulation of the potential U are called force fields (FF).

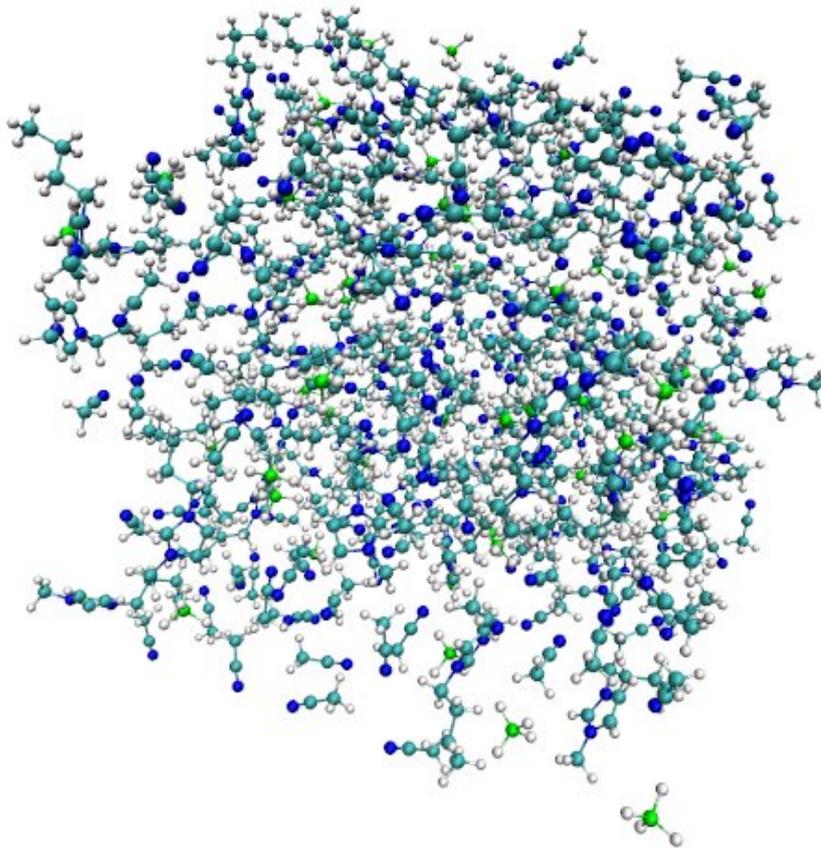


Figure 1.1: *Example of a MD simulation configuration for an ionic liquid pair consisting of hundreds of ions. Two different ions can be distinguished, a small anion and a more complex cation. Each colourful sphere represents the atom of a chemical element. For example, light blue spheres represent carbon atoms, while the white ones represent hydrogen atoms. The radius of each visualized atom is equal to the respective atomic radius of the chemical element (≈ 0.1 nm).*

MD relies on integration in time to compute successive kinetic states of the molecular system. However, the large size and complex nature of such systems leads always to a chaotic behaviour [9]. That means that the slightest change in the initial position of only one atom in a molecule could make the system evolve in a totally different way in respect with time. Thus, MD is based on the statistical mechanics theory [41]. The simulation needs to run long enough to allow to every possible configuration to appear. As happens in nature, the system tends to prefer the state of lowest possible energy. This state isn't known from the beginning of the simulation. In

fact, it isn't just one state but a collection of the most probable/low energy states. Consequently, a MD system is often considered to be equilibrated if the simulation time was long enough to allow it to reach the lowest possible **average** total energy value.

The term **average** is very important in MD because of the statistical nature of the phenomenon. All the macroscopic or microscopic properties measured in a MD systems are statistical variables. The properties of a microscopic system have real meaning only if the system has firstly reached an equilibrium and if the calculation was conducted for long enough [41]. The previous conditions are necessary for the calculated values to be considered statistically relevant. In practice, the user of a MD simulation software must check if important thermodynamic properties like pressure, density, volume, temperature and total energy have all converged on average to a final value. Sometimes, it is even necessary to run many simulations starting with different initial configurations. In that case, all the different simulations must reach similar equilibrium points, for the results to be considered safe.

In statistical mechanics, ensemble is an idealization that represents all the possible states of a mechanical system. In MD, three categories of ensembles exist. In a NVE ensemble the system is isolated from changes in moles (N), volume (V) and energy (E). For example this would be the case of a gas inside a tank with perfectly insulated walls. The same example without insulated walls should be modelled as an NVT ensemble. In that case, the system is isolated from changes in moles (N), volume (V) and temperature (T). Last but not least, in an NPT ensemble the quantities that are conserved are the number of molecules, pressure and temperature [42]. An example of this is the molecular system inside a glass of water.

To enforce a temperature or a pressure value to a system, a coupling algorithm is used. These algorithms are usually called thermostats if they impose the a temperature value or barostats if they impose a pressure value. A simple thermostat can operate by scaling the velocities and thus the kinetic energies of the particles in the system. Other methods define additional friction or source terms in the equation 1.1 to decrease or increase the kinetic energy of the particles. The algorithms for controlling pressure are more complex and they usually involve extra artificial degrees of freedom to be able to control the pressure [41]. It is worth noting, that the selection of the coupling method can affect significantly the convergence speed and the stability of the simulation. In the context of this thesis, Nose-Hoover [45], Berendsen [46] and V-rescale [48] thermostats were used. For the pressure coupling, Parrinello-Rahman [47] and Berendsen [46] barostats were used. The analysis of each thermostat and barostat used is outside of the purposes of this thesis. However, some important characteristics of the previous coupling methods will be explained in chapters 2 and 3.

Another important aspect of MD simulations is the scale of the system. The important parameter in MD which increases the computational cost is the number of particles N in the box. Thus, the size of a system is governed only by the total

number of atoms and not by the box size. The size of the box can even be changing during the simulation, according to the momentary pressure of the system (e.g. when an NPT ensemble is converging to an equilibrium). MD simulation can be conducted for different system sizes. The total number of molecules can range from 10^2 for simple mono-atomic gas systems to 10^5 for complex macromolecular systems. The timestep size can range from 1 fs (10^{-15} s) in accurate simulations to 50 fs in more approximate CG simulations. This isn't the only difference between CG and AA simulations. The accurate definition and unique characteristics for AA and CG simulations will be explained in next paragraphs. The total number of steps required before equilibrium is reached is heavily depended on the size of the system. For small systems a duration of a few ns is often sufficient [1], while bigger systems can even require a few μs to reach equilibrium [22]. The simulation box edge size can range from 5 nm in small systems to $1\text{ }\mu m$ in large systems.

1.2 Introduction to PEM fuel cells

A PEM fuel cell is a system that converts chemical energy stored in the fuel (hydrogen) directly to electric energy. This energy then powers the electric motor of the vehicle. The anode side is supplied with hydrogen, while the cathode side is supplied with filtered air. Fuel and air flow through the gas flow channels (GFCs). These gases are then been diffused through the gas diffusion layers (GDL) on each side. The hydrogen reaches the catalyst layer (CL) and it releases 2 electrons and 2 protons per hydrogen molecule. The electrons released flow back to the anode bipolar plate (BPP) and, then, to the external circuit to reach the cathode CL. There, oxygen is reduced, which means that it acquires the free electrons that have just reached the cathode CL. In chemistry, reduction is the process in which a chemical substance acquires one or more electrons. The membrane of the fuel cell allows only protons to pass through it. After reaching the cathode side, the protons react with the reduced oxygen to form water and heat. The previously described processes are better shown in figure 1.2. Reactions are given by:



1.3 The PEM of a fuel cell

PEMs are typically made from nafionTM. This material has proton conductivity that can range up to 20 S/m , depending on the level of hydration (relative humidity

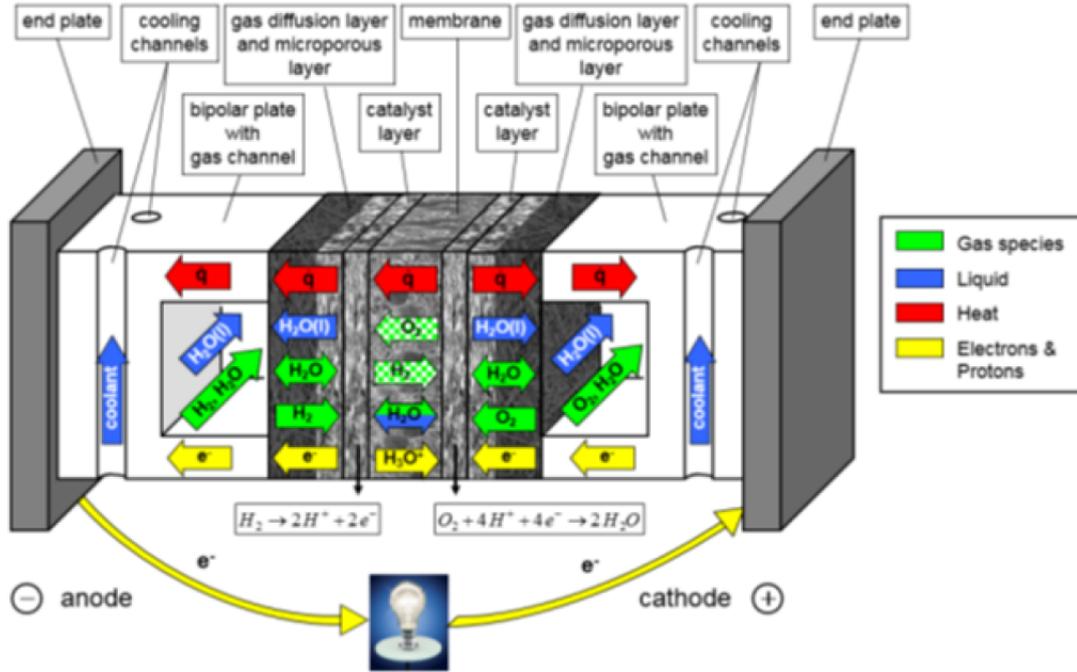


Figure 1.2: Schematic of the transport properties and components in a PEM fuel cell [34].

conditions). The derived unit of conductance $S = kg^{-1}m^{-2}s^3A^2$ (Siemens) is often used in materials science. NafionTM is a material made by the company DuPont [43], [56], [57] and comes in a number of different grades, varying according to thickness, hydration and permeability properties. Being chemical inert and having high permeability (conductivity) to cations are some of the main advantages of nafionTM.

In order to operate properly, nafionTM requires temperatures close to 80 °C and high relative humidity. These conditions can be maintained only by using expensive and heavy cooling systems. In addition, temperatures close to 80 °C have a negative impact on the efficiency of the catalytic processes that take place in the CL component of the fuel cell. This is why another material, resistant at higher temperatures, ideally close to 120 °C, is required for achieving higher efficiency in the PEMFC. Many researchers have proposed polymerized or poly-ionic liquids (PILs) as substances that could replace nafionTM [36, 37, 38, 39, 44].

1.4 Polymerized ionic liquid membranes

Ionic liquids (ILs) are organic salts in which the ions are poorly coordinated and melt below 100 °C [35]. When the ions in a salt are well coordinated, they form a crystal

structure even in very high temperatures, like NaCl which is solid even at 800 °C. Thus, most of the ILs are liquid in room temperature. They are electrolytes, which means that they conduct electricity when they are dissolved in polar substances like water. Of course, pure ionic liquids have the ability to conduct electricity too. The vapour pressure of these substances isn't sufficient to allow IL vapours to form in atmospheric pressure. This means that ILs cannot be evaporated or be burned at atmospheric conditions. Because of the previous property, they are considered as environmental friendly substances, since they cannot pollute the atmosphere.

PILs is a subclass of poly-electrolytes (polymers that conduct electricity), that feature an IL species in each monomer of the polymeric chain. Some of the properties of ILs are incorporated into the polymer chains giving rise to a very promising for the PEM technology material. PILs can then be combined with a crosslinker and a catalyst to create a PIL membrane (PILM). The crosslinker is a substance used in small weight ratios (*wt.%*) to connect the PIL chains into a PILM. The catalyst is also used in small *wt.%* ratios to accelerate the chemical reaction. The mixture is then placed in a mold to be polymerized. The polymerization is accomplished either by UV radiation or by thermal processes.

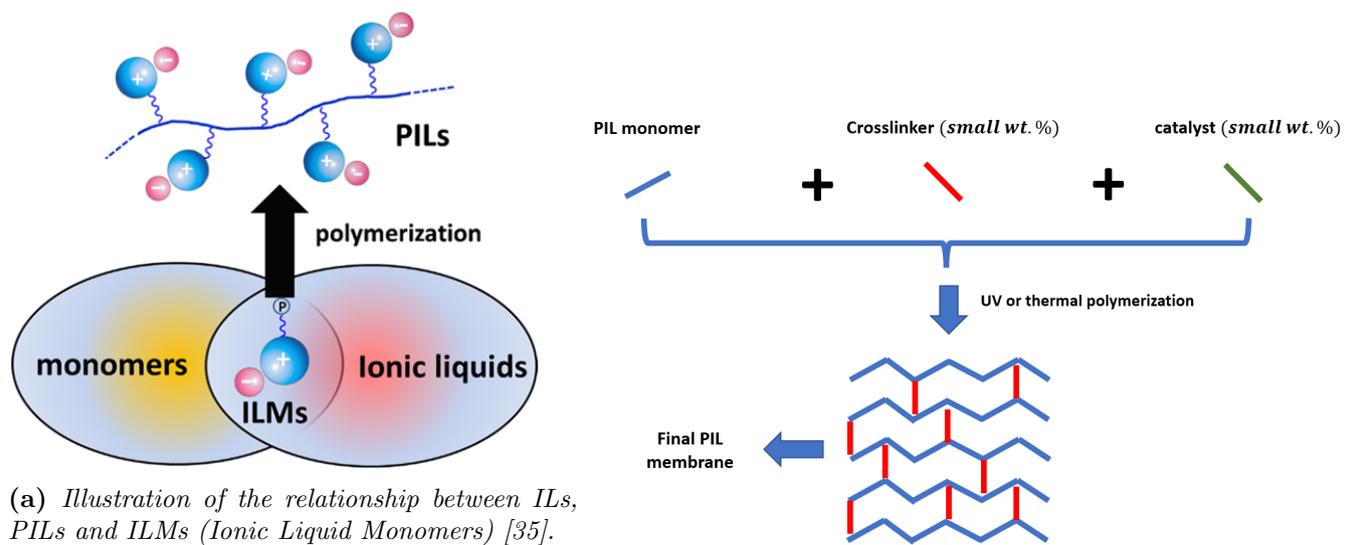


Figure 1.3: Schematics about ILs, PILs, ILMs, and PILMs.

Before proceeding any further, it is necessary to explain that proton and ionic conductivity are two different, but directly proportional properties. By using Nernst-Einstein Equations of the transport theory [40], under some hypotheses, it is possible to convert ionic to proton conductivity and vice versa. In this work all the conductivity calculations are referring to ionic conductivity.

Even though, PILMs can maintain their advantageous properties at higher temperatures, in contrary to nafionTM, they have one important issue to overcome. The

main problem prohibiting PILMs to be used in a commercial PEM is their insufficient ionic conductivity value. Researchers efforts have been focused to find IL pairs with sufficiently high conductivity value based on empirical data, chemical intuition and knowledge from similar substances. In the context of this thesis, the evolutionary algorithm based optimization software EASY of NTUA will be used to an attempt to find a high conductivity ionic liquid pair. Before any optimization is performed, a fast and sufficiently accurate MD model needs to be created, both for ILs and PILs.

The development of the models will be supported by available experimental data for about 5 IL pairs and their respective PILMs (see section 2.2.1 for more information about these 5 IL pairs). These experimental data will be used to a certain degree to validate the models before they are used for the optimization. It is worth noting, that PIL experimental data are not available. Thus PILM experimental data will be used for the validation of the PIL models. This is approximately correct by the assumption that the ionic conductivity of PILs and PILMs is similar due to their similar structure.

Of course, the first step before developing any MD model, couldn't be other than conducting a literature survey about all the available methods that exist for IL and PIL MD simulations.

1.5 Literature survey on the force fields in MD

Maybe the most important parameter in a MD simulation is the force field which is used. Force field is called a technique used to model the interactions between the simulation spheres. The force fields can be categorized in four main categories, according to Salanne [1]. The simplest force field is the Non-polarizable all atom force field, some examples are OPLS, AMBER, CL&P (see section 1.7.1). The term AA means that the interacting spheres of the simulation are referring to atoms. Thus, each atom in a molecule is an interacting sphere in a AA model. Non-polarizable means that the force field doesn't take into account the polarization effects. All forces and phenomena caused by the non-uniform distribution of charge in a molecule are considered as polarization effects. The non-uniform distribution of charge creates electric dipoles which then produce additional interaction forces between the atoms.

Polarizable all atom force fields (PAAFF) exist too and they are much more accurate than NPAAFFs, when substances such ILs or PILs are simulated. The polarization phenomena are not negligible in ionic substances such as ILs and PILs. Some of the most popular are, PIM [2], Borodin's method [3] and SAPT [4] (see section 1.8). Another category of fields are charge-scale force fields. These are non-polarizable AA force fields which try to mimic the polarization mechanism through a simple scale of charge in the ions (see section 1.9). Last but not least, when complex molecules are simulated, it is essential from the computational prospective to use coarse-grained

(CG) force fields. In these force fields the interacting spheres of the simulation are referring to a group of atoms in each molecule which is called bead. In other words, coarse graining is a process in which a molecule is divided in groups of atoms (see section 1.10.2).

Unfortunately, CG force fields are mostly case specific, when used in ILs, which means that the force field needs to be reparameterized (changing the definitions and interaction properties for some or all the beads) in order to be used to a different IL. However, some models like the one proposed by Voth [5], [6] partially tackles the lack of transferability problem. In general, it is believed that the prediction of physico-chemical properties of ILs is an ambitious but feasible objective [1].

It should be pointed out that in the context of this thesis only two out of all the previous categories will be implemented for performing ionic conductivity calculations. Only NPAAFFs and CG force fields will be used. To justify the previous selection, a detailed review for each force field is necessary. The advantages and disadvantages of the previous 4 categories in ILs simulations are summarized in table 1.1. The information contained in this table is explained in more detail in the following sections. The purposes of those are to describe:

- What a classical MD force field is and why it is advantageous over other computational tools (see paragraph 1.6).
- The basic structure of AA and CG force fields (see paragraphs 1.7.1 and 1.10.1 respectively)
- The different categories and distinct characteristics of force fields.

FF Category	Transferability	Accuracy	Complexity	Computational Cost
NPAAFF	Very high	Adequate	High	High
PAAFF	Low	High	Very high	Very high
Charge-Scale	Low	High	Low	High
CG	Adequate	Low	Very low	Very low

Table 1.1: *Advantages and disadvantages for each force field category used in IL simulations.*

1.6 Force fields in classical MD simulations

According to Leimkuhler [9] the most accurate method for calculating the behavior of a N atom system is by using the probabilistic Schrodinger equation, which gives the probability for a specific configuration of particles to exist at time t_0 . In spite of its accuracy, the previous equation is currently impractical to be applied for atomic calculations in large systems. This equation is very different from the classical laws of motion, where the input is time and the output is the coordinates of the particles. In Schrodinger equation both time and coordinates are required to obtain how probable is for this specific input pair to exist in reality.

A simple example will be used to better explain the challenges of applying Schroedinger equation. In order to describe the 3D motion of a simple molecule, for example the motion of a water molecule, 13 variables are required in each spatial dimension. The reason for this is that a water molecule consist of 10 electrons and 3 nuclei. So, the corresponding coordinates will be $q_{1,x}, q_{1,y}, q_{1,z}, q_{2,x}, q_{2,y}, q_{2,z}, \dots, q_{13,x}, q_{13,y}, q_{13,z}$. The probabilistic equation is a partial differential equation of the following form:

$$i\hbar \frac{\partial \Phi}{\partial t} = -\hbar^2 \sum_{j=1}^{13} \frac{1}{2\mu_j} \left(\frac{\partial^2 \Phi}{\partial q_{j,x}^2} + \frac{\partial^2 \Phi}{\partial q_{j,y}^2} + \frac{\partial^2 \Phi}{\partial q_{j,z}^2} \right) + U_P(q_{1,x}, q_{1,y}, \dots, q_{13,z}) \Phi \quad (1.3)$$

Here, \hbar is Plank's constant, U_p is the primitive atomic potential energy function, i is the square root of -1 , μ_j is the mass of the j_{th} particle (electron or nuclei) and Φ is the unknown wave function. The primitive potential energy function U_p is calculated by applying models about the repulsion of electrons due to their similar charge and the attraction of electrons to the positively charged nucleus of each atom.

The solution of Schroedinger equation is the wave function $\Phi = \Phi(q_{1,x}, q_{1,y}, \dots, q_{13,z}, t)$. The squared wave function $\Phi(q_{1,x}, q_{1,y}, \dots, q_{13,z}, t)^2$ tells how probable is for the $\mathbf{S}^0 = (q_{1,x}^0, q_{1,y}^0, \dots, q_{13,z}^0, t^0)$ particle configuration to exist at time t^0 . In other words, the wave function Φ has 40 arguments for this simple system and the output is a single number that gives the probability that the system at time t^0 is at the state S^0 . Therefore, it is obvious that the problem with the quantum mechanical approach isn't only very computationally complex and intensive, but is also very difficult to visualize the results for systems of thousand atoms, like the ones MD attempts to solve. The large number of independent spatial dimensions rules out any straight-forward attack on the problem.

In order to reduce the complexity of the problem, classical Newtonian mechanics theory is used. By using Newton's second law, we can describe the coordinates of the i_{th} nucleus of an N-atom system denoted by $q_{i,x}, q_{i,y}, q_{i,z}$.

$$m_i \frac{d^2 q_{i,x}}{dt^2} = -\frac{\partial U}{q_{i,x}}, \quad m_i \frac{d^2 q_{i,y}}{dt^2} = -\frac{\partial U}{q_{i,y}}, \quad m_i \frac{d^2 q_{i,z}}{dt^2} = -\frac{\partial U}{q_{i,z}} \quad (1.4)$$

Here m_i is the atomic mass and U is the potential energy function. Of course equation 1.4 must be supplemented by the proper initial conditions. In classical MD simulations, function U is almost always approximated by empirical force fields methods. The usage of these methods simplifies significantly the simulated system permitting the conduction of MD simulations in ordinary computers. A great deal of chemical insight, experimental data and simulation work is necessary for the production of a new force field.

Equation 1.3 is a linear partial differential equation. In contrast, 1.4 is a nonlinear system of ordinary differential equations. From one point of view it seems like the two above equations have similar complexity. However Newton's equations involve differentiating solutions only in time. Therefore, when we discretize the equations to solve them, we only have to introduce a grid in this single direction. This way the total number of equations increases linearly with N . The same is true for the computational cost which is much smaller compared to the exponentially dependent cost observed in the treatment of quantum mechanics. The cost is exponentially dependent on quantum mechanics methods because they use not only the total number of nuclei like classical methods do, but also the total number of electrons.

1.7 Non-polarizable all atom force fields (NPAAFF)

1.7.1 The typical structure of a NPAAFF

NPAAFFs are the simplest methods to approximate the energy function U appeared in equations 1.1 and 1.4. In chapter 2, NPAAFFs (OPLS-AA and Gromos) will be used in IL simulations. Therefore, it is important to firstly describe the characteristics and structure of these force fields. The principles and the basic terms of NPAAFFS are the basis for understanding all the other force fields and MD models in general.

The energy terms of the force field are divided in two main categories. The first category of terms of the force field are called bonded terms. They model the energy stored in the chemical bonds during the oscillation of the atoms in a molecule. This energy should not be confused with the energy stored in the chemical bonds that can be released by a chemical reaction, like combustion. The bonded energy is the sum of all the potential energies stored in the chemical bonds as a result of the atomic oscillations. Therefore, the bonded potential is the energy stored in different kind of equivalent springs.

E_{bond} is the energy stored during the elongation of a bond between two atoms and it is usually modelled as a linear spring. E_{angle} is the energy stored during the change of the angle between a triplet of atoms and it is usually approximated as a torsional spring. $E_{torsion}$ is the energy stored in a formation of 4 atoms without branches, because of the change in the proper dihedral or more simply torsion angle. Torsion angle is called the angle between two intersecting planes. In this case the first plane is formed by the first 3 atoms of the previous formation and the second one by the last 3 atoms. $E_{torsion}$ is usually modelled by using Fourier series. Finally, $E_{improper}$ is the energy stored in a configurations like the lower-right one of figure 1.4, because of the change in the improper dihedral angle. This is the angle between the plane formed by atoms j, i, k and the line formed by atoms i, l .

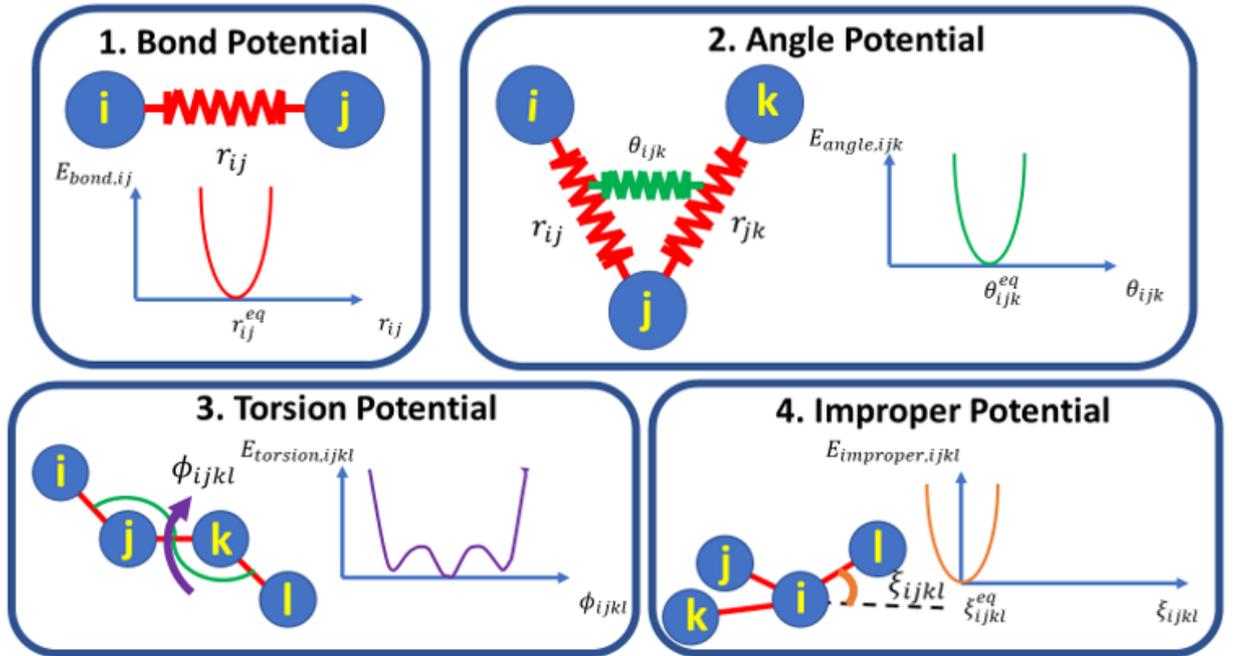


Figure 1.4: Schematic about the potential of the bonded terms of a NPAAFF.

The bonded terms can be further explained by the equations

$$E_{bonded} = E_{bonds} + E_{angles} + E_{torsions} + E_{impropers} \quad (1.5a)$$

$$E_{bonds} = \sum_{ij}^{bonds} \frac{K_{ij}^b}{2} (r_{ij} - r_{ij}^{eq})^2 \quad (1.5b)$$

$$E_{angles} = \sum_{ijk}^{angles} \frac{K_{ijk}^a}{2} (\theta_{ijk} - \theta_{ijk}^{eq})^2 \quad (1.5c)$$

$$E_{torsions} = \sum_{ijkl}^{dihedrals} \sum_{m=1}^4 \frac{V_{ijkl}^m}{2} [1 + (-1)^m \cos m\phi_{ijkl}] \quad (1.5d)$$

$$E_{impropers} = \frac{1}{2} k_{\xi} (\xi_{ijkl} - \xi_{eq})^2 \quad (1.5e)$$

From the previous equations one can observe that the energy of the bonds is modeled like a linear spring connecting 2 atoms. K_{ij}^b is the spring constant, r_{ij} is the interatomic distance between the atom i and the atom j that they are bonded and the r_{ij}^{eq} is the same distance as before but in the equilibrium state of the molecule. The same logic applies to the potential energy of the angles which are formed between 3 atoms i, j, k . In this case, a torsion spring model is used. The potential energy of the torsion angles ϕ_{ijkl} which are formed between 4 atoms are modelled by using Fourier series with constants V_{ijkl}^m . Finally, the improper dihedrals are approximated as a harmonic function of the angle ξ_{ijkl} . Therefore the quantities $K_{ij}^b, r_{ij}^{eq}, K_{ijk}^a, \theta_{ijk}^{eq}, V_{ijkl}^m$ and k_{ξ} are parameters of the force field. The values for these parameters for a specific molecule are usually acquired by experiments or by quantum mechanics simulations.

The second category of energy terms are called non-bonded terms. As the name implies, they model all the other potentials except from those caused because of the bonds. The non-bonded terms consist of the Van der Waals and Coulomb forces. Van der Waals forces are the interatomic forces caused by local electrostatic phenomena in the atoms and they consist of two main potentials. The first potential is the one regarding Pauli repulsion forces and the second one is the dispersion forces potential. The potential field which is caused by the Pauli repulsion forces [11] and London dispersion attraction forces [12] is usually modeled as a Lennard-Jones Potential. As shown in figure 1.5, for small interatomic distances, Pauli repulsive forces are dominant. The opposite is true for bigger distances between the atoms where the dispersion mechanism is dominant. The final term of the non-bonded potential is the electrostatic one. In this case the Coulomb law is followed as shown in equation 1.6d. Thus, the non-bonded terms can be explained by the equations

$$E_{non-bonded} = E_{repulsion} + E_{dispersion} + E_{electrostatics} \quad (1.6a)$$

$$E_{Lennard-Jones} = E_{repulsion} + E_{dispersion} \quad (1.6b)$$

$$E_{Lennard-Jones} = \sum_i \sum_{j>i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (1.6c)$$

$$E_{electrostatic} = \frac{1}{4\pi\epsilon_0} \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}} \quad (1.6d)$$

Therefore, the total potential energy $U \equiv E_{tot}$ of a NPAAFF is given by

$$E_{tot} = E_{bonded} + E_{non-bonded} \quad (1.7)$$

where E_{bonded} are the bonded terms and $E_{non-bonded}$ are the non-bonded terms of the potential equation. The model previously described has been the basis for almost every NPAAFF that has been developed in the previous 30 years. In the next paragraphs, the most popular NPAAFFs will be presented.

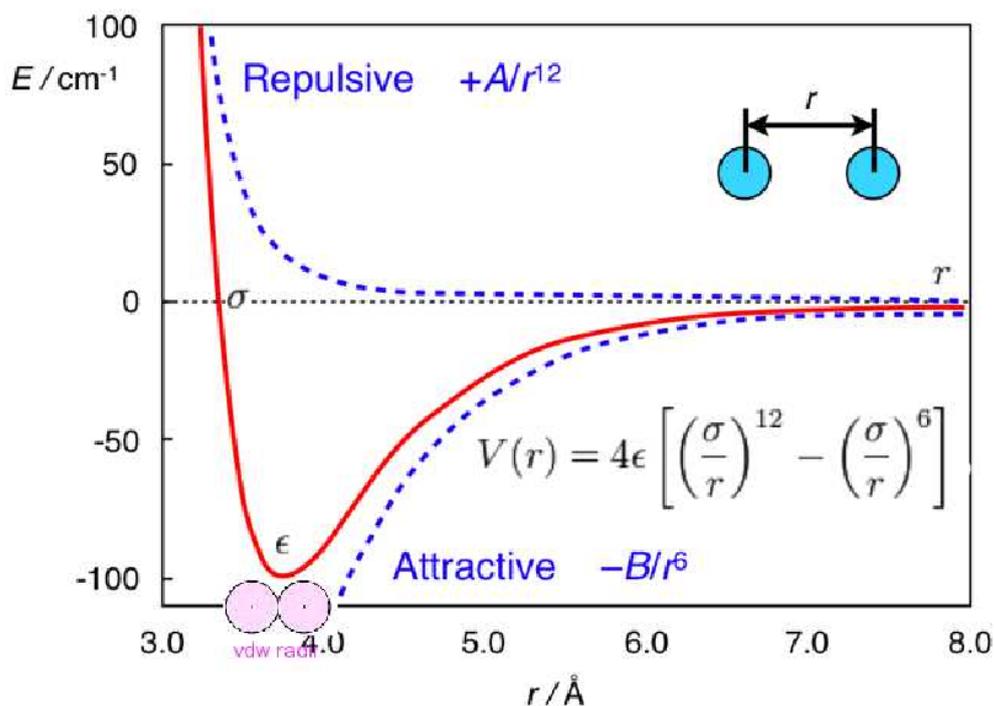


Figure 1.5: *Lennard-Jones potential visualization [13]*

1.7.2 AMBER force field

AMBER is one of the first general purpose NPAAFFs. Even though it will not be used in for any simulation in this thesis, its simple form will be briefly explained for the sake of completeness. The potential function of AMBER force field can be described by the following equation [8].

$$E_{tot} = \sum_{bonds} K_b(b - b_{eq})^2 + \sum_{angle} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}(1 + \cos(n\phi - \gamma))^2 + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] \quad (1.8)$$

where K_b and K_θ are the force constants for the bond and bond angles, respectively; b and θ are bond length and bond angle; b_{eq} and θ_{eq} are the equilibrium bond length and bond angle; ϕ is the dihedral angle and V_n is the corresponding force constant; The phase angle γ takes values of either 0° or 180° . The nonbonded part of the potential is represented by the van der Waals (A_{ij}), the London dispersion terms (B_{ij}) and interactions between partial atomic charges (q_i and q_j). The rest of the terms are explained in [7] pp 18. For example, in this paper, terms A_{ij} and B_{ij} are given by the equations: $A_{ij} = \epsilon_{ij}(R_{ij})^{12}$, $B_{ij} = 2\epsilon_{ij}(R_{ij})^6$, where the values of ϵ_{ij} , R_{ij} can be found in tables of [7]. These values and many other necessary constants have been calculated using Monte Carlo simulations or experimental methods in well known and often used bonds.

1.7.3 OPLS All Atom (OPLS-AA) force field

In contrary to AMBER, OPLS-AA will be used in the next chapter for IL simulations. OPLS-AA is very popular and an obvious choice for everyone who try to simulate liquid substances. Compared to other generalist force fields, special attention was devoted in OPLS-AA to the simulation of liquid-state thermodynamic properties [14] is applied. Therefore, an overview of its most important aspects is considered necessary.

The functional form of the OPLS force field is very similar to that of AMBER. The OPLS-AA potential is given by the following equations [19].

$$E_{tot} = \sum_{bonds} K_b(b - b_{eq})^2 + \sum_{angle} K_\theta(\theta - \theta_{eq})^2 + E_{torsion} + E_{non-bonded} \quad (1.9a)$$

$$E_{torsion} = \frac{V_1}{2} [1 + \cos(\phi + f_1)] + \frac{V_2}{2} [1 - \cos(2\phi + f_2)] + \frac{V_3}{2} [1 + \cos(3\phi + f_3)] \quad (1.9b)$$

$$E_{non-bonded} = \sum_i^{on\ a} \sum_j^{on\ b} \left[\frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right] f_{ij} \quad (1.9c)$$

$$f_{ij} = \begin{cases} f_{ij} = 0.5, & i, j \in \{1, 4\} \\ f_{ij} = 1.0, & i, j \notin \{1, 4\} \end{cases} \quad (1.9d)$$

Here, e is the elementary charge. All the other input parameters of the field, like $V_1, V_2, V_3, f_1, f_2, f_3$ can be obtained for each specific case in the original OPLS-AA force field paper [19], in other supplementary papers or in OPLS-AA parameter databases.

1.7.4 CL&P force field

CL&P force field is currently the most famous force field for the simulation of ILs. It was build in the functional form of the OPLS-AA force field. The equations are very similar and they can be found in the original paper [14]. This field was parameterized for a large set of IL compounds by using different methods than the ones used in the OPLS-AA. The bonded and Lennard-Jones parameters were in general the same as in the OPLS-AA. However some of them were reparameterized by reproducing the molecular geometries and energy torsion profiles for isolated molecules using quantum chemistry methods. For the proper reproduction of the electrostatic field generated by the molecule the CHelpG method [23] was applied. Consequently, the high complexity and lack of transferability of CL&P are the main reasons for not using it in the next chapter, despite of being specialized in ILs.

1.8 Polarizable all atom force field (PAAFF)

1.8.1 Advantages of PAAFF over NPAAFF

Although NPAAFF methods are simple, well understood and popular, they often fail to predict the transport properties of many ionic liquid pairs, like the diffusion constant, the ionic conductivity and the viscosity. NPAAFF methods often underestimate by an order of magnitude the first two properties and overestimate

by an order of magnitude the latter [1]. Therefore NPAAFF can be used mainly for qualitative conclusions about these very important properties in ILs. This problem was solved by Yan et al. [61]. In this work, the polarization effects that play an important role in the cations and ions interactions were included in the force field equations. The PAAFF aren't just some NPAAFF in which some polarization terms are added. Although they use the basic structure that was described in paragraph 1.7.1, they often need a partial or a total reparameterization. This is usually done by Ab initio methods which are quantum mechanics simulations performed in a very small number of molecules. In conclusion these relatively new methods are much more accurate than NPAAFF methods. However, they haven't been implemented in a big scale yet, because of the complexity in producing input parameters for each molecule. In the following 2 sections, a description of the most popular PAAFFs for IL simulations will be given. This process is necessary for drawing conclusions about PAAFFs and for justifying their absence in all the following IL simulations.

1.8.2 The Polarized Ion Model (PIM)

PIM is a popular PAAFF for IL simulation according to [1]. In the PIM force field, the repulsion and dispersion terms are modified. An extra electrostatic term is added compared to a NPAAFF. Supposing that all the bonded terms have the form of equations 1.5 the non-bonded terms are given by [2].

$$E_{non-bonded}^{PIM} = E_{vdW}^{PIM} + E_{electrostatic}^{PIM} = (E_{repulsion} + E_{dispersion}) + (E_{Coulomb} + E_{polarization}) \quad (1.10a)$$

$$E_{repulsion} = \sum_i \sum_{j>i} B_{ij} e^{-a_{ij} r_{ij}} \quad (1.10b)$$

$$E_{dispersion} = \sum_i \sum_{j>i} \left(-f_6^{ij}(r_{ij}) \frac{C_6^{ij}}{r_{ij}^6} - f_8^{ij}(r_{ij}) \frac{C_8^{ij}}{r_{ij}^8} \right) \quad (1.10c)$$

$$f_n^{ij}(r_{ij}) = 1 - e^{-b_n^{ij} r_{ij}} \sum_{k=0}^n \frac{(b_n^{ij} r_{ij})^k}{k!} \quad (1.10d)$$

$$E_{Coulomb} = \frac{1}{4\pi\epsilon_0} \sum_i \sum_{i>j} \frac{q_i q_j}{r_{ij}} \quad (1.10e)$$

$$E_{polarization} = \sum_i \sum_{i>j} (-g_{Dij}(r_{ij}) q_i T_1^{ij} \mu_j + g_{Dij}(r_{ij}) q_j T_1^{ij} \mu_i - \mu_i T_2^{ij} \mu_j) + \sum_i \frac{\mu_i^2}{2a_i} \quad (1.10f)$$

$$T_1^{ij} = \nabla_i \frac{1}{r_{ij}} = -\frac{r_{ij}}{r_{ij}^3} \quad (1.10g)$$

$$T_2^{ij} = \nabla_i T_1^{ij} = \frac{3r_{ij} \times r_{ij}}{r_{ij}^5} - \frac{1}{r_{ij}^3} I \quad (1.10h)$$

$$g_{Dij}(r_{ij}) = 1 - c_{ij} e^{-b_{Dij} r_{ij}} \sum_{k=0}^4 \frac{(b_{Dij} r_{ij})^k}{k!} \quad (1.10i)$$

Here, C_6^{ij} , C_8^{ij} is the dipole-dipole or dipole-quadrupole dispersion coefficients, f_n^{ij} are the Tang-Toennies dispersion damping functions [15], b_n^{ij} is a parameter that sets the range of the damping effect, a_i is the polarizability of ion i , which is assumed to be isotropic. μ_i is the induced dipole of ion i , while T_1 , T_2 are the charge-dipole and dipole-dipole interaction tensors. Also $g_{Dij}(r_{ij})$ is a Tang-Toennies function too and c_{ij} , b_{Dij} , B_{ij} , a_{ij} are parameters. In general, all these parameters are calculated for a specific molecule using an electronic structure density functional theory (DFT), which is a quantum mechanics, ab initio method.

PIM is very complex and it needs many more procedures than the ones described previously. For example, the induced dipole values μ_i are obtained by performing a single minimization procedure in the polarization potential as shown in equation $\left(\frac{\partial V_{PIM}^{polarization}}{\partial \mu_\alpha^i}\right)_{\mu_\alpha^N} = 0$ [2]. Generally, this potential is much more accurate than every other NPAFF potential. However, the development of such force fields has begin relatively recently and they often include very complex procedures for obtaining the force field parameters.

1.8.3 Other Polarizable all atom force fields

Another specialized force field has been developed by Borodin [3] for a variety of different ionic liquids as well for other popular chemical substances like alkanes, fluoroalkanes, propylene carbonate, etc. This force field is capable of predicting with accuracy important properties such as density, heat of vaporization, self-diffusion coefficients, **ionic conductivity** and viscosity. The bonded terms have a similar analytical form as in the CL&P force field, while the non-bonded terms resemble the PIM ones. A difference between Borodin's force field and PIM force field is

that the Tang-Toennies functions are replaced by a different repulsive term for the dispersion. The parameterization of the force field was achieved by a combination of quantum chemistry data and experimental results.

Finally, it is worth to briefly describe the SAPT method [4], [16]. The method is quite accurate in predicting the transport properties of ionic liquids. The parameters are usually obtained by *ab initio* calculations. This model has many different terms compared to the previously described force fields. The method and the fundamental equations of the force field can be found in detail in the study by McDaniel and Schmidt [17].

1.8.4 Conclusions on Polarizable all atom force fields

PAAFFs are without any doubt much more complex and accurate than the NPAAFFs. This new generation of force fields implements modern physics and sophisticated mathematics to best describe the physical mechanisms of molecular interactions. However, it is important to mention that this kind of complexity in the model has a direct impact to the computational cost. Molecular dynamics simulation using NPAAFFs is already a very computationally intense process. The implementation of PAAFFs for the same simulations increases the computational cost by about 10 times. In addition, these methods are very difficult to implement in many simulation packages if not possible at all. Therefore, one usually has to sacrifice computational and development time for gaining more accurate simulation results. For the previous reasons, it was decided to use NPAAFFs instead of PAAFFs in this thesis.

1.9 Charge scaling methods

In recent years, many papers have proposed a relatively inexpensive way to approximately account for polarizability and charge-transfer effects. In these methods, the charges of ions are scaled down from $|e|$ to γe through multiplying the charge of each electron in the ions by a scaling factor γ . As a result, the cation will have a bigger absolute charge by $(1 - \gamma)e$ than the anion. In addition, with this approach the charge transfer is uniform and there is no change in the relative distribution of charge in each chemical species. Therefore, charge scaling mimics the average effect of polarization and charge transfer in ILs, without the need of significant changes in the NPAAFF which is used [18].

Justification of the charge-scaling approach has been provided by quantum mechanics calculations. The study of Young and Hardacre [20] showed that for 1,3-dimethylimidazolium chloride IL a γ factor of 0.6 – 0.7 provided excellent agreement with the *ab initio* MD results. Another way for determining the optimal scale factor is through comparisons with experimental data [18]. It is important to note, that

there is no single optimal value for the factor γ . For example, for obtaining optimal results in ionic conductivity, a different γ value is necessary compared to the value needed for optimal results in viscosity.

In conclusion, charge scaling does not differ substantially than the less accurate NPAAFFs. They usually predict with decent accuracy the transport properties of ILs, while remaining simple to apply. A great disadvantage of these methods is that they are purely empirical, because they always require ab initio calculations or experimental data for determining the optimal value of γ . For this reason, this method will not be applied for any simulations in the next chapters.

1.10 Coarse-grained models

1.10.1 Basic principles of coarse grained models

CG models are used extensively in the context of this thesis for creating IL models for optimization purposes and for creating PIL models. Thus, it is important to describe their unique characteristics. In a coarse-grained model, the molecule for simulation is divided into groups of atoms which called beads. It is easier to imagine beads like bigger atoms which interact with each other with similar mechanisms as simple atoms do in a NPAAFF. Most of the coarse grained force fields include about 4 heavy atoms (C, N, S, O, ... but not H). The process of assigning a group of atoms in a molecule to a specific bead is called mapping and the inverse process is usually referred as backward mapping.

The main reason for splitting the molecule to simpler units, the beads, is to reduce the total number of atoms. In general, the computational cost increases at least linearly by the number of atoms N . For the electrostatic interactions, the total number of atoms N affects the computational cost by $O(n^2)$ when classic Ewald Summation is used, by $O(n \log(n))$ when PME method is used or by $O(n)$ when more sophisticated methods like P3M are used [21]. Thus, the computational cost of large systems, like lipid membranes, polymer chains, RNA chains or DNA chains, is prohibitively expensive, if an AA method is applied. In these cases, the only viable option is to use coarse-grained models. When coarse-grained models are used the integration step of the simulation can be increased by almost 10 times or more. The increased time step allows performing simulations for longer time scales, which is useful in the replication of phenomena that need more time to complete, like the lipid bilayer formation. The basic idea of coarse grained models can be better understood by figure 1.6, which originates from the original paper of the coarsened model MARTINI 2.0 [22].

Nevertheless, coarse grained models have some important drawbacks. After the mapping is finished a significant amount of information about the molecule is lost.

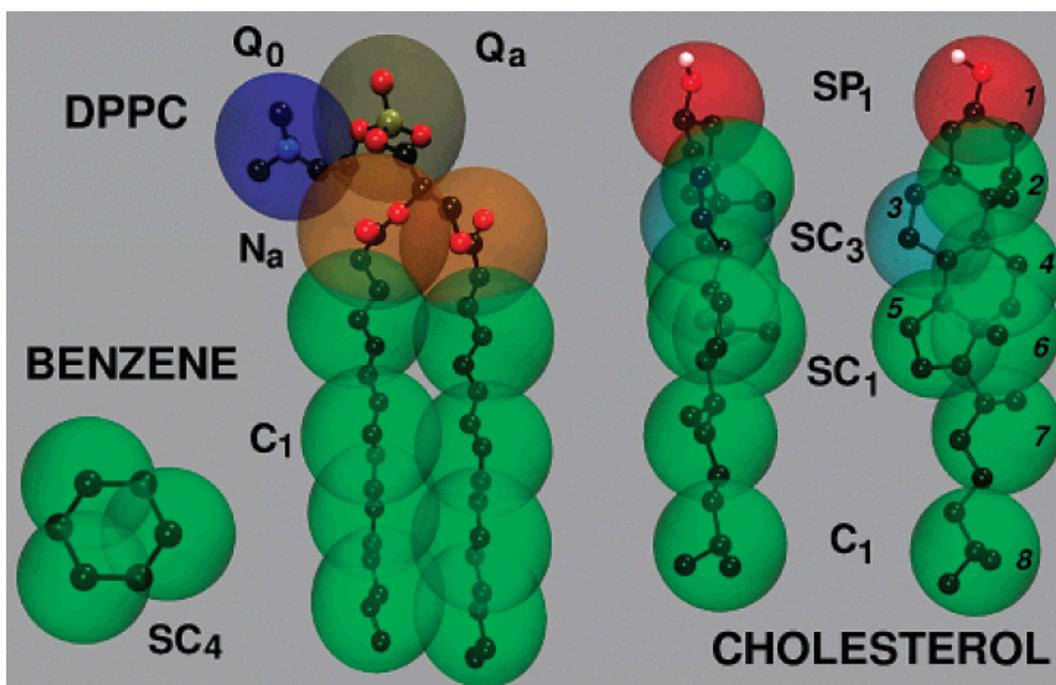


Figure 1.6: Mapping between the chemical structure and the coarse grained model for DPPC, cholesterol and benzene [22]

This can affect significantly the accuracy of post processing calculations, like those regarding ionic conductivity. Another often problem of coarse grained models is the lack of transferability. Mapping is a special process for each molecule which usually requires a great amount of chemical intuition as well as some experience in coarse graining. Therefore, an automated mapping algorithm for mapping all atom models to the corresponding coarse grained ones is considered, in general, difficult to develop.

1.10.2 Brief introduction to the MARTINI 2.0 force field

MARTINI is a general purpose open source CG force field [22]. It will be used for creating every CG model in the following chapters. The main reasons for selecting the specific force field were its rich documentation, backward mapping capabilities (necessary for the IL optimization) and high standardization. Although, version 3.0 is implemented in the next chapters, because of some necessary unique features, version 2.0 is very similar to version 3.0 and easier to explain. Therefore, in the next paragraphs the main features of MARTINI 2.0 will be explained.

MARTINI 2.0 force field includes four different types of interacting sites (beads): polar (P), nonpolar (N), apolar (C) and charged (Q). Each bead type is followed by a subscript which gives more information either for the degree of polarity (from 1, low polarity, to 5, high polarity) or for the hydrogen-bonding capabilities (d=donor,

a=acceptor, da=donor&acceptor, 0=none).

The non bonded interactions are described by a typical Lennard-Jones potential and the electrostatic interactions by a slightly different Coulomb potential. The Lennard-Jones function are exactly the same as in equations 1.6. However, the Coulombic function of equations 1.6 has been divided by an extra relative dielectric constant $\epsilon_r = 15$. In almost every type of bead the parameter σ_{ij} is equal to 0.47 nm , except for some special classes of rings and antifreeze particles. In table 1.2 Marrink et al [22] show all the possible types of interactions between the different types of particles.

The bonded parameters are very similar to those of a typical NPAAFF (Equations 1.5). For the sake of completeness, the bonded terms of the MARTINI 2.0 force field are shown below. All the terms have the same physical meaning as explained in equations 1.5. However, the equation for the improper dihedral angle is simpler than the one in equations 1.5.

$$V_{bond}(R) = \frac{1}{2}K_{bond}(R - R_{bond})^2 \quad (1.11a)$$

$$v_{angle}(\theta) = \frac{1}{2}K_{angle}[\cos \theta - \cos \theta_0]^2 \quad (1.11b)$$

$$V_{id}(\theta) = K_{id}(\theta - \theta_{id})^2 \quad (1.11c)$$

	Q				P					N				C					
sub	da	d	a	0	5	4	3	2	1	da	d	a	0	5	4	3	2	1	
Q	da	O	O	O	II	O	O	O	I	I	I	I	IV	V	VI	VII	IX	IX	
	d	O	I	O	II	O	O	O	I	I	I	III	I	IV	V	VI	VII	IX	IX
	a	O	O	I	II	O	O	O	I	I	I	I	III	IV	V	VI	VII	IX	IX
	0	II	II	II	IV	I	O	I	II	III	III	III	III	IV	V	VI	VII	IX	IX
P	5	O	O	O	I	O	O	O	O	I	I	I	IV	V	VI	VII	IX	IX	
	4	O	O	O	O	O	I	I	II	II	III	III	III	IV	V	VI	VI	VII	VII
	3	O	O	O	I	O	I	I	II	II	II	II	II	IV	V	V	VI	VII	
	2	I	I	I	II	O	II	II	II	II	II	II	II	III	IV	IV	V	VI	VII
	1	I	I	I	III	O	II	II	II	II	II	II	II	III	IV	IV	IV	V	VI
N	da	I	I	I	III	I	III	II	II	II	II	II	IV	IV	V	VI	VI	VI	
	d	I	III	I	III	I	III	II	II	II	II	III	II	IV	IV	V	VI	VI	
	a	I	I	III	III	I	III	II	II	II	II	II	III	IV	IV	V	VI	VI	
	0	IV	IV	IV	IV	IV	IV	IV	III	III	IV	IV	IV	IV	IV	IV	IV	V	VI
C	5	V	V	V	V	V	V	IV	IV	IV	V	V							
	4	VI	VI	VI	VI	VI	VI	V	IV	IV	V	V	V	IV	IV	IV	IV	V	V
	3	VII	VII	VII	VII	VI	VI	V	IV	VI	VI	VI	IV	IV	IV	IV	IV	IV	
	2	IX	IX	IX	IX	VII	VII	VI	VI	V	VI	VI	V	V	V	IV	IV	IV	
	1	IX	IX	IX	IX	VIII	VIII	VII	VII	VI	VI	VI	VI	V	V	IV	IV	IV	

Table 1.2: Table of interactions from the original MARTINI 2.0 paper [22]. The level of interaction indicates the well depth of the Lennard-Jones potential: O: $\epsilon = 5.6\text{kJ/mol}$; I: $\epsilon = 5.0\text{kJ/mol}$; II: $\epsilon = 4.5\text{kJ/mol}$; III: $\epsilon = 4.0\text{kJ/mol}$; IV: $\epsilon = 3.5\text{kJ/mol}$; V: $\epsilon = 3.1\text{kJ/mol}$; VI: $\epsilon = 2.7\text{kJ/mol}$; VII: $\epsilon = 2.3\text{kJ/mol}$; VIII: $\epsilon = 2.0\text{kJ/mol}$; IX: $\epsilon = 2.0\text{kJ/mol}$; The Lennard-Jones parameter $\sigma = 0.47\text{nm}$ for all the interactions except level IX for which $\sigma = 0.62\text{nm}$

1.11 Scope and structure of the thesis

The main purpose of this thesis is to search for new ILs that have ionic conductivity as high as possible. In order to calculate the ionic conductivity, AA and CG models will be used with the open source software GROMACS. For the creation of the AA models, OPLS-AA and the similar GROMOS force field are applied. For the CG models MARTINI 3.0 force field is applied. The optimization is performed by using CG MARTINI 3.0 models and then by using OPLS-AA models. The evolutionary algorithm based software EASY of NTUA is used for all the optimization processes.

The contents of this diploma thesis are outlined as follows:

- Chapter 2: The structure and necessary software for performing AA MD simulations is explained. The AA model is created and validated, firstly by conducting some convergence studies and then by comparing with available experimental data for specific ILs. The results that were obtained by applying OPLS-AA force field are compared with the results of GROMOS force field.
- Chapter 3: The same process as in AA model is repeated initially by using CG MARTINI 2.0 and afterwards CG MARTINI 3.0. Additionally, CG MARTINI 3.0 models are created and validated by experimental data for the $Rmim^+TFSI^+$ IL family.

- Chapter 4: An optimization is performed by using EASY to find an optimal IL, regarding the conductivity, that consist of a 3 bead cation and a 3 bead anion. The optimal CG pairs are then translated to the AA ones in order to be validated with more accurate AA models. The results and the process is discussed.
- Chapter 5: The process of producing input data for AA simulations from the optimization parameters is described. The results of the AA optimization are presented and then validated by more accurate AA models. New, previously unknown, ionic liquids with high conductivity that emerged from the process are presented and discussed.
- Chapter 6: The work is summarized and conclusions are drawn.

Chapter 2

AA simulations of ionic liquids

In this chapter, the structure of ionic conductivity calculations by using MD combined with an AA model will be explained in detail. For the rest of the thesis, the term AA model will mean that a NPAAFF is used, as explained in the paragraph 1.7.1. The simulation parameters are properly calibrated by performing convergence studies to an OPLS-AA model (see paragraph 1.7.3). The results from the previous study are then used as simulation parameters both in an OPLS-AA model and in a GROMOS model, which is a very similar force field to OPLS-AA. The models are used for 5 different ILs for which experimental data are available. The results from the two different force fields are compared and discussed.

2.1 The structure of a typical AA MD simulation

2.1.1 Input files and prerequisites of the simulation

In order to run a MD simulation in Gromacs, geometry, topology and option files are required. The geometry file, usually a .pdb file (protein data bank), contains all the necessary information about the positions and the connections of the atoms in a molecule. Each different molecule of the simulation must have its own geometry file. The most important file for each molecule is the .itp topology file. This file contains all the information about the molecular properties that are required for the simulation. Such properties are the partial charge and mass of each atom, the atom type, the force field atom type, the bonded and non-bonded constants described in section 1.7.1. The force field atom type is a translation of the real atom type in the vocabulary that it is used in the force field. For example, in the GROMOS force field, a CH₃ chemical group is often replaced by a single force field atom type to reduce the total number of atoms and thus the computational cost.

The topology files are difficult to be obtained, because they often require excellent knowledge of the using force field, experiments or ab initio simulations. Therefore, a trustworthy source for these files is necessary. For the OPLS—AA geometry and topology files the LigParGen web server [24], [25], [26] of Yale university was used. Regarding the topology and geometry files for the Gromos force field, they were obtained from the Automated Topology Builder (ATB) [27] by Australian Research Council, University of Queensland and Q.C.I.F. Firstly, a proper simulation box must be created (Initialization). For this purpose, PACKMOL software is used [28]. PACKMOL simply replicates the required number and types of molecules from the .pdb files in a box with dimensions specified by the user. The .pdb file of the box that was generated is then translated by a Gromacs command to a .gro file, which includes the same information as the previous file, but in the language that Gromacs understands.

Before any kind of simulation is started, a simulation parameter .mdp file must be created for each simulation part. For example, in the .mdp file regarding the minimization process, information exists about the method (steepest descent, conjugate gradient, etc), the integration step, the output frequency and other important simulation parameters. The different necessary files and processes before the launch of the simulation are shown in the figure 2.1.

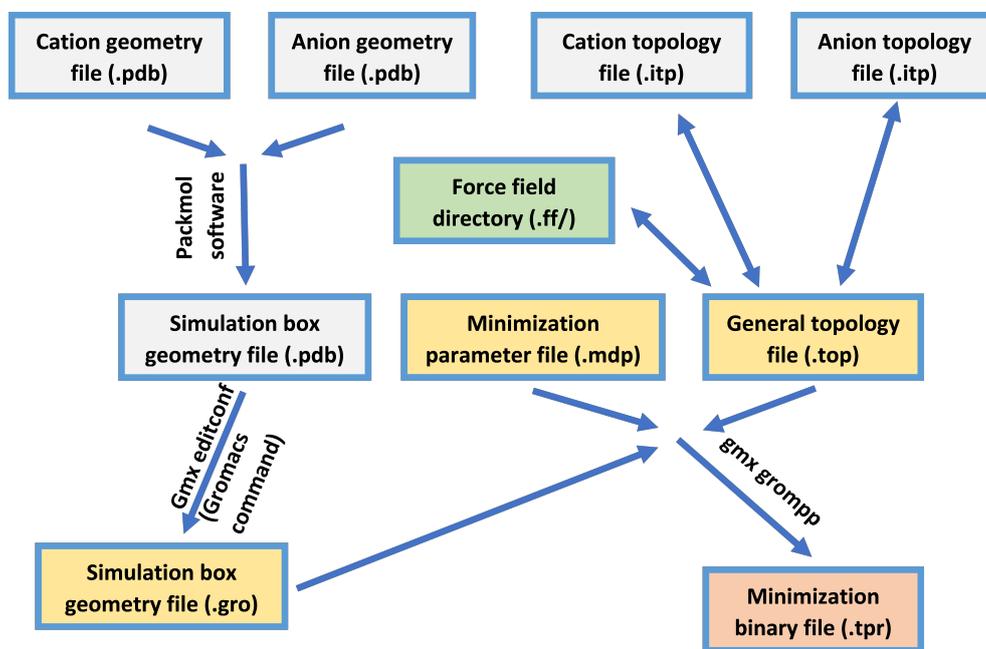


Figure 2.1: Flow diagram of the preprocessing procedure

2.1.2 The main section of the simulation

As mentioned in the section 1.1, in order to extract thermodynamic or transport properties, like density or conductivity, the energy values of the system and the trajectories of every particle at all the time moments are required. In this section, the process of acquiring the necessary information for such calculations is explained. This process is referred here as the main simulation process and it is necessary to be conducted, before any thermodynamic or transport properties are extracted from the MD simulation. The simulation process applied in this work includes four sections; the minimization, the 1st equilibration, the second equilibration and the main simulation section.

After using the Packmol software [28] to create the initial simulation box the molecules have initial positions and velocities which very rarely can be observed in nature, because of the high total energy of the system. In order to decrease the total energy of the molecules, a minimization process is required. The most popular minimization algorithms implemented in Gromacs are the steepest descent and the conjugate gradient methods. The first one is slower and more stable than the second one. In this work, the steepest descent algorithm is used as safer and more robust method. The options for the minimization (.mdp file) are shown in the following table. Some of these options should be changed according to the needs of each case. For example, bigger systems of molecules often require a bigger number of total steps to converge.

After the minimization has finished, the first equilibration run is launched. The purpose of this simulation stage is to couple properly the system temperature to the desired value. The temperature coupling is often achieved by using an artificial thermostat for the system. From the wide variety of thermostats, that Gromacs provides, only the Berendsen and Nose-Hoover thermostats are tested. After performing tests in different all atom cases, the Nose-Hoover thermostat had on average about 5 times faster convergence than the Berendsen thermostat. However, the last one had smaller fluctuations and thus it is considered to be more stable. In all the following AA simulations the faster Nose-Hoover thermostat is used.

After the temperature coupling process has converged successfully a pressure coupling is performed in the 2nd equilibration run. For this reason, an barostat is needed, as explained in the 7th paragraph of section 1.1. After performing tests, comparing the Parrinello-Rahman barostat with the Berendsen barostat, the second one is much slower (≈ 5 times slower) and it requires Berendsen thermostat in order to function properly. Therefore Parrinello-Rahman was considered a better option for running the simulations, although it results in bigger pressure fluctuations.

The final step is the main simulation run. This step has the same options as the 2nd equilibration step, but it can be smaller in duration, because the simulation box is already fully equilibrated. The duration of this run depends on the time required by the postprocessing methods to give accurate results. In theory, it could be avoided by applying the postprocessing methods at the end of the previous run.

However, a great amount of caution is needed to be sure that the 2nd equilibration run has converged before the postprocessing is started. In the figure 2.2 the simulation process flow chart is drawn. It is important to note that the postprocessing files shown inside the red ellipsis are printed out for every execution of the “gmxdrun” command, which is the terminal command of GROMACS for running a stage of the simulation. That means that the minimization, 1st equilibration and 2nd equilibration have postprocessing files too. In the table A.2 of the appendix, a typical list for the .mdp file options in AA systems is presented, regarding the 1st, 2nd and main simulation run.

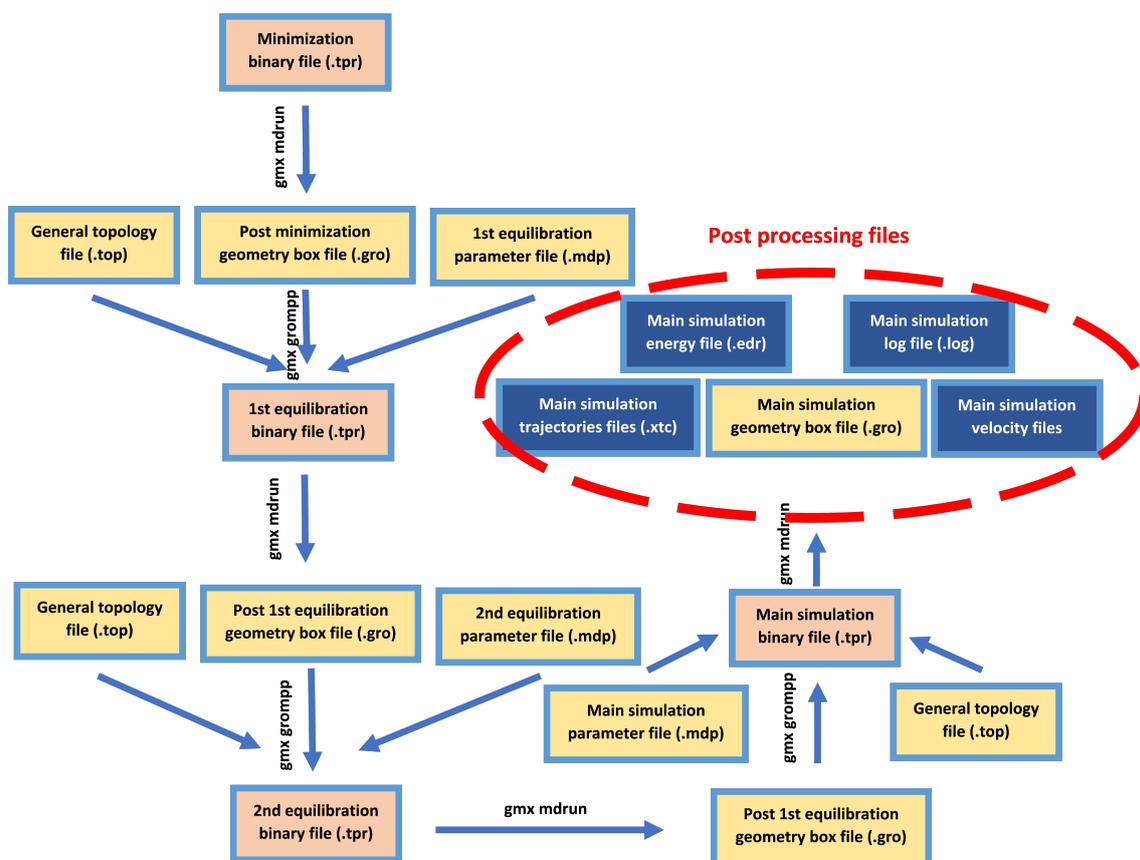


Figure 2.2: Flow diagram of the simulation procedure

2.1.3 Post-processing and Convergence

MD simulations are arithmetic processes and thus they need some kind of convergence verification, before extracting any results. For the minimization, the most important index for convergence is the potential of the system. The curve of the potential is usually close to a hyperbola. As the timesteps are increasing, the potential value should converge to a big negative value, for the minimization to be considered successful.

In the same way, the 1st equilibration process is considered successful, if the average value of temperature has converged to the desired value with relatively small fluctuations. Also, the mean value of the total energy must reach a minimized equilibrium value. The moving averaged value of this curve is similar to the one of the minimization.

For the 2nd equilibration process the most important indices for convergence are the average pressure and the total energy. The fluctuations in this simulation procedure are increased dramatically, because the Parrinello-Rahman barostat is used (see 7th paragraph of section 1.1). This is considered normal as long as the average value of the curve is remaining constant.

Finally, as mentioned before the main simulation run has the same parameters as the 2nd equilibration run. Thus, the convergence indices are the same, though they are not actually required, because the system has been already equilibrated from the 2nd equilibration run. Furthermore, it is important to check if the average density of the main simulation run is close to the experimental value of the simulated substance. In all the next chapters, a case will be considered converged, only if the average values of temperatures, pressure and density are close to the desired or the experimental values. This is a practical assumption needed in order to perform further simulations and extract results. In figures A.1, typical curves are shown, like the previously described.

All the thermodynamic properties of the ensemble are easily extracted by using the "gmx energy" command of GROMACS. The transport properties, such as ionic conductivity are more difficult to extract. In the bibliography, there are two methods for calculating the ionic conductivity. The first one is the MSD (Mean Square Displacement) method. In this case, the self diffusion coefficients are calculated from the mean-square displacement of the center of mass of each ion using the Einstein relation

$$D = \frac{1}{6} \lim_{t \rightarrow +\infty} \frac{d}{dt} \langle [\vec{r}_i(t) - \vec{r}_i(0)]^2 \rangle \quad (2.1)$$

where the quantity in $\langle \dots \rangle$ is the ensemble-averaged MSD of the center of mass of ion i over time interval t and $\vec{r}_i(t)$ is the location of the corresponding ion. The locations of the molecules are stored every 1 *ps* as suggested from the bibliography [18]. The slope of the MSD curve (time derivative) is approximated in every main simulation run by a linear regression.

The same coefficient was calculated also by using the Green-Kubo relation, in which the time integral of the velocity autocorrelation function (VACF)

$$D = \frac{1}{3N} \int_0^\infty \sum_{i=1}^N \langle \vec{v}_i(t) \cdot \vec{v}_i(0) \rangle dt \quad (2.2)$$

is used. Here N is the number of molecules for which the self diffusion constant is calculated. A great effort was made to implement both of methods in the calculation process. However, only the MSD method gave results with similar trend and values to the available experimental data. Therefore, all the ionic conductivity results are calculated by using the MSD method. It is important to note that Gromacs can calculate separately the diffusion constants for the cation and for the anion of the system, using different .ndx index files. After calculating the self diffusion constants of the ions, the ionic conductivity value can be calculated by applying the Nerst-Einstein equation

$$\sigma_{MSD} = \frac{N_i q^2}{V k_B T} (D_{MSD}^+ + D_{MSD}^-) \quad (2.3)$$

Here N_i is the total number of ion pairs, V is the volume, T is the temperature, q is the effective net charge on the ions, k_B is the Boltzmann constant, and D_{MSD}^+ and D_{MSD}^- are the self-diffusion coefficients of cations and anions, respectively. The procedures followed for the MSD and Green-Kubo calculations of is shown in figures 2.3 and A.2 respectively.

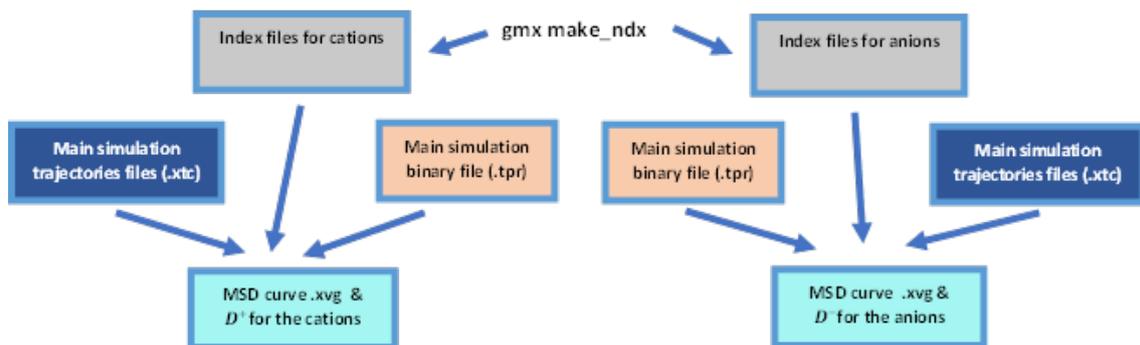


Figure 2.3: Flow diagram of the self diffusion constant calculation procedure, using the MSD method.

2.2 Convergence studies for the simulation parameters

2.2.1 Information about the simulated ionic liquid pairs

In the context of this thesis, 5 ionic liquid pairs will be examined for the ionic conductivity. The available experimental data will be used for the validation of the models created. For confidentiality reasons, the cations, anions and the respective IL will be referred as CAT_i^+ , AN_i^- , IL_i . In each IL pair the CAT_1^+ is always the

cation. The different anions used in each pair are AN_1^- , AN_2^- , AN_3^- , AN_4^- and AN_5^- . For confidentiality reasons, all the simulation results presented in this thesis will be normalized by dividing with the experimental values.

2.2.2 Determining the proper size for the simulation box

The size of the simulation box is an important parameter for the stability of the simulation. A very small box size can lead to very high interactions forces between the molecules, especially before the minimization is performed. Such high forces can lead to rapid changes in the kinetic state of some molecules and thus in molecules shooting across the system in an uncontrolled way in the next timestep. This kind of simulation failure is often called "Blowing Up".

On the other hand if the box is very big, the system will eventually reach a small equilibration volume, after oscillating for some time. However, using a very big box is computationally more expensive. The system will take more time to reach an equilibrium and the parallelization process will be probably less efficient. Various geometries exist for the simulation box, such as dodecahedron, octahedron, etc. In all the next simulations a cubic simulation box is used.

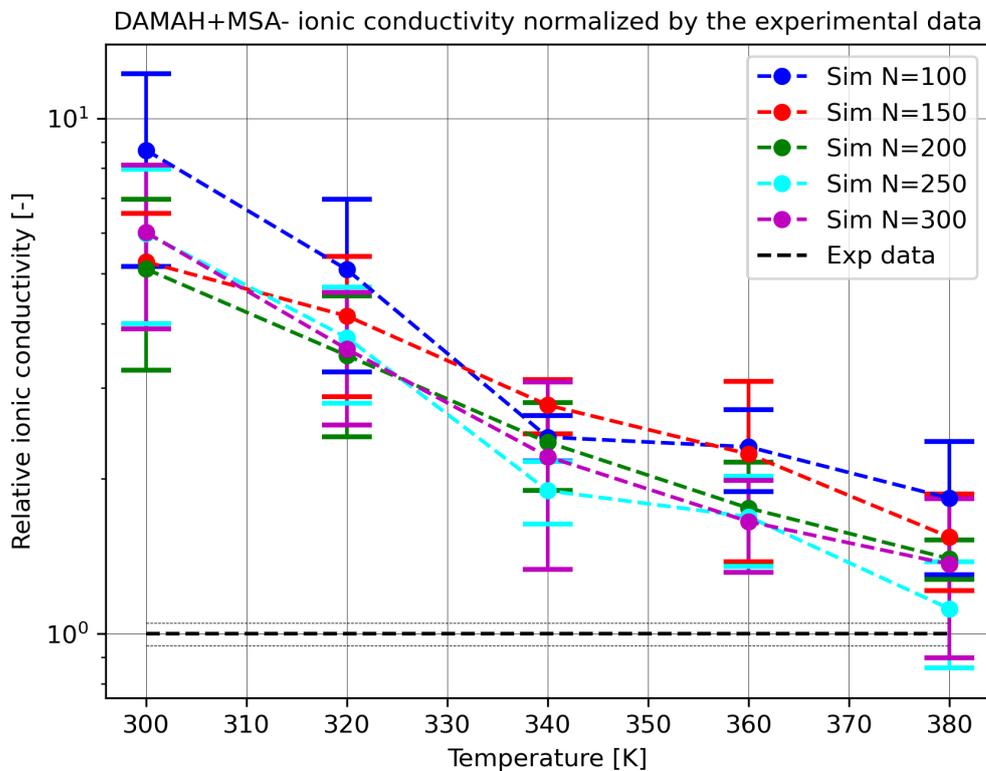
For determining the size of the simulation box, a test run is performed. The test system contains 150 molecules of CAT_1^+ and 150 molecules of AN_3^- , at 300 K and 1 bar. After the end of the simulation, the average volume is $V_{avg} \approx 47 \text{ nm}^3$. From the literature, the final average volume of a system is proportional to the number of molecules in it. Starting from the previous fact, the following empirical relation can be derived for the approximation of the final simulation edge size of the box.

$$a_f \approx \sqrt[3]{\frac{47}{300}N} \quad (2.4)$$

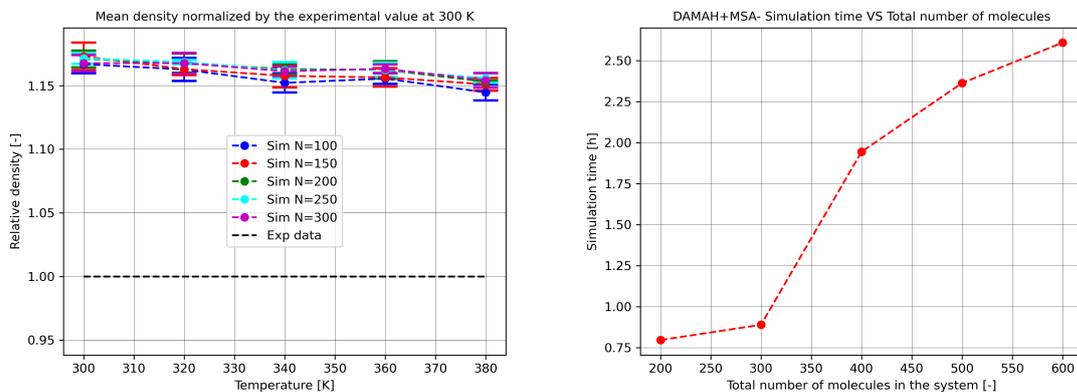
Where a_f is the final edge length of the cubic box in nm and N is the total number of molecules. The a_f parameter is very important for determining the cut-off distances for the Coulomb and Van der Waals interactions (see section 2.2.4). However, the previous formula gives the final edge size of the system and not the initial one. A safe, empirical option for avoiding "Blowing Up" systems is to use a simulation box at least 3 times bigger in volume than the final average volume.

2.2.3 Determining the total number of molecules in the system

For the determination of the total number of molecules in the system, a convergence study is performed. In each case all the parameters are the same except from the total number of molecules N . The $CAT_1^+AN_3^-$ ionic liquid pair is simulated at the temperature range of 300 K to 380 K and at pressure equal to 1 bar using the OPLS-AA force field. The duration of the 1st equilibration run is 2 ns, while the duration of the second one is 8 ns. For the ionic conductivity calculations, the MSD method has been applied. In figures 2.4 the results of the convergence study are shown.



(a) Ionic conductivity-temperature curves for different N values. All the simulation values have been normalized by the experimental data in the corresponding temperature.



(b) Mean density normalized by the experimental (c) Average computational time in hours for different density value at 300 K for different N values. ferent N values.

Figure 2.4: Results of the convergence study for different total number of molecules. The symbol N means the number of cations in the system for the first two figures and the total number of molecules in the last one. The computational resources used include 48 Intel Xeon logical processors and a Nvidia 2080 Ti at 100% and 30% utilization respectively (used in parallel)

The density values are close to the experimental ones and almost constant for different temperatures. Increasing the number of molecules seems to have no effect in density. The conductivity results are improving as long as the total number of molecules in the system is less or equal to 200 for each residue. After that point there is no significant improvement in the conductivity results. Although the $N = 200$ value seems the obvious choice for the simulations, the $N = 150$ case is selected to be used in all the next runs. By using $N = 150$ the computational cost is decreased by approximately 50%, as shown in the figure 2.7c. This is why the last decision was made.

2.2.4 Determining the cut-off distances for the non-bonded interactions

Ideally the non-bonded interactions are calculated between each atom pair in the system. However, the total number of atom pairs in a system is usually a very big number, hence the previous ideal method is computationally expensive. Furthermore, the intensity of Van der Waals and Coulomb forces is decreasing rapidly with the distance ($\propto r^2$ for the Coulomb, $\propto r^{13}$ for the Pauli repulsion and $\propto r^7$ for the dispersion attraction forces). Therefore, the non-bonded forces aren't usually calculated after a specific cut-off distance. For this reason, another study is performed to find the proper cut-off parameters for the simulations. The ionic liquid pair used is the $CAT_1^+ AN_3^-$ and it is assumed that the other IL pairs will follow a similar trend. All the simulation parameters are very similar to those presented in paragraph 2.1.2.

The R_{vdw} and R_{clb} values are changing in the range of 0.8 nm to 1.6 nm with step of 0.1 nm. The conductivity mean and standard deviation 2D maps are presented in the figure 2.5. To make easier the extraction of conclusions from these maps, a bicubic interpolation has been applied to the data. For the specific configuration of options used, it's impossible to have $R_{vdw} > R_{clb}$. That is the reason for not presenting the values of the upper-left section in each map.

From the previous colormaps, the mean density is about 20% closer on average to the experimental value around the pair $(R_{clb}, R_{vdw}) = (1.6, 1.3)$. Also, in this region of the plot, the standard deviation is almost 4 times lower than the average value of the colormap. As a result, if (R_{clb}, R_{vdw}) is close to this area, the total amount of simulations required to have good statistics about the case is reduced by approximately 4 times. Consequently, the $(R_{clb}, R_{vdw}) = (1.6, 1.4)$ combination is selected for the all the next simulations.

The simulation box is usually divided in smaller unit cells. Each cell is a mirror of one master cell. The purpose is to simulate bigger systems without increasing dramatically the computational cost. This is a set of boundary conditions and their basic idea is explained in figure 2.6. As mentioned in the section 2.2.2, using cut-off distances bigger than the half of the final box edge size isn't possible. This is

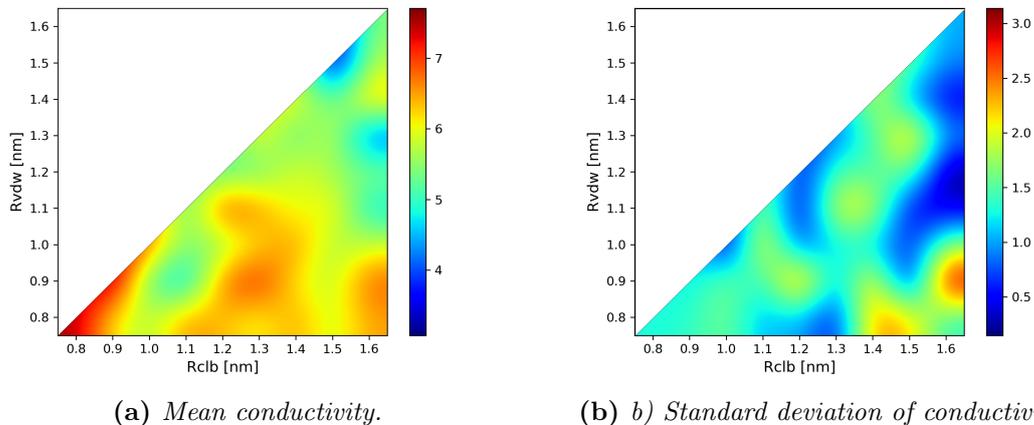


Figure 2.5: Maps of conductivity for the $CAT_1^+AN_3^-$ pair ($N = 300$). All the results have been normalized by the experimental value at 300 K. For each (R_{clb}, R_{vdw}) combination, 5 cases were ran to obtain meaningful statistics.

happening, because of the periodic boundary conditions applied in the simulation box. If that was possible, the molecule could interact with itself resulting to a totally wrong simulation. Consequently, using cut-off distances bigger than 1.6 would require increasing the number of the molecules which will be further increase the computational cost. This is why, the previous option will be avoided in the simulations following.

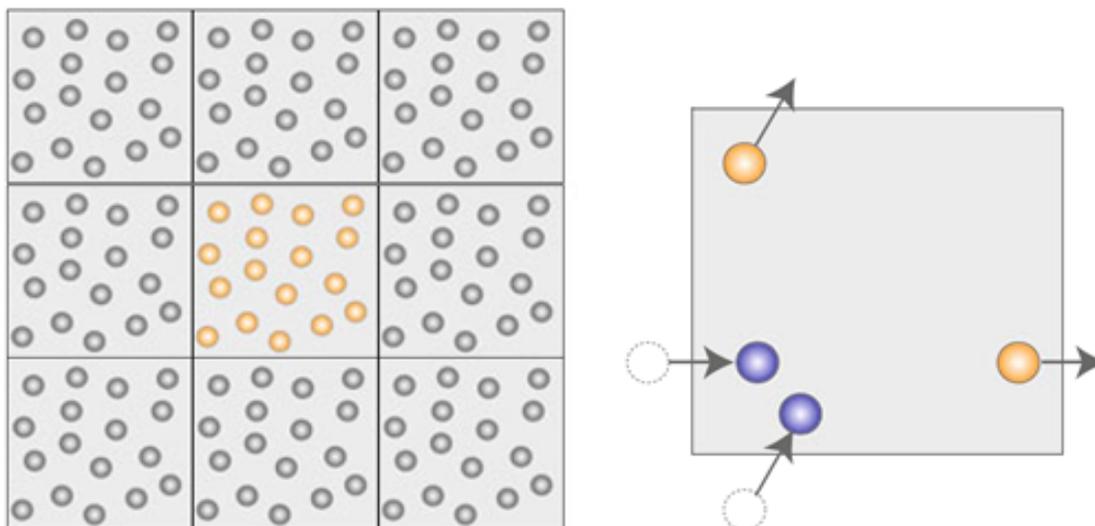


Figure 2.6: Schematic representation of the idea of periodic boundary conditions [9].

2.2.5 Determining the proper frequency and time duration for the MSD method

Two more important parameters for the calculation of the ionic conductivity using the MSD method are the total sampling time as well as the total amount of data points used for the MSD fitting. To find proper values for these parameters, another study is performed for the $CAT_1^+AN_3^-$ pair as previously done. The cut-off distances are equal to those previously selected, while the rest of the options are still the same as in section 2.2.4. In figure 2.7, the 2D maps for mean conductivity, conductivity standard deviation and total simulation time are shown.

Using the previous maps an optimal set of parameters can be selected. From the first two maps it is clear that the simulation time (main simulation time) required for obtaining good results is at least 500 ps. At the same time, a decent percentage of computational cost can be saved if a proper combination is selected. Consequently, the combination $(N_d, t) = (1000, 500)$ is considered a good compromise between accuracy and cost and will be used in all the next simulations.

2.3 Results and discussion about the AA MD simulations

After applying the options discussed in the previous sections, AA simulations are performed in order to find the temperature curves and compare with the available experimental results. For all the ionic liquids described in the section 2.2.1, the Gromos57A7 force field topology is applied. For comparison reasons the OPLS-AA force field is applied for the $CAT_1^+AN_3^-$ and $CAT_1^+AN_4^-$ pairs. It is worth to note that, the OPLS-AA force field topology files were available only for these two ionic liquids.

Each pair has been simulated in 5 different temperatures in the range of 300 K to 380 K. Each data point shown in the following plots have been acquired by running 5 times each case with different RNG number each time to obtain meaningful statistics. All the topology files are including all the atoms, even the hydrogen atoms, except from the Gromos topology file for CAT_1^+ . In this file all the apolar hydrogen atoms in the molecule have been merged with the corresponding carbon atom. The Gromos force field is able to cope with this new merged atomtype and produce meaningful results. On the other hand the OPLS-AA models include all the atoms. Therefore, a comparison between the OPLS-AA model and the Gromos model is important. The density-temperature curves have been drawn in the figure 2.8a, while the conductivity-temperature curves have been drawn in the figure 2.8b.

All the density curves are in a $\pm 15\%$ area around the experimental value. This

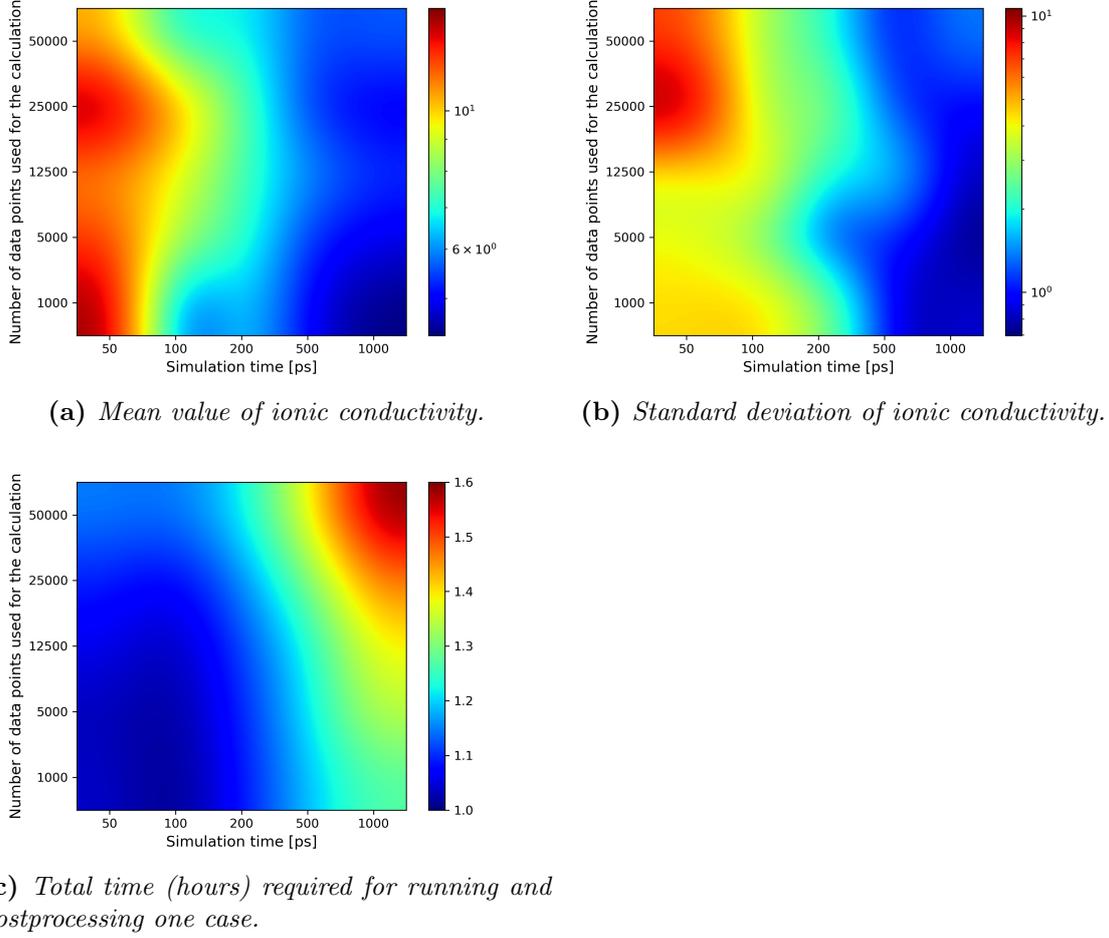
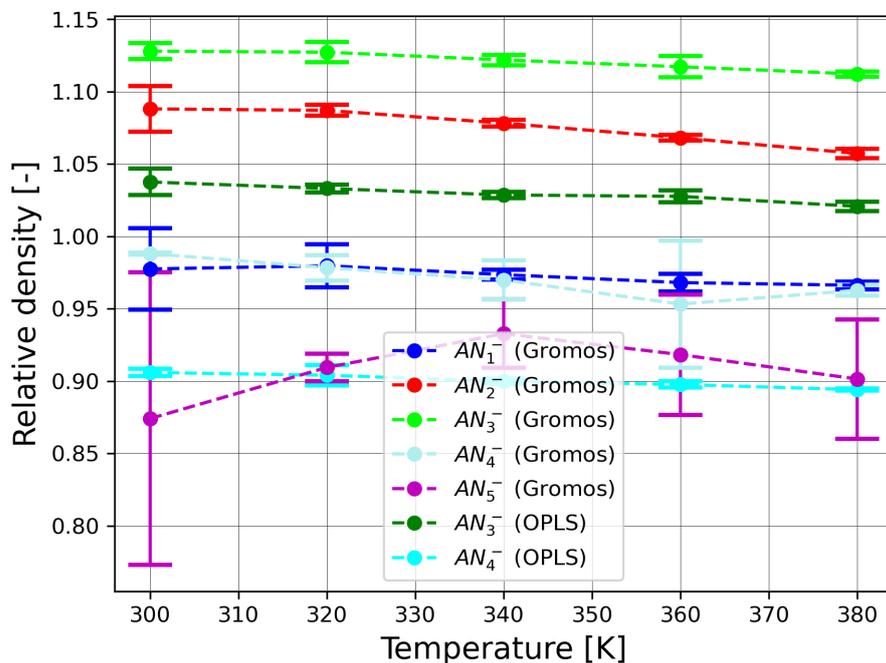


Figure 2.7: Maps of conductivity for the $CAT_1^+AN_3^-$ pair ($N_{total} = 300$) and total simulation time. All the conductivity results have been normalized by the experimental value at 300K. For each parameter combination, 5 cases were ran to obtain meaningful statistics. The total simulation time includes the minimization, the 1st and the 2nd equilibration, the main simulation (noted as “Simulation time” in the plots) and the postprocessing time. The computational resources used for the simulation part include 48 Intel Xeon logical processors and a Nvidia 2080 Ti at 100% and 30% utilization respectively (used in parallel). For the postprocessing part 1 Intel Xeon logical processor at 2.6 GHz was used (Parallelization wasn’t possible).

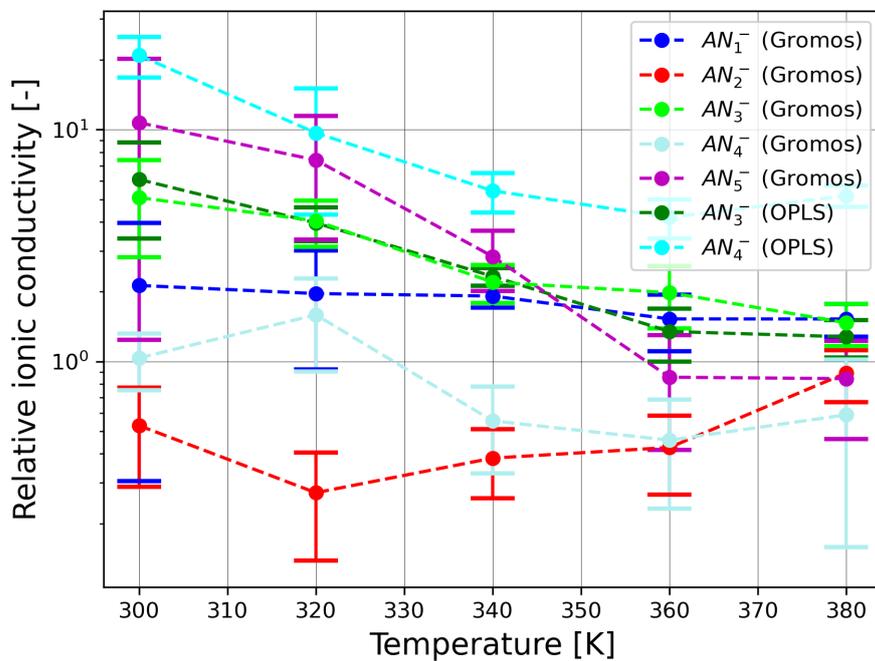
is a good indication that the model is capable of predicting some basic physical properties of the real model. For almost every ionic system, the density is slightly decreasing with temperature, as it is observed in real ionic liquid systems. However, the $CAT_1^+AN_5^-$ pair has a different behaviour until 340 K. In the majority of the density data points, the standard deviation is very small, indicating that these cases have totally converged. That means that both the temperature and the pressure coupling were successful.

The shape of the conductivity-temperature curve depends strongly on the ionic liquid pair. Although, most of the pairs have overestimated simulation values, the $CAT_1^+AN_1^-$ (Gromos) and $CAT_1^+AN_4^-$ (Gromos) have underestimated conductivity results. Generally in literature, getting results of the same order of magnitude for the ionic conductivity using NPAAFF is considered a rare and difficult task [1]. For that reason, all the models except from those regarding $CAT_1^+AN_4^-$ (OPLS) and $CAT_1^+AN_5^-$ (Gromos) are considered successful. Furthermore, as the temperature increases, the conductivity simulation values converge more and more to the experimental ones. The previous fact is true for all the models.

The $CAT_1^+AN_3^-$ (OPLS) curve is very close to the one of $CAT_1^+AN_3^-$ (Gromos). The previous fact probably verifies that the assumption about the apolar hydrogen atoms in the Gromos force field is correct. Therefore, the CAT_1^+ topology file used for all the Gromos models is considered safe to be applied and not very different from the one of the OPLS-AA. On the other hand the Gromos force field gives much more accurate results than the OPLS-AA in the case of the $CAT_1^+AN_4^-$ ionic liquid.



(a) Density-temperature curves for the 5 ionic liquids.



(b) Ionic conductivity-temperature curves for the 5 ionic liquids.

Figure 2.8: Results of all atoms simulations. The conductivity results have been normalized by the corresponding, available, experimental value at each temperature. The density results have been normalized by the experimental density value at 300 K.

Chapter 3

Coarse grained simulations of ionic liquids

The purpose of this chapter is to create and validate a robust CG model for performing conductivity calculations by using the CG MARTINI force field. Of course, the results of CG calculations should have lower accuracy than the AA calculations of chapter 2. However, they could be used in an optimization process if the conductivity values for different ILs have the correct trend. As mentioned before, CG models are significantly cheaper than AA models. Therefore, it is important to examine the possibility of using CG models for an optimization process.

First of all, a simplified CG model is created by using MARTINI 2.0 for an IL for which experimental data are available. Then, a parametric study regarding the bonded terms is performed to see the behavior and flexibility of the model. Afterwards the simulation parameters are modified to better fit the CG model conditions and the force field is changed from MARTINI 2.0 to MARTINI 3.0, because of the limitations presented in the previous model. The new model is applied for the five available ILs presented in the section 2.2.1 and then for the $Rmim^+AN_1^-$ IL family. The results are discussed and conclusions are drawn about using the previous CG model for an optimization.

3.1 Coarse grained models based on Martini 2.0

3.1.1 A simplified coarse grained model

Mapping separately the cation and the anion is a necessary process to create the IL coarse grained model. The different groups of atoms in each molecule should be approximated by the proper bead type of the force field. In this process the shape, charge and molecular weight distribution should be as close as possible to the corresponding property of the real molecule. In Martini 2.0, only 4 to 1 mapping is possible, except when ring configurations need to be approximated. For this simple test case, only 4 to 1 mapping will be used, even if the model could be more accurate by using ring specialized bead types. The target of this paragraph is to extract qualitative results from this simplified reference model. The mapping process and the respective topological configuration of the beads are shown for each ion in figure 3.1.



Figure 3.1: *Simplified coarse grained model for $CAT_1^+ AN_3^-$ IL pair.*

CAT_1^+ is modelled by using 3 beads and the charge is located in the central Q_d bead, while AN_3^- is modelled by using only 1 Q_a charged bead. The capital letter is denoting the type of the bead, the number is denoting the interaction intensity and the letters “a” or “d” that follow are denoting the hydrogen bonding capabilities (see paragraph 1.10.2 for more information about the beads). Every bead used, has a molecular weight of 72 amu . This is the equivalent weight of 4 water molecules. Therefore, the molecular weight of each group differs significantly from the real one. The same is also true for the distribution of molecular weight in CAT_1^+ ion. In table 3.1 the exact molecular weight for each group of atoms is compared with the corresponding molecular weight of the bead used in the model.

Group/Bead	Exact Molecular Weight	Bead Molecular Weight	Relative Difference
$C1$	27 amu	72 amu	+166 %
Q_d	55 amu	72 amu	+30 %
Q_a	95 amu	72 amu	-24 %

Table 3.1: *Molecular weight differences between the real and the modelled sites*

3.1.2 Parametric studies regarding the bonded terms

Based on the previous model, a parametric study was performed to determine if any correlation exists between the bonded parameters and the physical properties of the ionic liquid pair. The bonded parameters examined were the angle and the bond length value of the CAT_1^+ model. Because of the symmetry, the length for each bond is the same. In addition, different values for bond and angle spring constants were tested, as shown in table 3.2. Then, the average of the different values of table 3.2 was calculated for each (L, θ) combination. The simulation parameters were the same with those of the AA model described in paragraph 2.1.2 and in table A.2.

Bond spring constant values $\left[\frac{kJ}{mol\ nm^2}\right]$	10^3	$5 \cdot 10^3$	$2 \cdot 10^4$	$4 \cdot 10^4$
Angle spring constant values $\left[\frac{kJ}{mol\ rad^2}\right]$	5	25	10^2	$4 \cdot 10^2$

Table 3.2: Spring constant values tested in the parametric study.

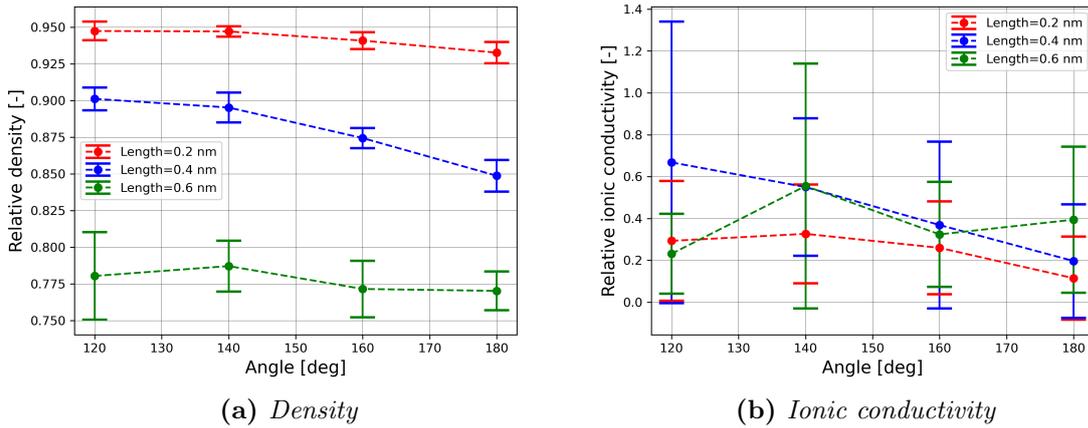


Figure 3.2: Physical properties of a simple $CAT_1^+ AN_3^-$ CG model for different bond length L and angle θ values. All the results have been normalized by the corresponding experimental value.

Density is affected mainly by the bond length and less by the angle value. Increasing the bond length leads to a lower density value. This was expected, since bigger distances between the beads mean that there is more empty space between them. Also, configurations with angle value closer to 180° have slightly lower density. An ion with an equilibrium angle equal to 180° is considered linear. It seems that linear molecules can be packed slightly more easily than molecules with equilibrium angles closer to 90° .

The correlation between ionic conductivity and angle or length isn't so clear, because of the high standard deviation in the results. Increasing the bond length seems to lead to higher conductivity, but only for lengths below $0.4\ nm$. A similar trend is observed when the angle value is closer to 180° .

3.1.3 Problems and restrictions of the model

Every bead of the force field is a combination of 3 parameters. The first one is the molecular weight of the bead. The rest are used to determine the shape of the Lennard-Jones potential curve (σ and ϵ). These affects the non-bonded behaviour of the molecule. The previous parameters are usually obtained by applying a fitting method to data about physical properties for specific substances. For example, P_4 bead type has been created specifically to imitate the properties of 4 water molecules. Consequently, higher variety in the force field bead types can lead to higher flexibility in modelling different molecules.

A model like the one described in paragraph 3.1.1 is barely adequate to describe the qualitative effect of the bonded parameters to the conductivity. The molecular weight of a bead can either be 72 amu or 45 amu , if ring bead types are used. In the original paper of the force field [22], isn't clear if ring bead types can be used for modelling non ring groups of atoms. The previous facts make difficult to create a model with molecular weight distribution close to the one of the real substance. Therefore, the resolution of Martini 2.0 force field doesn't seem to be adequate for describing properly an ion like CAT_1^+ .

One last problem of the previous configuration is that the parameters of the simulation like the integration step, the cut-off schemes and the thermostat haven't been changed to fit the coarse grained model philosophy. At the same time, the ionic conductivity in these models doesn't scale up enough when the temperature is increased. This isn't in agreement with the real phenomenon, since the ionic conductivity can often be increased by 10 times in the temperature range which is examined.

3.2 An improved coarse grained model

3.2.1 Modifying the simulation parameter files

As mentioned before, the simulation parameters used previously were adapted to the AA model philosophy. In order to improve the accuracy of the coarse grained model, these parameters were changed to be better adapted to the CG philosophy. The modifications applied were suggested by relative papers about the Martini 2.0 and 3.0 force fields [22], [29], [30], [31].

The first and most important change was in the integration timestep which changed from 2 fs to 10 fs . This modification alone leads to 500% reduction in the simulation cost. The cost is lower, because the total number of steps required is inversely proportional to the timestep value. Furthermore, the inertia of the moving particles is greater in a CG model, thus the time step required for achieving the same level of accuracy as before is usually bigger. The postprocessing phase will be faster as

well, since less data about trajectories need to be processed for calculating the ionic conductivity.

The method for calculating the Coulomb interactions was changed from “PME” to “Reaction Field”. An extra parameter called “Verlet-Buffer-Tolerance” was introduced, regarding the convergence of the simulation. The constraints about the hydrogen atom bonds were removed, since no hydrogen atoms are present in a coarse grained model. Moreover, $\epsilon_r = 15$ value described in [22] was added in the parameter files. The thermostat was changed from “Nose-Hoover” to the more stable “V-rescale” one. Lastly, the relaxation value for pressure was increased significantly to 10 *ps* from 3 *ps* to ensure stable convergence. The simulation parameters are shown in greater detail in tables B.1, B.2 of the appendix.

3.2.2 The Martini 3.0 force field

Martini 3.0 is an evolution of Martini 2.0. Its purpose is to offer higher flexibility in modelling materials while improving the model accuracy. Ionic liquids, (poly)aromatic rings and organic solvents have been added to the list of feasible models. The respective table of interactions (see table 1.2) has been broadened significantly as shown in table B.3. Moreover, a higher variety of bead sizes is provided as shown in table 3.3.

Size name	Mapping capabilities	Molecular weight	Description
N	4-1 mapping	72 <i>amu</i>	The writing of this is omitted
S	3-1 mapping	54 <i>amu</i>	-
T	2-1 mapping	36 <i>amu</i>	-

Table 3.3: *Bead size types in the Martini 3.0 force field.*

Martini 3.0 provides a solid algorithm for parameterizing new molecules. The flow diagram of this algorithm is shown in figure B.1. Further more table B.4 can be used both for mapping and for inverse-mapping purposes. This is very important especially if a coarse grained model is used for optimization. In the end, it will be necessary for the optimal solution to be translated in a real world molecule. Finally, some relative papers have been published, regarding the modelling process of small molecules and ionic liquids [30], [29]. Thus, the available amount of documentation about ionic liquid like models is more plentiful for this force field. For this reason, all the next CG models will be developed based on Martini 3.0.

3.2.3 Mapping process for the $CAT_1^+ AN_i^-$ pairs

In order to mimic as best as possible, the physical properties of the $CAT_1^+ AN_i^-$ pairs, the cation and the anion will be modelled separately by using the density of the respective neutral molecule as reference guide for the design. The bonded parameters

were obtained from the AA topology files by assuming that each group of atoms is represented by the center of mass of the corresponding bead. After performing some initial tests, an important problem considering the molecular weight was emerged.

During the modelling of the AN_3^- respective acid (AN_3^- with an extra hydrogen atom), it was impossible to achieve density higher than $\approx 1200 \text{ kg/m}^3$. The model was including only one normal size bead. The actual molecular weight of the residue is 96 amu , thus the closest bead size in terms of molecular weight to use is the N size type that weights 72 amu . Alternative configurations were rejected, because they lacked in physical meaning. However, the actual density of the residue is close to 1480 kg/m^3 . Therefore, a more accurate model is possible only if the molecular weight of the bead is changed to the real one. Indeed, changing just this simple parameter in a $P5$ bead was enough to achieve $\rho = 1460 \pm 21 \text{ kg/m}^3$. This could be considered a good model, since the substance is very polar as well.

Proceeding to the modelling of AN_4^- and AN_5^- acids (AN_4^- and AN_5^- with an extra hydrogen atom), a similar approach was applied. However, the melting point for these substances is very close to 300 K . These model types often fail to converge in conditions like the previous ones. There is also an upper limit to the density that can be achieved by using N size beads ($\rho_{max} \approx 1540 \text{ kg/m}^3$ by using a single $P5$ bead). These problems can be bypassed by using S bead size, while maintaining the correct value for the molecular weight. The reason for using a smaller bead is that decreasing the size of the bead also decreases σ value in the Lennard-Jones potential (see figure 1.5). Consequently, the repulsive forces between the beads will become more significant in shorter distances than before, permitting to them to be more densely packed in the simulation box. Finally, $SP3$ and $SP5$ beads were selected for modelling AN_4^- and AN_5^- acid respectively.

The parent residue of CAT_i^+ was approximated by using 3 beads. In the center, $TC6$ bead was used. The size T was selected because of its proximity in terms of molecular weight and the type was selected based on table B.4. The side beads should be $SC3$ by following table B.4. The density of the last configuration is around 1200 kg/m^3 which is very different from the experimental density $\rho_{exp} = 790 \text{ kg/m}^3$. Such difference couldn't be ignored, hence the same approach used in sulphuric and phosphoric acid was followed. By converting the side bead to $C3$ from $SC3$ the acceptable density of $\rho = 860 \text{ kg/m}^3$ was achieved.

Regarding the parent residue of AN_2^- , table B.4 doesn't have any useful information. Therefore, $P3$ and $C2$ beads were used as shown in figure ??, based on empirical data and on a brief manual optimization. The molecular weight of each bead was again modified based on the AA model. A similar process, was applied for the parent residue of AN_1^- . A comparison between the experimental and the CG model density is shown in table 3.4 for each molecule. It is important to note, that in every ionic liquid binary system, each ion has different bead types compared to its pair. This was done to secure that no conflict will be occurred between the molecular weight value used in a cation bead with the respective value in another anion bead of the

same type.

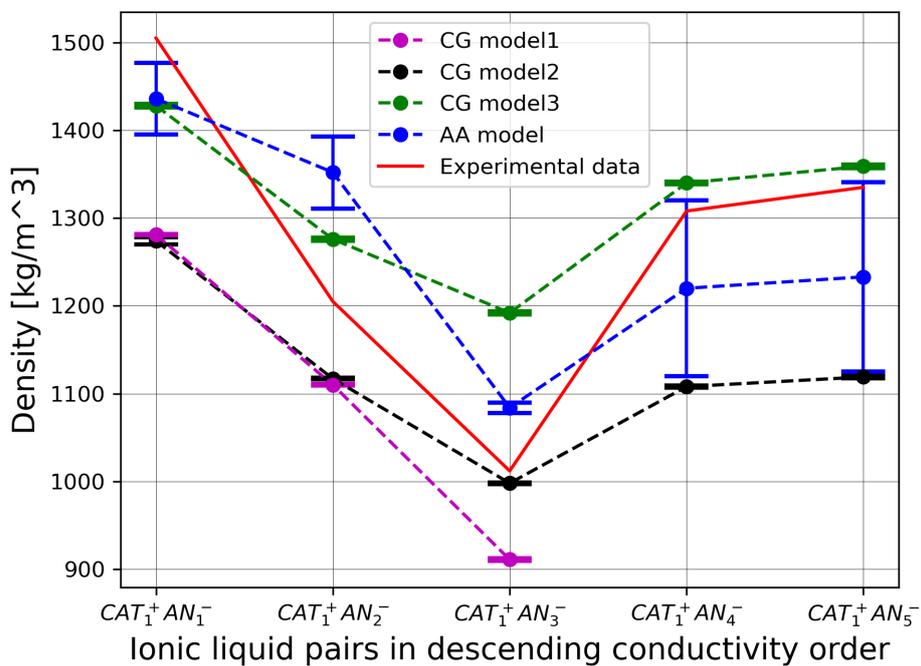
Residue	CAT_i^+	AN_1^-	AN_2^+	AN_3^+	AN_4^+	AN_5^-
$\rho_{exp} [kg/m^3]$	790	1940	1700	1480	1830	1880
$\rho_{CG} [kg/m^3]$	860	1790	1760	1460	2010	2170
ϵ_{rel}	+8.9%	-7.7%	+3.5%	-1.4%	+9.8%	+15.4%

Table 3.4: Comparison between the experimental and CG simulation density for each neutral residue. The experimental data were obtained from the online database [54].

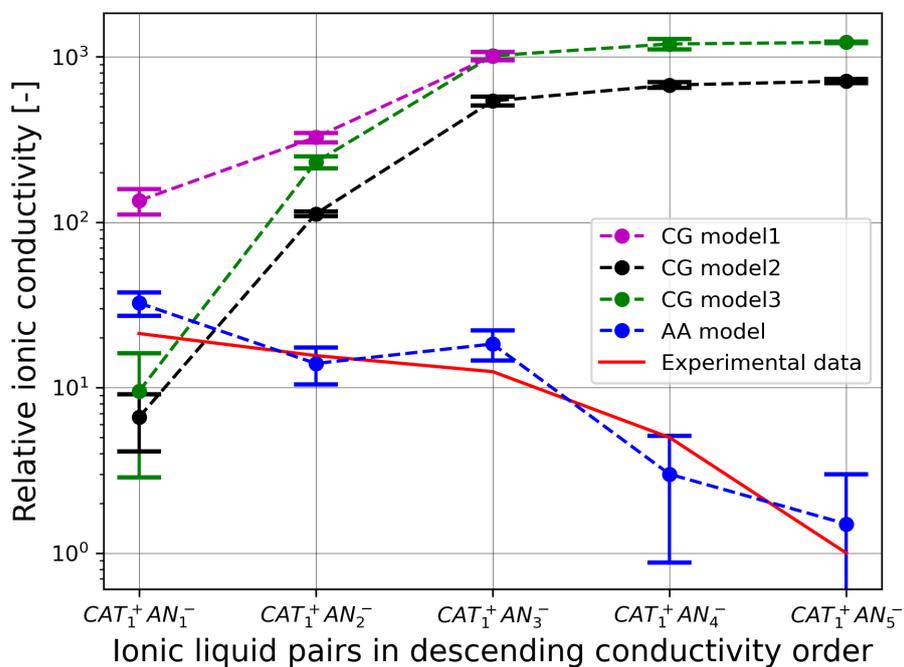
3.2.4 Results for the $CAT_1^+ AN_i^-$ ionic liquid pairs

The IL pair models were created based on the models previously described by assigning the proper partial charge in each bead. This was done by summing all the partial charges for each group of atoms and for each ion. The partial charge value for each atom was obtained by the respective AA model. After running the simulations, the 1st CG model curve was obtained, as it is shown in figures 3.3. The pairs $CAT_1^+ AN_4^-$ and $CAT_1^+ AN_5^-$ didn't converge. The density of the mixture follows the correct trend for different ILs. However, the conductivity results follow exactly the opposite trend compared to the experimental data. Furthermore, assigning the charges only to the protonated/deprotonated bead had almost no effect on the ionic conductivity curve.

For this reason, a second version of the previous model was created, in which the protonated/deprotonated beads were replaced by charged bead types. The approach for these replacements was similar to the one used for the 1st model. A list with all the ions of the second model is shown in table 3.5. The relative interaction intensity between the replacing beads is similar as previously. However, the interactions are stronger and thus an increase in density and a decrease in conductivity is expected. Indeed CG model curve 2 is behaving as expected, but the model is once more unable to predict correctly the trend in conductivity. On the other hand the density trend is in perfect agreement with the one of the experimental data.



(a) Density.



(b) Ionic conductivity normalized by the experimental value of $CAT_1^+ AN_5^-$.

Figure 3.3: Trend of physical properties for the $CAT_1^+ AN_i^-$ pairs. The pairs are listed in descending order regarding conductivity. The results have been obtained from simulations and experiments at 380 K, except from the experimental density which was measured in 300 K.

An alternative model could have everything being the same as before, but with the C3 side beads of CAT_1^+ replaced by SC3 beads. In terms of molecular weight this is the most accurate bead for describing this molecule. The reason for not using this bead type in the CG model 1 have been explained in paragraph 3.2.3. The 3rd model is in good agreement with the trend of the density curve, while is slightly worse, regarding the conductivity, than CG model curve 2.

During the previous modelling process a lot of different combinations were tested manually for maximizing the proximity of the density and conductivity curve to the experimental ones. From these tests, many qualitative conclusions about the bonded parameters of martini 3.0 were drawn as shown in table 3.6.

It is important to observe that not even the AA model curve of figures 3.3 is able to predict correctly the trend of conductivity between $CAT_1^+AN_2^-$ and $CAT_1^+AN_3^-$. These two substances have similar conductivities, thus the resolution of the model isn't adequate for distinguishing which of them has the biggest value. From table B.4, one can see that Martini has been built mainly by using carbon based substances. These chemical groups contain other elements like O, S, P, F and N but they aren't rich in them like the previous ionic liquids. Therefore, it is probable that more accurate models could be created for ions based mainly on carbon atoms.

CG model	CAT_1^+	AN_1^-	AN_2^-	AN_3^-	AN_4^-	AN_5^-
1st	C3-TC6-C3	C2-SN1-TC5-SN1-C2	P3-C2	P5	SP5	SP3
2nd	C3-TQ _p -C3	C2-SN1-TQ _n -SN1-C2	Q _n -C2	Q _n	SQ _n	SQ1
3rd	SC3-TQ _p -SC3	C2-SN1-TQ _n -SN1-C2	Q _n -C2	Q _n	SQ _n	SQ1

Table 3.5: List of all the CG models that are presented in figure 3.3.

Most affected parameter	Type of change	Example	Effect on density
σ [nm]	Bead size increases	C1 to T1	+3
ϵ [kJ/mol]	Interactions intensity increase	N1 to N3	+1
Molecular weight [amu]	increasing the M_w	72 amu to 95 amu	+2

Table 3.6: Qualitative effect on the IL density for changing different model parameters.

3.3 Simulation of the $Rmim^+AN_1^-$ ionic liquid pairs

3.3.1 Mapping process for the $Rmim^+AN_1^-$ pairs

As mentioned in the previous paragraph, the $CAT_1^+AN_i^-$ pairs were particularly difficult to model. Replacing the cation by a substance more close to the Martini 3.0 philosophy and by maintaining the same anion for the different pairs, better trend results should be expected. To validate the previous claim, two different CG models of $Rmim^+AN_1^-$ pairs will be simulated. The results will then be compared with experimental data obtained by Tokuda et al. [33].

Contrary to the CAT_1^+ cation, $Rmim^+$ cations can be very easily coarse grained by following figure B.1 and table B.4. In addition, Marrink et al. [29] have already proposed CG models for $C2mim^+$, $C4mim^+$ and $C6mim^+$ by using Martini 3.0. The CG models of figure 3.4 were created by applying the principles described in the previous sources. The angles and bond lengths were estimated as previously, from the respective all atom topology files. The spring constants were set equal to $2 \cdot 10^4 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ for the bonds and $25 \text{ kJ mol}^{-1} \text{ rad}^{-2}$ for the angles as proposed in [22]. Lastly, the charge has been distributed equally between the two beads that contain nitrogen atoms as suggested in [29].

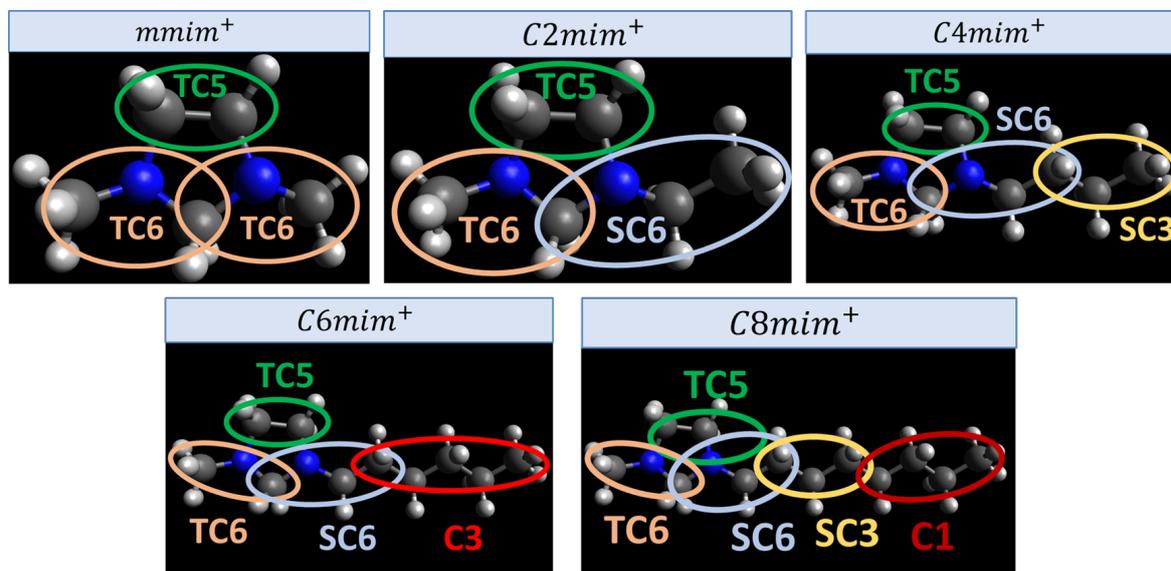


Figure 3.4: Visualization of the $Rmim^+$ CG models by using the Avogadro software [32].

Regarding the molecular weight of the cations, Marrink et al. didn't modify the molecular weight of any bead while modelling any of the imidazolium-based cations. To examine the importance of molecular weight, a second model will be created, in which each bead will have exactly the same weight as the respective group of atoms have in the real ion.

3.3.2 Results for the $Rmim^+AN_1^-$ ionic liquid pairs

In figures 3.5, the density and the conductivity values for each model are compared with the experimental data [33]. By looking in figure 3.5a, it is clear that molecular weight has a great impact on the accuracy of the calculated density. The trend of the first model curve is partially correct, while the second one is in good agreement with the experimental results. The 2nd model is able to correctly predict the trend and the value of density, since the maximum density difference compared to the experimental value is smaller than 10 %.

Regarding the conductivity, the 2nd model is again more accurate. The CG models have on average an offset of about 150% from the experimental values. The offset is greater for cations with higher molecular weight values. However, the trend in these models is more accurate than the trend that the lighter cation models have. This is considered to be a result of the simpler chemical configurations that these ions have. Alkyl tails are very simple chemical groups that Martini is able to model with great accuracy, in contrary to rings that have been added recently as a modelling capability in the force field.

As explained in paragraph 3.2.4, the model for AN_1^- isn't very accurate. This probably has an impact to the accuracy of the $Rmim^+AN_1^-$ CG models as well. Replacing AN_1^- with another Martini friendly anion could further improve the predictive ability of the CG model. Nevertheless, the second coarse grained model was able to capture the effect that alkyl chain length has to the ionic conductivity. Thus, Martini 3.0 seems to have some basic prediction capabilities about the ionic conductivity, provided that the modelled substances have similar chemical structure with the bead types of the force field used for the mapping process. This important assumption will be the basis for the conductivity optimization that will be performed in the next chapter.

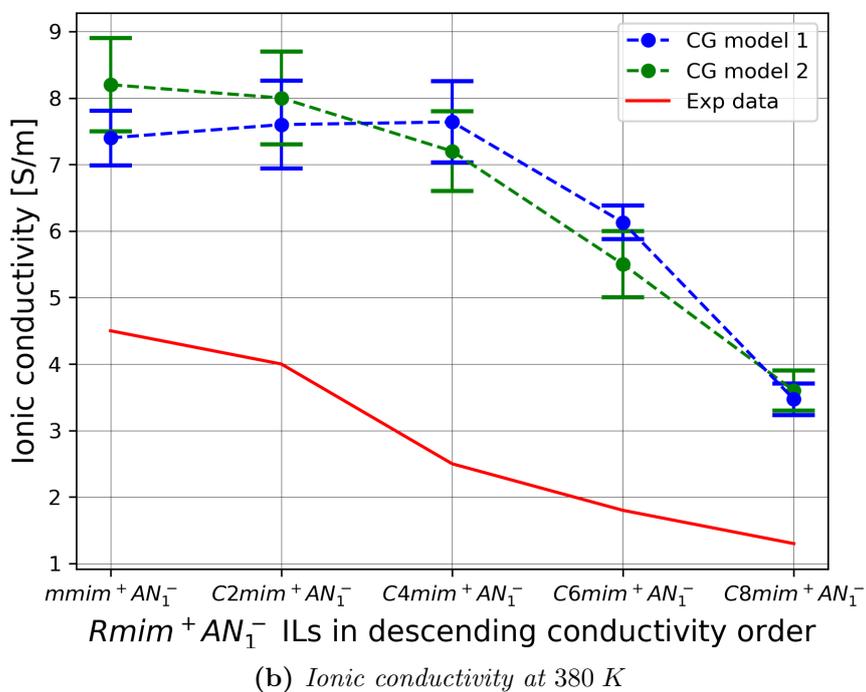
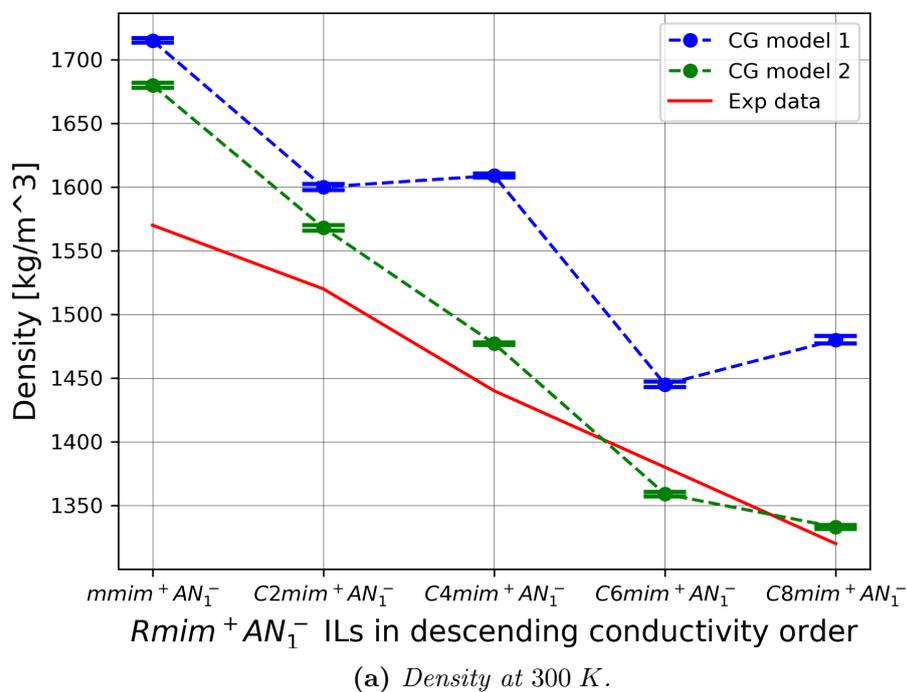


Figure 3.5: Trend of physical properties for the $Rmim^+AN_1^-$ pairs. The pairs are listed in descending order regarding conductivity.

Chapter 4

Optimization by using CG models

After the creation and validation of the CG models, in the previous chapter, a CG optimization is performed by using the evolutionary algorithm based software EASY. The basic principles of evolutionary algorithms and the process of translating CG ions to design variables will be explained. The optimization results are discussed and compared with the respective AA models.

4.1 Evolutionary Algorithms

Optimization methods can be split into two main categories, stochastic methods and deterministic or gradient based methods [49]. In the first one, statistics are used in combination with random processes to search for the optimal solution. Usually, they don't require any extra information about the problem besides the value of the objective function. In the contrary, gradient based methods always require the derivatives of the objective function with respect to the design variables.

The computation of the previous derivatives is often a tedious process and requires modification of the evaluation tool. In this case the evaluation tool is the MD simulation procedure for calculating the ionic conductivity (objective function), that was described in section 2.1. The development time for these methods is usually long. Moreover, sophisticated methods, like the Adjoint method [50], should be applied in order to keep the additional computational cost low. If programmed properly, deterministic methods converge fast to a local or global minimum with the proper initialization.

Evolutionary algorithms (EAs) are one of the main representatives of the stochastic optimization methods. As their name implies, these algorithms are inspired by the basic principles of natural selection in biology, like mutation, recombination and

survival of the fittest (elitism). The previous principles can be translated into mathematical operators to be used in a optimization process. EAs are population based methods, which means that in each generation a population of different individual exists. The collection of all the unique characteristics (design variables) of an individual is called genotype. The design variables are the inputs of the evaluation tool which will calculate the respective objective function value. In a single objective optimization (SOO) problem, the reproduction probability of an individual is proportional to the respective objective function value. With this process, a parent population is created. Then, the parents combine their design variables (genes) between them in a process called crossover to create the new generation of individuals. Afterwards, a mutation can happen with low probability in an individual, changing slightly its genotype. The previous process is repeated for many generations until an individual with high enough objective function value emerges (convergence criterion). Finally, inserting elitism to some degree can accelerate or stabilize the process by boosting more the survival of the elites (best individuals).

Unfortunately, EAs have some disadvantages, in their standard form. The computational cost of these algorithms is often very high, because of the large number of evaluations required in order to obtain many generations of individuals during the evolution process. The evaluation here refers to the computationally expensive MD simulation that is necessary for the calculation of the conductivity of an individual. This is also the most expensive part of the optimization procedure. In addition, the computational cost increases significantly, if the number of design variables is increased, because more evaluations are required for convergence of the algorithm. However, some additional methods like metamodels and distributed search schemes can accelerate the convergence. Metamodels are interpolation-like methods which implement machine learning and statistical methods to reduce the total number of evaluations. Distributed schemes divide the population to a few isolated islands. The exchange of gene information between the islands is governed by other rules which respect the laws of probability theory and statistics.

On the flip side, EAs can operate like a black box. They don't require any extra information, except from the objective function value for each individual. It isn't necessary to modify the simulation tool as it is in the gradient based methods. Furthermore, they are able to always find the global optimum, provided that an "infinite" number of evaluations is performed.

In this case, the simulation software is very difficult to be modified. Moreover, the optimization will be performed in two different types models (CG and AA models). Therefore, a more versatile optimization tool like EAs fits better in this specific optimization problem.

4.2 EASY

EASY is a general purpose optimization platform developed by the PCOpt/NTUA [55]. It can be used both for SOO and multi-objective (MOO), constrained or unconstrained optimization problems. EASY is evolutionary algorithm based software, but it also supports hybrid optimization methods, including gradient-based techniques. A variety of options for decreasing the computational cost is provided, such as machine learning assisted metamodels, distributed schemes and hierarchical optimization techniques.

The software can easily be coupled with any evaluation tool. The sole necessities for this purpose are a pre-processor and a post-processor. The pre-processor should translate the design variables of an individual to proper input data for the evaluation tool (MD simulation). On the contrary, the post-processor should translate properly the results of the evaluation tool to value(s) of the objective function(s). In this case the value of the objective function is always the ionic conductivity. The design variables are different in the CG and AA optimization. In the next paragraph, the details about the CG optimization design variables will be explained.

4.3 Parameterization of CG Ions

4.3.1 Computational Cost and Configuration Selection

In reality, the total number of possible IL pairs is enormous. Coarse-graining reduces significantly the previous number, because it greatly reduces the total number of particles in the system. In figure 4.1, different topology configurations regarding a CG neutral molecules are presented. To examine which configuration is both simple and versatile enough, an approximate calculation is performed about the total number of possible configurations.

In this simple calculation, the total number of beads is always less than four. So, only the first row of figure 4.1 is examined. The different types of beads that can be used is limited to 12. Each molecule can have only one charged beads, meaning that the charge value should either be 1 or -1. The angle value of the scenario 3a is discretized to only three values. All the bond, angle and dihedral constants (see paragraph 1.10.2) have fixed values. The length of each bond is automatically derived from the types of the connected beads.

For this simple example, 9408 different possible configurations exist. This result was obtained by an algorithm which applies all the previous rules. Therefore, the total number of possible IL pairs should be

$$N_{tot} = \binom{9408}{2} \approx 4.4 \cdot 10^7 \quad (4.1)$$

The evaluation cost of a CG simulation is about 2 minutes when 16 processors are used. To obtain meaningful results, each simulation should run 5 times with different random number generator seeds. In a realistic situation, the evaluation tool can run for 5 days in 80 logical processors. In this case, only 3600 evaluations will be performed. As one can see, the ratio between the possible states in the design space and the total number of evaluations is very small ($\approx 10^{-4}$). By making the same assumptions for a 4 bead configuration, the previous number is increased exponentially. Also, CG MARTINI 3.0 offers more than 36 different bead types. In conclusion, using more than 3 beads for the CG optimization is prohibitive, because of the computational cost.

For the sake of simplicity, the CG optimization will be performed, only for the case 3a of figure 4.1. In the same time, this choice is considered to give the necessary flexibility for a wide range of results to emerge from the optimization. Therefore, each ion of the IL pair will consist of three beads and one free angle.

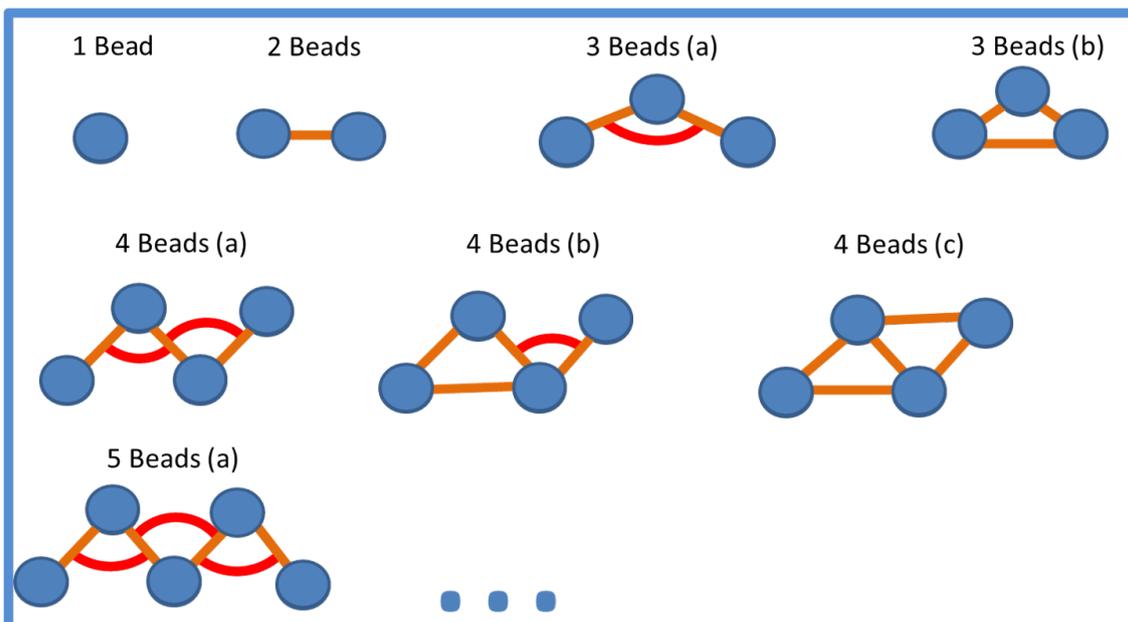


Figure 4.1: Different topology configurations for a CG neutral molecule. The orange lines denote the bonds between the beads and the red arcs the angular degree of freedom.

4.3.2 Design Variables

As mentioned in the previous paragraph, 8 design variables are used to describe each IL pair, as shown in figure 4.2. The variables p_1 and p_5 can range from 90° to 180° and they describe the angle formed between the 3 beads. The rest of the design variables can take any value in the set of beads allowed. The set of the beads used is shown below.

$$\begin{aligned} \text{Beads} = \{ & C1, C3, C4, C5, C6, N0, N2, N3, P1, P2, P3, P4, P5, Q1, Qp, \\ & SC1, SC2, SC3, SC4, SC5, SC6, SN0, SN2, SN3, SP1, SP2, SP3, SP4, SQ1, SQp, \\ & TC1, TC2, TC3, TC4, TC5, TC6, TN0, TP1, TQ1, TQp \} \quad (4.2) \end{aligned}$$

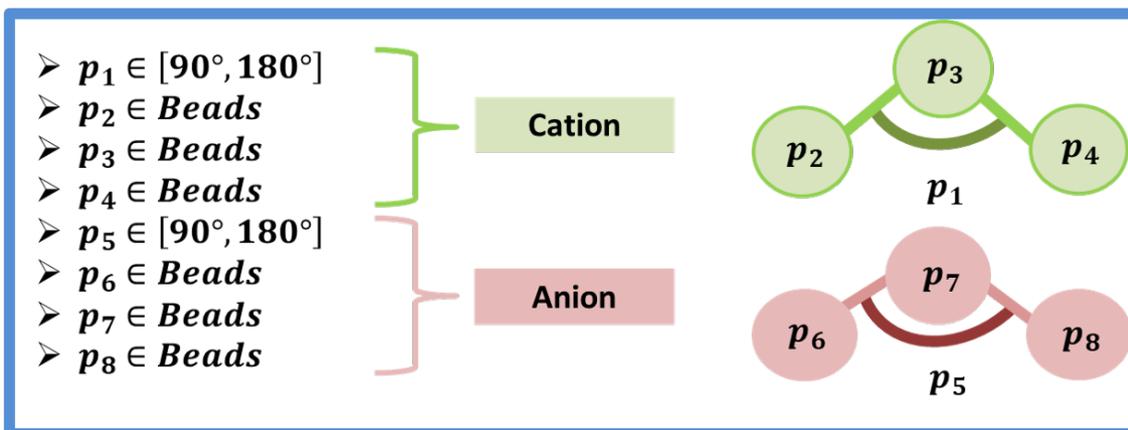


Figure 4.2: Simple explanatory schematic about the design variables of the CG optimization.

The previous bead types were selected according to table B.4. The idea is to only use beads for which a translation to AA chemistry exists. By doing so, every CG ion can be translated to a real ion (reverse mapping). Translating the CG elites of the optimization to AA elites is necessary for the validation of the final/optimal result. Table B.4 is very important, because it is a standardized and objective tool to give physical meaning to any CG model. Of course, the table itself is created by chemical intuition and empirical data and, thus, reverse mapping is always possible without it. However, in a complex optimization process, the need for standardization cannot be ignored. Finally, the spring constants for the bonds and the angles were set equal to the typical values, $20000 \frac{\text{kJ}}{\text{mol nm}^2}$ and $25 \frac{\text{kJ}}{\text{mol}}$ respectively, suggested in [22].

Before proceeding to the results, it is important to explain how the bond lengths change based on the size of the connected beads. As mentioned in paragraph 3.2.2, in the MARTINI 3.0 force field, different sizes of beads exist (Normal, Small, Tiny). To derive the bond length, each one of them is translated to an approximate AA model (see table 3.3), consisting of carbon and hydrogen atoms. The AA beads can be connected in different ways, as shown in the example of figure 4.3. After connecting the AA beads, the MMFF94 AA force field is applied to optimize the molecular geometry. This optimization is necessary to create a realistic AA geometry. The next step is to calculate the center of mass for each AA bead. All the previous procedures were performed by using the Avogadro software [32].

Finally, the length of each configuration is considered to be equal to the distance

between the two center of masses. The final length is approximated as the average of all the previous lengths. For example, if a normal size bead is connected with a small size bead, the length is derived as described in the following figure. The above procedure is repeated for all the possible combinations of bead sizes (6 in total) and the results are presented in table 4.1.

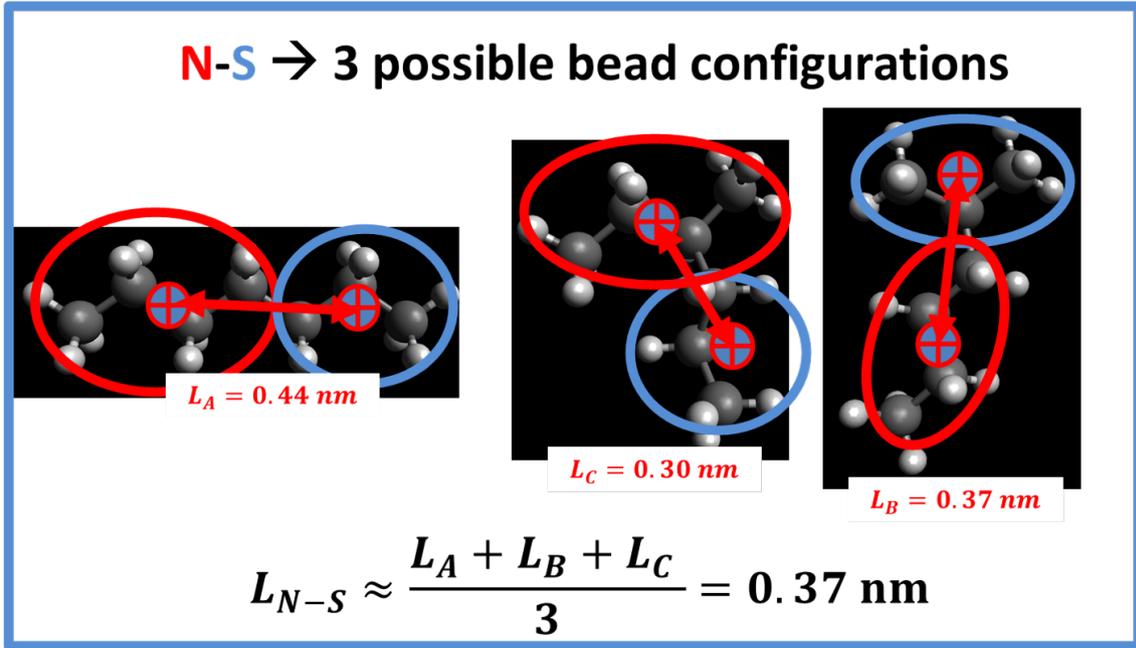


Figure 4.3: Schematic about how the bond lengths are derived from the bead size of the connected beads.

Connection of bead size i with bead size j	Approximation of the bond length (nm)
$N_i - N_j$	0.445
$N_i - S_j$	0.370
$N_i - T_j$	0.330
$S_i - S_j$	0.325
$S_i - T_j$	0.280
$T_i - T_j$	0.250

Table 4.1: Bond length derivation from the size of the connected beads.

4.4 Results of the CG Optimization

After the end of the optimization, all the individuals are arranged in descending order, in terms of the objective function value (ionic conductivity). The convergence history is shown in figure 4.4. The x axis displays the total number of evaluations. Only 1073 individuals were evaluated, the rest of them were rejected before being

evaluated, because they didn't correspond to feasible solutions. The computational cost of those were negligible, since they weren't evaluated.

As far as the search engine of EASY is concerned, the most important configuration settings are the following. Three demes were used with the population for each one consisting of 10 parents (μ) and 40 offspring (λ). Three parents were required to produce an offspring. The gradually decreasing mutation probability was initially 0.2 for achieving better exploration of the design space. One elite was kept and forced as an offspring in each generation. Radial distribution function metamodels were used. The minimum and maximum training patterns were 300 and 500 respectively, from which at least 300 were required to be successfully evaluated. The evolution was allowed to expand for about 1000 evaluations, without any other convergence criterion. Lastly, table 4.2 contains in more detail the settings of EASY that were used for the CG optimization.

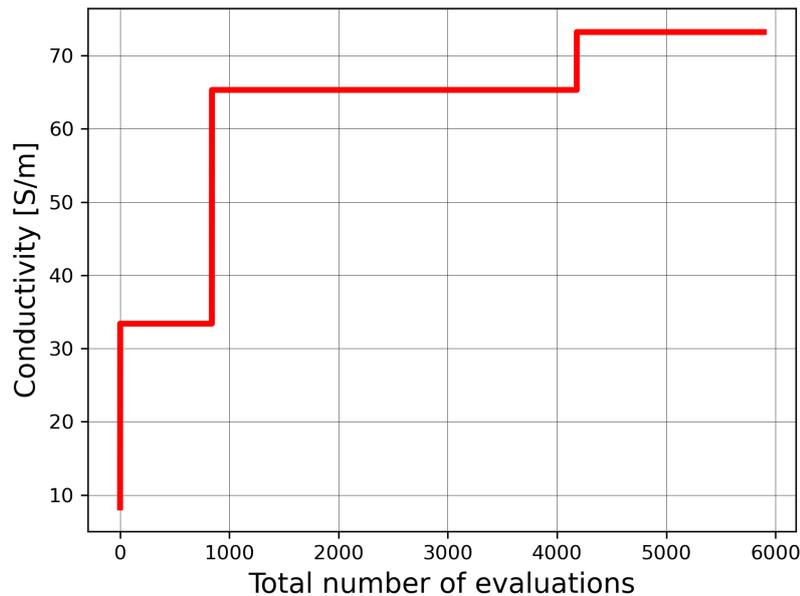


Figure 4.4: *CG optimization convergence curve by using Distributed Metamodel Assisted Evolutionary Algorithm (DMAEA).*

At this point, it is important to point out that all the simulations were performed at 380 K, because closer to these temperatures all the models are more accurate. The previous fact has been explained in more details in chapters 2 and 3 (see figure 2.8b).

Table 4.3 contains the design variables and the conductivity value of the top 5 elites. As shown in it, the angle values don't seem to affect significantly the conductivity result. However, table 4.3 doesn't contain any angle value in the range of 150° to 180°. On the other hand, the bead type is the main force of conductivity

Population options		Metamodel settings (ON)		Distributed scheme (ON: 3 demes)	
Parent population size	10	Metamodel type	RBF	Inter-deme communication	
Offspring population size	40	Exact evaluations Min	5	Migration frequency	5
Max life span	0	Exact evaluations Max	15	Maximum migrations	Infinite
Parents of one offspring	3	IPE pause gen.	10	Emigrants (best)	2
Elite archive size	20	Min. DB entries	300	Emigrants (random)	3
Elite indiv. to force as new offsp.	1	DB ent. not failed	120	Immigrants	3
Probability to select an elite	0.05	Training patterns Min	300	Migration mode	Rnd-replace if better
Tournament size	3	Training patterns Max	500	Migration graph	Each to all
Tournament probability	0.9	Proximity factor	1.2	Sharing frequency	75
Operators options		IF relaxation	0.3	Mutation multiplier	3.3
Coding	Binary-Gray	RBF-Radius	Auto	Max mutation prob.	0.2
Probability (Crossover)	0.95	Use failed patterns	Yes	Infection radius factor	0.3
Mode (Crossover)	Two point/var.	Use PCA for ifs	No	Sharing duration	7
Probability (Mutation)	0.2	Not failed patterns	10	Maximum penalty	90.0
Multiplier (Mutation)	0.8	Allow extrapolation	Auto		
Idle generations	10	Non dimensionalize	Yes		
		Failed obj. multiplier	10		
		Max DB percentage	1.0		
		Prediction mode	Auto		
		CRBFN			
		Min number of Centers	2		
		Max number of Centers	120		
		Radius multiplier	0.5		
		Test-to-total ratio	0.3		
		Idle iterations	15		
		Learn rate ratio	0.1		

Table 4.2: Configuration settings of the EASY search engine that were used for the CG optimization.

diversification in this model. Both the cations and anions prefer to have charged beads (Q beads) of normal size at the one side and apolar beads (C beads) of tiny size at the center and at the other side.

Similar conclusions can be derived from figures 4.5, in which the bead type distribution for each design variable is shown, regarding the 100 top elites. For comparing with the rest of the population, figures 4.6 show the same distribution but for all the ranking classes of the population. The previous comparison is essential, because some bead types have been promoted more than similar bead types purely out of chance. A great example of this is shown in figure 4.5b, where there is a clear preference for TC5 beads over TC4 beads. The bead definitions are very similar in the force field and this means that they shouldn't differ significantly. Indeed, by observing the respective figure 4.6b, it is clear that the whole population has a lack of TC4 beads compared with TC5 beads. The ratios of TC4 and TC5 beads

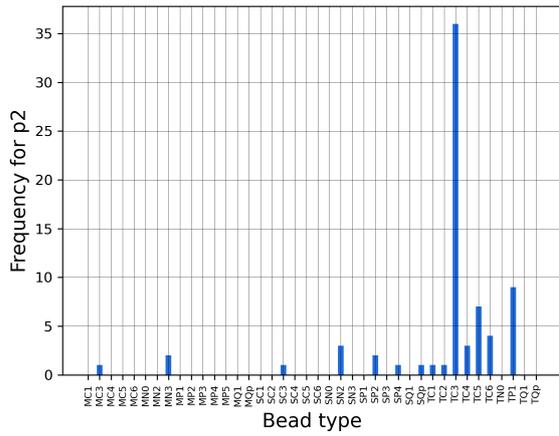
between the elite and the total population is almost the same.

Another conclusion that can be drawn from these figures is that the initial, chance driven, high conductivity results can influence the course of the evolution as well as the final solution. This means that some solutions appeared more frequently than others in the optimization process, because they were the first high conductivity results. For this reason, they survived throughout the optimization with small changes (mutations). The preference of the algorithm for TC5 over TC4 central beads (figure 4.5b) could be an example of how initially good solutions can affect the population evolution. This is usually the outcome of applying high elitism and low mutation probability in the evolutionary algorithm settings.

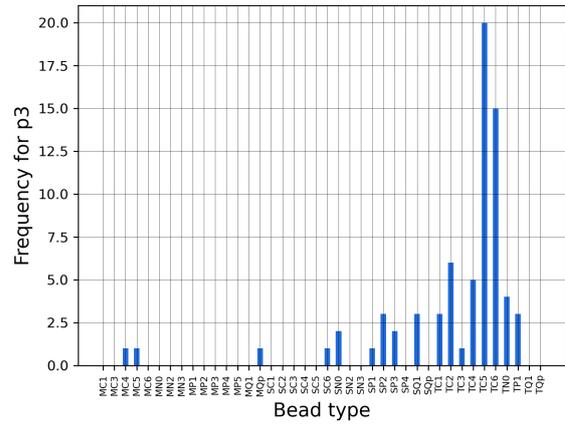
The most interesting result is the large magnitude of the conductivity values presented in the elite table 4.3. These values are about 3 times higher than the conductivity value of nafionTM. Unfortunately, this is the conductivity of the CG model and not the conductivity of the more accurate AA model. For this reason, it is necessary to validate the previous CG results by using AA models.

Elite ID	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	σ [S/m]
1	101°	TC3	TC6	Qp	93°	TC5	TC1	Qn	73 ± 5
2	140°	TC3	TC5	Qp	131°	TC5	TC3	Qn	66 ± 1
3	131°	TC3	TC5	Qp	123°	TC6	TC5	Qn	65 ± 1
4	141°	TC3	TC5	Qp	116°	TC5	TC3	Qn	65 ± 1
5	96°	TC3	TC6	Qp	126°	TC5	TC6	Qn	64 ± 1

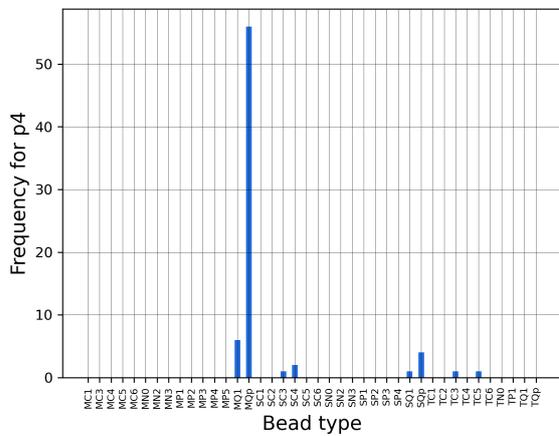
Table 4.3: *The design variables ($p_1...p_8$) and the objective function value (ionic conductivity) for the top 5 elites. All the MD simulations were performed at 380 K for increased accuracy.*



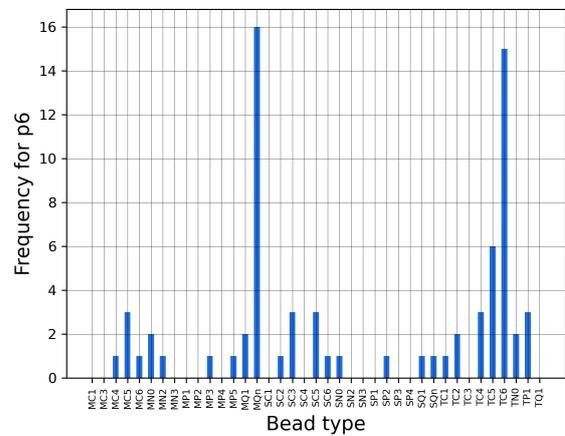
(a) Distribution for the p_2 parameter



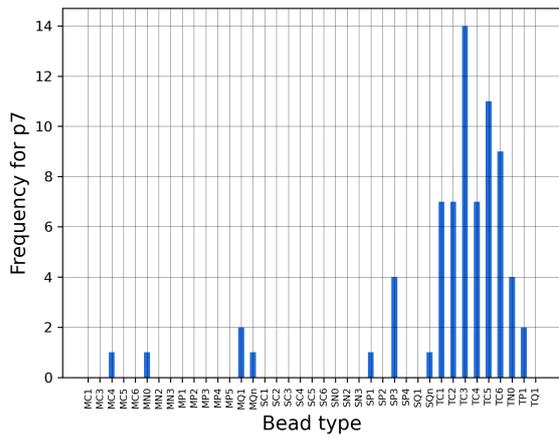
(b) Distribution for the p_3 parameter



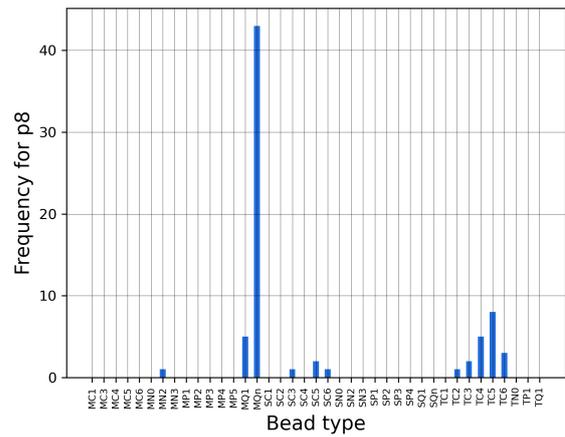
(c) Distribution for the p_4 parameter



(d) Distribution for the p_6 parameter



(e) Distribution for the p_7 parameter



(f) Distribution for the p_8 parameter

Figure 4.5: Bead distribution for the design variables p_2 , p_3 , p_4 , p_6 , p_7 and p_8 , regarding the top 100 elites of the optimization.

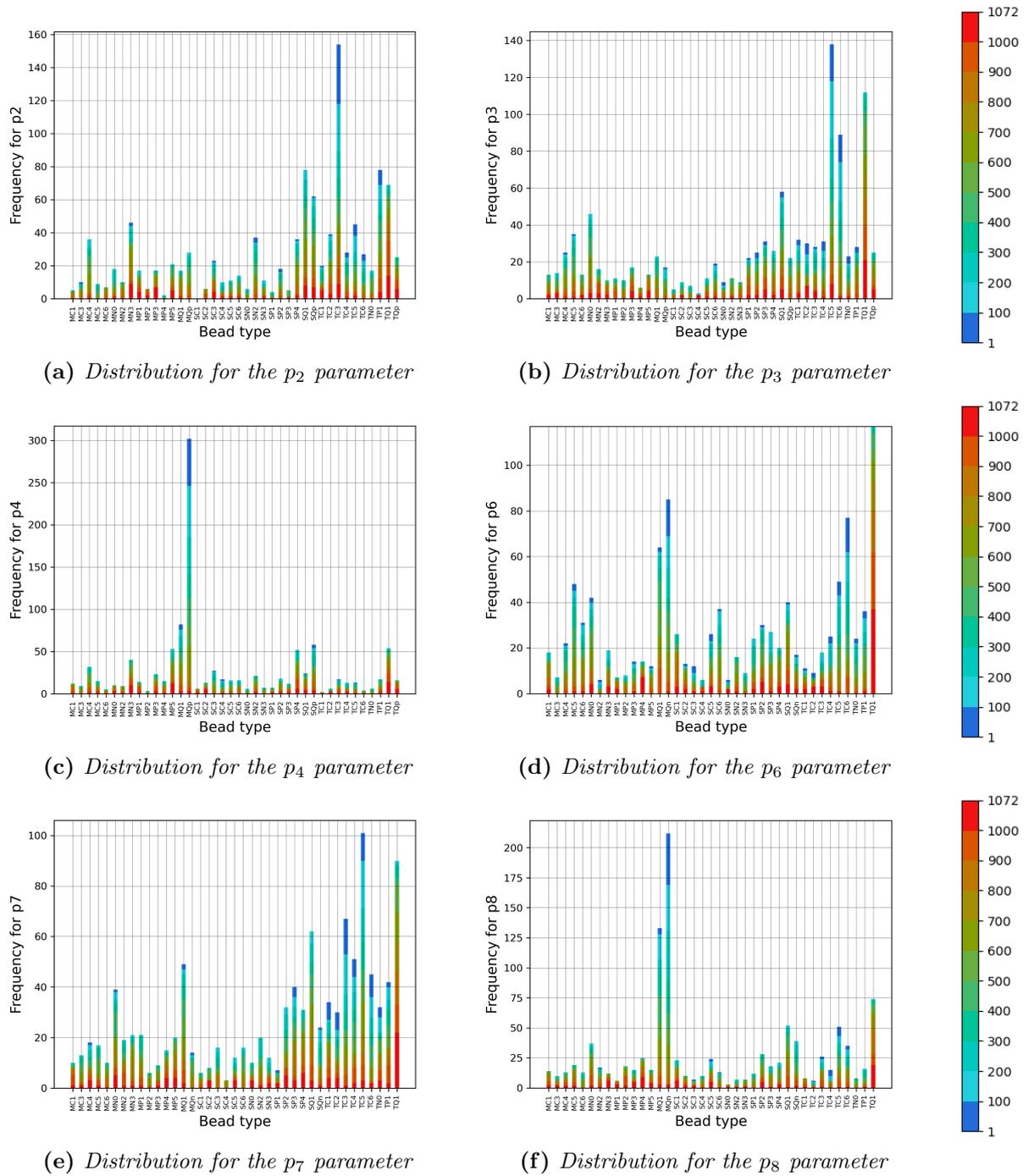


Figure 4.6: Bead distribution for the design variables p_2 , p_3 , p_4 , p_6 , p_7 and p_8 , regarding all the individuals of the optimization. Each color denotes a ranking class of individuals. For example, the dark blue color refers to the top 100 individuals.

4.5 Validation by using AA models

Each CG configuration produced by the optimization process can be translated to the AA equivalent by using table B.4 which the MARTINI 3.0 guide provides. For example, the translation of the optimal CG IL (1st elite of table 4.3) is shown in figure 4.7. These molecules can then be inserted as input data to the Ligpargen web server [24], [25], [26] for acquiring the respective AA topology in order to run OPLS-AA MD simulations.

The previous process was followed for the individuals shown in table 4.4. Each one of them has a representative conductivity value in the CG conductivity space. The AA conductivity results that were obtained by the OPLS-AA simulations for the translated molecules are shown in figure 4.8 in blue color. The ideal, one-to-one, correlation between CG and AA conductivity is shown by the purple dashed line. The light blue dashed curve is a linear fit regarding the optimization data points (blue points). The correlation magnitude is shown by the factor $R^2 = 0.295$. Therefore, the correlation between the conductivity of the CG optimization model and the respective AA model is weak. It worth noting that, figure 4.8 validates the conclusions drawn in chapter 3. It is clear that the CAT_1 CG model (red points) isn't accurate, while the one for Rmim gives a curve (green curve) which is close to the ideal line.

In conclusion, using a CG force field like MARTINI 3.0 for performing an optimization regarding ionic conductivity presents significant challenges. Even though there is a weak correlation between the CG and AA conductivity, the best IL that emerged from the previous process has almost 4 times lower conductivity than the reference IL ($CAT_1^+AN_3^-$). Accuracy and reverse mapping limitations of CG MARTINI 3.0 are the main reason for that.

In the next chapter, an alternative approach will be followed. An AA optimization will be performed, in spite of the high complexity and computational cost. This chapter proved that in ionic conductivity calculations, accuracy is really important.

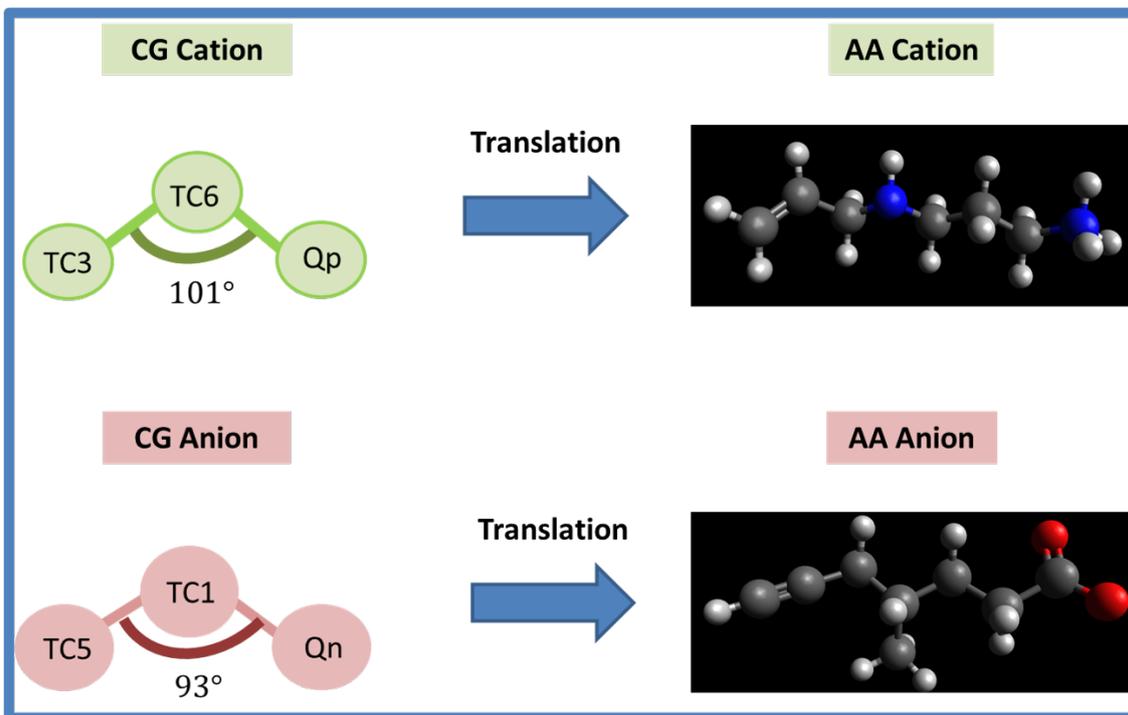


Figure 4.7: Translation of the CG model to an AA model for the best elite of the optimization. For translating the model, table B.4 was used. Gray, white, red and blue spheres represent carbon, hydrogen, oxygen and nitrogen atoms respectively.

Elite ID	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	σ [S/m]
1	101°	TC3	TC6	Qp	93°	TC5	TC1	Qn	73 ± 5
5	96°	TC3	TC6	Qp	126°	TC5	TC6	Qn	64 ± 1
11	157°	TC3	TC6	SQp	100°	SC3	TC4	Qn	50 ± 3
18	101°	SN2	TC6	Qp	93°	Qn	TC1	TC3	42 ± 2
37	133°	TC3	TN0	Qp	130°	TC6	SP3	Qn	35 ± 1
59	100°	N3	Qp	SC4	103°	Qn	TC1	TC5	29 ± 2
100	90°	TC4	SP4	Qp	126°	SC6	TP1	Qn	21 ± 1
183	94°	N3	TC6	Qp	92°	TC6	SN2	Qn	14 ± 1
395	106°	TQp	C4	TN0	96°	SP3	N2	Qn	7 ± 1

Table 4.4: The design variables ($p_1 \dots p_8$) and the objective function value (ionic conductivity) for 9 individuals that represent the objective function space (ionic conductivity range). All the MD simulations were performed at 380 K for increased accuracy.

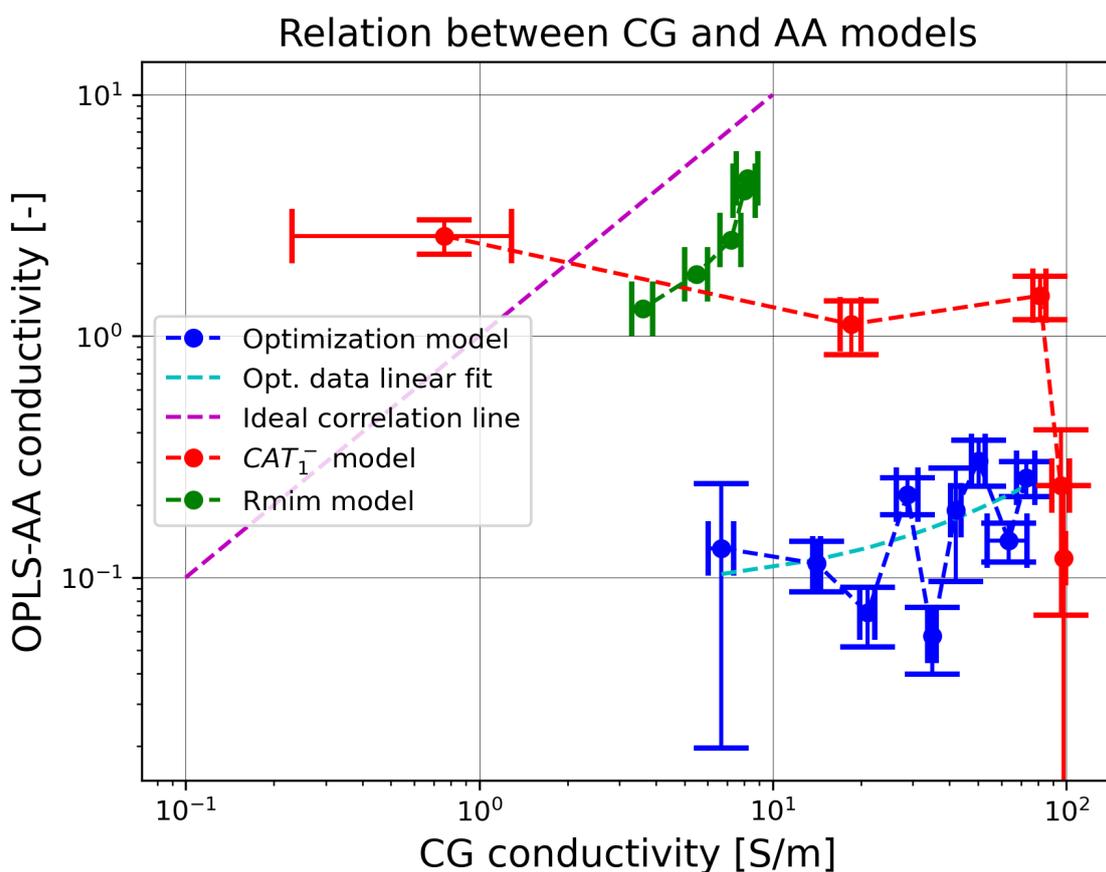


Figure 4.8: Correlation between CG and AA conductivity for various models. For the AA results regarding the CAT_1 ILs, GROMOS force field was used (see paragraph 3.2.4). Considering the Rmim ILs, the data of the y axis are the respective experimental values (see paragraph 3.3.2). Finally, the AA validation of the CG optimization model was performed by applying the OPLS-AA force field. For the last model, there is a weak correlation between the CG and AA conductivity results. All the MD simulations were performed at 380 K for increased accuracy.

Chapter 5

Optimization using AA Models

In this chapter, an AA optimization is performed using EASY. The most difficult part of the process is to create a robust and automated method for translating the optimization parameters (design variables) to anions and cations (ILs). For this purpose, the standardized SMILES strings are used. Each SMILES string is a sequence of character, which can be converted to geometry and topology input data for an AA MD simulation. To accomplish that, many opensource topology conversion tools and scripts are used. An AA optimization is then performed based on this "rough" geometry input data model. Finally, the best results of the optimization are validated by using more accurate input data.

5.1 The challenges of AA optimization

Producing ion geometry directly from parameters is prohibitive, because a relatively small number of different atoms can give a huge number of possible chemical configuration. The complexity of combining different particles is increasing exponentially by their number. Furthermore, organic ions in ILs usually contain many hydrogen atoms, which further increase the number of design variables required to describe them.

Moreover, chemistry sets limitations about what ions can exist in reality. A parameterization process could produce a configuration that doesn't exist in the real world. Supposing that an ion is described by using N design variables and that the process for producing it is stochastic (as happens in EAs), only a small percentage of the produced ions $S\%$ will be a feasible configuration. For the rest of the thesis, $S\%$ will be the success rate of the parameterization process, which is the procedure of producing AA ion MD input data from design variables. The success rate should ideally be equal to 100%. A very small $S\%$ will lead the evolutionary algorithm

to prefer not necessarily the highest conductivity individuals, but the individuals that correspond to a real chemical configuration. By doing so, the EA will help in keeping the number of valid produced ions high. Unfortunately, in the same time, the exploration rate will be decreased and the necessary time for convergence will significantly be increased.

For example, if the preprocessing procedure has 1% success rate instead of 10%, the optimization will be much slower, because less information about the conductivity value will be transferred to the EA. Also, in the first case, each generation will need to include significantly more individuals, slowing down the total process even more. If producing an ion has 1% chance to succeed, then producing an IL pair will have 10^{-2} % chance to do so. Therefore, the production of both random and valid IL pairs is very difficult, if the success rate is low.

In order to reduce the impact of the previous issues, the parameterization is conducted by using SMILES strings. SMILES strings require only heavy atoms (e.g. C, O, N and S) to describe a molecule. Thus, the hydrogen atoms will automatically be generated, reducing the complexity of the search. SMILES strings have many other advantages too, as it will be explained in the next section. Last but not least, the cation CAT_1^+ is kept the same throughout the whole optimization process. Parameterizing only the anion of each IL will keep the success rate in acceptable levels.

5.2 Design Variables

As mentioned before, SMILES strings are sequences of characters used to describe a chemical substance. In the context of this thesis, complex chemical configurations like rings are omitted. A simple molecule, like ethanol, can be described easily as shown in figure 5.1. In this case, **CCO** is the SMILES string, including only the heavy atoms of the molecule (**C**arbon and **O**xygen). The hydrogen atoms are assumed to exist in every case, because each chemical element has a limited number of free electrons to share in a covalent bond. A simple rule which can help to create the initial topology of a molecule is the following. Each carbon atom must have 4 connections (bonds), oxygen must have 2, nitrogen 3 and sulfur 2 or 6. Of course, in many cases these guidelines don't apply. Nevertheless, the majority of organic molecules usually follow the previous rules. Figure 5.1 contains two more examples of molecules that can be described by using SMILES strings.

By following the previous principles, a simple and robust method is developed for producing anions from design variables. Each anion produced consist of 5 heavy atoms. Parameters p_1 to p_5 control the element of each heavy atom. The design variables p_6 and p_7 control which bonds or branches will be used in the new anion by using the respective symbols of the SMILES code. To determine the position of these in the anion, parameters p_8 and p_9 are used. At last, parameter p_{10} is used to

denote the position from which a hydrogen atom will be removed. This is necessary for converting the neutral molecule to an anion. The parameterization process is shown in the figure 5.2.

A dedicated python script was developed for the implementation of the previous procedure. For the sake of simplicity, each event is equally probable to happen. This means that, the probability of choosing a specific atom type (e.g. oxygen) is equal to 0.25 in any case ($P_C = P_O = P_N = P_S = P_a$ and $P_a = \frac{1}{4}$). The same is true for the parameters p_6 and p_7 ($P_- = P_# = P_{()} = P_{(=)} = P_b$ and $P_b = \frac{1}{5}$). The SMILES string generation script applies other rules too in order to ensure that the emerged anion will have high chances to exist in reality. It is worth considering that, all these rules affect indirectly the outcome of parameterization process. A SMILES sequence for an anion and information about the validity of the produced string are the final outputs of the script. For example, the string: CC#O is invalid, because an oxygen atom would never have a triple bond, hence the respective set of design variables is considered infeasible. This information is automatically recognized by the script.

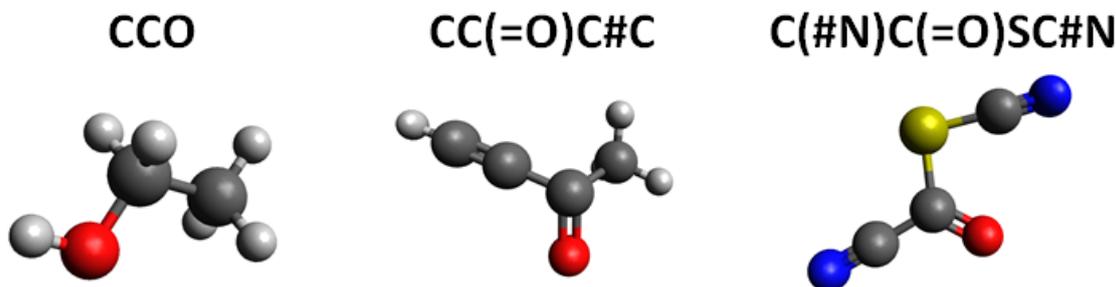


Figure 5.1: Example of describing molecules by using SMILES strings.

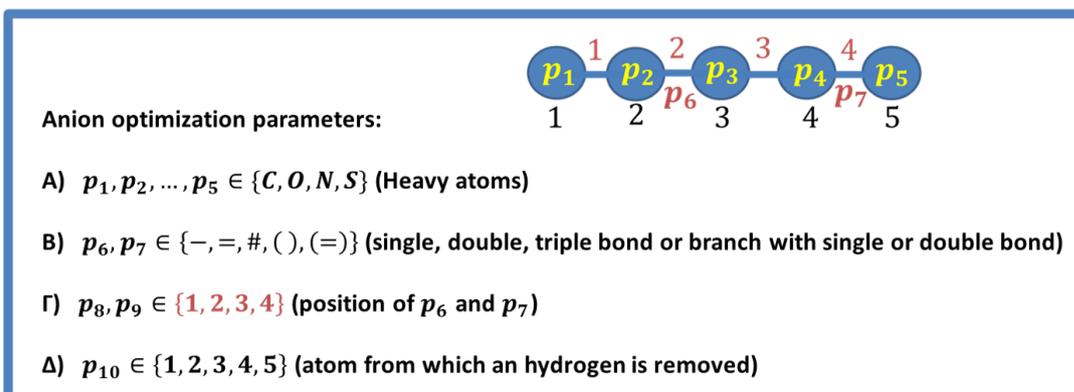


Figure 5.2: Simple explanatory schematic about the design variables of the AA optimization.

5.3 AA Topology Generation from SMILES Strings

After the creation of the SMILES string, a proper geometry .pdb file needs to be created. This file contains all the necessary information about the positions and the connections of the atoms in a molecule.

Open Babel is an opensource software which operates as a chemical toolbox designed to search, convert and analyze molecular modeling data [51]. All the following procedures were performed by using the Open Babel toolbox. The corresponding .pdb file was generated directly from the SMILES string. Initially, an energy minimization was performed by using the MMFF94 force field. This is necessary in order to obtain a more stable configuration than the initialized one. This process will change the initial bond lengths, angles and dihedrals to better fit the real conditions. To further increase the accuracy of the anion model, a partial charge calculation is performed by using the MMFF94 force field for once more. This step is very important, because the partial charges govern the electrostatic interactions of the simulation, which influence significantly the IL transport properties. During the previous step, the .pdb file is converted to .mol2 format, because the first one doesn't support any information about the charges.

Apart from the .pdb file, a topology (.itp) file is also necessary for conducting an MD simulation. This is the reason for which the partial charges calculation was performed. Therefore, a tool for converting the .mol2 that was previously acquired to a .itp file was required. Thankfully, another open source software, topolbuild, exist to serve exactly the previous purpose [53]. Topolbuild supports the AMBER, GAFF, GLYCAM and OPLS-AA force fields. OPLS-AA was used extensively in chapter 2 in AA models. For this reason, topolbuild was used for converting the .mol2 produced files to OPLS-AA .itp files.

Even though topolbuild was able to produce good quality OPLS-AA topology files for many cases, some serious issues emerged in the process in the first place. Some atomic combinations in randomly generated anions couldn't be recognized by the software, because the available OPLS-AA tables didn't contain them. Furthermore, most of the times, many bond lengths, bond constants, angles, angle constants and dihedrals were missing from the final topology file. Thankfully, topolbuild measures the bond lengths and angles for each atomic connection (probably from the .mol2 file). Therefore, completing the final topology file was possible, on condition that the values for the bond and angle constants are somehow assumed/approximated.

Unfortunately, a great piece of information about proper and improper dihedrals is usually missing. For that reason, it was decided that, these rough geometry AA models will not include any dihedral information in their topology files. Of course omitting the dihedrals has an impact on the accuracy of the simulation. Nevertheless, the relative energy percentage of proper and improper dihedrals compared to the total system's energy E_{tot} (see equation 1.7) is usually lower than 10 % in a typical simulation. As a result, omitting dihedrals can be considered a reasonable

trade-off, between accuracy and practicality.

To complete the OPLS-AA .itp file produced by topolbuild, some python scripts were developed. The basic philosophy of these scripts is to produce the final .itp and .pdb that are required for the simulation in a very robust way. In case that the previous task is impossible, the scripts provide information about the location and the cause of the error. It is worth reminding that, AA simulations are very computationally expensive. Avoiding running simulations with wrong initial data is a high priority. In addition, robustness is very important in order to keep the success rate of the process as high as possible (see paragraph 5.1). The whole pre-processing procedure described is shown in the following figure.

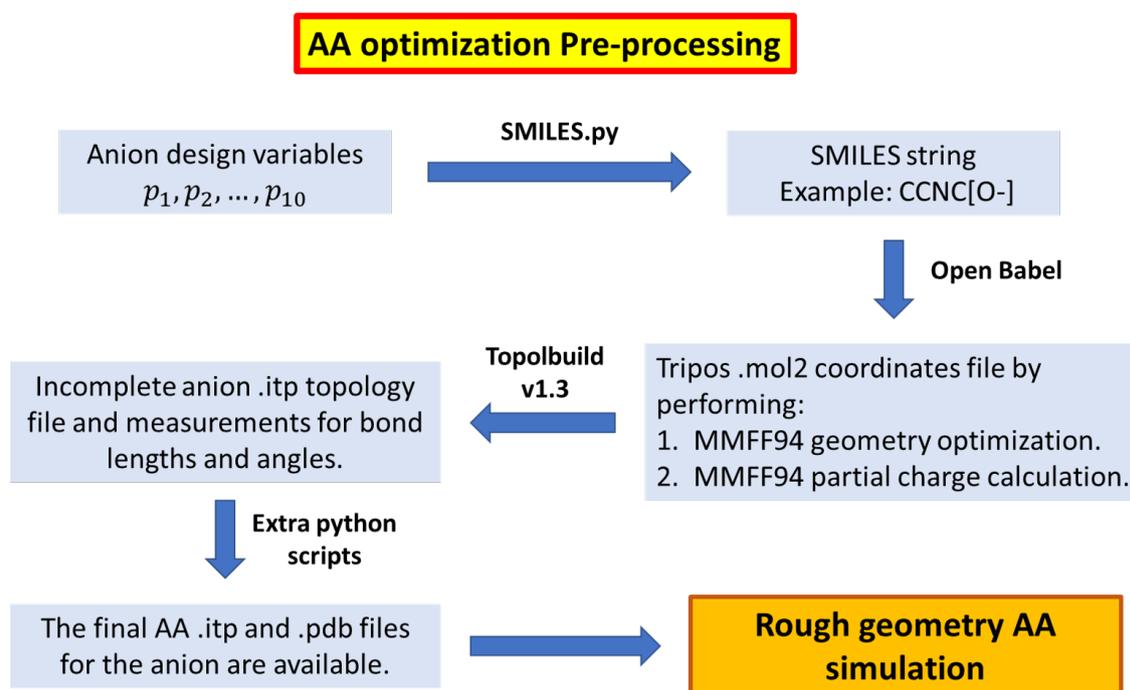


Figure 5.3: Schematic about the pre-processing procedure followed in the rough geometry AA optimization.

5.4 The Optimization Setup

Having developed the previous pre-processing structure, the rough geometry AA optimization is almost ready to run. Because of the high computational cost, EASY options must be modified to better fit the AA optimization conditions. The total cost was about 3 days on 144 Processors with gpu acceleration from one NVIDIA 2080 Ti operating at 100 %. The evolution was allowed to expand for about 10000 evaluations, without any other convergence criterion. From those, only 64 anions were valid solutions. Thus, the success rate of the process was equal to 0.64 %.

The most important configuration settings of EASY search engine were the following. The number of demes were reduced from three to one. The population was consisting of 500 parents (μ) and 2000 offspring (λ). Three parents were required to produce an offspring. The mutation probability was always equal to 0.15 for achieving better exploration of the design space. One elite was kept and forced as an offspring in each generation. Radial distribution function metamodels were used. The minimum and maximum training patterns were 300 and 500 respectively. Lastly, table 5.1 contains in more detail the settings of EASY that were used in the AA optimization.

Choosing properly the optimization settings in a search engine like EASY isn't by any mean a simple task. Furthermore, EASY setup depends heavily on the unique characteristics of each problem. In this case, a huge population is required, because the low success rate of the pre-processing stage creates high evolutionary pressure to the population. This means that the vast majority of individuals will score very low, while a tiny percentage will have a desirable fitness score. This isn't a usual case, thus different tests could be conducted for finding the optimal setup. Nevertheless, it was decided to mainly scale up the number of parents, offspring and training patterns compared to the more usual setup of section 4.4.

Indeed, running the previous setup for only 10000 individuals is far from optimal in this case. It's even possible that EASY operated more as a random number generator than as an evolutionary algorithm, because of the low success rate of the pre-processing procedure. The total number of evaluations would need to be at least 10 times higher in order to allow EASY to operate optimally. However, this would require the allocation of the same computing resources for about a month. Unfortunately, such high computational resources weren't available at the time. A higher success rate would improve the process too. Some possible solutions on the previous issues were examined and they will be discussed in section 6.3.

Population options		Metamodel settings (ON)		Distributed scheme (OFF)	
Parent population size	500	Metamodel type	RBF	Inter-deme communication	
Offspring population size	2000	Exact evaluations Min	5	Migration frequency	-
Max life span	0	Exact evaluations Max	15	Maximum migrations	-
Parents of one offspring	3	IPE pause gen.	5	Emigrants (best)	-
Elite archive size	20	Min. DB entries	5000	Emigrants (random)	-
Elite indiv. to force as new offsp.	1	DB ent. not failed	1000	Immigrants	-
Probability to select an elite	0.05	Training patterns Min	300	Migration mode	-
Tournament size	3	Training patterns Max	500	Migration graph	-
Tournament probability	0.9	Proximity factor	1.2	Sharing frequency	-
Operators options		IF relaxation	0.3	Mutation multiplier	-
Coding	Binary-Gray	RBF-Radius	Auto	Max mutation prob.	-
Probability (Crossover)	0.95	Use failed patterns	Yes	Infection radius factor	-
Mode (Crossover)	Two point/var.	Use PCA for ifs	No	Sharing duration	-
Probability (Mutation)	0.15	Not failed patterns	1000	Maximum penalty	-
Multiplier (Mutation)	1.0	Allow extrapolation	Auto		
Idle generations	10	Non dimensionalize	Yes		
		Failed obj. multiplier	10		
		Max DB percentage	0.5		
		Prediction mode	Auto		
		CRBFN			
		Min number of Centers	2		
		Max number of Centers	120		
		Radius multiplier	0.5		
		Test-to-total ratio	0.3		
		Idle iterations	10		
		Learn rate ratio	0.1		

Table 5.1: Configuration settings of the EASY search engine that were used for the AA optimization.

5.5 Results of the AA Optimization

Despite the above challenges, EASY was able to find an IL pair with significantly improved ionic conductivity as shown in figure 5.4. It worth noting that, $CAT_1^+AN_3^-$ gives the PILM with the highest ionic conductivity at 380 K compared to the other available ILs. For this reason, it will be used as reference for any comparative purposes.

The highest conductivity results of the AA optimization are shown in table 5.2. In this table, only the anion, which changes each time, is presented by its SMILES string. The cation is CAT_1^+ in every case. For reference, one of the best available ionic liquids, regarding conductivity is $CAT_1^+AN_3^-$ (see figure 2.8b). The exper-

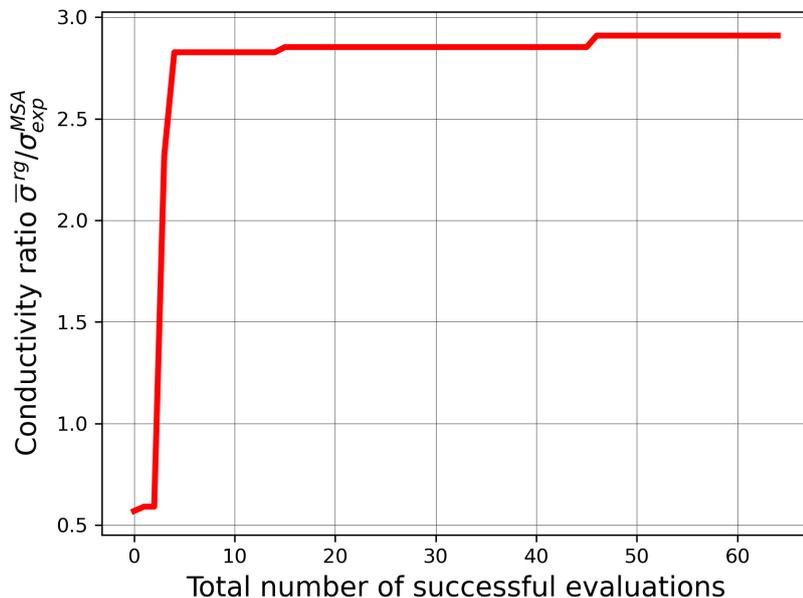


Figure 5.4: AA optimization convergence curve using Metamodel Assisted Evolutionary Algorithm (MAEA). $\bar{\sigma}^{rg}$ is the mean conductivity of the rough geometry model. $\sigma_{exp}^{AN_3}$ is the experimental conductivity of $CAT_1^+ AN_3^-$ IL pair.

imental and AA simulation conductivity is respectively equal to $\sigma_{exp}^{ref} = 1.0$ and $\sigma_{simAA}^{ref} = 1.1$ (these results are dimensionless). All the MD simulations were performed at 380 K for obtaining more accurate results (see section 2.3). The four top elites from table 5.2 have higher conductivity than the reference IL. The best IL pair that was obtained from the optimization has a conductivity improvement of about 157 % compared to the reference AA simulated IL.

Elite ID	SMILES	$\bar{\rho}$ [kg/m^3]	ρ_{std} [kg/m^3]	$\bar{\sigma}/\sigma_{exp}^{AN_3}$	$\sigma_{std}/\sigma_{exp}^{AN_3}$
1	[O-]COC#C	$9.33 \cdot 10^2$	4.65	2.83	0.24
2	CC(=S)[N-]C	$9.26 \cdot 10^2$	0.36	2.32	0.07
3	SO[N-]OS	$1.15 \cdot 10^3$	0.53	1.47	0.09
4	SO[N-]S#S	$1.18 \cdot 10^3$	1.47	1.19	0.33
5	CCSC[O-]	$9.54 \cdot 10^2$	1.65	0.94	0.20
6	NC(C)([O-])S	$1.01 \cdot 10^3$	2.30	0.68	1.44
7	SC(C)N[O-]	$9.94 \cdot 10^2$	2.23	0.59	0.13
8	NC(C)([O-])C	$9.03 \cdot 10^3$	1.04	0.57	0.08
9	SNO[N-]N	$1.08 \cdot 10^2$	1.07	0.56	0.13
10	[O-]CC(N)N	$9.58 \cdot 10^2$	2.33	0.37	0.05

Table 5.2: Rough geometry AA optimization results for the top 10 elites. All the MD simulations were performed at 380 K for increased accuracy.

5.6 Validation using more accurate AA input data

The input data for the previous simulations were missing information about the dihedral angles and various assumptions were made. Although the previous values seem promising, it is necessary to validate the rough geometry results by using more accurate AA models.

In order to do so, the geometry .pdb and the topology .itp files considering the top 5 ILs from table 5.2 are acquired, for once more, from the Ligpargen web server [24], [25], [26]. Therefore, all the simulations options are identical to those of chapter 2. This is considered to be the most accurate AA model in the context of this thesis. Consequently, a direct comparison with the reference IL $CAT_1^+ AN_3^-$ is now possible.

In table 5.3, the 5 highest conductivity ILs that emerged from the AA optimization are shown and compared with the reference IL. In the fourth column, the average density of the validation models is shown. In fifth and sixth columns, the mean value and the standard deviation of conductivity is presented, regarding the validation model. In the seventh column, the average conductivity is presented for the rough geometry model. Finally, the relative conductivity improvement is shown with respect to the AA model for the reference IL. It is worth noting that, all the conductivity results have been divided by the experimental value of the reference IL.

The conductivity standard deviation values are between reasonable levels, indicating the proper convergence of all the validation simulations. For the first, third and fourth ILs the rough geometry model doesn't deviate significantly from the validation model (less than $\pm 50\%$ relative difference). However, the other two ILs have a noticeable difference.

Maybe, the most important conclusion that can be drawn from table 5.3 is that 3 out of the 5 new IL pairs present improved ionic conductivity, when compared with the reference IL. The improvement presented in the first (+140%) and in the third (+101%) is not negligible at all. These alternative ILs were directly derived from a stochastic process. The respective neutral molecule for all of the following anions, except from $SO[N-]S\#S$, are commercially available or they have been synthesized, according to the open chemistry database of the National Institutes of Health (USA) [52]. Synthesizing these anions could be possible if the respective neutral molecules are deprotonated (removing the proper hydrogen atom from the molecule). The accomplishment and explanation of the deprotonation process is out of the scope of this thesis.

It is worth noting that, having alternative IL options with higher or similar ionic conductivity could be important even from an economic or environmental standpoint. For example, if an IL has a very high conductivity, but it is very expensive to be produced, an alternative IL that has slightly lower conductivity, but significantly lower cost could be a favorable option. The previous rough AA optimization process could run for more time, producing even more previously unknown ILs, potentially with

higher ionic conductivity. Furthermore, the process is far from optimized itself and it could be improved by a lot, regarding both the computational cost and the accuracy. Despite the challenges and limitations presented in the preceding sections, it seems that the rough geometry AA optimization procedure worked according to the expectations.

ID	Cation	Anion	$\bar{\rho}^v$ [kg/m^3]	$\bar{\sigma}^v/\sigma_{exp}^{AN_3}$	$\sigma_{std}^v/\sigma_{exp}^{AN_3}$	$\bar{\sigma}^{rg}/\sigma_{exp}^{AN_3}$	$\frac{\bar{\sigma}^v - \sigma_{simAA}^{ref}}{\sigma_{simAA}^{ref}}$
1	CAT_1^+	[O-]COC#C	$9.42 \cdot 10^2$	2.64	0.17	2.83	+140 %
2	CAT_1^+	CC(=S)[N-]C	$9.67 \cdot 10^2$	0.38	0.05	2.32	-65 %
3	CAT_1^+	SO[N-]OS	$1.11 \cdot 10^3$	2.21	0.44	1.47	+101 %
4	CAT_1^+	SO[N-]S#S	$1.15 \cdot 10^3$	1.12	0.23	1.19	+8 %
5	CAT_1^+	CCSC[O-]	$9.28 \cdot 10^2$	1.75	0.43	0.94	+59 %
<i>Ref</i>	CAT_1^+	AN_3^-	—	1.10	0.19	0.83	+0 %

Table 5.3: Top 5 IL pairs that have been obtained from the AA optimization process. $\bar{\sigma}^v$ is the mean conductivity for the validated models (topology from Ligpargen), while $\bar{\sigma}^{rg}$ is the conductivity of the rough geometry model. The final column of the table shows the relative improvement in conductivity compared to the respective AA model for the reference IL. All the MD simulations were performed at 380 K.

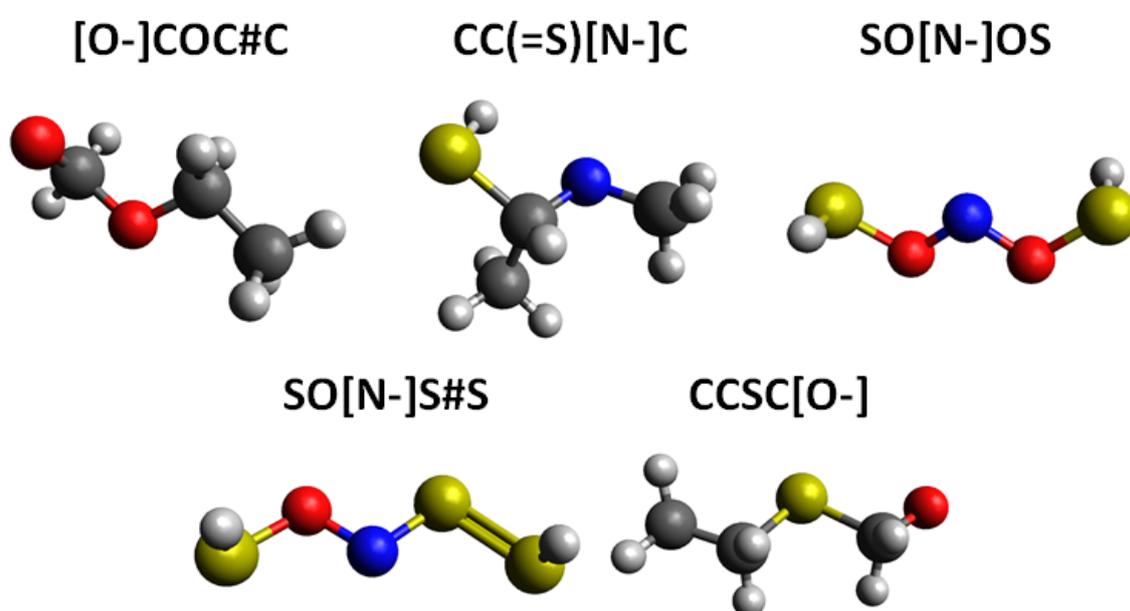


Figure 5.5: 3D representation of the top 5 anions that emerged from the AA optimization.

Chapter 6

Overview and Conclusions

6.1 Overview

In this diploma thesis, AA and CG MD models were created and validated for calculating the ionic conductivity of ILs. Both models were used for searching new ILs with high ionic conductivity, in order to be proposed as alternative/improved raw materials for a new class of PEMs in fuel cells. The search was conducted both in CG and AA level by using the evolutionary algorithm software EASY. The AA optimization indicated previously unknown pairs of ILs, which could be synthesized, while having high ionic conductivity.

The basics of MD simulation and PEMs were presented. Various MD simulation methods (force fields) from literature, like NPAAFFs, PAAFFs, charge scaling and coarse graining, were discussed and compared. NPAAFFs and CG models were selected to be applied for optimization purposes.

A general simulation structure for performing ionic conductivity calculations was developed and explained. The OPLS-AA and GROMOS force fields were used to create an appropriate AA model. During the development phase, various convergence studies were conducted and available experimental data were used for the model validation. The ionic conductivity results were more accurate at higher temperatures (close to 380 K).

A CG model for ILs was created in order to be used for optimization, because of its low computational cost. The general purpose and well standardized CG MARTINI 3.0 force field was used as basis for the model. Various ILs, including the $CAT_1^+ AN_i^-$ and $Rmim^+ AN_1^-$ pairs were mapped into CG models. The ionic conductivity trend of the first group were inaccurate for different ILs, while the opposite was true for the second group.

Following the rich MARTINI 3.0 documentation that was available, a CG optimization was performed by using the evolutionary algorithm software EASY. The obtained results were discussed and validated by using more accurate AA models. The correlation between the CG and AA models was weak, indicating that a more accurate optimization procedure should be followed.

Lastly, a random anion generation procedure was created by using molecular modeling data processing software (Open Babel and topolbuild) as well as other linking python scripts. The anion generation software was coupled with the MD AA model of chapter 2 and with EASY to perform a rough geometry AA optimization. The rough geometry results were validated by the same model presented in chapter 2, which was even more accurate. The final validated new ILs acquired from the process presented relative improvement as high as +140 %. At the same time, the majority of them can be synthesized.

6.2 Conclusions

Upon completion of all the previous studies, the following conclusions are drawn:

1. An NPAAFF is capable of predicting with good accuracy the density of the examined ILs ($|\frac{\rho_{AA}-\rho_{exp}}{\rho_{exp}}| < 15 \%$). The respective conductivity predictions present higher uncertainty ($|\frac{\sigma_{AA}-\sigma_{exp}}{\sigma_{exp}}| < 300 \%$). In higher temperatures the same models are much more accurate ($|\frac{\sigma_{AA}-\sigma_{exp}}{\sigma_{exp}}| < 90 \%$). From the comparisons that took place in chapter 2, the GROMOS force field proven to be more accurate than OPLS-AA for the specific ILs. Nevertheless, both of them could be considered accurate enough for optimization purposes.
2. All the CG MARTINI 3.0 models were able to give correct trend regarding density. However, only the models considering the $Rmim^+AN_1^-$ IL family presented correct conductivity trend, which is necessary for the optimization. MARTINI 3.0 is very good at capturing the effect of simple molecular changes, like changes in the molecular weight of the substance. Unfortunately, when totally different organic ions are involved, transferrability limitations are clearly visible. A CG model could never match the accuracy of an AA model, because important information is always lost during the mapping process. This doesn't seem that its comparative accuracy couldn't be used in a CG optimization.
3. The computational cost of a CG optimization is low enough to allow a meaningful optimization to be performed. The ionic conductivity correlation between the CG MARTINI and OPLS-AA models is weak. The accuracy and reverse mapping limitations of CG MARTINI is the main reason for the weak correlation.
4. A rough geometry AA optimization is able to produce new IL pairs with high

ionic conductivity. The low success rate $S\%$ of the process is prohibiting changing simultaneously the cation and the anion. The computational cost of this process is very high, forcing the evolutionary algorithm EASY to operate more as a random number generator. These problems would be less important, if neutral molecules, instead of ions, were involved. The correlation between the rough geometry AA model and the model presented in chapter 2 seems to be adequate for the purposes of this thesis.

6.3 Future Work Proposals

Based on the previous findings, the following future works are proposed:

1. The existing AA optimization structure could run for more time to examine if even higher conductivity IL pairs can emerge. The new IL pairs that have been obtained could be further validated by performing lab experiments, since the respective neutral molecules have already been synthesized.
2. The existing AA optimization structure could be used with some small modifications for optimizing other molecular systems too. In fact, using molecules instead of ions in the previous process will, surely, increase significantly the success rate $S\%$, because of the lower complexity involved. The most frequent cause of rejection in the anion generation procedure was related to the fact that ions weren't available in the databases of topobuild.
3. More sophisticated techniques could be used for increasing the success rate $S\%$ of the AA optimization process. Using databases of frequently appeared ionic groups (e.g. $O[N-]$, $N[O-]$ or $S[O-]$) by applying conventional or machine learning algorithms could increase significantly the success rate. A huge database is relatively easy to be created for the machine learning algorithm, since it just requires to randomly run the pre-processing procedure described in chapter 5. Even the design variables could be changed to better and more efficiently describe the ions. If the success rate is high enough, this process could be used for conductivity optimization in ILs without having to keep the cation or the anion constant.
4. The AA models of chapter 2 could be further improved by using alternative force fields or simulation settings. The same is true for the CG models presented in chapter 3.
5. CG models could be created for PILs. This would be useful for validating that the high ionic conductivity in some ILs translates directly in high conductivity in the respective PILs.

Bibliography

- [1] M. Salanne, “Simulations of room temperature ionic liquids: from polarizable to coarse-grained force fields,” *Phys. Chem. Chem. Phys.*, vol. 17, no. 22, pp. 14270–14279, 2015, doi: 10.1039/C4CP05550K.
- [2] M. Salanne, B. Rotenberg, S. Jahn, R. Vuilleumier, C. Simon, and P. A. Madden, “Including many-body effects in models for ionic liquids,” *Theor. Chem. Acc.*, vol. 131, no. 3, p. 1143, Mar. 2012, doi: 10.1007/s00214-012-1143-9.
- [3] O. Borodin, “Polarizable Force Field Development and Molecular Dynamics Simulations of Ionic Liquids,” *J. Phys. Chem. B*, vol. 113, no. 33, pp. 11463–11478, Aug. 2009, doi: 10.1021/jp905220k.
- [4] E. Choi, J. G. McDaniel, J. R. Schmidt, and A. Yethiraj, “First-Principles, Physically Motivated Force Field for the Ionic Liquid [BMIM][BF₄],” *J. Phys. Chem. Lett.*, vol. 5, no. 15, pp. 2670–2674, Aug. 2014, doi: 10.1021/jz5010945.
- [5] Y. Wang, S. Feng, and G. A. Voth, “Transferable Coarse-Grained Models for Ionic Liquids,” *J. Chem. Theory Comput.*, vol. 5, no. 4, pp. 1091–1098, Apr. 2009, doi: 10.1021/ct800548t.
- [6] Y. Wang, W. G. Noid, P. Liu, and G. A. Voth, “Effective force coarse-graining,” *Phys. Chem. Chem. Phys.*, vol. 11, no. 12, p. 2002, 2009, doi: 10.1039/b819182d.
- [7] W. D. Cornell et al., “A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules,” *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, May 1995, doi: 10.1021/ja00124a002.
- [8] Y. Duan et al., “A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations,” *J. Comput. Chem.*, vol. 24, no. 16, pp. 1999–2012, Dec. 2003, doi: 10.1002/jcc.10349.
- [9] B. Leimkuhler and C. Matthews, *Molecular Dynamics*, vol. 39. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-16375-8.
- [10] W. L. Jorgensen and J. Tirado-Rives, “The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin,” p. 10.

- [11] J. A. Rackers and J. W. Ponder, "Classical Pauli repulsion: An anisotropic, atomic multipole model," *J. Chem. Phys.*, vol. 150, no. 8, p. 084104, Feb. 2019, doi: 10.1063/1.5081060.
- [12] F. London, "Zur Theorie und Systematik der Molekularkrafte," *Z. Fur Phys.*, vol. 63, no. 3–4, pp. 245–279, Mar. 1930, doi: 10.1007/BF01421741.
- [13] J. Zhang, "Canonical dual theory applied to a Lennard-Jones potential minimization problem," *ArXiv11075146 Phys. Q-Bio*, Dec. 2012, Accessed: Aug. 25, 2021. [Online]. Available: <http://arxiv.org/abs/1107.5146>
- [14] C. Lopes, J.N., Pádua, A.A.H. CL&P: A generic and systematic force field for ionic liquids modeling. *Theor Chem Acc* 131, 1129 (2012). <https://doi.org/10.1007/s00214-012-1129-7>
- [15] K. T. Tang and J. P. Toennies, "An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients," *J. Chem. Phys.*, vol. 80, no. 8, pp. 3726–3741, Apr. 1984, doi: 10.1063/1.447150.
- [16] B. Jeziorski, R. Moszynski, and K. Szalewicz, "Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes," *Chem. Rev.*, vol. 94, no. 7, pp. 1887–1930, Nov. 1994, doi: 10.1021/cr00031a008.
- [17] J. G. McDaniel and J. R. Schmidt, "Physically-Motivated Force Fields from Symmetry-Adapted Perturbation Theory," *J. Phys. Chem. A*, vol. 117, no. 10, pp. 2053–2066, Mar. 2013, doi: 10.1021/jp3108182.
- [18] A. T. Nasrabadi and L. D. Gelb, "Structural and Transport Properties of Tertiary Ammonium Triflate Ionic Liquids: A Molecular Dynamics Study," *J. Phys. Chem. B*, vol. 121, no. 8, pp. 1908–1921, Mar. 2017, doi: 10.1021/acs.jpcc.6b12418.
- [19] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids," *J. Am. Chem. Soc.*, vol. 118, no. 45, pp. 11225–11236, Jan. 1996, doi: 10.1021/ja9621760.
- [20] T. G. A. Youngs and C. Hardacre, "Application of Static Charge Transfer within an Ionic-Liquid Force Field and Its Effect on Structure and Dynamics," *ChemPhysChem*, vol. 9, no. 11, pp. 1548–1558, Aug. 2008, doi: 10.1002/cphc.200800200.
- [21] Lindahl, Abraham, Hess, & van der Spoel. (2020). GROMACS 2020.2 Manual (2020.2). Zenodo. <https://doi.org/10.5281/zenodo.3773799>
- [22] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI Force Field: Coarse Grained Model for Biomolecular

- Simulations,” *J. Phys. Chem. B*, vol. 111, no. 27, pp. 7812–7824, Jul. 2007, doi: 10.1021/jp071097f.
- [23] C. M. Breneman and K. B. Wiberg, “Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis,” *J. Comput. Chem.*, vol. 11, no. 3, pp. 361–373, Apr. 1990, doi: 10.1002/jcc.540110311.
- [24] Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. Jorgensen, W. L.; Tirado-Rives, J. *Proc. Nat. Acad. Sci. USA* 2005, 102, 6665-6670
- [25] 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B*, 2017, 121 (15), pp 3864-3870
- [26] LigParGen web server: An automatic OPLS-AA parameter generator for organic ligands. Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. *Nucleic Acids Research*, Volume 45, Issue W1, 3 July 2017, Pages W331-W336
- [27] A. K. Malde et al., “An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0,” *J. Chem. Theory Comput.*, vol. 7, no. 12, pp. 4026–4037, Dec. 2011, doi: 10.1021/ct200196m.
- [28] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, “PACKMOL: A package for building initial configurations for molecular dynamics simulations,” *J. Comput. Chem.*, vol. 30, no. 13, pp. 2157–2164, Oct. 2009, doi: 10.1002/jcc.21224.
- [29] L. I. Vazquez-Salazar, M. Selle, A. H. de Vries, S. J. Marrink, and P. C. T. Souza, “Martini coarse-grained models of imidazolium-based ionic liquids: from nanostructural organization to liquid–liquid extraction,” *Green Chem.*, vol. 22, no. 21, pp. 7376–7386, 2020, doi: 10.1039/D0GC01823F.
- [30] R. Alessandri et al., “Martini 3 Coarse-Grained Force Field: Small Molecules,” *Chemistry*, preprint, Jul. 2021. doi: 10.33774/chemrxiv-2021-1qmq9.
- [31] S. J. Marrink et al., “The Martini 3.0 CG Force Field: Open beta (version 3.0.b.3.2), 2018
- [32] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, “Avogadro: an advanced semantic chemical editor, visualization, and analysis platform,” *J. Cheminformatics*, vol. 4, no. 1, p. 17, Dec. 2012, doi: 10.1186/1758-2946-4-17.
- [33] H. Tokuda, K. Hayamizu, K. Ishii, Md. A. B. H. Susan, and M. Watanabe, “Physicochemical Properties and Structures of Room Temperature Ionic Liquids. 2. Variation of Alkyl Chain Length in Imidazolium Cation,” *J. Phys. Chem. B*, vol. 109, no. 13, pp. 6103–6110, Apr. 2005, doi: 10.1021/jp044626d.

- [34] P. P. Lamprinidis, “Modeling & Optimization of Gas Diffusion Layer in Fuel Cells, in the OpenFOAM environment,” thesis.
- [35] J. Yuan, D. Mecerreyes, and M. Antonietti, “Poly(ionic liquid)s: An update,” *Prog. Polym. Sci.*, vol. 38, no. 7, pp. 1009–1036, Jul. 2013, doi: 10.1016/j.progpolymsci.2013.04.002.
- [36] Lin B, Qiu L, Lu J, Yan F. Cross-linked alkaline ionic liquid-based polymer electrolytes for alkaline fuel cell applications. *Chemical Materials* 2010;22:6718–25.
- [37] Ran J, Wu L, Varcoe JR, Ong AL, Poynton SD, Xu T. Development of imidazolium-type alkaline anion exchange membranes for fuel cell application. *Journal of Membrane Science* 2012;415–416:242–9.
- [38] Qiu B, Lin B, Qiu L, Yan F. Alkaline imidazolium- and quaternary ammonium-functionalized anion exchange membranes for alkaline fuel cell applications. *Journal of Material Chemistry* 2012;22:1040–5.
- [39] O. I. Deavin, S. Murphy, A. L. Ong, S. D. Poynton, R. Zeng, H. Herman, J. R. Varcoe. Anion-exchange membranes for alkaline polymer electrolyte fuel cells: comparison of pendent benzyltrimethylammonium- and benzylmethylimidazolium-head-groups. *Energy and Environmental Science* 2012;5:8584–97.
- [40] J. Pozuelo, E. Riande, E. Saiz, and V. Compañ, “Molecular Dynamics Simulations of Proton Conduction in Sulfonated Poly(phenyl sulfone)s,” *Macromolecules*, vol. 39, no. 25, pp. 8862–8866, Dec. 2006, doi: 10.1021/ma062070h.
- [41] M. Griebel, S. Knapek, and G. W. Zumbusch, *Numerical simulation in molecular dynamics: numerics, algorithms, parallelization, applications*. Berlin: Springer, 2007.
- [42] Rapaport, D. (2004). *The Art of Molecular Dynamics Simulation* (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511816581
- [43] S. Banerjee and D. E. Curtin, “Nafion® perfluorinated membranes in fuel cells,” *J. Fluor. Chem.*, vol. 125, no. 8, pp. 1211–1216, Aug. 2004, doi: 10.1016/j.jfluchem.2004.05.018.
- [44] T. M. Chang, L. X. Dang, R. Devanathan, and M. Dupuis, “Structure and Dynamics of N,N-Diethyl-N-methylammonium Triflate Ionic Liquid, Neat and with Water, from Molecular Dynamics Simulations,” *J. Phys. Chem. A*, vol. 114, no. 48, pp. 12764–12774, Dec. 2010, doi: 10.1021/jp108189z.
- [45] H. A. Posch, W. G. Hoover, and F. J. Vesely, “Canonical dynamics of the Nosé oscillator: Stability, order, and chaos,” *Phys. Rev. A*, vol. 33, no. 6, pp. 4253–4265, Jun. 1986, doi: 10.1103/PhysRevA.33.4253.

- [46] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, “Molecular dynamics with coupling to an external bath,” *J. Chem. Phys.*, vol. 81, no. 8, pp. 3684–3690, Oct. 1984, doi: 10.1063/1.448118.
- [47] A. Rahman, “Correlations in the Motion of Atoms in Liquid Argon,” *Phys. Rev.*, vol. 136, no. 2A, pp. A405–A411, Oct. 1964, doi: 10.1103/PhysRev.136.A405.
- [48] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *J. Chem. Phys.*, vol. 126, no. 1, p. 014101, Jan. 2007, doi: 10.1063/1.2408420.
- [49] K. C. Giannakoglou *Optimization Methods in Aerodynamics*. Notes. Laboratory of Thermal Turbomachines, 2005; NTUA, Athens, 2005
- [50] E. M. Papoutsis-Kiachagias and K. C. Giannakoglou, “Continuous Adjoint Methods for Turbulent Flows, Applied to Shape and Topology Optimization: Industrial Applications,” *Arch. Comput. Methods Eng.*, vol. 23, no. 2, pp. 255–299, Jun. 2016, doi: 10.1007/s11831-014-9141-9.
- [51] N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, “Open Babel: An open chemical toolbox,” *J. Cheminformatics*, vol. 3, no. 1, p. 33, Dec. 2011, doi: 10.1186/1758-2946-3-33.
- [52] S. Kim, J. Chen, T. Cheng, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2021;49(D1):D1388–D1395. doi:10.1093/nar/gkaa971
- [53] B. D. Ray, *topolbuild*, IUPUI Physics Dept. 402 N. Blackford St. Indianapolis, IN 46202-3273.
- [54] A Smart Chem-Search Engine, www.chemsrc.com
- [55] EASY webpage. <http://velos0.ltt.mech.ntua.gr/EASY>.
- [56] DuPont Fuel Cells, Chestnut Run Plaza CRP701/213, Wilmington, DE 19805-0701, USA
- [57] DuPont Fuel Cells, 22828 NC Highway 87W, Fayetteville, NC 28306, USA
- [58] A. P. Sunda, A. Mondal, and S. Balasubramanian, “Atomistic simulations of ammonium-based protic ionic liquids: steric effects on structure, low frequency vibrational modes and electrical conductivity,” *Phys. Chem. Chem. Phys.*, vol. 17, no. 6, pp. 4625–4633, 2015, doi: 10.1039/C4CP05353B.
- [59] R. Pant, S. Sengupta, A. V. Lyulin, and A. Venkatnathan, “Computational investigation of a protic ionic liquid doped poly-benzimidazole fuel cell electrolyte,” *J. Mol. Liq.*, vol. 314, p. 113686, Sep. 2020, doi: 10.1016/j.molliq.2020.113686.

- [60] S. Liu et al., “Ionic-Liquid-Based Proton Conducting Membranes for Anhydrous H₂/Cl₂ Fuel-Cell Applications,” *ACS Appl. Mater. Interfaces*, vol. 6, no. 5, pp. 3195–3200, Mar. 2014, doi: 10.1021/am404645c.
- [61] T. Yan, C. J. Burnham, M. G. Del Pópolo, and G. A. Voth, “Molecular Dynamics Simulation of Ionic Liquids: The Effect of Electronic Polarizability,” *J. Phys. Chem. B*, vol. 108, no. 32, pp. 11877–11881, Aug. 2004, doi: 10.1021/jp047619y.
- [62] D. Bedrov, J.-P. Piquemal, O. Borodin, A. D. MacKerell, B. Roux, and C. Schröder, “Molecular Dynamics Simulations of Ionic Liquids and Electrolytes Using Polarizable Force Fields,” *Chem. Rev.*, vol. 119, no. 13, pp. 7940–7995, Jul. 2019, doi: 10.1021/acs.chemrev.8b00763.
- [63] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, “A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model,” *J. Phys. Chem.*, vol. 97, no. 40, pp. 10269–10280, Oct. 1993, doi: 10.1021/j100142a004.
- [64] R. Bader, *Atoms in Molecules: a quantum theory*, Oxford University Press, 1990.
- [65] P. E. Blochl, *J. Chem. Phys.*, 1995, 103, 7422–7428.

Appendix A

Appendix-All atom simulations of ionic liquids

Parameter name	Parameter value	Description
integrator	steep	Minimization algorithm
emtol	10	Minimization stops when $F_{max} < emtol$
emstep	0.001	Minimization step size
nsteps	80000	Total number of steps
nstlog	2000	Frequency of printing values in the log file
nstenergy	1000	Frequency of printing energy values in the .edr file
constraint-algorithm	lincs	Method used for constrains
constraints	h-bonds	Convert the bonds with H-atoms to constrains
nstlist	20	Frequency to update the neighbour list and long range forces
cutoff-scheme	Verlet	The method of cutoff scheme used
ns_type	grid	Method to determine the neighbour list
coulombtype	PME	Method for long range electrostatic interactions
vdwtype	cut-off	Method used for the Van der Waals forces
pbc	xyz	Periodic boundary conditions in all 3 dimensions
DispCorr	EnerPres	Applies long range dispersion corrections for Energy and Pressure

Table A.1: *Typical parameters for the minimization .mdp file in AA simulations*

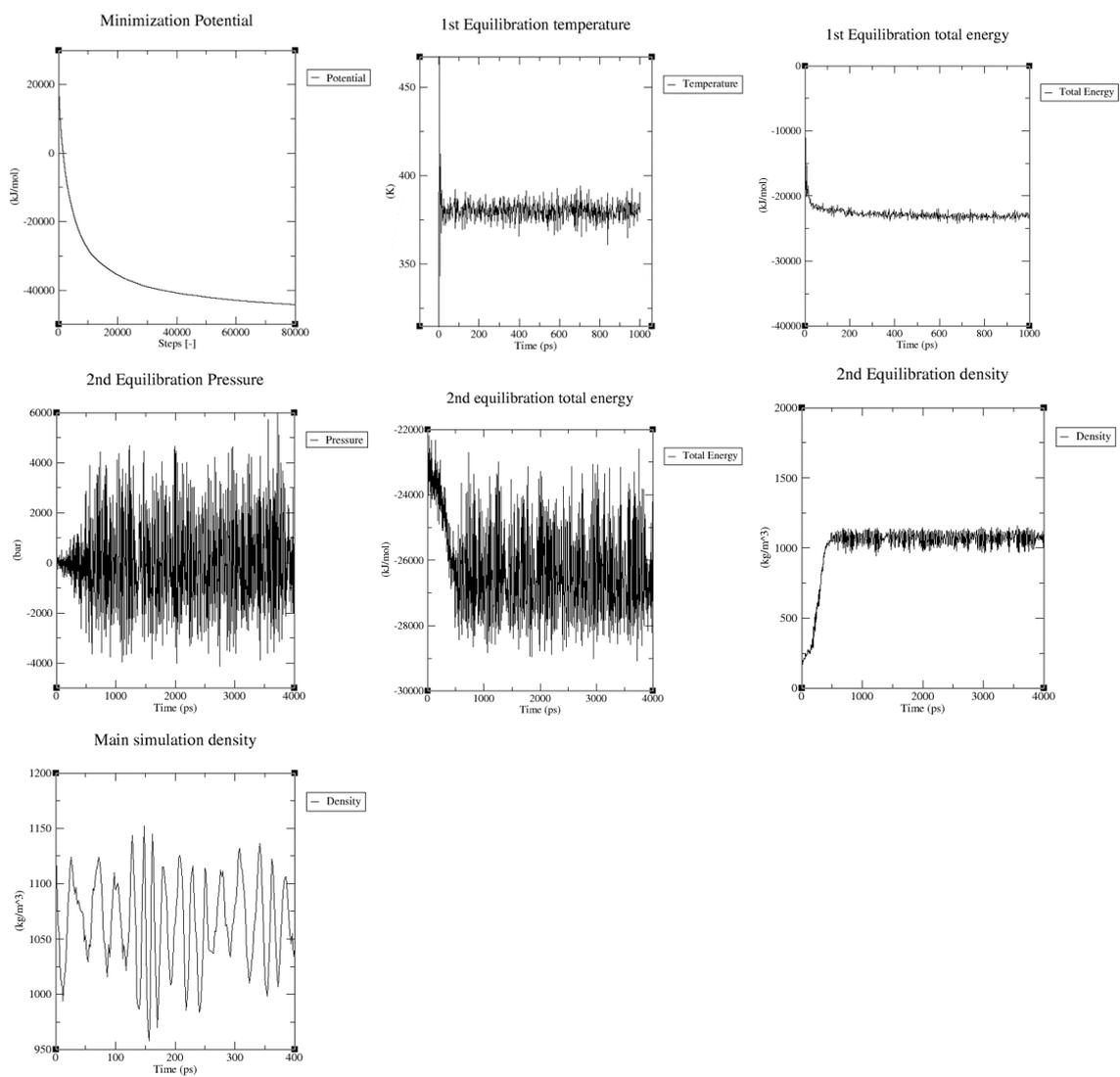


Figure A.1: *Indicative plots for convergence verification*

Parameter name	Parameter value	Description
integrator	md	Leap-grog algorithm for integrating Newton's equation of motion
dt	0.002	Timestep of integration [ps]
nsteps	$5.0E + 5$	Total number of steps for the 1st equilibration
nstlog	2000	Frequency of printing values in the log file
nstenergy	1000	Frequency of printing energy values in the .edr file
gen-vel	yes	Generate velocities for molecules according to Maxwell distribution
gen-temp	300	Temperature for the gen-vel option
constraint-algorithm	lincs	Method used for constrains
constraints	h-bonds	Convert the bonds with H-atoms to constrains
cutoff-scheme	Verlet	The method of cutoff scheme used
coulombtype	PME	Method for long range electrostatic interactions
vdwtype	cut-off	Method used for the Van der Waals forces
rcoulomb	1.6	The distance for the Coulomb cut-off [nm]
rvdw	1.4	The distance for the Lennard-Jones cut-off
DispCorr	EnerPres	Applies long range dispersion corrections for Energy and Pressure
tcoupl	Nose-Hoover	Type of thermostat used
tc-grps	System	Groups to couple to separate temperature baths
tau-t	1.5	Time constant for temperature coupling [ps]
ref-t	300	Reference-desired temperature value for coupling [K]
nsteps	$2.0E + 6$	Total number of steps for the 2nd equilibration
pcoupl	Parrinello-Rahman	Type of barostat used
tau_p	3.01	Time constant for temperature coupling [ps]
compressibility	$4.46E - 5$	Compressibility value for pressure coupling [bar^{-1}]
ref_p	1.0	Reference-desired pressure value for coupling [bar]
nsteps	$2.0E + 5$	Total number of steps for the main simulation run
nstvout	4	Step frequency for printing velocities for each molecule in the system
nstxout-compressed	5	Step frequency for printing coordinates for each molecule in the system
pcoupl	Parrinello-Rahman	Type of barostat used
tau_p	3.01	Time constant for temperature coupling [ps]
compressibility	$4.46E - 5$	Compressibility value for pressure coupling [bar^{-1}]
ref_p	1.0	Reference-desired pressure value for coupling [bar]

Table A.2: Typical .mdp file parameters for the 1st equilibration, the 2nd equilibration and the main simulation run in AA simulations. The highlighted parameters are unique for each type of simulation run. Red for the 1st equilibration, blue for the 2nd equilibration and green for the main simulation run.

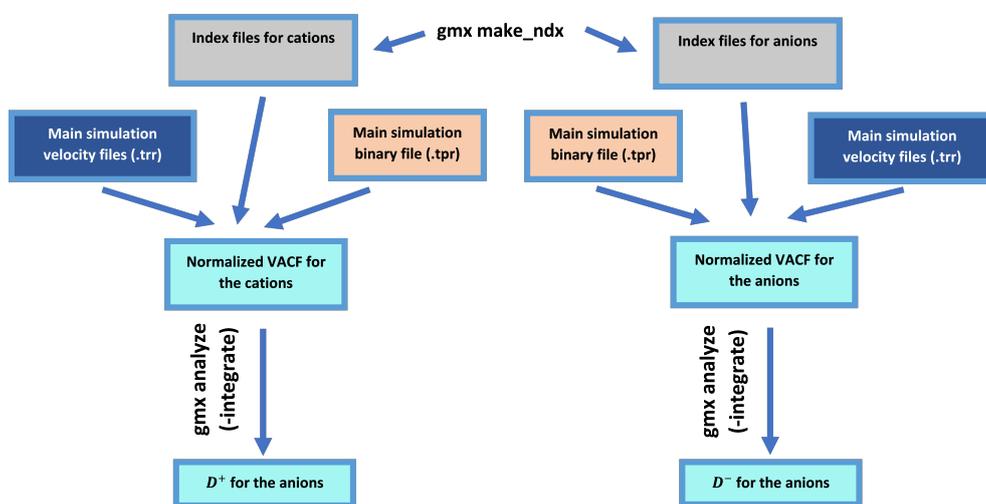


Figure A.2: Flow diagram of the self diffusion constant calculation procedure, using the Green-Kubo method.

Appendix B

Appendix-Coarse grained simulations of ionic liquids

Parameter name	Parameter value	Description
integrator	steep	Minimization algorithm
emtol	10	Minimization stops when $F_{max} < emtol$
emstep	0.001	Minimization step size
nsteps	50000	Total number of steps
nstlog	100	Frequency of printing values in the log file
nstenergy	100	Frequency of printing energy values in the .edr file
constraint-algorithm	lincs	Method used for constrains
constraints	none	Convert the bonds with H-atoms to constrains
nstlist	20	Frequency to update the neighbour list and long range forces
cutoff-scheme	Verlet	The method of cutoff scheme used
ns_type	grid	Method to determine the neighbour list
coulombtype	reaction-field	Method for long range electrostatic interactions
vdwtype	cut-off	Method used for the Van der Waals forces
pbc	xyz	Periodic boundary conditions in all 3 dimensions
DispCorr	EnerPres	Applies long range dispersion corrections for Energy and Pressure

Table B.1: *The parameters for the minimization .mdp file in CG simulations*

Parameter name	Parameter value	Description
integrator	md	Leap-grog algorithm for integrating Newton's equation of motion
dt	0.01	Timestep of integration [ps]
nsteps	$5.0E + 4$	Total number of steps for the 1st equilibration
nstlog	1000	Frequency of printing values in the log file
nstenergy	100	Frequency of printing energy values in the .edr file
gen-vel	yes	Generate velocities for molecules according to Maxwell distribution
gen-temp	300	Temperature for the gen-vel option
constraint-algorithm	lincs	Method used for constrains
constraints	none	Convert the bonds with H-atoms to constrains
cutoff-scheme	Verlet	The method of cutoff scheme used
coulombtype	reaction-field	Method for long range electrostatic interactions
vdwtype	cut-off	Method used for the Van der Waals forces
rcoulomb	1.6	The distance for the Coulomb cut-off [nm]
rvdw	1.4	The distance for the Lennard-Jones cut-off
tcoupl	v-rescale	Type of thermostat used
tc-grps	System	Groups to couple to separate temperature baths
tau-t	1.0	Time constant for temperature coupling [ps]
ref-t	300	Reference-desired temperature value for coupling [K]
nsteps	$4.5E + 5$	Total number of steps for the 2nd equilibration
pcoupl	Parrinello-Rahman	Type of barostat used
tau_p	10.0	Time constant for temperature coupling [ps]
compressibility	$1.0E - 4$	Compressibility value for pressure coupling [bar^{-1}]
ref_p	1.0	Reference-desired pressure value for coupling [bar]
nsteps	$5.0E + 4$	Total number of steps for the main simulation run
nstvout	10	Step frequency for printing velocities for each molecule in the system
nstxout-compressed	10	Step frequency for printing coordinates for each molecule in the system
pcoupl	Parrinello-Rahman	Type of barostat used
tau_p	10.0	Time constant for temperature coupling [ps]
compressibility	$1.0E - 4$	Compressibility value for pressure coupling [bar^{-1}]
ref_p	1.0	Reference-desired pressure value for coupling [bar]

Table B.2: *The .mdp file parameters for the 1st equilibration, the 2nd equilibration and the main simulation run in the coarse grained models. The highlighted parameters are unique for each type of simulation run. Red for the 1st equilibration, blue for the 2nd equilibration and green for the main simulation run.*

-Blocks with different interaction levels

organic
water

ion
ion-organic/water

	Q2	Q1	Qn	Qp	Q0		P5	P4	P3	P2	P1	N3	N2	N1	N0	C6	C5	C4	C3	C2	C1		W	
Q2	13	12	12	12	12		1	2	3	3	4	4	5	6	7	14	16	18	20	21	21		0	Q2
Q1	12	0	1	1	10		2	3	4	4	4	5	6	7	8	13	15	17	18	21	21		1	Q1
Qn	12	1	8	4	10		4	5	5	5	5	6	7	8	9	12	14	16	17	20	20		2	Qn
Qp	12	1	4	8	10		4	5	5	5	5	6	7	8	9	11	12	14	15	20	20		2	Qp
Q0	12	10	10	10	9		5	5	5	5	5	6	7	8	9	10	12	14	19	19		3	Q0	
P5	1	2	4	4	5		4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		4	P5
P4	2	3	5	5	5		5	5	6	7	8	9	10	11	12	12	13	14	15	16	17		5	P4
P3	3	4	5	5	5		6	6	7	7	8	9	10	11	12	12	13	14	15	16	17		6	P3
P2	3	4	5	5	5		7	7	7	7	8	8	9	10	11	11	12	13	14	15	16		6	P2
P1	4	4	5	5	5		8	8	8	8	8	8	9	10	10	11	12	13	14	15	16		7	P1
N3	4	5	6	6	6		9	9	9	8	8	8	9	10	10	11	11	12	13	14	15		7	N3
N2	5	6	7	7	7		10	10	10	9	9	9	9	10	10	11	11	12	13	14	15		8	N2
N1	6	7	8	8	8		11	11	11	10	10	10	10	10	10	11	11	12	12	13	14		9	N1
N0	7	8	9	9	9		12	12	12	11	10	10	10	10	10	11	11	12	12	13	13		10	N0
C6	14	13	12	11	9		13	12	12	11	11	11	11	11	11	10	10	11	12	13	13		12	C6
C5	16	15	14	12	10		14	13	13	12	12	11	11	11	11	10	10	11	11	13	13		13	C5
C4	18	17	16	14	12		15	14	14	13	13	12	12	12	12	11	11	11	11	13	13		15	C4
C3	20	18	17	15	14		16	15	15	14	14	13	13	12	12	12	11	11	11	12	12		16	C3
C2	21	21	20	20	19		17	16	16	15	15	14	14	13	13	13	13	13	12	11	11		17	C2
C1	21	21	20	20	19		18	17	17	16	16	15	15	14	13	13	13	13	12	11	11		18	C1
W	0	1	2	2	3		4	5	6	6	7	7	8	9	10	12	13	15	16	17	18		4	W
	Q2	Q1	Qn	Qp	Q0		P5	P4	P3	P2	P1	N3	N2	N1	N0	C6	C5	C4	C3	C2	C1		W	

v2.2 levels	0		1		2		3		4		5		6		7		8		9			
v3.0 levels	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	hyper attractive	supra attractive	attractive	almost attractive	semi attractive	intermediate	almost intermediate	semi repulsive	almost repulsive	repulsive	super repulsive											

Table B.3: Levels of interactions between all the different bead types in the Martini 3.0 force field [31].

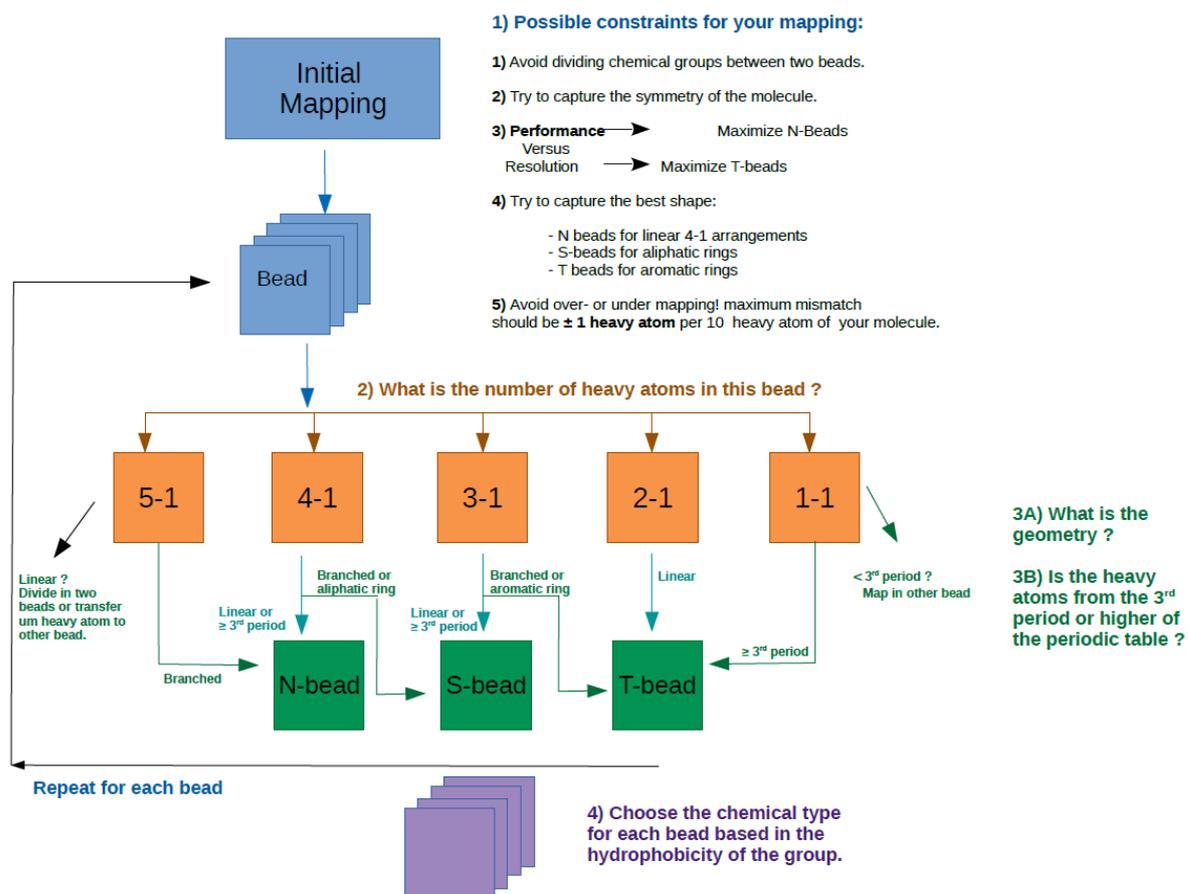


Figure B.1: Flow diagram of the parametrization algorithm for new molecules in Martini 3.0 [31].

Oil-water partitions - kJ/mol				Examples of possible usage							
Normal	HD->WN	OCOS->WN		name	structure	name	structure	name	structure	name	structure
C1	18.2	18.3		butane	C-C-C-C	neopentane	C-(CH3)4 *	2-methyl-butane	C-C-(CH3)-C-C *	2-methyl-butane	C-C-(CH3)-C-C *
C2	16.5	17.5									
C3	12.8	15.2		butene	C=C-C-C						
C4	8.9	12.8		chloro propane	C-C-C-Cl	propanethiol	C-C-C-SH	buta-1,3-diene	C=C-C=C	ethyl methyl sulfide	C-C-S-C
C5	5.2	10.1		butyne	C≡C-C-C						
C6	3.5	9.1		imine	C-C=N(CH3)-C *						
N0	-0.2	6.2		diethyl ether	C-C-O-C-C						
N1	-4.7	4.0									
N2	-8.3	1.9		ethylmethylamine	C-NH-C-C						
N3	-10.5	1.0		propanol	C-C-C-OH	2-methyl-propanol	C-C(CH3)-C-OH				
P1	-12.7	-1.1		prop 2-en-1-ol	-C=C-C-OH						
P2	-14.7	-2.6		propanoic acid	-C-C-COOH	N-methylacetamide					
P3	-16.9	-4.2		propylene glycol	-C(OH)-C(OH)-C						
P4	-19.4	-6.4		propanamide	-C-C-CONH2						
P5	-23.9	-10.6		amino acid	NH3+-C-COO-						
Q0	-34.5	-12.4		choline	C-C-N-(C)3						
Qp	-46.5	-15.5		propanoate	-C-C-COO(-)						
Qn	-46.4	-14.7		propyl-ammonium	-C-C-C-NH2(+)						
Q1	-60.6	-16.5		phosphate -1	-PO4(-1)-						
Q2	-63.2	-16.9		phosphate -2	-PO4(-2)	sulfate	-SO4(-2)				
N0d/a	1.4	7.8						diethyl ether	C-C-O-C-C		
N1d/a	-2.1	6.4		butanone (a)	-C-C(=O)-C-C-			ethyl methyl ether	C-O-C-C		
N2d/a	-6.3	3.4		propanal (a)	-C-C-C=O	n-propylamine (d)	-C-C-C-NH2	methyl acetate (a)	-C-O-C(=O)-		
N3d/a	-8.2	2.5									
P1d/a	-10.6	1.0									
P2d/a	-12.8	-0.9		1,3-dicarbonyl	-C(=O)-C-C(=O)-						
P3d/a	-14.7	-2.0		n,n-dimethylformamide (a)	C(=O)-N-(CH3)2						
P4d/a	-16.4	-3.7									
P5d/a	-21.0	-7.5									
Small	HD->WN	OCOS->WN		name	structure	name	structure	name	structure	name	structure
SC1	16.2	14.0		2-methyl-propane	C-C(CH3)C*						
SC2	14.5	13.0		propane	C-C-C	cyclohexane ?	(-C-C-C)-				
SC3	10.7	10.3		propene	C=C-C	cyclopropane	C-C-C	cyclohexane ?	(-C-C-C)-		
SC4	6.8	8.0		chloro ethane	C-C-Cl	ethane-thiol	C-C-SH	3-1 - conjugated	-(C=C-C)=C	dimethyl sulfide	C-S-C
SC5	3.4	5.3		propyne	C≡C-C						
SC6	1.5	4.4		imine - conjugated	C-C-N-C *						
SN0	-2.0	1.4		dimethyl ether?	-C-O-C-						
SN1	-5.7	0.2									
SN2	-9.9	-2.2		dimethylamine	C-NH-C						
SN3	-12.2	-3.2		2-propanol	C-C(CH3)-C-OH*						
SP1	-14.0	-4.2		ethanol	C-C-OH						
SP2	-16.0	-6.1		acetic acid	C-C-COOH*						
SP3	-17.7	-6.8		ethylene glycol	-C(OH)-C(OH)*						
SP4	-20.5	-9.2		acetamide	-C-CONH2						
SP5	-24.5	-12.5									
SQ0	-47.1	-18.8		trimethyl amonium							
SQp	-58.0	-21.8		acetate	-C-COO(-)						
SQn	-58.7	-21.8		ethyl-ammonium	-C-C-NH2(+)						
SQ1	-70.8	-21.9		hydrated chloride	Cl(-) (H2O)2	hydrated sodium	Na(+)(H2O)2				
SQ2	-112.4	-33.7		hydrated calcium	Ca(2+) (H2O)2						
SN0d/a	-0.3	3.3									
SN1d/a	-4.2	1.4		propanone (a)	-C-C(=O)-C-	methyl formate (a)	-O-C(=O)-	dimethyl ether in polyethers?	-C-O-C-	dimethyl ether in polyethers?	-C-O-C-
SN2d/a	-7.9	-0.3		ethanal (a)	-C-C=O			dimethyl ether in rings	-C-O-C-		
SN3d/a	-9.9	-1.3		ethyl amine (d)	-C-C-NH2						
SP1d/a	-12.0	-2.5									
SP2d/a	-13.9	-4.1									
SP3d/a	-15.9	-5.1									
SP4d/a	-18.1	-7.1									
SP5d/a	-21.9	-9.7									
Tiny	HD->WN	OCOS->WN		name	structure	name	structure	name	structure	name	structure
TC1	14.4	12.3		isopropyl group	-C(CH3)-CH3						
TC2	11.9	10.5		ethane	-C-C-						
TC3	8.2	8.0		ethene	-C=C-	ethyl near to polar group	-C-C-				
TC4	4.3	6.2		chloromethane	-C-Cl	thiol- group	-C-SH	2-1 - conjugated/aromatic	-(C=C)-C=C	sulfide group	-C-S-
TC5	1.2	3.7		ethyne	-C≡C-	thiol- conjugated/aromatic	-C-SH				
TC6	-0.5	2.8		imine - conjugated	C-N-C *						
TN0	-3.5	0.5		ether group	-C-O-						
TN1	-7.4	-1.5									
TN2	-10.9	-3.8									
TN3	-13.2	-5.1									
TP1	-15.7	-5.9		methanol	C-OH						
TP2	-17.7	-7.6									
TP3	-19.4	-8.3									
TP4	-22.3	-10.8									
TP5	-26.6	-13.4									
TQ0	-62.2	-17.4		heavy metals complexes	M(+)						
TQp	-76.4	-18.8		methylammonium	-C-NH2(+)						
TQn	-76.5	-18.5									
TQ1	-93.5	-22.5		dehydrated chloride	Cl(-)	dehydrated sodium	Na(+)				
TQ2	-198.5	-29.1		dehydrated calcium	Ca(2+)						
TN0d/a	-1.7	2.0									
TN1d/a	-5.6	0.2		ether -conjugated -aromatic	-C-O-						
TN2d/a	-8.9	-1.8		carbonyl group	-C=O						
TN3d/a	-11.2	-3.4		methyl-amine	C-NH2						
TP1d/a	-13.5	-3.9		carbonyl group in nucleot.	-C=O						
TP2d/a	-15.6	-5.8									
TP3d/a	-17.2	-6.4									
TP4d/a	-19.5	-8.4									
TP5d/a	-23.9	-10.9									

* overlapping but branched or part of ring
(a) and (d) indicate that you should use an acceptor or a donor version of the head, respectively.
OBS: Most of the examples correspond to chemical groups attached to aliphatic molecules.
The head types can change depending of the situation (for example, chemical groups attached to aromatic rings).

Table B.4: Table with suggestions about coarse grained mapping in the Martini 3.0 force field [31].



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Μηχανολόγων Μηχανικών

Τομέας Ρευστών

Μονάδα Παράλληλης Υπολογιστικής Ρευστοδυναμικής
& Βελτιστοποίησης

Μοντελοποίηση με χρήση Μοριακής Δυναμικής και
Βελτιστοποίηση της Πρωτονιακής Αγωγιμότητας σε
(Πολυμερισμένα) Ιονικά Υγρά για Κυψέλες Καυσίμου
Μεμβράνης Ανταλλαγής Πρωτονίων.

Διπλωματική Εργασία

Ματθαίος Ν. Χατζόπουλος

Επιβλέπων :

Κυριάκος Χ. Γιαννάκογλου, Καθηγητής ΕΜΠ

Βιομηχανικός Επιβλέπων :

Δρ. Κωνσταντίνος Γκαγκάς, Toyota Motor Europe

Αθήνα, Φεβρουάριος 2022

Εκτενής Περίληψη στα Ελληνικά

Εισαγωγή

Στην προσπάθειά της να μειώσει τις εκπομπές διοξειδίου του άνθρακα, η αυτοκινητοβιομηχανία αναπτύσσει κυψέλες καυσίμου με την τεχνολογία μεμβράνης ανταλλαγής πρωτονίων. Η μεμβράνη είναι ένα από τα βασικότερα εξαρτήματα σε μία κυψέλη καυσίμου και το υλικό το οποίο έχει επικρατήσει να χρησιμοποιείται στις μεμβράνες είναι το nafionTM. Ανεξαρτήτως του υλικού της μεμβράνης, μία κυψέλη καυσίμου είναι επιθυμητό να λειτουργεί σε υψηλές θερμοκρασίες, άνω των 120 °C, εξαιτίας του αυξημένου βαθμού απόδοσης της μηχανής σε αυτές. Ωστόσο, αυτό είναι αδύνατον να

πραγματοποιηθεί με χρήση του nafionTM, καθώς απαιτεί συνθήκες υψηλής υγρασίας για να λειτουργήσει και, επομένως, θερμοκρασίες μικρότερες από 100 °C.

Για να λυθεί το προηγούμενο πρόβλημα, εξετάζεται η αντικατάσταση της μεμβράνης από nafionTM με μία άλλη μεμβράνη που θα προκύψει πολυμερίζοντας το ανιόν ή το κατιόν σε ένα ιονικό υγρό. Τα ιονικά υγρά είναι οργανικά άλατα με σημείο τήξης μικρότερο των 100 °C. Αν και οι τρέχουσες μεμβράνες αυτού του τύπου μπορούν να λειτουργούν αποτελεσματικά σε πολύ υψηλές θερμοκρασίες, έχουν περίπου δύο τάξεις μεγέθους μικρότερη πρωτονιακή αγωγιμότητα σε σχέση με τις μεμβράνες από nafionTM. Για τον λόγο αυτό, οι μεμβράνες που προκύπτουν από πολυμερισμό ιονικών υγρών δεν είναι ακόμη σε θέση να αντικαταστήσουν τις μεμβράνες από nafionTM και απαιτείται βελτίωση/βελτιστοποίησή τους με την αγωγιμότητα ως συνάρτησή-στόχο.

Στη διπλωματική αυτή εργασία αναζητούνται εναλλακτικά ζεύγη κατιόντων και ανιόντων (ιονικά υγρά) με υψηλότερη πρωτονιακή αγωγιμότητα, εφαρμόζοντας στοχαστικές μεθόδους βελτιστοποίησης, ώστε να λυθεί το προηγούμενο πρόβλημα. Κατά τη διαδικασία αυτή θεωρείται πως η αγωγιμότητα ενός ιονικού υγρού και της αντίστοιχης μεμβράνης που παράγεται από αυτό είναι μεγέθη ανάλογα. Η ρεαλιστική αυτή παραδοχή γίνεται λόγω του απαγορευτικά υψηλού υπολογιστικού κόστους που έχει η προσομοίωση μεμβρανών. Ο υπολογισμός της αγωγιμότητας πραγματοποιείται μέσω προσομοίωσης μοριακής δυναμικής (Molecular Dynamics ή MD) και με χρήση του λογισμικού ανοικτού κώδικα GROMACS. Το λογισμικό αυτό συνοδεύει ένα πλήθος λογισμικών, καθώς και συνδυαστικών κωδικών, για να είναι εφικτή η αξιολόγηση του κάθε ιονικού υγρού, αλλά και η παραγωγή γεωμετρίας και τοπολογίας ιόντων από τις παραμέτρους της βελτιστοποίησης.

Μοριακή δυναμική είναι η επιστήμη που μελετά κινήσεις μορίων ώστε να εξαχθούν μικροσκοπικές ή μακροσκοπικές ιδιότητες του εκάστοτε υλικού που μελετάται. Η έννοια της μοριακής δυναμικής συχνά ταυτίζεται με την τεχνική προσομοίωσης κατά την οποία εισάγονται πολλά μόρια του υπό εξέταση υλικού εντός ενός κύβου προσομοίωσης.

Σε ένα μοντέλο που συμπεριλαμβάνει όλα τα άτομα (All atom ή AA model), το κάθε μόριο αποτελείται από άτομα υπό τη μορφή σφαιριδίων τα οποία κινούνται σύμφωνα με τις κινηματικές εξισώσεις του Νεύτωνα. Τα σφαιρίδια αλληλεπιδρούν μόνο εξ αποστάσεως μέσω ηλεκτροστατικών και διαμοριακών δυνάμεων Van der Waals. Οι κάθε λογής ταλαντώσεις οι οποίες οφείλονται στη σύνδεση των ατόμων με χημικούς δεσμούς μοντελοποιούνται με τη βοήθεια ελατηρίων σε μία θεώρηση κλασικής μηχανικής. Οι προηγούμενες αλληλεπιδράσεις συνοψίζονται στην παράγραφο 1.7.1 του πλήρους κειμένου. Οι μαθηματικές εξισώσεις μοντελοποίησης των προηγούμενων δυνάμεων, καθώς και των τεχνασμάτων επιβολής των εξωτερικών συνθηκών πίεσης και θερμοκρασίας εισάγονται στον 2ο νόμο του Νεύτωνα ως επιπλέον όροι.

Σε ένα αδρομερές μοντέλο (Coarse-grained ή CG model), η λογική της μοριακής δυναμικής δεν αλλάζει καθόλου. Ωστόσο, τα διάφορα ατομικά συμπλέγματα, εντός των μορίων που προμοιώνονται, ομαδοποιούνται στις λεγόμενες "χάντρες" (beads).

Τα σφαιρίδια της προσομοίωσης δεν είναι πλέον μεμονωμένα άτομα αλλά οι χάντρες. Η θεώρηση αυτή μειώνει το υπολογιστικό κόστος κατά 1 με 2 τάξεις μεγέθους.

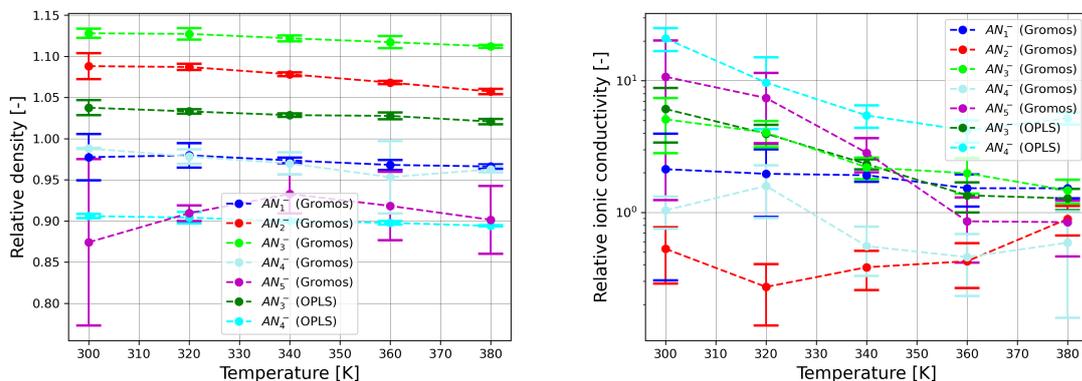
Στο 1ο κεφάλαιο της διπλωματικής, εξετάζονται αρκετά διαφορετικά μοντέλα, για να διαπιστωθεί πιο ταιριάζει περισσότερο στις απαιτήσεις του τρέχοντος προβλήματος. Εν τέλει, επιλέγονται τα μοντέλα OPLS-AA και GROMOS ως AA μοντέλα, και το MARTINI ως CG μοντέλο.

AA Προσομοιώσεις

Δημιουργείται μία ολόκληρη δομή προσομοίωσης, με επίκεντρο το ανοικτού κώδικα λογισμικό μοριακής δυναμικής GROMACS και με στόχο τον υπολογισμό της ιονικής αγωγιμότητας με τη μεγαλύτερη δυνατή ακρίβεια, που ένα μοντέλο AA μπορεί να παρέχει. Η διαδικασία και οι διάφορες ρυθμίσεις της περιγράφονται λεπτομερώς. Περιγράφονται και δοκιμάζονται δύο μέθοδοι υπολογισμού ιονικής αγωγιμότητας (MSD και Green-Kubo) από τη βιβλιογραφία. Τελικώς επιλέγεται η μέθοδος MSD (σχέση 2.3), επειδή έδωσε πιο ακριβή αποτελέσματα με τις συγκεκριμένες ρυθμίσεις προσομοίωσης.

Στη συνέχεια, περιγράφονται τα 5 ιονικά υγρά για τα οποία υπήρχαν διαθέσιμα πειραματικά δεδομένα. Αυτά χρησιμοποιούνται ως αναφορά για σύγκριση με τα αποτελέσματα του κάθε κεφαλαίου. Πειραματικά δεδομένα για την ιονική αγωγιμότητα υπάρχουν τόσο για τα εν λόγω υγρά όσο και για τις αντίστοιχες μεμβράνες που προκύπτουν από αυτά. Συνεπώς, όλα τα μοντέλα αναπτύσσονται γύρω από αυτά τα πειραματικά δεδομένα. Στην ενότητα 2.2 πραγματοποιούνται παραμετρικές μελέτες για να διερευνηθούν οι διάφορες τιμές των ρυθμίσεων για τις οποίες τα μοντέλα δίνουν ακριβέστερα αποτελέσματα. Αυτές αφορούν την απόσταση αποκοπής των διαμοριακών και ηλεκτροστατικών δυνάμεων, το συνολικό χρόνο και τη συχνότητα δειγματοληψίας για τη μέθοδο MSD, καθώς και το συνολικό αριθμό μορίων που απαιτούνται να προσομοιωθούν για να υπάρχει αφενός ακρίβεια και αφετέρου χαμηλό υπολογιστικό κόστος.

Με χρήση των παραμέτρων που προέκυψαν, διεξάγονται προσομοιώσεις για τα 5 δεδομένα ιονικά υγρά στο εύρος θερμοκρασιών 300 K με 380 K. Για δύο εξ αυτών πραγματοποιείται σύγκριση μεταξύ των μοντέλων OPLS-AA και GROMOS. Το τελευταίο δίνει ελαφρώς καλύτερα αποτελέσματα. Δεδομένα εισόδου και για τα 5 ιονικά υγρά ήταν διαθέσιμα σε διαδικτυακές βάσεις δεδομένων μόνο στην περίπτωση του μοντέλου GROMOS. Η απόλυτη απόκλιση της πυκνότητας από τις πειραματικές τιμές είναι πάντα μικρότερη από 15 % (σχήμα 1i). Η μέση απόλυτη απόκλιση της αγωγιμότητας είναι μικρότερη από 300 %, το οποίο θεωρείται ιδιαίτερα ακριβές αποτέλεσμα με βάση τη βιβλιογραφία [1]. Στη θερμοκρασία των 380 K τα αποτελέσματα είναι αρκετά πιο ακριβή, δίνοντας μέση απόλυτη απόκλιση μικρότερη από 90 %. Για το λόγο αυτό όλες οι επόμενες AA προσομοιώσεις διεξάγονται σε αυτήν τη θερμοκρασία.



(i) Καμπύλες πυκνότητας-θερμοκρασίας για τα 5 (ii) Καμπύλες αγωγιμότητας-θερμοκρασίας για τα 5 ιονικά υγρά αναφοράς.

Σχήμα 1: Αποτελέσματα των AA προσομοιώσεων. Η αγωγιμότητα είναι αδιαστατοποιημένη με τα αντίστοιχα πειραματικά δεδομένα για κάθε επιμέρους θερμοκρασία. Η πυκνότητα έχει αδιαστατοποιηθεί με την πειραματική τιμή για τη θερμοκρασία 300 K.

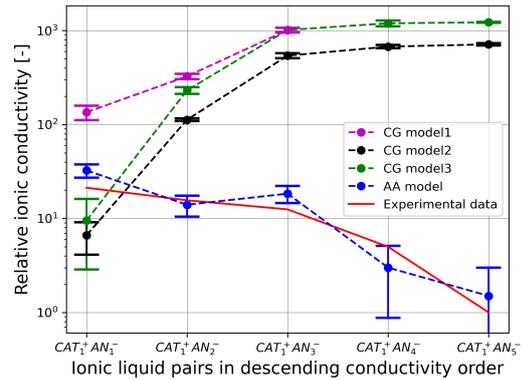
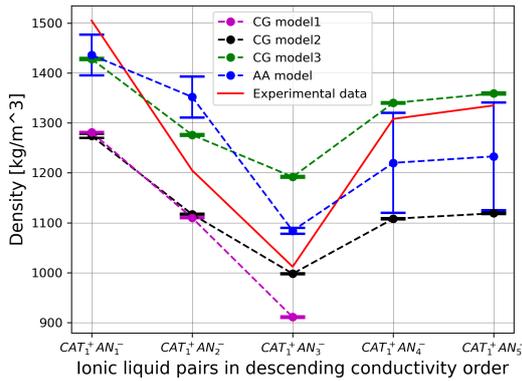
CG Προσομοιώσεις

Στο 3ο κεφάλαιο του πλήρους κειμένου διερευνάται εάν είναι εφικτό να χρησιμοποιηθούν CG μοντέλα για τη βελτιστοποίηση ιονικών υγρών. Για τον σκοπό αυτόν, πραγματοποιείται παραμετρική μελέτη με χρήση του μοντέλου CG MARTINI 2.0. Οι περιορισμοί αυτού είναι εμφανείς από τα πρώτα κίολας αποτελέσματα και έτσι κρίνεται απαραίτητο αφενός να μεταβληθούν οι συνθήκες προσομοίωσης, ώστε να ταιριάζουν περισσότερο σε ένα CG μοντέλο και αφετέρου να χρησιμοποιηθεί η έκδοση 3.0 του μοντέλου CG MARTINI, η οποία προσφέρει αυξημένες δυνατότητες μοντελοποίησης.

Αρχικά, αναπτύσσονται CG μοντέλα για τα πέντε ιονικά υγρά που αναφέρθηκαν στο προηγούμενο κεφάλαιο. Παρά τις πολλές και διαφορετικές προσεγγίσεις που πραγματοποιήθηκαν στη διαδικασία μοντελοποίησης η τάση της ιονικής αγωγιμότητας για τα διαφορετικά ιονικά υγρά δεν ήταν ορθή. Ο στόχος ήταν το υγρό με τη μεγαλύτερη πειραματική τιμή αγωγιμότητας να έχει και τη μεγαλύτερη τιμή ως CG μοντέλο. Δηλαδή να υπάρχει μια έστω ποιοτική τάση με αλλαγή της χημείας. Ωστόσο, αυτό δεν μπόρεσε να επιτευχθεί, παρά το γεγονός ότι η πυκνότητα που προέκυπτε από το μοντέλο ήταν εντός ικανοποιητικών ορίων απόκλισης από τα πειραματικά δεδομένα. Στα σχήματα 2i και 2ii φαίνονται καθαρά όλα τα προαναφερθέντα.

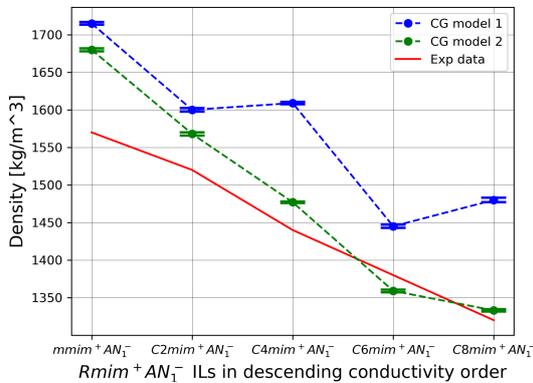
Προκειμένου να μην απορριφθεί κατευθείαν η ιδέα της χρήσης CG μοντέλων για τη βελτιστοποίηση, προσομοιώνεται μία εναλλακτική οικογένεια ιονικών ρευστών υπό τις ίδιες συνθήκες. Τα αποτελέσματα φαίνονται στα σχήματα 2iii, 2iv και δείχνουν πως η αλλαγή της χημείας μεταβάλλει σταδιακά και μονοσήμαντα την τιμή της ιονικής αγωγιμότητας, παρά τη σημαντική ποσοτική απόκλιση από τις πειραματικές τιμές. Αν τυχαία κάθε φορά CG ιονικά υγρά συμπεριφέρονται με έναν παρόμοιο τρόπο, τότε μια διαδικασία βελτιστοποίησης θα είχε νόημα, επειδή το υπολογιστικό κόστος των μον-

τέλων αυτών είναι ιδιαίτερα χαμηλό. Επιπλέον, ο πίνακας του παραρτήματος Β.4 είναι ένα χρήσιμο εργαλείο που περιέχεται στους οδηγούς του MARTINI 3.0 και παρέχει έναν αντικειμενικό και εμπειρικό τρόπο για τη μετατροπή των τυχαίων CG ιονικών υγρών σε πραγματικά AA ιονικά υγρά. Για τους δύο τελευταίους λόγους, ελήφθη η απόφαση να πραγματοποιηθεί η CG βελτιστοποίηση.

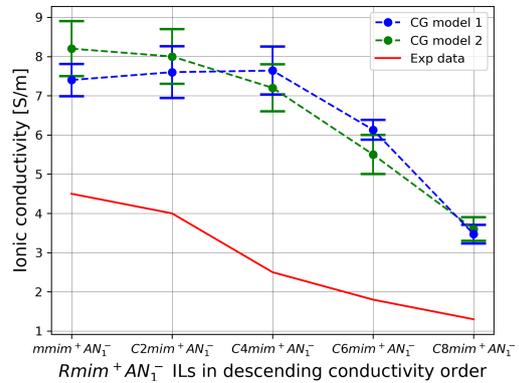


(i) Πυκνότητα για την ομάδα ιονικών υγρών $CAT_1^+ X^-$. Οι προσομοιώσεις έγιναν στους 380 K, η πειραματική πυκνότητα μετρήθηκε στους 300 K.

(ii) Ιονική αγωγιμότητα για την ομάδα ιονικών υγρών $CAT_1^+ X^-$. Τα ζεύγη είναι διατεταγμένα κατά σειρά φθίνουσας ιονικής αγωγιμότητας.



(iii) Πυκνότητα για την ομάδα ιονικών υγρών $Rmim^+ AN_1^-$. Οι προσομοιώσεις και οι πειραματικές μετρήσεις έγιναν στους 380 K.



(iv) Ιονική αγωγιμότητα για την ομάδα ιονικών υγρών $Rmim^+ AN_1^-$. Τα ζεύγη είναι διατεταγμένα κατά σειρά φθίνουσας ιονικής αγωγιμότητας. Όλα τα αποτελέσματα αναφέρονται σε θερμοκρασία 380 K.

Σχήμα 2: Σχήματα και διαγράμματα που αφορούν τις CG προσομοιώσεις.

CG Βελτιστοποίηση

Η CG βελτιστοποίηση πραγματοποιείται με χρήση του λογισμικού εξελικτικών αλγορίθμων EASY της ΜΠΤΡΒ του ΕΜΠ. Ένας από τους κυριότερους λόγους για τους

οποίους επιλέχθηκε, ήταν η ευελιξία που προσφέρει στη διαδικασία της παραμετροποίησης. Η εφαρμογή αιτιοκρατικών μεθόδων βελτιστοποίησης θα ήταν εξαιρετικά περίπλοκη στην εφαρμογή και στην ανάπτυξη για το δεδομένο πρόβλημα.

Αρχικά, διερευνάται το πως διαφορετικές τοπολογίες ιόντων, όπως αυτές που φαίνονται στο σχήμα 3i επηρεάζουν το συνολικό αριθμό πιθανών συνδυασμών για ένα ιονικό ζεύγος. Διαπιστώνεται πως ήδη με χρήση μόλις τριών σωματιδίων (χαντρών), ο χώρος σχεδιασμού έχει τεράστιες διαστάσεις για λίγες σχετικά παραμέτρους σχεδιασμού. Συνεπώς, λαμβάνεται η απόφαση να βελτιστοποιηθεί ένα ιονικό ζεύγος στο οποίο τόσο το ανιόν όσο και το κατιόν θα είναι τοπολογικά όμοια με το σχηματισμό 3a του σχήματος 3i. Θεωρείται πως ο συνδυασμός αυτός αποτελεί έναν δίκαιο συμβιβασμό μεταξύ ευελιξίας και απλότητας.

Ο τρόπος παραμετροποίησης παρουσιάζεται στο σχήμα 3ii. Σε κάθε ιόν μεταβάλλεται μόνο η γωνία και ο τύπος της χρησιμοποιούμενης χάντρας (bead) για κάθε διαθέσιμη θέση του σχηματισμού 3a. Τα μήκη των χημικών δεσμών προσδιορίζονται με μία αυτοματοποιημένη προσεγγιστική διαδικασία που βασίζεται στον τύπο των συνδεόμενων κάθε φορά χαντρών και η οποία περιγράφεται με μεγαλύτερη λεπτομέρεια στην υπό-ενότητα 4.3.2 (βλ. σχήμα 4.3 και πίνακα 4.1 της Δ.Ε.). Οι σταθερές ελατηρίου που αφορούν τα μήκη και τις γωνίες στροφής των δεσμών λαμβάνουν μία τυπική τιμή που προτείνεται από τη βιβλιογραφία.

Η καμπύλη σύγκλισης της βελτιστοποίησης φαίνεται στο σχήμα 3iii από την οποία φαίνεται πως προέκυψαν ιονικά υγρά με εξαιρετικά υψηλές τιμές αγωγιμότητας, έως και 3 φορές μεγαλύτερες από την αντίστοιχη του nafionTM. Εντούτοις, τα αποτελέσματα αυτά είναι απαραίτητο να επαληθευτούν με χρήση AA μοντέλων τα οποία έχουν μεγαλύτερη ακρίβεια. Τα προκύπτοντα ιονικά υγρά μετατρέπονται από το CG πεδίο στο AA με τη βοήθεια του πίνακα του παραρτήματος Β.4, όπως φαίνεται για παράδειγμα στο σχήμα 3iv. Πραγματοποιείται μία γραμμική παρεμβολή μεταξύ των δεδομένων της CG και της AA αγωγιμότητας (βλ. σχήμα 3v). Ο συντελεστής συσχέτισης είναι χαμηλός, υποδεικνύοντας τους περιορισμούς του CG MARTINI 3.0, όσον αφορά τη δυνατότητα μετάφρασης από χημικούς σχηματισμούς CG σε AA. Τα αποτελέσματα αυτά οδηγούν στην αναζήτηση ενός εναλλακτικού, ακριβέστερου τρόπου βελτιστοποίησης. Για τον σκοπό αυτό, επιλέγεται η AA βελτιστοποίηση παρά τη μεγάλη περιπλοκότητα της και το τεράστιο υπολογιστικό κόστος της.

AA Βελτιστοποίηση

Τα γεωμετρικά και τοπολογικά δεδομένα εισόδου για μία AA προσομοίωση μοριακής δυναμικής είναι εξαιρετικά δύσκολο να βρεθούν και συνήθως προκύπτουν από ειδικές δι-αδικτυακές βάσεις δεδομένων. Για να είναι εφικτή μία διαδικασία AA βελτιστοποίησης, πρέπει να παραχθεί ένα λογισμικό το οποίο από κάποιες παράμετρους (τυχαίοι αριθμοί) να παράγονται τα γεωμετρικά και τοπολογικά δεδομένα εισόδου για το εκάστοτε ιόν με τη μέγιστη εφικτή ακρίβεια. Επειδή η διαδικασία είναι ήδη ιδιαίτερα περίπλοκη, το

κατιόν διατηρείται πάντα το ίδιο (CAT_1^+). Προκειμένου να αρθρούν τα προηγούμενα προβλήματα αποφασίζεται να χρησιμοποιηθούν οι ακολουθίες χαρακτήρων SMILES ως μέσο επικοινωνίας των παραμέτρων με τη χημεία. Οι μεταβλητές σχεδιασμού φαίνονται στο σχήμα 4i. Με χρήση ενός κώδικα *rython* οι παράμετροι p_1 έως p_{10} μετααφράζονται σε ένα ανιόν. Το ανιόν αυτό μπορεί να μην υφίσταται στην πραγματικότητα, αν για παράδειγμα σε αυτό ένα άτομο οξυγόνου συνδέεται με τρία άλλα άτομα. Τέτοιες περιπτώσεις ανιχνεύονται αυτόματα από τον εν λόγω κώδικα και εμφανίζονται τα απαραίτητα ενημερωτικά μηνύματα, έτσι ώστε το ανιόν αυτό να απορριφθεί πριν φτάσει στο ακριβό στάδιο της προσομοίωσης.

Έπειτα, με χρήση του λογισμικού ανοικτού κώδικα επεξεργασίας δεδομένων μοριακής μοντελοποίησης *Open Babel* [2], η ακολουθία SMILES που αφορά το ανιόν μεταφράζεται στο γεωμετρικό αρχείο της μορφής *.mol2*. Προτού αυτό παραχθεί, έχει πραγματοποιηθεί ελαχιστοποίηση της ενέργειας στο σχηματισμό, ώστε αυτός να λάβει μια μορφή κοντά στην πραγματική. Επιπλέον έχει πραγματοποιηθεί υπολογισμός που αφορά τα μερικά φορτία που έχει κάθε άτομο του ανιόντος. Οι υπολογισμοί αυτοί πραγματοποιούνται με το *Open Babel*, εφαρμόζοντας το πεδίο δυνάμεων (μοντέλο) *MMFF94*. Ο υπολογισμός των μερικών φορτίων είναι απαραίτητος για να προκύψει το αρχείο τοπολογίας το οποίο έχει την πληροφορία που απαιτείται για την αναπαράσταση των ηλεκτροστατικών και διαμοριακών δυνάμεων, καθώς και των ελαστικών δυνάμεων που οφείλονται στην ύπαρξη χημικών δεσμών.

Στη συνέχεια, το αρχείο *.mol2*, που παράχθηκε, μετατρέπεται με χρήση του λογισμικού ανοικτού κώδικα *topolbuild* στο αντίστοιχο αρχείο *OPLS-AA* τοπολογίας *.itp*. Το αρχείο αυτό έχει πολλές ελλείψεις στην πλειοψηφία των περιπτώσεων και δεν είναι έτοιμο να εισαχθεί ως δεδομένο σε μία προσομοίωση μοριακής δυναμικής. Αρκετή πληροφορία απουσιάζει σχετικά με τα μήκη, τις γωνίες και τις ελαστικές σταθερές των χημικών δεσμών. Επιπλέον, οι πληροφορίες σχετικά με μία ειδική κατηγορία ελαστικών συνιστωσών (*dihedrals*, βλ. ενότητα 1.7.1 της Δ.Ε.) λείπει σχεδόν εξ ολοκλήρου. Οι ελλείψεις αυτές αντιμετωπίζονται με χρήση αρκετών επιπλέον κωδίκων *rython* οι οποίοι συμπληρώνουν τα δεδομένα που λείπουν με χρήση μετρήσεων από το αρχείο *.mol2* που περιέχει γεωμετρικά δεδομένα, δεδομένων από το ελλειπές αρχείο *.itp* και ποικίλων προσεγγίσεων. Τα τελικά αρχεία τοπολογίας είναι πλέον έτοιμα να χρησιμοποιηθούν για προσομοίωση.

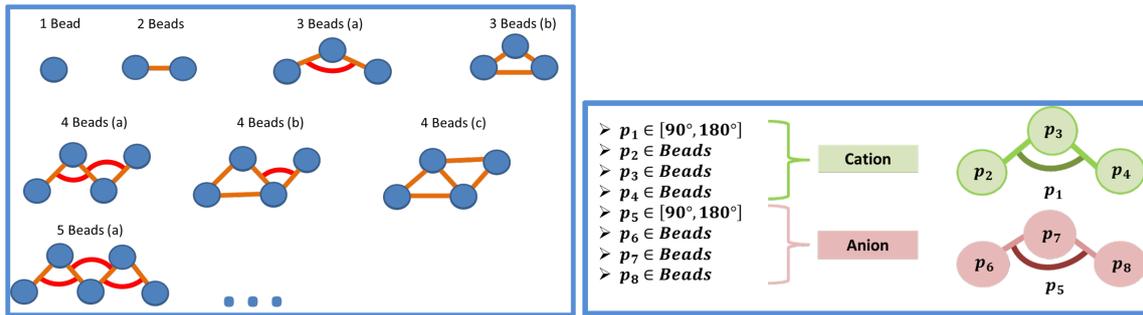
Εξαιτίας της περιπλοκότητας της παραπάνω διαδικασίας και επειδή δεν είναι επιθυμητό να τρέξουν προσομοιώσεις με λάθος δεδομένα εισόδου, λόγω του τεράστιου υπολογιστικού κόστους των *AA* προσομοιώσεων, μόνο ένα πολύ μικρό ποσοστό $S\%$ των ιόντων προσομοιώνονται με επιτυχία. Το γεγονός αυτό καθιστά την επιλογή ρυθμίσεων για τον *EASY* μία περίπλοκη διαδικασία. Οι ρυθμίσεις του *EASY* που επιλέχθηκαν θα ήταν πολύ πιο εύστοχες αν ο συνολικός χρόνος της βελτιστοποίησης ήταν δεκαπλάσιος του διαθέσιμου. Παρά τον περιορισμένο χρόνο βελτιστοποίησης, ο *EASY* βρήκε ιονικά υγρά με έως και 157 % μεγαλύτερη αγωγιμότητα σε σχέση με τα αποτελέσματα του *AA* μοντέλου του 2ου κεφαλαίου για το ιονικό υγρό αναφοράς $CAT_1^+ AN_3^-$. Ακριβώς επειδή πραγματοποιήθηκαν αρκετές παραδοχές έτσι ώστε να είναι εφικτή η προηγούμενη

διαδικασία, γίνεται επαναξιολόγηση των καλύτερων αποτελεσμάτων με χρήση ακριβώς του ίδιου OPLS-AA μοντέλου που εφαρμόστηκε στο κεφάλαιο 2. Τα αποτελέσματα της διαδικασίας αυτής φαίνονται στον πίνακα 1. Τρία από τα πέντε καλύτερα ιονικά υγρά έχουν υψηλότερη ιονική αγωγιμότητα σε σχέση με το $CAT_1^+AN_3^-$. Για καθένα από αυτά, εκτός του 4ου, το αντίστοιχο ουδέτερο μόριο του εκάστοτε ανιόντος έχει παραχθεί σε εργαστήριο σύμφωνα με το NIH των ΗΠΑ [4]. Με ορισμένες μεθόδους μπορεί θεωρητικά να αφαιρεθεί το κατάλληλο πρωτόνιο, να παραχθεί το ανιόν και επομένως το αντίστοιχο ιονικό υγρό.

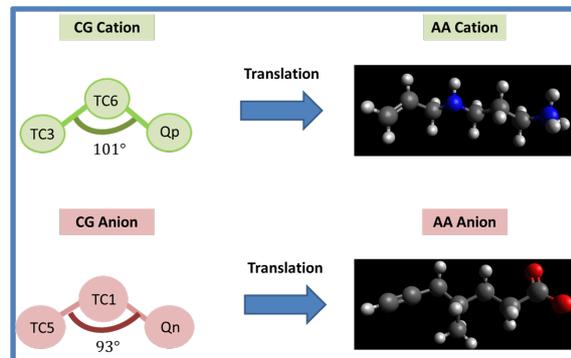
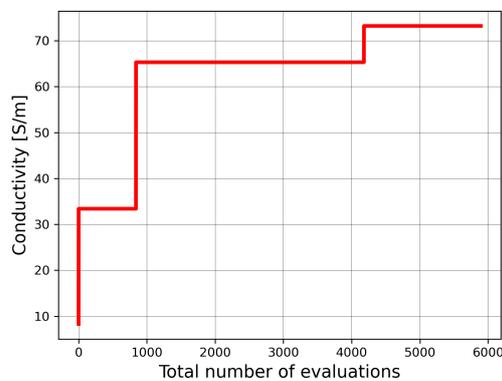
Σύνοψη-Συμπεράσματα

Συνοψίζοντας, η διπλωματική εργασία ασχολήθηκε με την αναζήτηση εναλλακτικών ιονικών υγρών, ώστε να αποτελέσουν την πρώτη ύλη για έναν νέο βελτιωμένο τύπο μεμβράνης ανταλλαγής πρωτονίων σε κυψέλες καυσίμου. Για τον σκοπό αυτόν, δημιουργήθηκαν και επαληθεύτηκαν AA και CG μοντέλα ιονικών υγρών, ώστε χρησιμοποιηθούν σε AA και CG βελτιστοποιήσεις της ιονικής αγωγιμότητας. Και στις δύο περιπτώσεις χρησιμοποιήθηκε το λογισμικό εξελικτικών αλγορίθμων EASY της ΜΠΥΡΒ του ΕΜΠ για τη βελτιστοποίηση. Έπειτα από επαλήθευση, η CG βελτιστοποίηση δεν έδωσε κάποιο βελτιωμένο αποτέλεσμα. Αντίθετα, η AA βελτιστοποίηση κατέληξε σε ιονικά υγρά με αρκετά βελτιωμένη ιονική αγωγιμότητα τα οποία μάλιστα φαίνεται πως μπορούν να παραχθούν εργαστηριακά. Με βάση τα αποτελέσματα που παρουσιάστηκαν προηγουμένως, προκύπτουν τα ακόλουθα συμπεράσματα.

1. Τα AA μοντέλα έχουν πολύ καλές επιδόσεις όσον αφορά την πυκνότητα και μέτριες ως προς την ιονική αγωγιμότητα. Σε υψηλότερες θερμοκρασίες είναι πιο ακριβή και φαίνεται πως μπορούν να χρησιμοποιηθούν σε μία διαδικασία βελτιστοποίησης αλλά και για επαλήθευση άλλων μοντέλων.
2. Η βελτιστοποίηση με χρήση του CG MARTINI 3.0, με συνάρτηση στόχο την ιονική αγωγιμότητα, παρουσιάζει περιορισμούς, παρά το χαμηλό υπολογιστικό κόστος της. Δεν υπάρχει ισχυρή και μονοσήμαντη συσχέτιση μεταξύ της CG και της AA ιονικής αγωγιμότητας, εξαιτίας των περιορισμών στην αντίστροφη μοντελοποίηση (από CG σε AA μοντέλο).
3. Η προσεγγιστική AA βελτιστοποίηση ήταν σε θέση να υποδείξει ιονικά ζεύγη με σημαντικά αυξημένη ιονική αγωγιμότητα σε σχέση με το ζεύγος αναφοράς. Η αντίστοιχη συσχέτιση μεταξύ της αγωγιμότητας του προσεγγιστικού και του ακριβέστερου AA μοντέλου φαίνεται ικανοποιητική.
4. Η προηγούμενη διαδικασία θα μπορούσε να βελτιωθεί αισθητά μειώνοντας το τεράστιο υπολογιστικό της κόστος, με χρήση μεθόδων τεχνητής νοημοσύνης. Ο στόχος είναι να υπάρχουν μεγαλύτερα ποσοστά επιτυχίας $S\%$ κατά την παραγωγή των ιόντων στη διαδικασία προ-επεξεργασίας. Με λίγες μετατροπές, η ίδια διαδικασία θα μπορούσε να εφαρμοστεί και για βελτιστοποίηση άλλων συστημάτων μορίων.

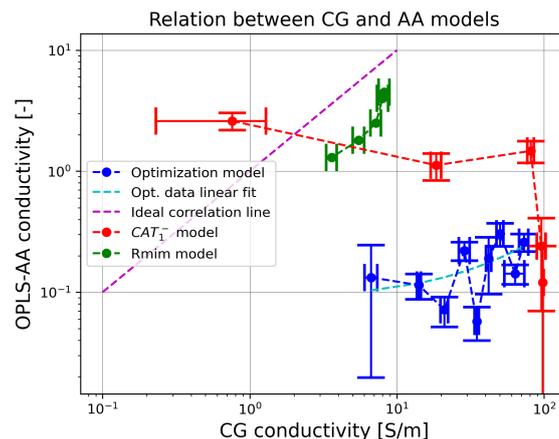


(i) Διαφορετικοί τοπολογικοί σχηματισμοί σε (ii) Επεξηγηματικό σχήμα σχετικά με τις ένα ουδέτερο μόριο. Οι πορτοκαλί γραμμές παραμέτρους σχεδιασμού της βελτιστοποίησης αναφέρονται στους δεσμούς μεταξύ των χαντρών και τα κόκκινα τόξα στους γωνιακούς βαθμούς ελευθερίας.



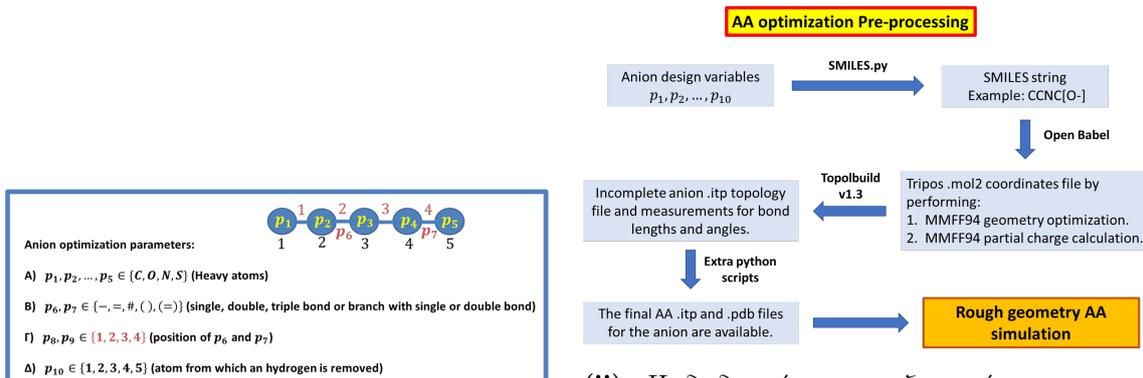
(iv) Μετάφραση του καλύτερου ιονικού υγρού που προέκυψε, από CG μοντέλο σε AA μοντέλο. Οι γκρι, άσπρες, κόκκινες και μπλέ σφαίρες αναπαριστούν αντίστοιχα τα άτομα του άνθρακα, του υδρογόνου, του οξυγόνου και του αζώτου.

(iii) Καμπύλη σύγκλισης για τη CG βελτιστοποίηση με χρήση DMAEA.



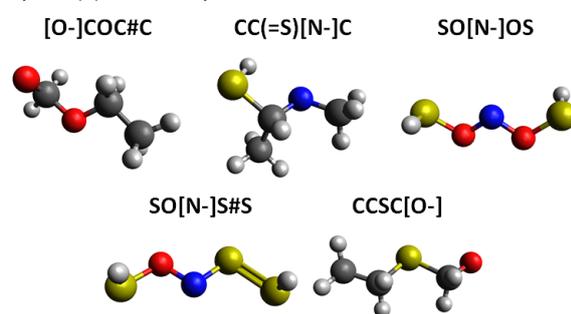
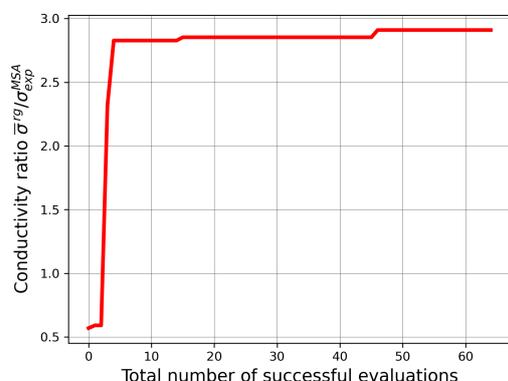
(v) Συσχέτιση της CG και της AA αγωγιμότητας για διάφορα μοντέλα. Τα AA αποτελέσματα για τα ιονικά υγρά που είναι βασισμένα στο κατιόν CAT_1^+ έχουν προκύψει έπειτα από εφαρμογή του μοντέλου GROMOS. Για την οικογένεια ιονικών υγρών $Rmim^+AN_1^-$ χρησιμοποιήθηκαν πειραματικά δεδομένα. Η επαλήθευση των CG μοντέλων που χρησιμοποιήθηκαν στη βελτιστοποίηση πραγματοποιήθηκε με χρήση του μοντέλου OPLS-AA. Παρατηρείται ασθενής συσχέτιση μεταξύ της CG και της AA αγωγιμότητας. Όλα τα αποτελέσματα αφορούν τους 380 K.

Σχήμα 3: Σχήματα και διαγράμματα που αφορούν την CG βελτιστοποίηση.



(i) Επεξηγηματικό σχήμα σχετικά με τις μεταβλητές σχεδιασμού της AA βελτιστοποίησης.

(ii) Η διαδικασία προ-επεξεργασίας που εφαρμόστηκε κατά την AA βελτιστοποίηση με χρήση προσεγγιστικών μοντέλων.



(iv) Τριδιάστατη αναπαράσταση των πέντε καλύτερων ανιόντων που προέκυψαν από την AA βελτιστοποίηση.

(iii) Καμπύλη σύγκλισης για την AA βελτιστοποίηση με χρήση MAEA. $\bar{\sigma}^{rg}$ είναι η μέση αγωγιμότητα του προσεγγιστικού AA μοντέλου. $\sigma_{exp}^{AN_3}$ είναι η πειραματική αγωγιμότητα του ιονικού υγρού $CAT_1^+ AN_3^-$.

Σχήμα 4: Σχήματα και διαγράμματα που αφορούν την AA βελτιστοποίηση.

ID	Cation	Anion	$\bar{\rho}^v$ [kg/m ³]	$\bar{\sigma}^v / \sigma_{exp}^{AN_3}$	$\sigma_{std}^v / \sigma_{exp}^{AN_3}$	$\bar{\sigma}^{rg} / \sigma_{exp}^{AN_3}$	$\frac{\bar{\sigma}^v - \sigma_{std}^{ref}}{\sigma_{std}^{ref}} \frac{\sigma_{exp}^{ref}}{\sigma_{std}^{ref}}$
1	CAT_1^+	[O-]COC#C	$9.42 \cdot 10^2$	2.64	0.17	2.83	+140 %
2	CAT_1^+	CC(=S)[N-]C	$9.67 \cdot 10^2$	0.38	0.05	2.32	-65 %
3	CAT_1^+	SO[N-]OS	$1.11 \cdot 10^3$	2.21	0.44	1.47	+101 %
4	CAT_1^+	SO[N-]S#S	$1.15 \cdot 10^3$	1.12	0.23	1.19	+8 %
5	CAT_1^+	CCSC[O-]	$9.28 \cdot 10^2$	1.75	0.43	0.94	+59 %
Ref	CAT_1^+	AN_3^-	–	1.10	0.19	0.83	+0 %

Table 1: Συγκεντρωτικά αποτελέσματα για τα 5 καλύτερα ιονικά υγρά. Περιέχονται αγωγιμότητες από προσεγγιστικά μοντέλα (rg), μοντέλα επαλήθευσης (v) και πειραματικά δεδομένα (exp).

Επιλεγμένη βιβλιογραφία

- [1] M. Salanne, “Simulations of room temperature ionic liquids: from polarizable to coarse-grained force fields,” *Phys. Chem. Chem. Phys.*, vol. 17, no. 22, pp. 14270–14279, 2015, doi: 10.1039/C4CP05550K.
- [2] N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, “Open Babel: An open chemical toolbox,” *J. Cheminformatics*, vol. 3, no. 1, p. 33, Dec. 2011, doi: 10.1186/1758-2946-3-33.
- [3] B. D. Ray, *topolbuild*, IUPUI Physics Dept. 402 N. Blackford St. Indianapolis, IN 46202-3273.
- [4] S. Kim, J. Chen, T. Cheng, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2021;49(D1):D1388–D1395. doi:10.1093/nar/gkaa971