1-1-2017

Classical And Quantum Mechanical Simulations Of Condensed Systems And Biomolecules

Hedieh Torabifard
Wayne State University,

Recommended Citation

Torabifard, Hedieh, "Classical And Quantum Mechanical Simulations Of Condensed Systems And Biomolecules" (2017). *Wayne State University Dissertations*. 1884.
https://digitalcommons.wayne.edu/oa_dissertations/1884

CLASSICAL AND QUANTUM MECHANICAL SIMULATIONS OF CONDENSED SYSTEMS AND BIOMOLECULES

by

HEDIEH TORABIFARD

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2017

MAJOR: CHEMISTRY

Approved By:

Advisor Date

DEDICATION

To my family

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Professor G. Andres Cisneros, who was always very supportive and taught me a lot during my Ph.D Program. He gave me the opportunity to work with him in his laboratory. I was always inspired by his hard work, deep knowledge, motivations and the time that he dedicated to his students. I really appreciate all the time he spent with me during this program and all the constructive advices. Professor Cisneros gave me freedom to choose my direction in my research while encouraging me to grow as a scientist. I was really lucky that I did my Ph.D under his supervision and I will be grateful forever.

Besides my advisor, I would like to thank the rest of my committee: Professor H. Bernhard Schlegel, Professor Christine Chow, Professor Wen Li and Professor Jeff Potoff for their support and constructive advices during this program. Dr. Schlegel and Dr. Potoff and their lab members were always helpful by sharing knowledge and experiences. In addition, I would like to thank Dr. Vladimir Chernyak for all the help and company that I received from him.

I thank all my fellow lab mates in Professor Cisneros' lab: Dr. Robert Duke, Dr. Oleg Starovoytov, Dr. Eric Kratz, Dr. Rebecca Swett, Dr. Dong Fang, Dr. Sajeewa Dewage, Pavel Silvestrov, Alice Walker, Yi-Jung Tu, Erik Vazquez-Montelongo. I have learnt a lot from them and I enjoyed my time in the lab. I also thank my friends in Wayne State University, Majeed Nader, Benyamin Dehkan, Hoda Sharifi, Azam Peyvandipour who were supportive and helping me toward my Ph.D. I would like to especially thank Dr. Ali Shahini for all the support, love and patience.

I want to thank the Staff in the department at Wayne State University, Mrs. Melissa Barton, Jacqueline Baldyga, Diana Kudla, Bernadette Miesik, Kellie Lauder, Lisa Smith, Nestor Ocampo for their great help. I am also grateful for all the financial support from Department of Chemistry, Wayne State University and National Institute of Health.

Finally, I would like to thank my family: my parents and my sister for supporting me spiritually throughout my Ph.D program. They were always helping me towards my goals in my life and encouraged me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION AND REVIEW

Portions of the text in this chapter were reprinted or adapted with permission from: *Nat. Chem. Biol., 13, 181 (2017), Theor. Chem. Acc., 134, 101:1 (2015) and J. Phys. Chem. B., 118, 7156 (2014)*. All rights to the work are retained by the authors and any reuse requires permission of the authors.

Computer simulations have become a powerful tool to study chemical and biological problems. This thesis will cover both development and application of theoretical methods for the computational simulations of condensed and biological systems. Force field development using distributed multipoles derived from the Gaussian Electrostatic Model (GEM) and its amenability for use in the AMOEBA force field will be discussed here. Molecular Dynamics (MD) simulations and Quantum Mechanical/Molecular Mechanical (QM/MM) are used to study energetics, structural and thermodynamic properties as well as the reaction mechanism pathways for these systems. This thesis describes various types of studies carried out on the Fe(II)/α-ketoglutarate(α-KG)-dependent family of dioxygenases (TET2 and AlkB) as well as water, imidazolium- and pyrrolidinium-based ionic liquids. This chapter gives a brief introduction to each system in terms of the literature review and the goal of the research for each studied system.

Chapter 2 and 3 focus on TET2 demethylation activity using MD and QM/MM studies. Chapter 4 describes the intermolecular tunnels for the oxygen transport within AlkB. The force field development, MD simulations and properties calculations of ionic liquids and water are covered in chapters 5, 6 and 7. A brief conclusion is provided in Chapter 8.

1.1 MD and QM/MM studies on Fe(II)/α-KG-dependent enzymes

1.1.1 TET family enzyme

The discovery of ten-eleven translocation (TET) enzymes transformed the known repertoire of epigenetic DNA modifications.[1] TET enzymes catalyze the oxidation of 5-methylcytosine (mC), the mainstay of the epigenome, into three additional bases: 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxylcytosine (caC).[1-6] Mounting evidence suggests that these oxidized mC (ox-mC) bases stably populate mammalian genomes, aid in DNA demethylation, and potentially encode unique epigenetic information.[7-10] The central questions now facing the field involve the functions of each individual base and the mechanisms governing their formation.

The overall catalytic mechanism of TET enzymes (TET1–3 in mammals) has been largely inferred from related proteins in the Fe(II)/α-KG-dependent family of dioxygenases, such as AlkB.[11] Enzymes in this family couple decarboxylation of α-KG with substrate oxidation via a transient Fe(IV)-oxo intermediate, with succinate and CO_2 as byproducts. TET enzymes apply this general mechanism to not one but three stepwise reactions, raising the question of whether these enzymes are specialized for one particular step of oxidation, or for three-step oxidation as a whole. Moreover, stepwise oxidation obscures the function of individual ox-mC's, creating a need to break the linkage between steps in order to study each base in isolation. The first step of oxidation, conversion of mC to hmC, has so far drawn the most attention, as it best explains the physiological levels of cytosine modifications: in the human genome, mC accounts for approximately 0.6-1% of all bases, hmC is typically 1–5% of mC, and fC and caC are at least 1–2 orders of magnitude rarer than hmC.[9] Consistent with these observations, biochemical studies have shown that mC substrate is preferred over hmC and fC, with 2- to 5-fold differences in K_M and K_{react} reported for human TET2.[12] Crystal structures did not reveal substrate-specific interactions that could explain these differences,[12-13] but computational modeling suggested that hydrogen abstraction is more efficient on mC than on hmC and fC, which adopt unfavorable conformations.[12, 14] Together, these studies portray TET enzymes as predominantly serving to generate hmC; in this case, decreased capacity for further oxidation would help to maintain stable levels of hmC for epigenetic functions. Indeed, most functional studies on ox-mC bases have focused on hmC in health and disease, with fC/caC considered as fairly negligible. However, this view does not explain why fC and caC are present at all, and it contrasts with evidence for the importance of higher-order oxidation. Apart from being intermediates in demethylation, fC/caC potentially also function as stable epigenetic marks. Genomic sequencing has mapped fC/caC to gene regulatory regions separate from hmC.[9] Biochemically, mouse TET2 is capable of iterative oxidation; it can catalyze multiple rounds of oxidation upon a single encounter with mC, without releasing the hmC-containing DNA strand.[15] Although the prevalence of genomic hmC implies that most encounters are not iterative, this mechanism could allow TET enzymes to generate fC and caC marks without first accumulating hmC. Together, these studies encourage the alternate view that TET enzymes are specialized for making not only hmC but fC and caC as well, even that conversion of hmC to fC could be the key “committed” step to DNA demethylation. To resolve these competing views of TET function, one question comes to the fore: whether TET enzymes are adapted to facilitate higher-order oxidation. The mC-to-hmC step is most favored, but if fC and caC serve important functions, mechanisms should be in place to permit their formation, yet these mechanisms remain largely unknown. They could be extrinsic to TET, e.g. other proteins could recruit TET enzymes or regulate their activity. However, intrinsic features, especially structure-function support for higher-order oxidation, would suggest an enzyme specifically shaped to generate not one but three epigenetic bases and would add value to stalling the hmC-to-fC reaction in particular, to test whether it is indeed a key transition.

We examined the active site of human TET2 for potential structure-function determinants of stepwise oxidation. The target nucleobase is everted out of the DNA duplex and occupies a tunnel-like space in the active site, with the 5-modified group pointing toward the α-KG analogue and Fe(II) (Figure 1.1). Although the residues that form this tunnel have no obvious interaction with the 5-modified groups.[12-13] We hypothesized that they could impact the progress of stepwise oxidation by hydrogen bonding or steric interactions. Our collaborators targeted the one conserved residue located closest to the 5-methyl group, Thr1372 (Figure 1.1). By substituting all 20 amino acids at these positions, Thr1372, they uncovered a relationship between the side chain properties and stepwise oxidation activity, including variants that stall oxidation at hmC, with little to no fC/caC formed. Molecular dynamics simulations, coupled with biochemical analyses, revealed that a conserved Thr1372-Tyr1902 active site scaffold is required for efficient fC/caC formation, providing the first evidence that wild type TET2 is specifically shaped to enable higher-order oxidation. We further show that mutations along this core scaffold can reconfigure active site interactions to stall oxidation at hmC, which opens opportunities to test the importance of hmC versus fC/caC in biological and pathological systems. Combined biochemical and computational methods are applied to elucidate an active site scaffold that is required for wild type stepwise oxidation and that, when perturbed, explains the mutants' "hmC-stalling" phenotype. We present the computational methods, analyses and results along with some experimental results in Chapter 2. Our results suggest that the TET2 active site is shaped to enable higher-order oxidation and provide the first TET variants that could be used to probe the biological functions of hmC separately from fC and caC.

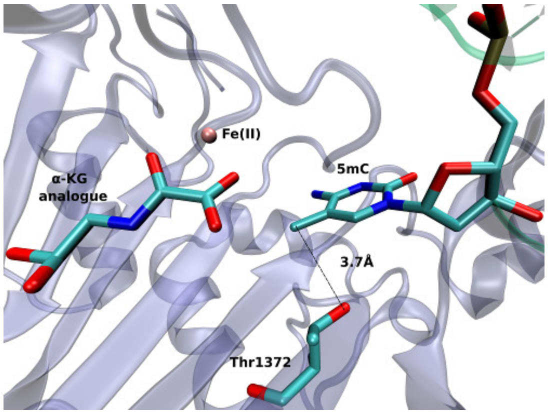

Figure 1.1 Structure of the hTET2-CS active site highlighting the targets for mutagenesis, Thr1372.

In Chapter 3, we turn our focus to the *ab initio* QM/MM study of the oxidation of 5hmC to 5fC to complement our MD studies and to investigate the source of the substrate preference in the wild type. As we discussed above, our MD results revealed that the active site of TET2 is shaped to enable higher-order oxidation states of the substrate.[16] This study provides a qualitative image of the stepwise oxidation path in TET2. However, it raises new questions such as whether oxidation of 5hmC to 5fC is kinetically favorable. To gain further insight into the oxidation of 5hmC to 5fC, we performed a hybrid quantum mechanical/molecular mechanical (QM/MM) studies on wild type TET2 and T1372E mutant as an example of “hmC-stalling" mutants.

1.1.2 AlkB family enzyme

E. coli AlkB is an Fe(II)/α-KG dependent enzyme that repairs alkylated DNA and RNA bases by catalyzing an oxidative dealkylation mechanism.[11,17-18] The oxidative dealkylation mechanism catalyzed by AlkB, has been investigated extensively.[19-31] The reaction mechanism can be separated into two stages. The first stage involves the formation of a ferryl (Fe(IV)-oxo) intermediate, followed by the oxidation of the alkyl moiety on the damaged base. To initiate the oxidation of Fe, an O_2 molecule needs to displace a water molecule from the primary coordination sphere of Fe. The diffusion of O_2 into the active site is an essential process in this mechanism.

A key question in this first stage of the reaction is how O_2 molecules are transported into the active site. Various enzymes that use molecular Oxygen have been shown to employ transient intra-molecular tunnels formed by flexible hydrophobic residues to transport O_2 from the surface of the protein to the active site.[32-35]

A number of computational studies have been reported on the diffusion of O_2 through intra-molecular tunnels.[35-42] One possible approach to investigate the transport of these types of molecules involves standard MD simulation of ligand diffusion, which ideally requires a large number of independent replicate runs of several ns to attain adequate sampling (flooding simulation).[32,35] A second approach involves the determination of the potential of mean force (PMF).[35 43] Extensive sampling is essential to obtain accurate

PMFs since incomplete sampling can result in large errors in the calculated barriers. In recent years, GPU computing has resulted in a large increase in the timescales of MD simulations, which make it possible to perform the sufficient sampling that is necessary to achieve a realistic description of ligand diffusion.[39, 44]

Yu *et al*. proposed the possibility of O_2 transport through an intra-molecular tunnel in AlkB based on the first AlkB crystal structures.[45] They showed that there is little unoccupied volume in the binding cavity, in which the target base directly contacts its molecular surface. This study portrays open and closed states of a tunnel putatively gating O_2 diffusion into the active site in the presence of iron(II) and cobalt(II), respectively. The open and closed states of this tunnel are due to the structural variation of the side chain of residue W178, which is located at the entrance of a binding cavity.[46]

Based on these observations, we present results from various computational approaches to determine the likelihood of the existence of this putative intra-molecular tunnel, and whether it is amenable for the passive transport of O_2 into the active sit of AlkB. In addition, given that molecular oxygen is a neutral albeit highly polarizable molecule, we have employed both non–polarizable fixed-charge (AMBER) and multipolar/polarizable (AMOEBA) potentials to investigate the role of electronic polarization. Explicit simulations on three W178 mutants have also been performed to investigate the role of this particular residue on AlkB structure and the structure of the tunnel. In Chapter 4, I present the details for the various methods employed to simulate the transport of the substrate, calculate the energy for this transport along the tunnels, and various analysis. Subsequently, the results and discussion for the calculated tunnels and transport/energetics are described in this chapter, followed by concluding remarks.

1.2 AMOEBA/GEM force field development

1.2.1 Immidazolium-based ionic liquids

The first ionic liquids (ILs) were introduced in the 1950s, with a significant increase in interest in the 1980s.[47] Room temperature ionic liquids (RTILs) are organic salts with melting points below room temperature. Ionic liquids are a combination of asymmetric organic and inorganic ions that stay in a liquid phase over a wide temperature range. These ionic liquids are usually composed of organic cations and inorganic anions. The most studied are imidazolium$^+$, pyridinium$^+$, and pyrrolidinium$^+$ based cations in a combination with a broad number of anions. These cations are relatively easy to functionalize, and various types of side groups can be attached to obtain particular ionic liquid properties. Thermodynamic properties such as low vapor pressure, excellent thermal and electrochemical stability, and good solvation properties make ionic liquids attractive for a wide range of applications. Practical applications can include materials for electrodeposition,[48] battery electrolytes,[49] and applications for catalysis.[50] The desired properties can be achieved by choosing a particular cation/anion combination. Therefore, there is an essential need to determine the physical and chemical properties of ILs in order to optimize their usage. There are a large number of cation/anion combinations, which makes experimental determination of properties an expensive and time-consuming process.[51] Thus, molecular modeling has become a powerful tool to predict ionic liquid properties without conducting an experiment. It is also practical to carry out molecular dynamics simulations to get some fundamental understanding on ionic liquids' behavior and to improve predictive capabilities of computational models. However, accurate simulations are needed to facilitate investigations of ionic liquids. There have been many computational studies performed to model and simulate a wide range of ionic liquids.[52-61]

A systematic improvement of the force fields for molecular dynamics simulations of various compounds is a viable process, since molecular dynamics simulations are becoming an essential tool for scientific research and investigations. A broad variety of molecular dynamics simulation packages are available for researchers.[62-67] However, predictive capabilities of that software mostly depend on the quality of the employed force fields. Most conventional force fields for ILs use partial atomic charges to describe the electrostatic interactions of molecules. A number of non-polarizable[68-71] and polarizable[72-74] force fields that are based on partial atomic charges have been successfully developed and employed for the simulation of various compounds. Despite the good agreement with the experiment for a number of properties, those force fields show some inaccuracies for certain cases and do not fully account for the anisotropy of the electrostatic interactions or for the charge density penetration effects.[75]

In order to improve the description of electrostatic interactions, higher order multipoles can be employed. The gaussian electrostatic model (GEM)[76-77] uses Hermite Gaussians to describe the molecular charge density. In addition, we have shown recently that it is possible to obtain distributed multipoles, which we call GEM-DM multipoles, to describe electrostatic interactions in the atomic multipole optimized energetics for biomolecular applications (AMOEBA) potential.[78-79] Several force fields have been developed using the distributed multipole approach.[80-82] The advantage of distributed multipoles over partial atomic charges is that point multipoles provide a better description of the charge density anisotropy, compared with point charge models.[83-84] However, these models do not account for the charge penetration effects at short interatomic distances.[85] One way to take into account anisotropy of electrostatic interactions and charge penetration effects is to employ damping functions.[86-88] An alternative is to use a continuous description of the molecular charge densities. Recently, a continuous description of the charge densities was successfully implemented for molecular dynamics simulations of water.[89]

In Chapter 5, we present the development of the multipolar/polarizable AMOEBA force field for 1,3-dimethylimidazolium based ILs using GEM-DM multipoles. We present the methods employed for the parametrization, including energy decomposition analysis (EDA) and comparison of the GEM-DM multipole results to the Gaussian distributed multipole analysis (GDMA) results. Subsequently, we describe the methodology of the force field development and the details for the molecular dynamics simulations based on the new parameters. Finally, we discuss the results and MD simulations on these ILs.

1.2.2 Pyrrolidinium-based ionic liquids

Lithium ion batteries are widely used in various areas including portable electronics, electric vehicles and aerospace applications, among others due to their high energy density, low self-discharge, low maintenance and small size and weight. Lithium ion battery technology shows great promise as power sources that can revolutionize the electric vehicles industry.

Typically, organic electrolytes such as ethylene carbonate (EC) have been used in batteries; however, these types of electrolytes are volatile and amiable, raising safety concerns especially for high temperature applications. Recently, ionic liquids have been introduced as new electrolytes for lithium ion batteries due to their unique properties. However, these electrolytes are most commonly identified as poor electrolytes in batteries.[90-92] To address this issue, a detailed, atomic level understanding of ion transport processes and redox stability would help to improve the design of electrolyte-electrode couples in batteries. Our combined computational and experimental approach was inspired by this principle to design a new electrolyte for Li-ion batteries.

The wide variety of cation-anion combinations make the synthesis and experimental determination of ILs properties an expensive and time consuming process. The synthesis of highly pure ILs places particular demands on the preparative work. Moreover, predicting which ILs are well suited for a given application remains an important pre-synthesis pursuit.

In Chapter 6, we present force field development and MD simulation along with some experimental study for spirocyclic pyrrolidinium tetraflorouborate [sPyr$^+$][BF$_4^-$]. In addition, a computational study on the properties of this IL pair with added Li is presented. This chapter describes a brief overview of the parametrization details followed by a description of the experimental characterization and the details of the MD simulations. Subsequently, it details the results and discussion of some thermodynamical and transport properties, followed by concluding remarks.

1.2.3 Water

Water is arguably the most important and interesting liquid on earth. It has many interesting physical and chemical properties largely due to the ability of water molecules to form multiple hydrogen bonds. Unique properties such as the temperature of maximum density, large heat capacity, expansion on freezing, and unique solvation properties make water very important in several areas including biology and geology. The manner in which water interacts with ions, organic molecules, and biomolecules holds the key for many questions about the foundation of biomolecular structure and function. The investigation of microscopic properties of liquid water remains a subject of intense research.[101-107] Several theoretical and computational models have been developed to attempt to describe the detailed microscopic properties of water. These models can be useful tools to guide and supplement experimental work.[78, 108-110] There are many non-polarizable and polarizable water models as has been reviewed extensively.[111-112] Some of these potentials are very popular in computational studies; for instance, point charge-based TIP3P water model is widely used for biomolecular simulations.[113] This model was parameterized to reproduce a number of properties at ambient conditions, including heat of vaporization, liquid density, and isobaric heat capacity at 298 K.[114] However, this model has been shown to fail to reproduce thermodynamic properties at elevated temperatures.[112, 115-116] Improvements based on TIP3P have been developed in several improved models such as TIP4P and TIP4P-Ew.[114] One of the main reasons for the issues observed from these simple models involves the reduced accuracy of the non-bonded interactions. This reduced accuracy arises from several factors, including the inaccurate representation of the charge density anisotropy, the failure to account for penetration effects,[75] and the neglect of many-body interactions.[117] There are several ways to address these shortcomings; for instance, the error due to charge density anisotropy can be reduced by using higher-order multipoles; however, these models do not account for penetration errors.[85, 118] The penetration effect may be accounted by the use of damping functions at close distances.[85-86, 88, 119-121] In addition, various explicit polarization models have been introduced to account for the change in the charge distribution due to an external electric field.[122-123] The inclusion of polarizabilities also results in (partial) accounting of many-body effects.[120] For instance, the POL3 model[124] is an atomic polarizable water model that employs undamped isotropic dipoles.[125] Some polarizable models such as AMOEBA,[82] SIBFA,[80] EFP,[126] XPOL[127-128] and NEMO[129] rely on distributed multipoles and explicit polarization for the electrostatic component of the non-bonded interactions. The improved description of the charge density anisotropy results in an improved reproduction of the electrostatic interactions due to the use of multipoles.[83-84, 118, 130-131] One of the most accurate current models to describe water is the MB-pol potential developed by Babin *et al.*[132-133] MB-pol relies on the many-body expansion at medium and short range combined with a multipolar polarizable description at long range and inclusion of nuclear quantum effects.[132-133] MB-Pol is able to reproduce structural, thermodynamic, and dynamical properties of the liquid phase as well as dimer vibration-rotation tunneling spectrum, second and third virial coefficients, and cluster structures and energies in the gas phase.[133-134] AMOEBA is a multipolar polarizable potential that relies on distributed atomic multipoles up to the quadrupole and inducible point dipoles. The original AMOEBA water model, referred to as AMOEBA03 herein,[78, 135] employs mutual induction of dipoles at atomic centers for the reproduction of the polarization component. AMOEBA03 also employs a Tholè damping function to avoid the so-called polarization catastrophe[136] at short range. This potential showed improvement in the description of several gas-phase properties of water, which was attributed to the use of multipole moments beyond the monopole.[137] The AMOEBA03 model also shows good agreement with various thermodynamic properties compared to experiment.[78] A scaled-down version of the AMOEBA water model, termed iAMOEBA (inexpensive AMOEBA),[138] was introduced in 2013. The computational cost is reduced by a factor of two by implementing a method of direct polarization,[138] that is, in iAMOEBA only the electric fields that arise from the permanent multipoles are taken into account for the calculation of the polarization. The distinctive feature of iAMOEBA compared with AMOEBA03 is the use of ForceBalance for its parametrization.[139] Parametrization by means of ForceBalance involves the use of an objective minimization procedure to optimize the fitting parameters by combining gas-phase quantum data and computed thermodynamic properties.[138] A large and diverse data set with different weights and scaling factors such as experimental liquid properties and *ab initio* quantum mechanics (QM) gas-phase dimer results are used to optimize the parameters. We have recently developed a similar method for optimization of vdW parameters by fitting the dimer energies, density, and heat of vaporization to *ab initio* and experimental data.[140]

Recently, a new AMOEBA water model has been introduced, which also relied on the ForceBalance approach (AMOEBA14).[141] AMOEBA14 showed significant improvement over both AMOEBA03 and iAMOEBA for a number of liquid properties across a range of temperatures and pressures, as well as gas-phase properties. An alternative way to improve the accuracy of the electrostatic interactions is to employ a continuous description of the charge density. One method that relies on this approach is the Gaussian electrostatic model (GEM),[87, 89, 142-143] which uses Hermite Gaussian functions to describe the molecular charge density. An added advantage of the fitting procedure for GEM is that it enables the calculation of explicitly finite distributed multipoles (GEM-DM).[143] These multipoles have been employed to describe the electrostatic interactions in the AMOEBA potential.[76, 79] The original AMOEBA parametrization with GEM-DM was focused on several systems including liquid water. The electronic density for this water model was fitted by employing a basis set that included Gaussian Hermite functions up to $l = 2$ for the oxygen and $l = 0$ (only s-type functions) on the hydrogen atoms.[79] Therefore, the final distributed multipoles for the water model involve monopoles, dipoles, and quadrupoles on the oxygens and only monopoles on the hydrogen atoms. This initial AMOEBA model with GEM-DM showed very good agreement for liquid structure, density, and heat of vaporization compared with experimental data at 298 K. However, it failed to predict the temperature of maximum density of water.

In Chapter 7, we report the parametrization of a new AMOEBA water model using GEM-DM by considering moments up to quadrupoles on both oxygen and hydrogen atoms. The multipoles, which are fitted to QM data, describe the electrostatic interaction very well at long and medium range. The van der Waals parameters are fitted to experimental density and heat of vaporization at ambient condition. The accuracy of the AMOEBA/GEM-DM force field is tested by running a series of simulations and comparing thermodynamic and transport properties to results from a published benchmark of water properties reported by Vega *et al.*[114] The calculated results are shown to reproduce liquid- and gas-phase properties very well for a broad range of temperatures including the temperature of the maximum density of water as well as other thermodynamic and transport properties. In this chapter, we also describe the methodology and results for the parametrization of the force field and molecular dynamics simulations.

CHAPTER 2 MUTATIONS AT A SINGLE ACTIVE SITE RESIDUE IN TET2 STALL OXIDATION AT 5-HYDROXY METHYL CYTOSINE AND REVEAL REQUIREMENTS FOR CATALYSIS

Portions of the text in this chapter were reprinted or adapted with permission from: *Nat. Chem. Biol., 13, 181 (2017).* All rights to the work are retained by the authors and any reuse requires permission of the authors.

2.1 Introduction

Ten-eleven translocation (TET) enzymes catalyze stepwise oxidation of 5-methylcytosine (mC) to yield 5-hydroxymethylcytosine (hmC) and the rarer bases 5-formylcytosine (fC) and 5-carboxylcytosine (caC). Stepwise oxidation obscures how each individual base forms and functions in epigenetic regulation and prompts the question of whether TET enzymes primarily serve to generate hmC, or whether they are adapted to produce fC variants that permit oxidation to hmC but largely eliminate fC/caC. Our molecular dynamics simulations along with some biochemical analyses, elucidated an active site scaffold that is required for wild type stepwise oxidation and that, when perturbed, explains the mutants' hmC-stalling phenotype. Our results suggest that the TET2 active site is shaped to enable higher-order oxidation and provide the first TET variants that could be used to probe the biological functions of hmC separately from fC and caC.

2.2 Computational methods

Forty-four molecular dynamics (MD) simulations were carried out on wild type and all experimentally tested mutants (T1372S/C/A/E/Q/N/D/V, Y1902F, T1372A/Y1902F) with all four cytosine derivatives (mC/hmC/fC/caC), α-KG, and Mg(II). All structures were modeled based on wild type hTET2-CS bound to mC containing DNA (PDB 4NM6).[13] Initially, the PDB structure was evaluated with MOLPROBITY[144] to check all possible rotamers, followed by hydrogen atom addition to every system with the Leap program[145] using the ff99SB parameter set[146] and solvation in a truncated octahedral box of TIP3P water.[113] In addition, protonation states of titratable residues were tested with Propka, which confirmed that the default ionization at pH 7 was correct for all residues.[147-148] All systems were explicitly neutralized with potassium counter ions, which were added to the system using the LEaP program. The final system size was ~60,000 total atoms with 17-21 counter ions. All structures were minimized with 3,000 steps of conjugate gradient, followed by gradual warm-up to 300 K using Langevin dynamics with a collision frequency of 1.0 ps^{-1} in the NVT ensemble for 100 ps. All simulations were performed with the GPU version of the pmemd program in AMBER12.[146] The force field parameters for all cytosine derivatives α-KG, Mg(II), and Zn that are not available in the default ff99SB set are developed in house. The iron cation was approximated by using Mg(II) parameters based on the precedent established by previous studies on AlkB.[20-21] Once the systems achieved the target temperature, production MD simulations were performed using Langevin dynamics with a collision frequency of 1.0 ps^{-1} in the NPT (Canonical) ensemble with the Berendsen barostat using a 2 ps relaxation time at 300 K. The production length for each of the simulations was 50 ns, and snapshots were saved every 10 ps, and all snapshots were subjected to subsequent analysis. Values reported are generally a time average over calculations from all snapshots. The most relevant simulations were performed 2-5 times for 50 ns each, with the results averaged across all simulations. All systems were simulated using the Amberff99SB force field with a 1 fs step size and a 9 Å cutoff for non-bonded interactions. SHAKE was used for all the simulations, and the smooth particle mesh Ewald (PME) method[149] was employed to treat long-range Coulomb interactions. Hydrogen bond, root mean square deviation (RMSD), and distance analysis on trajectories were carried out using the CCPTRAJ module[150] available in the AMBER 12 suite, and the trajectories were visualized with the VMD program.[151] Hydrogen bond analysis criteria were 1) angles over 120 degrees and 2) O-H distances less than 3 Å (default cpptraj settings). Additional analyses to investigate intermolecular interactions in the active site were carried out by non-covalent interaction analysis (NCI) and energy decomposition analysis (EDA).

NCI is a visualization tool to identify non-covalent interactions between molecules.[152] The results obtained from the NCI analysis consist of surfaces between the interacting molecules. These surfaces are assigned specific colors to denote the strength and characteristic of the interactions: green surfaces denote weak interactions (e.g. van der Waals), blue surfaces strong attractive interactions (e.g. hydrogen bonds), and red surfaces strong repulsive interactions. The NCI calculations were performed with the NCI-Plot program.[153] We focused on the hmC systems, and a representative snapshot from every system was subjected to NCI analysis. In all cases, the hmC substrate was considered as a ligand interacting with a spherical region of 10 Å around the binding site. All calculations were obtained with a step size of 0.2 Å for the cube and a cutoff of 5 Å for the calculation of the interactions between the nucleotides and the active site. The snapshots for NCI plots have been selected to highlight the most frequent interactions relevant to the underlying mechanism.

All EDA calculations were carried out with an in-house FORTRAN90 program to determine the non-bonded interactions (Coulomb and VdW interactions) for all the residues.[43, 154-155] The average non-bonded interaction between a particular cytosine derivative and every other residue, ΔE_{int}, is approximated by $\Delta E_{int} = \langle \Delta E_i \rangle$, where i represents an individual residue, ΔE_i represents the non-bonded interaction (Coulomb or VdW) between residue i and the particular cytosine derivative, and the broken brackets represent averages over the complete production ensemble obtained from the MD simulations. This analysis has been previously employed for QM/MM and MD simulations to study a number of protein systems.[20-21, 156-159] As noted, the above-described analyses were performed on each individual snapshot over each individual simulation, and the reported data consist of the averages over all the simulations for each system.

2.3 Results and discussion

2.3.1 Summary of experimental results

Saturation mutagenesis at Thr1372 perturbs stepwise oxidation according to side chain properties. The experimental screening on mutant activities were performed by our experimental collaborators at University of Pennsylvania. Dot blotting showed that mutations at Thr1372 produced distinctive patterns of cytosine oxidation, which cluster based upon the biochemical properties of the side chain (Figure 2.1 a). Replacing Thr1372 with a proline, positively charged (H, K, R), or bulkier hydrophobic residue (I, F, L, M, W, Y) renders TET2 inactive. Only the T1372S mutant, which preserves the side chain hydroxyl group, exhibits wild type-like activity. Smaller residues (A, C, G) are proficient at oxidation to fC and caC, but at reduced levels compared to wild type. Most remarkably, the acidic or related polar residues (D, E, N, Q) and the nearly isosteric valine permit wild type-like formation of hmC but no fC or caC, as detected by dot blot. Given this stalling of oxidation at hmC, Thr1372 appeared to play a unique role in stepwise oxidation.

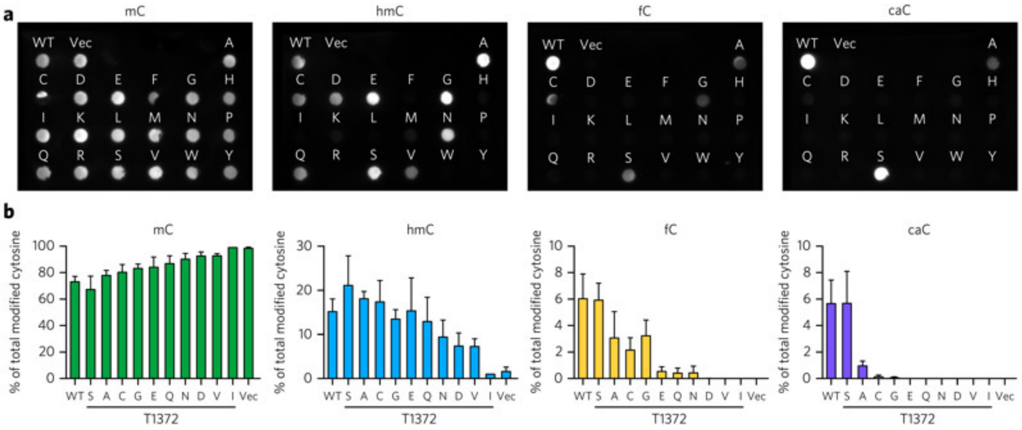

Figure 2.1 Experimental screen for mutant activity. (a) Dot blots for mC, hmC, fC, and caC. (b) Genomic levels of mC, hmC, fC, and caC modifications quantified by LC-MS/MS.

Nucleoside LC-MS/MS quantifies range of mutant activity. Fig 2.1 b shows the quantified cellular activity of all Thr1372 mutants capable of oxidizing at least to hmC. While the total modified cytosine bases (mC + ox-mCs) were similar across all conditions, the distribution of specific modifications differed significantly. The mutants exhibit a gradient of activity reflected in the fraction of genomic ox-mC. T1372S is the only mutant with wild type-like levels of fC and caC, and hmC levels slightly higher than wild type. A step below, T1372C/A/G mutants generate wild type-like levels of hmC but only one-third to one-half as much fC and barely detectable caC. Further down the activity gradient, another group consisting of the E/Q/N/D/V mutants produce hmC at levels at least half that of wild type, but fC and caC are below detection limits, consistent with the dot blotting results. Among this group, T1372E appears to have the highest activity with wild type-like hmC levels and <1% fC, while T1372V is lowest, generating half as much hmC but no fC. Finally, the slightly bulkier T1372I mutant resembles the vector control, underscoring the steric constraints at this position. Thus, the LC-MS/MS results more clearly elucidated the patterns seen on dot blot, showing a spectrum of activity among the Thr1372 mutants correlating with the side chain properties, with E/Q/N/D/V mutants stalling oxidation at hmC.

2.3.2 Modeling reveals the mechanism behind wild type and mutants' behavior

Computational modeling of mutants reveals a critical Thr1372-Tyr1902 scaffold. To probe potential mechanisms behind the mutants' effects, we turned to molecular dynamics (MD) simulations of all the active Thr1372 variants. We drew from our experience with modeling AlkB[20-21] to perform classical MD simulations of wild type hTET2-CS and the Thr1372 mutants bound to each of the four cytosine derivatives.

Our wild type results with modeled hmC and fC proved mostly consistent with the more recently published structures of TET2 with hmC and fC,[13] as the key interactions between the enzyme, α-KG, the active site metal ion, and the DNA substrate are observed for varying durations across our simulations.

Table 2.1 RMSDs for protein backbone in wild type and mutants with hmC-containing DNA.

RMSD-protein (5hmC)		RMSD-Protein (5hmC)	
T1372A	2.10 ± 0.37	Y1902F	2.53 ± 0.26
T1372A-Y1902F	2.39 ± 0.23	T1372C	2.40
T1372E	2.85 ± 0.18	T1372S	2.43
T1372N	2.29 ± 0.34	T1372V	2.79 ± 0.06
T1372D	2.45 ± 0.26		

Furthermore, energy decomposition analysis (EDA) and the root-mean-square deviation (RSMD) comparing the simulations to the reference crystal structure show that the cytosine bases stably occupy the active site across time in all our models (Table 2.1, 2.2, 2.3). The hmC models in particular revealed distinct patterns of active site interactions in wild type, C/A/G, and E/Q/N/D/V mutants, consistent with hmC being the fulcrum of the observed stalling effect. Most prominently, these distinctions highlight a key structural scaffold in the wild type enzyme that is required for efficient stepwise oxidation. The wild type active site scaffold consists of two interactions.

First, Thr1372 forms a hydrogen bond with Tyr1902 (Figure 2.2a, 2.4a). This hydrogen bond is observed in 65% of the simulation time (average over five runs of 50 ns each), and the total non-bonded interaction energy between these residues is –3.37 kcal/mol (Figure 2.2b).

Second, the Thr1372-Tyr1902 interaction orients Tyr1902 to promote non-bonded interactions with hmC (total -6.10 kcal/mol) (Figs. 2.2b, 2.4a, Tables 2.3, 2.4). This core scaffold is present across all wild type models bound to mC/hmC/fC/caC and remains fully intact in the T1372S mutant (Figure 2.4b), consistent with this mutant's wild type-like activity in cells. All the other mutants eliminate the Thr1372-Tyr1902 hydrogen bond, perturbing the interaction between Y1902 and the substrate base, with a corresponding loss of enzymatic activity. For the C/A/G mutants, loss of the Thr1372-Tyr1902 scaffold appears to weaken interactions between misaligned active site components, as exemplified by T1372A (Figure 2.2a, b, 2.4c).

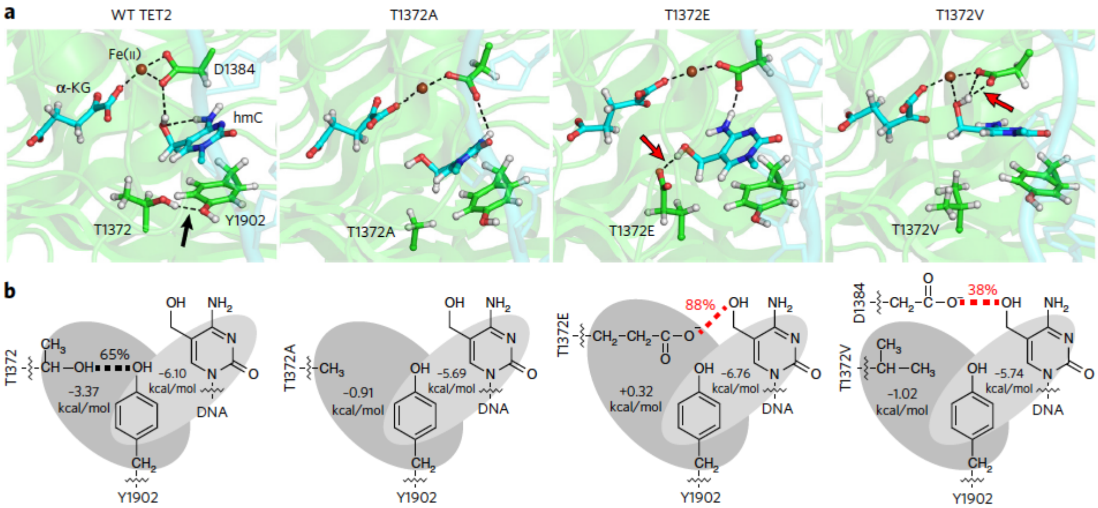

Figure 2.2 (a) Selected snapshots from MD simulations highlighting key active site components. (b) Scheme of interactions between key residues and hmC. Hydrogen bonds are shown in dashed lines. Non-bonded interactions are indicated in gray and in kcal/mol.

Combined with the DNA results, we propose that the C/A/G mutants represent a "low-efficiency" phenotype, which permit higher-order oxidation but at reduced levels compared to wild type. In our modeling, the E/Q/N/D/V mutants go a step further: they not only eliminate the Thr1372-Tyr1902 scaffold but also elicit new hydrogen bonds specifically with hmC. These new interactions, not present in wild type models, position hmC in a different orientation relative to Tyr1902 (Figure 2.2a, b, 2.4d-h). For instance, in T1372E, the Glu1372 hydrogen bonds directly with hmC for 88% of the simulation time (average over two runs of 50 ns each). Direct hydrogen bonding to hmC is also observed in T1372D and Q, whereas in T1372N and V, the new hydrogen bond is between hmC and other nearby residues (Figure 2.4d-h, Tables 2.3, 2.4). For example, T1372V elicits an hmC-Asp1384 hydrogen bond (38% of simulation time, average over two runs of 50 ns each). We suggest that the loss of the Thr1372-Tyr1902 scaffold, together with new interactions specific to hmC, could contribute to the unique stalling phenotype of T1372E/Q/N/D/V mutants, which we term “hmC dominant”.

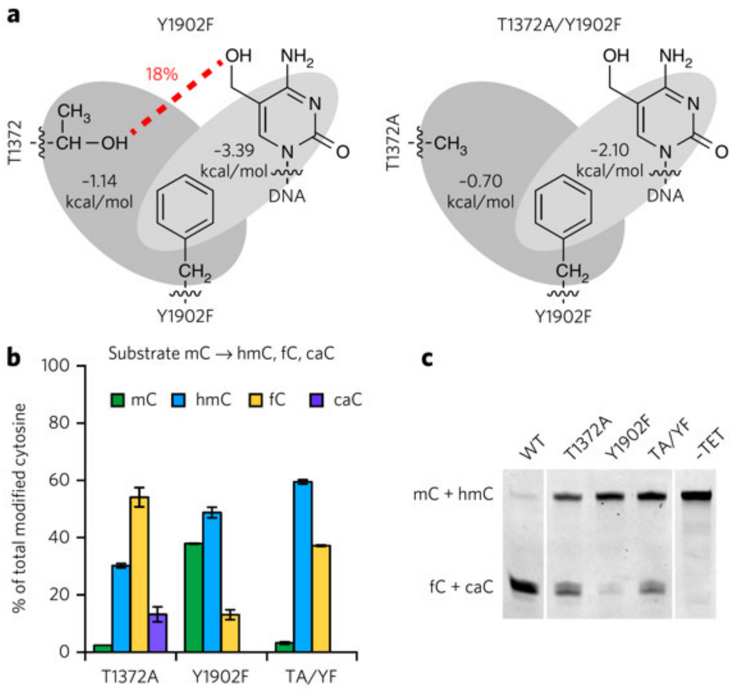

Figure 2.3 (a) Our modeling for Y1902F single mutant and T1372A/Y1902F double mutant. (b) Reaction of purified mutant on substrate, analyzed by LC-MS/MS. (c) To highlight fC and caC in the reaction products, the purified oligos were treated with recombinant TDG.

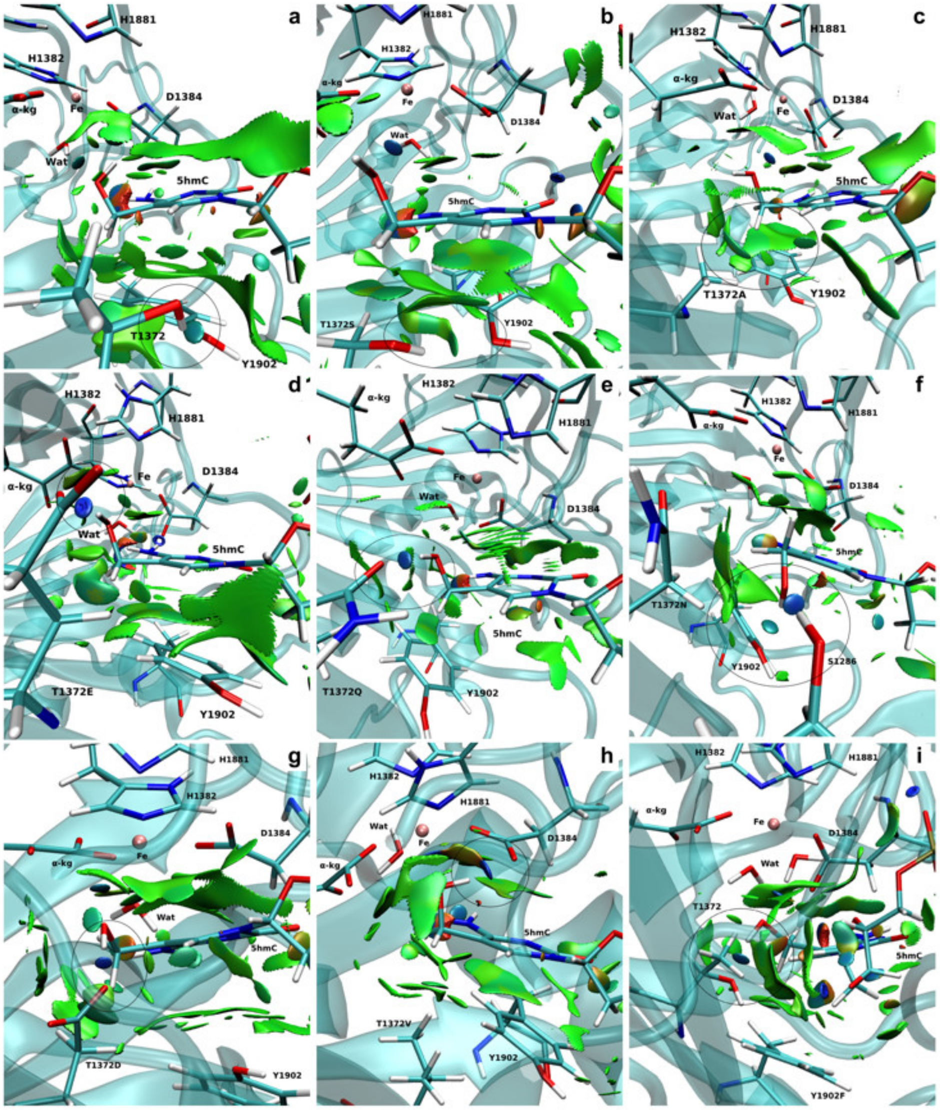

Figure 2.4 Non-covalent interaction (NCI) analysis on a representative snapshot for wild type hTET2-CS and mutants in the presence of hmC. (a) wild type (b) T1372S (c) T1372A (d) T1372E (e) T1372Q (f) T1372N (g) T1372D (h) T1372V (i) Y1902F. Key interactions are circled. The coordinating water occupying the sixth (equatorial) position is omitted for clarity. The isovalue for NCI is 0.3 au, and -0.2 au < sign$(\lambda_2)\rho < 0.2$ au.

Experimental results strongly support our model that the Thr1372-Tyr1902 scaffold is required for wild type TET2 activity. Loss of the active site scaffold decreases the activity of low-efficiency mutants and has a more severe effect on hmC-dominant mutants, which do not make significant fC/caC even under driving reaction conditions (Fig 2.1).

Table 2.2 EDA analysis between key residues and **(a)** 5mC, 5hmC and **(b)** 5fC, 5caC for all systems. X1372 denotes the residue at the 1372 position. Energies are given in kcal/mol, time averaged over the entire ensemble

(a)	Total Y1902-5mC	Total X1372-5mC	Total X1372-Y1902(5mC)	Total Y1902-5hmC	Total X1372-5hmC	Total X1372-Y1902 (5hmC)
WT	–5.18 ± 0.67	–2.76 ± 0.74	–3.55 ± 0.43	–6.10 ± 0.78	–3.59 ± 0.89	–3.37 ± 0.50
T1372S	–3.98	–3.02	–4.25	–6.59	–2.89	–3.02
T1372C	–6.03	–2.45	–1.54	–6.37	–3.94	–1.59
T1372A	–10.84	–1.50	–0.97	–5.69 ± 0.55	–1.40 ± 0.76	–0.91 ± 0.16
T1372E	–7.83	25.49	0.56	–6.76* ± 1.75	18.56 ± 0.20	0.32 ± 0.72
T1372Q	–8.36	–4.92	–1.48	–5.69 ± 1.56	–6.03 ± 2.37	–1.39 ± 0.38
T1372N	–6.22	–4.47	–1.81	–5.10 ± 0.59	–7.34 ± 4.14	–1.77 ± 0.33
T1372D	–5.42	22.59	–0.38	–6.25* ± 3.86	22.78 ± 0.82	–1.36 ± 3.89
T1372V	–4.40	–2.81	–0.94	–5.74 ± 0.26	–2.30 ± 0.29	–1.02 ± 0.08
Y1902F	–4.05	–2.35	–1.50	–3.39 ± 0.06	–4.84 ± 2.09	–1.14 ± 0.28
T1372A-Y1902F	–1.91	–2.37	–0.66	–2.10 ± 1.21	–1.85 ± 0.65	–0.70 ± 0.11

(b)	Total Y1902-5fC	Total X1372-5fC	Total X1372-Y1902 (5fC)	Total Y1902-5caC	Total X1372-5caC	Total X1372-Y1902 (5caC)
WT	–7.19 ± 0.59	–4.30 ± 0.17	–2.78 ± 0.26	–5.49 ± 0.57	–13.15 ± 5.56	–0.73 ± 0.91
T1372S	–6.98	–4.14	–3.01	–3.08	–10.98	–2.15
T1372C	–6.35	–3.69	–1.30	–5.05	–5.22	–1.85
T1372A	–7.57	–1.18	–0.98	–5.94	–3.44	–0.74
T1372E	–5.60	23.12	1.25	–6.18	85.40	–7.46
T1372Q	–6.22	–10.18	–1.46	–6.54	–23.42	–24.09
T1372N	–6.15	–9.29	–1.26	–6.41	–15.35	–1.24
T1372D	–4.61	20.58	–9.38	–6.67	85.10	–13.56
T1372V	–7.42	–2.28	–0.95	–5.42	–4.32	–1.16
Y1902F	–3.94	–2.95	–1.42	–6.38	–14.16	–1.45
T1372A-Y1902F	–3.80	–2.37	–0.74	–3.86	–2.26	–0.66

In addition, the hmC-dominant mutants exhibit decreased activity overall, but the usual mild preference for mC substrate is not sufficient to explain the larger loss of activity on hmC. Tyr1902 mutagenesis strongly supports model of active site interactions. Our MD simulations suggested that perturbations to the active site scaffold could introduce aberrant interactions that contribute to the additional loss of activity on hmC.

Table 2.3 Hydrogen bond analysis between key residues and **(a)** 5mC, **(b)** 5hmC, **(c)** 5fC, and **(d)** 5caC. X1372 denotes the residue at the 1372 position. Values are percentage of simulation time in which the hydrogen bond was observed. The values in parentheses are for the model with Fe(II) parameters. Only bonds observed in >10% of simulation time are included.

	X1372-Y1902 (5mC)	X1372-Y1902 (5hmC)	5hmC-X1372	5hmC-other	X1372-Y1902 (5fC)	X1372-Y1902	5caC-X1372	5caC-other
WT	77	65	-	16-D1384, 16-αKG	55	14	78	81-R1261
T1372S	85	56	-	-	53	39	49	186 R1261*
T1372C	-	-	-	-	-	-	-	-
T1372A	-	-	-	17-αKG	-	-	-	-
T1372E	-	-	88	-	-	27	-	38 R1261
T1372Q	-	-	21	42-D1384	-	-	98	-
T1372N	-	-	-	22-S1286, 16-Y1902, 14-αKG	-	-	57	21 R1261
T1372D	-	-	16)	16-αKG	59	80	-	27 R1261
T1372V	-	-	-	38-D1384	-	-	-	12 R1261
Y1902F	-	-	18	11-D1384	-	-	87	-
T1372A-Y1902F	-	-	-	57-R1261, 43-αKG, 15-T1259, 12-S1286, 11-S1284	-	-	-	-

We were cognizant of the challenges to modeling new interactions with classical MD and therefore subjected this model to an independent test: mutating the other scaffold residue, Tyr1902, to Phe. Our modeling predicts that Y1902F would liberate Thr1372 to form a hydrogen bond directly with hmC (18% of simulation time, average over two runs of 50 ns each), potentially favoring an hmC-dominant phenotype (Figure 2.3, 2.4i).

Taking the hypothesis one step further, by adding a T1372A mutation to Y1902F, our modeling predicts that the T1372A/Y1902F double mutant could rescue activity by alleviating the aberrant hydrogen bonding interaction.

Table 2.4 Energy decomposition analysis (EDA) for 5mC/hmC/fC/caC with all protein residues. The number of simulation runs are specified in parentheses. Values are kcal/mol, time averaged over the entire ensemble.

	5mC	5hmC	5fC	5caC
WT	-106.58±8.56(n=3)	-114.03±10.77(n=5)	-101.56±5.51(n=3)	-160.05±29.12(n=3)
T1372S	-75.83	-83.92	-99.14	-191.59
T1372C	-86.09	-105.90	-88.32	-116.01
T1372A	-69.25	-110.20 ±15.58(n=2)	-110.11	-134.67
T1372E	-92.54	-92.98 ±1.75(n=2)	-68.71	-18.46
T1372Q	-89.49	-114.93 ±38.25(n=2)	-119.55	-92.80
T1372N	-101.45	-113.70 ±5.25(n=2)	-102.18	-134.69
T1372D	-74.64	-63.41 ±51.68(n=2)	-71.82	-10.39
T1372V	-88.26	-76.21 ±16.39(n=2)	-64.76	-61.85
Y1902F	-78.21	-133.45 ±5.88(n=2)	-87.23	-123.85
T1372A-Y1902F	-90.50	-117.76 ±33.45(n=3)	-89.39	-75.13

To test these predictions, we compared the activities of purified T1372A, Y1902F, and T1372A/Y1902F enzymes in vitro. The results strikingly confirmed our predictions (Fig 2.3b, c). While Y1902F shows only trace generation of fC/caC, the addition of the second mutation in T1372A/Y1902F restores stepwise oxidation and mirrors the results for T1372A. Thus, our structural modeling correctly predicts the biochemical behavior of the Y1902F and T1372A/Y1902F mutants, strongly supporting both the requirement of the Thr1372-Tyr1902 scaffold for wild type stepwise oxidation and the contribution of aberrant active site interactions to the hmC-dominant phenotype.

2.4 Conclusions

TET-catalyzed stepwise oxidation populates the mammalian epigenome with three ox-mC bases, making it critical to dissect how each individual base forms and functions. Previous studies have elucidated various biases in favor of the first oxidation step, mC-to-hmC conversion, implying that TET enzymes may be primarily adapted for making hmC, with fC/caC as rare oxidative “overflow” products. However, in light of evidence for the importance of fC/caC in active DNA demethylation and as stable epigenetic marks, we asked whether TET enzymes bear structural features that specifically support fC/caC formation.

We have shown that a conserved Thr1372-Tyr1902 active site scaffold is required for efficient higher-order oxidation by human TET2, suggesting that the enzyme is shaped to enable production of not only hmC but fC/caC as well. We further uncover Thr1372 mutations that effectively abrogate higher-order oxidation by disrupting the active site scaffold; these are the first human TET variants that dissociate the steps of oxidation, providing a new tool to directly test the functions of hmC versus fC/caC. As a structure-function determinant in TET2, the Thr1372-Tyr1902 scaffold invites comparison to known TET homologues. In human TET2, the Ala-Phe double mutant only permits low-efficiency stepwise oxidation, suggesting that the Thr-Tyr dyad may have evolved to fine-tune efficient fC/caC generation. By leveraging this scaffold, our results offer the first variants that produce distinct stepwise oxidation patterns in human TET enzymes.

Our combined computational and biochemical approach shows how T1372E/Q/N/D/V mutants could reconfigure active site interactions to produce the hmC-dominant phenotype. As exemplified by T1372E, these mutants disrupt the active site scaffold, which results in moderate loss of overall catalytic activity.

To account for the additional loss of activity on hmC, our modeling most prominently implicates new hydrogen bonding to hmC in these mutants. While modeling is always subject to a dependence on input parameters, our calculations correctly predict the hmC-dominant behavior of the Y1902F single mutant, as well as rescued activity in the T1372A/Y1902F double mutant. Indeed, it is quite unusual that the addition of a second mutation rescues activity of the first, helping to bolster our mechanistic model. We note, however, that other related mechanisms could also play a role and are not mutually exclusive with this model. These possibilities reflect the complex dynamics of TET-DNA interactions, which remain priorities for future research.

Importantly, independent of the mechanism of action, the hmC-dominant Thr1372 mutants fill the need for experimental tools to dissect the individual steps of mC oxidation. If introduced into a suitable model system, these new TET variants potentially allow for the first direct studies of the epigenetic functions of hmC as distinct from fC and caC. Until now, functional studies have by necessity been all-or-none, showing that loss of one or more TET isozymes can produce diverse phenotypes. In many cases, reintroduction of a single active TET isozyme can fully rescue the phenotype. Such systems provide ideal opportunities to introduce low-efficiency and hmC-dominant TET variants to probe whether hmC alone is sufficient to rescue the defect, whether fC/caC are required, or whether interacting enzymes such as TDG are actually the key players. These direct functional applications, made possible by manipulation of the TET active site itself, would solidify whether TET enzymes evolved in favor of generating all three ox-mCs. These will bring new challenges as well, such as examining the mutants' activity under more physiological conditions. It will also be important to determine whether the mutant phenotypes in TET2 translate to other TET isoforms, which is needed both for applying the mutants in various biological systems and for helping to address whether TET1/2/3 have similar or distinct mechanisms of action. Finally, we envision that chemical biology approaches, including additional mutagenesis or unnatural modifications along the Thr1372-Tyr1902-cytosine scaffold, could further hone selectivity for particular bases and potentially uncover TET variants that stringently stall at fC as well or accelerate conversion to caC.

CHAPTER 3 INSIGHT INTO TET2-MEDIATED 5-HYDROXY METHYL CYTOSINE OXIDATION USING AB INITIO QM/MM CALCULATIONS

3.1 Introduction

Ten-eleven translocation 2 (TET2) converts 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) by a sequential oxidation mechanism. Our biochemical and computational studies discussed in Chapter 2, uncovered an active site scaffold that is required for wild type stepwise oxidation.[16] We showed that the mutation of a single residue, T1372, can impact the iterative oxidation steps and stop the oxidation of 5hmC to 5fC/caC.

However, the source of the stalling at the first oxidation step by some mutants still remains unclear. Here, we studied the catalytic mechanism of oxidation of 5hmC to 5fC mediated by TET2 using an *ab initio* quantum mechanical/molecular mechanical (QM/MM) approach.

Our results suggest that the first step, which is the rate-limiting step, involves a hydrogen atom abstraction from the hydroxyl group of 5hmC by the ferryl moiety in the wild type. Our results are in very good agreement with the results obtained for other enzymes of the Fe/α-KG dependent super family such as AlkB. Furthermore, our calculations for the T1372E mutant show very large barrier for the 5hmC to 5fC oxidation due to unfavorable orientation of 5hmC.

Combined electron localization function (ELF) and non-covalent interaction (NCI) analyses provides a qualitative description of the electron structure evolution along the reaction path. Energy decomposition analysis (EDA) has been performed to investigate the impact of each MM residues on catalytic activity. Our computational results demonstrate the substrate preference of TET2 and can provide guides for future experimental and theoretical research on TET2.

3.2 Computational methods

Two systems corresponding to TET2 with 5hmC and 5fC substrates were generated and equilibrated based on the TET2 crystal structure with 5mC (PDBID: 4NM6[160]) as discussed in Chapter 2.[16] Several stable structures were extracted from the MD simulation to perform QM/MM optimizations on the reactant state (with 5hmC). The water molecule coordinated to the iron atom (trans to H1881) was replaced by an oxyl moiety. The α-KG molecule was also replaced by a succinate molecule since the present work is focused on the mechanism after the formation of the ferryl moiety and succinate. The mechanism studied in this paper is presented in Figure 3.1. The snapshot with the lowest energy was selected for further QM/MM calculations. All the QM/MM calculations were performed with LICHEM,[161] which is able to interface Gaussian09[162] and TINKER[163] to perform additive QM/MM with electrostatic embedding.

For each model, the QM part contains the Fe atom, the side chain of the residues coordinated to the iron (H1382, D1384 and H1881), the oxyl moiety, the succinate molecule, the water molecule (coordinated to the iron at an equatorial position) and substrate base (5hmC or 5fC). As the residues T1372 and Y1902 were found to establish a scaffold which is essential for iterative oxidation, they are also included in the QM regions. Fang *et al.* showed that the existence of a water molecule which is located near Fe(IV)=O is very important for the mechanism of AlkB, another Fe/ α-KG dependent enzyme.[20] We observed from our MD simulations, one water molecule from the solvent diffuses into the active site and occupies the vacancy between D1384 and 5hmC for almost the entire simulation time.

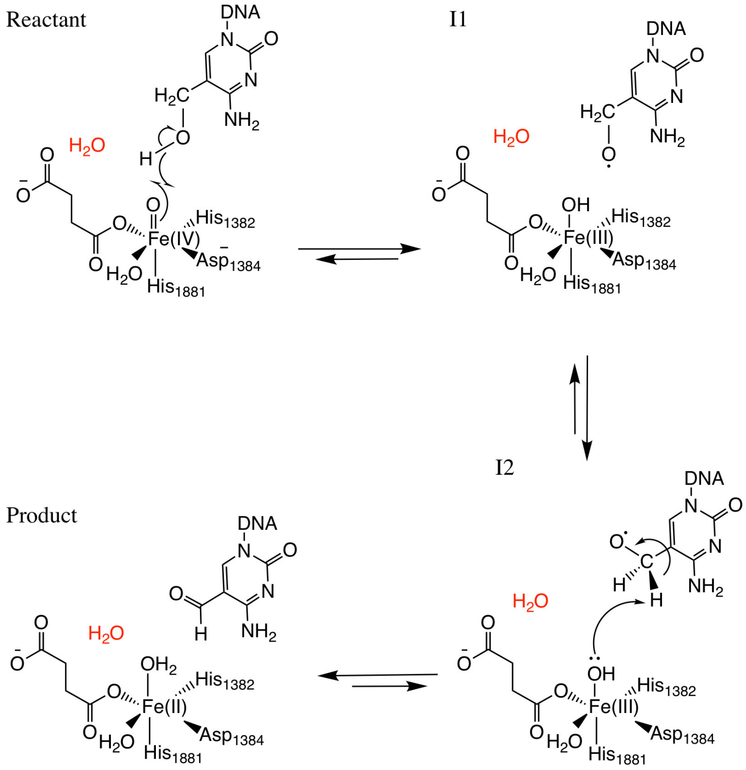

Figure 3.1 Overall mechanism for 5hmC to 5fC oxidation in TET2. The “2nd-shell" water molecule is shown in red.

Therefore, we included that water molecule in our QM region. This water will be named as the “2nd-shell” water molecule in the subsequent discussion to distinguish it from the coordinated water at the equatorial position. Figure 3.2 shows QM and pseudobond atoms in the reactant state and the “2nd-shell” water molecule is circled in red. All calculations were carried out in the quintet electronic state since previous spectroscopic [27] and computational [20, 164-165] studies found that mononuclear iron enzymes were in the high-spin (S=2) electronic ground state configurations and quintet Fe(IV)-oxo species was the most reactive toward C-H bond activation.

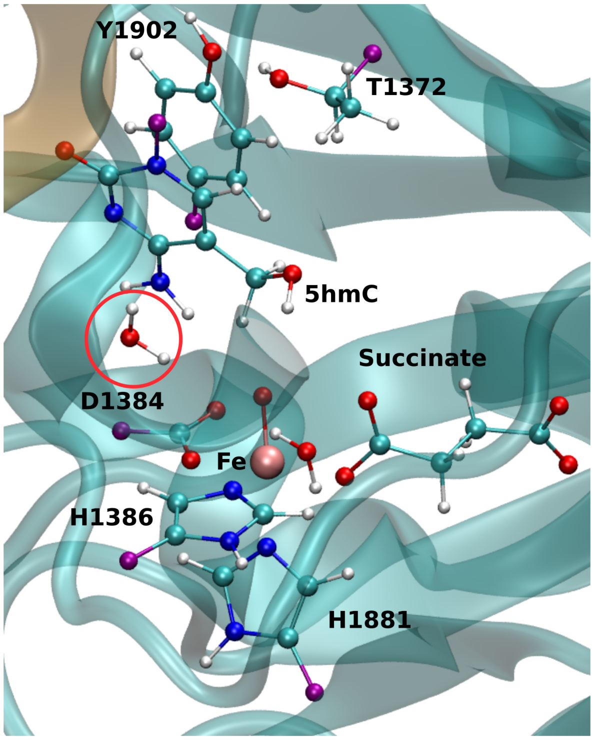

Figure 3.2 QM/MM active site in the wild type TET2. The protein (DNA) in MM region is shown in cyan (brown). The purple atoms are the pseudobond atoms. The “2^{nd}-shell" water molecule is shown by a red circle.

The QM region was modeled at the ωB97XD/6-311G(d,p) level of theory.[20, 166 167] The recent studies on the hydrogen abstraction mechanism by cytochrome P450 and AlkB indicate the importance of this functional in calculation of the barrier energy, and for the accurate description of the Fe(IV)=O intermediate.[168-169] The remainder of the system (MM region) was treated by the AMEBR99SB force field.[146] In LICHEM, the geometry optimizations are performed by adding the MM point-charges to the effective QM Hamiltonian,[170-171] where the QM and MM atoms are optimized separately.[161] To have a smooth connection at the QM/MM interface, the boundary atoms were modeled by the pseudobond approach.[172]

After the reactants and products were fully optimized, the Nudged Elastic Band (NEB) method[173-174] as implemented in LICHEM was used for path optimizations between critical structures. Frequency calculations were performed for all critical point structures, and it was found that reactants and intermediates have no imaginary frequencies. The transition states in NEB method are not necessarily optimized at a stationary point, because NEB algorithms converges the critical structures onto the minimum energy path (MEP). Therefore, the frequency calculations are not performed for TSs.

Energy decomposition analysis (EDA) was used to calculate non-bonded inter-molecular interaction energies (Coulomb and VdW interactions) between the QM subsystem (approximated by ESP fitted charges) and each residue in the MM environment for an ensemble of structures of 500 ps. The difference between the average interaction energies in the reactant and all critical structures (intermediates and transitions states) can qualitatively indicate the impact of each individual residue on the reaction barrier. All EDA calculations were carried out with an in-house FORTRAN90 program.[43, 154-155] This analysis has been previously employed for QM/MM and MD simulations to study a number of protein systems.[20-21, 156-159]

The electron localization function (ELF) analysis[175-177] is a topological analysis that can measure the electron localization in molecular systems. It was initially proposed on the basis of the Hartree-Fock (HF) approach[175] and then was extended for DFT. The details on the topological analysis of ELF and its calculation methods in enzyme systems have been discussed previously.[20-21, 169] All ELF calculations were performed with the TopMod package[178] with a 200^3 au grid with a step size of 0.1 au, and the wave functions are truncated, including only the QM atoms. The isovalue for the visualization is 0.5.

Non-covalent interaction (NCI) analysis is based on the analysis of the relation between the electronic density and the reduced density gradient (RDG) in regions of low electron density. The results obtained from the analysis consist of surfaces between the interacting molecules. These surfaces are assigned specific colors to denote the strength and characteristic of the interactions: green surfaces denote weak interactions (for example, van der Waals (VdW)), blue surfaces strong attractive interactions (for example, hydrogen bonds), and red surfaces strong repulsive interactions. All NCI calculations were performed with the NCI-Plot program[152-153] using the same truncated wave functions as for the ELF calculations.

3.3 Results and discussion

3.3.1 Optimization of the reactant structures

Fourteen and six stable structures from the MD simulation of wild type and T1372E mutant, respectively, were extracted to perform QM/MM optimizations on the reactant state (5hmC). Figure 3.3 shows the relative energies of QM/MM optimized snapshots for wild type and T1372E mutant. The lowest energy snapshots were selected for further calculations (snapshot number 2 in wild type and number 3 in T1372E mutant).

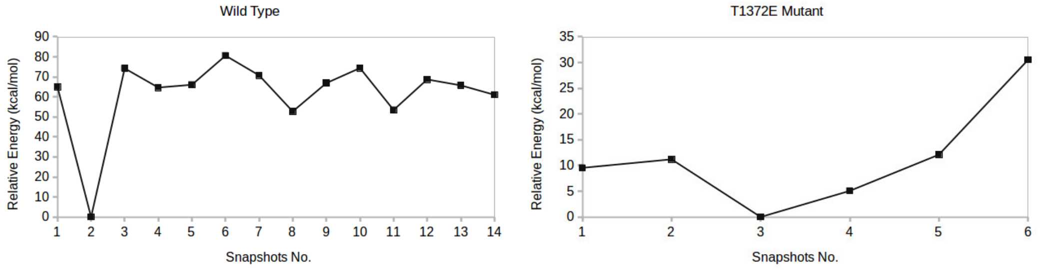

Figure 3.3 Relative QM/MM energies of the selected snapshots from the MD simulations in the wild type and the T1372E mutant.

The lowest energy structures for the TET2/5fC wild type system were used to model the product state for each selected snapshot. We also optimized the product state corresponding to the snapshot number 1 in the T1372E mutant due to its relatively low energy and different optimized geometry. The optimized geometries of the active site of the TET2-5hmC/fC complex in the wild type and the T1372E mutant along with NCI results are presented in Figures 3.4, 3.5 and 3.6.

In the wild type reactant state (Figure 3.4) the oxyl moiety forms a hydrogen bond with the “2nd-shell” water molecule (similar to AlkB[20]), and another hydrogen bond with the hydroxyl group of 5hmC, simultaneously. These hydrogen bonds align the ferryl moiety and 5hmC in the active site for further oxidation. In the product, the succinate forms a hydrogen bond with the newly formed water molecule (at the axial position) to stabilize the product.

Table 3.1 Mulliken spin densities and electron configurations for the reactant of the snapshot number 1 and 3 in theT1372E mutant and wild type.

Mulliken Spin density	snapshot number 1	snapshot number 3	Wild type
Fe (III)	3.28	4.32	3.25
Oxyl O	0.50	-0.69	0.54
Electron configuration	^{IS}Fe(III)-O$_F$	^{HS}Fe(III)-O$_{AF}$	^{IS}Fe(III)-O$_F$

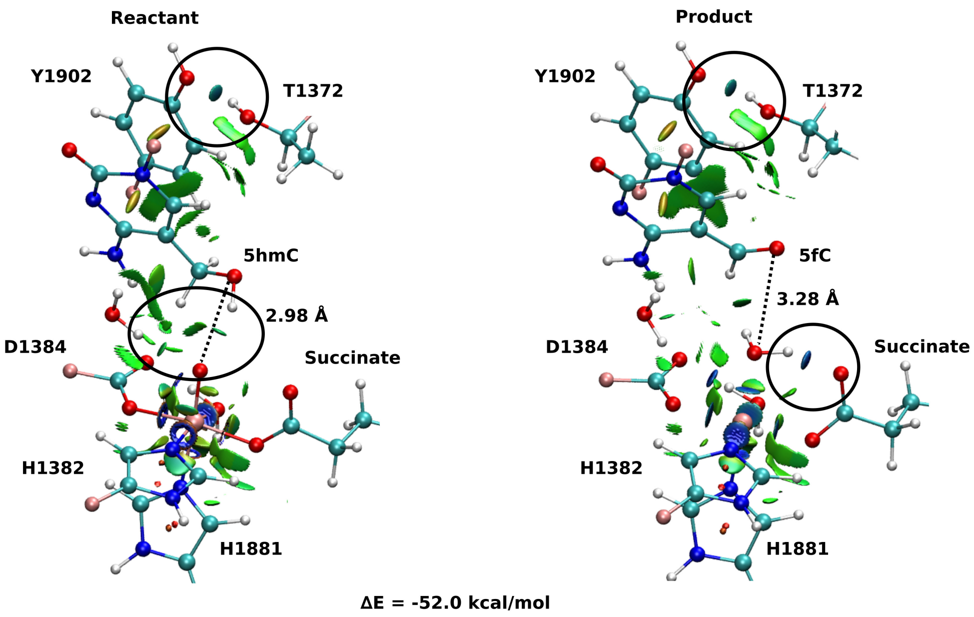

Figure 3.4 NCI plot for optimized geometries of the active site of the wild type with 5hmC/fC substrate. The hydrogen bonds are circled in black. The dashed line shows the distance between the oxygen atom of 5hmC/fC and the oxyl moiety. The MM subsystem is omitted for clarity.

The calculated reaction energy for the oxidation of 5hmC to 5fC in the wild type is -52.0 kcal/mol. The spin densities on the iron and the oxyl moiety in the reactant state are 3.25 and 0.54, respectively. This result suggests that there are 3 alpha electrons in the d orbitals of the iron, and an alpha electron in the p orbital of the oxyl moiety. Therefore, the electron configuration for the reactant is ^{IS}Fe(III)-O_F (Table 3.1 and Figure 3.7), which agrees well with the electron configuration of the iron-oxyl moiety in AlkB reactant.[20]

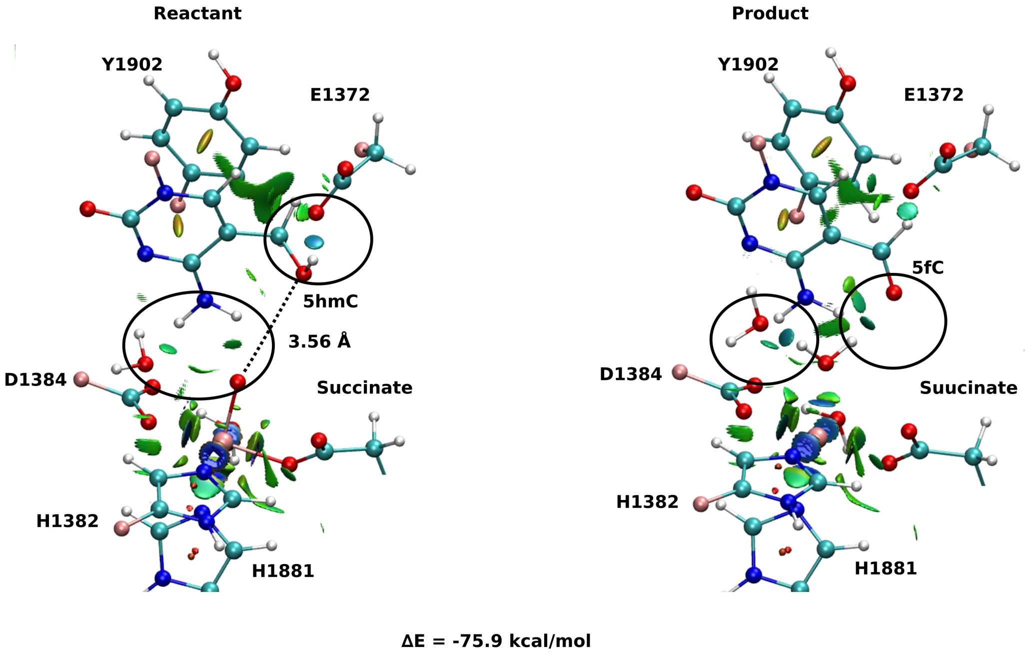

Figure 3.5 NCI plot for optimized geometries of the active site of the snapshot number 3 in the T1372E mutant with 5hmC/fC substrate. The hydrogen bonds are circled in black. The dashed line shows the distance between the oxygen atom of 5hmC/fC and the oxyl moiety. The MM subsystem is omitted for clarity.

In snapshot number 3 of the T1372E mutant, the oxygen atom of the “2nd-shell" water molecule is oriented toward the oxyl moiety (Figure 3.5). This is in contrast with the orientation of this water molecule in the wild type (Figure 3.4). Consequently, the spin densities on the iron and the oxyl moiety change and indicate the existence of 4 alpha electrons on the iron and a beta electron on the oxyl moiety. This spin densities correspond to an ^{HS}Fe(III)-O$_{AF}$ electron configuration (Table 3.1, Figure 3.7). The corresponding product is highly stabilized due to a hydrogen bond formed between the “2nd-shell" water molecule and the newly formed water molecule. In addition, the new water molecule forms a hydrogen bond with formyl group of 5fC. These new interactions provide a large stabilization of the product with a resulting reaction energy of -75.9 kcal/mol for the E mutant.

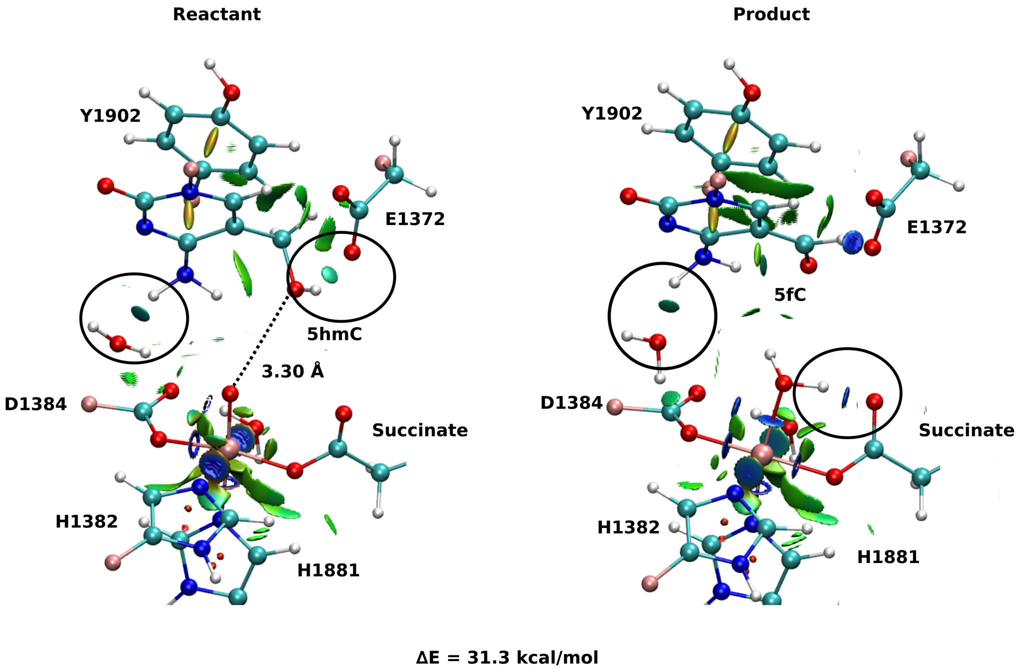

Figure 3.6 NCI plot for optimized geometries of the active site of the snapshot number 1 in the T1372E mutant with 5hmC/fC substrate. The hydrogen bonds are circled in black. The dashed line shows the distance between the oxygen atom of 5hmC/fC and the oxyl moiety. The MM subsystem is omitted for clarity.

As explained above, the product corresponding to snapshot number 1 of the T1372E mutant was also optimized to investigate whether there are other possible paths for the oxidation of this mutant. In contrast with snapshot number 3 of the E mutant, the hydrogen atom of the “2nd-shell" water molecule in this snapshot is pointed toward the oxyl moiety (similar to the orientation observed in the wild type), therefore, it can not form a hydrogen bond with the newly formed water molecule to stabilize the product

(Figure 3.6). The spin densities for the iron-oxyl moiety confirm the electron configuration of this system is similar to the wild type (^{IS}Fe(III)-O_F) (Table 3.1).

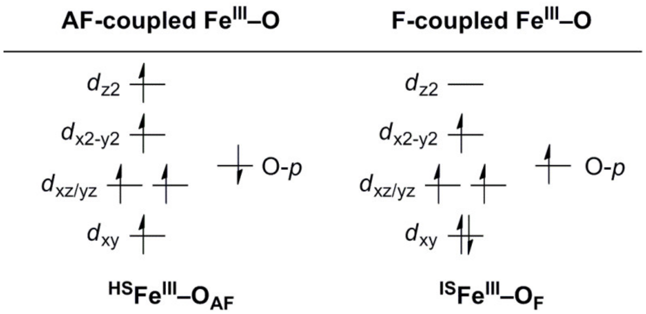

Figure 3.7 Electron configurations for the iron-oxyl moiety in quintet state.

These results support the important role of the “2nd-shell" water molecule in the oxidation mechanism. Although the electron configuration of the Fe-oxyl moiety is similar to the wild type, the reaction energy for this snapshot is very endothermic (31.3 kcal/mol). Thus, the snapshot number 1 of T1372E mutant was not used for subsequent reaction path calculations.

Our QM/MM optimized structure in both TET2-5hmC/fC complexes in the wild type show that T1372 forms a hydrogen bond with Y1902 (Figure 3.4). This hydrogen bond is eliminated by mutating T1372 to Glutamate. Instead, the mutated residue (E1372) forms a hydrogen bond directly with 5hmC in the active site (Figure 3.5). This new hydrogen bond stabilizes 5hmC and disrupts the non-bonded interaction between 5hmC and Y1902. Therefore, the 5hmC orientation toward the ferryl moiety changes and the distance between the hydroxyl group and the oxyl moiety increases in comparison with the with wild type (Figure 3.4, 3.5). The hydrogen bond formed between the amino group of 5hmC and water molecules in the reactant and the product of the T1372E mutant confirms the misalignment of 5hmC in the active site (Figure 3.5). These observations help to explain the large energy barrier calculated for the oxidation of 5hmC to 5fC in this mutant compared to the wild type reaction barrier as detailed below.

3.3.2 NEB path calculations for the oxidation of 5hmC to 5fC

Based on the optimized reactant and product structures, the potential energy surface of the whole reaction pathway of the TET2 mediated oxidation on 5hmC in the wild type and the T1372E mutant were obtained and compared. Fe(IV)-oxo in the reactant state in our study is generated after O-O heterolysis. The high valence Fe(IV)-oxo species is highly reactive and can therefore activate the O-H bond of the substrate and mediate the hydrogen abstraction reaction.

Table 3.2 Comparison of the reaction and the first barrier energies for TET2, AlkB and ALKBH2. [a] Data from A. R. Walker (unpublished)

	ΔE barrier (kcal/mol)	ΔE reaction (kcal/mol)
AlkB	23.2	-3.7
ALKBH2[a]	25.7	-3.5
TET2	20.1	5.4

The hydrogen abstraction step is known to be the rate-limiting step in the catalysis of the Fe(II)/α-KG dependent enzyme, AlkB. Figure 3.8 shows the NEB path for the 5hmC to 5fC oxidation in the wild type TET2. The calculated reaction path shows that 5hmC converts to 5fC via three steps. The first step is the rate-limiting step with a 20.1 kcal/mol barrier. The experimental barrier energy for the rate-limiting step of 1-methyladenine (1meA) oxidation mediated with AlkB is 19.8 kcal/mol.[179] In addition, a comparison of the calculated barrier and the reaction energies in the wild type TET2, AlkB and ABH2 is presented in Table 3.2.

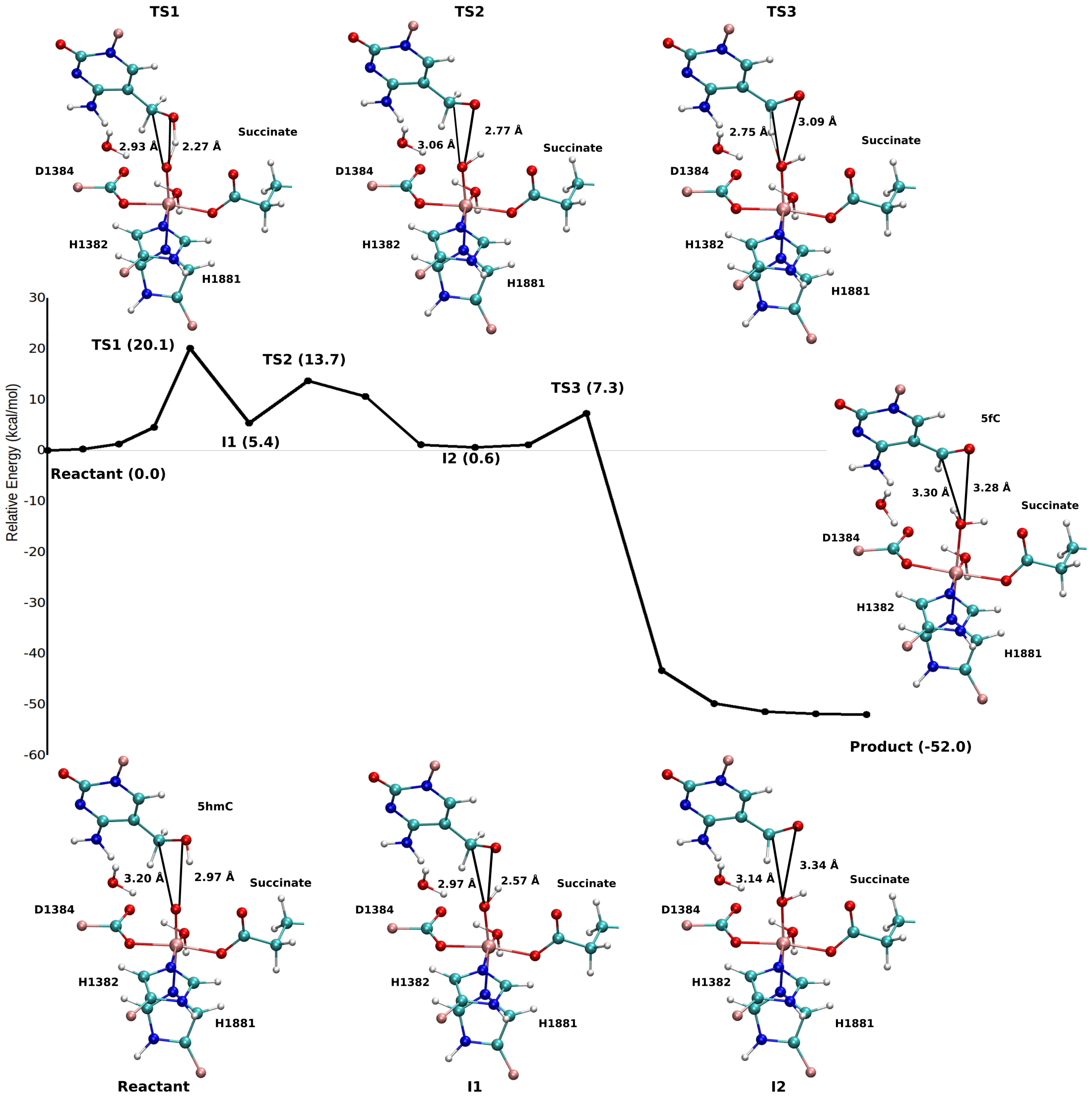

Figure 3.8 NEB path and the optimized geometries of critical structures for the wild type. The distance between the carbon and hydrogen atoms of the DNA substrate and the oxyl moiety is shown in solid black line.

In the second step, the distance between the substrate and the ferryl moiety changes to properly orient the substrate for the next step (Figure 3.8). The proton transfer from (-CH_2-) group of 5hmC occurs via a third transition state to form 5fC (Figure 3.8).

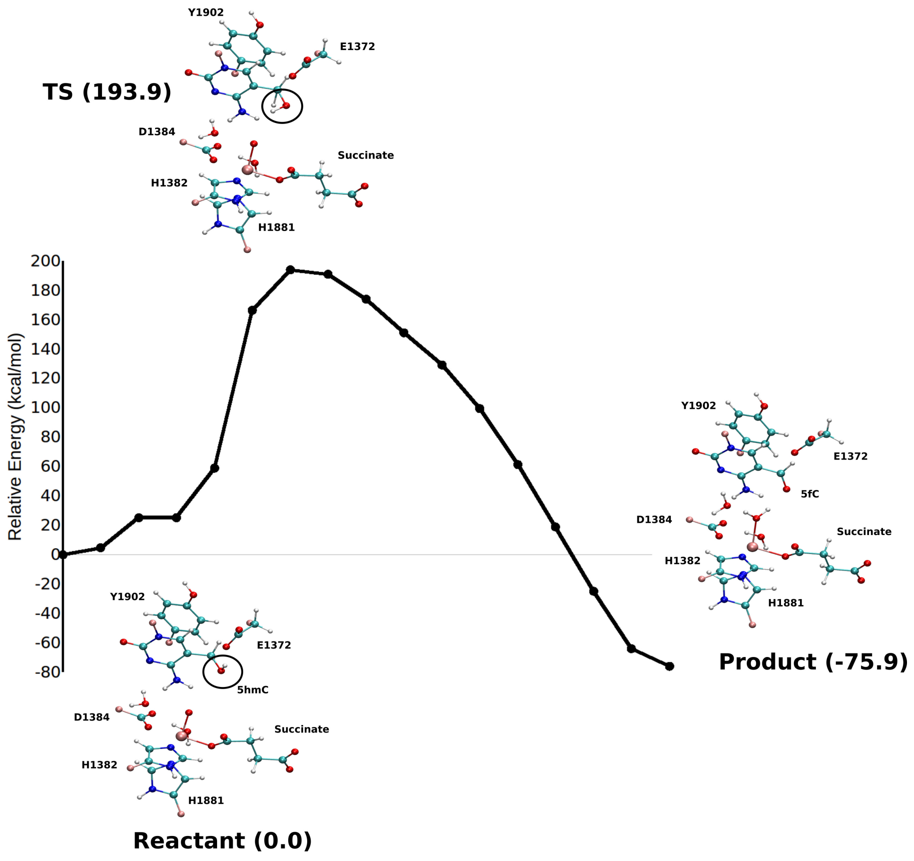

Figure 3.9 NEB path and the optimized geometries of critical structures for the T1372E mutant. The distance between the carbon and hydrogen atoms of the DNA substrate and the oxyl moiety is shown in solid black line.

The calculated path for the T1372E mutant (Figure 3.9) indicates that this reaction is a one-step reaction with a very large barrier (193.9 kcal/mol). This result shows that the oxidation of 5hmC to 5fC for this mutant is kinetically impossible. The hydroxyl group of 5hmC is pointed toward E1372 to form a hydrogen bond in the reactant, while it rotates toward the oxyl moiety at TS. Therefore, the hydroxyl group of 5hmC needs to break the hydrogen bond with E1372 and rotates toward the oxyl moiety to initiate the oxidation. Once it breaks the hydrogen bond, the oxidation occurs concomitantly. However, the large barrier for the rotation of 5hmC prohibits the oxidation of 5hmC to 5fC.

3.3.3 Spin states of critical structures in the path

According to the proposed reaction mechanism, the ferryl moiety bears two positive charges. Ideally, the charge of the iron should be 4+, and the charge of the oxygen should be 2-. However, an electron can transfer between Fe and O atoms, and both Fe(IV)=O and Fe(III)-O^- could be possible. In O^- the electron can be spin-down or spin-up, so there are two substrates for the quintet and triplet surfaces, respectively, corresponding to antiferro- or ferromagnetic coupling to the iron. Fang *et al* [20] studied the spin density of the ferryl moiety in AlkB extensively. They showed that the minimum energy path for AlkB oxidation involves a minimum energy crossing point between the reactant and the first transition state from anti-ferromagnetically coupled Fe(III)-O^- to ferromagnetically coupled Fe(III)-O^- in the quintet state.

Table 3.3 Mulliken spin densities and electron configurations for reactant, TS1 and I1 in TET2.

Mulliken spin density	Reactant	TS1	I1
Fe (III)	3.25	4.13	4.25
Oxyl O	0.54	0.01	0.25
Hydroxyl O	0	-0.43	-0.86
Hydroxyl H	0	0.04	0.01
Electron configuration	^{IS}Fe(III)-O_F	^{HS}Fe(III)-O_{AF}	^{HS}Fe(III)-O_{AF}

The Mulliken spin density on iron-oxyl moiety (Table 3.3) in the reactant confirms that the intermediate spin iron coupled to oxygen ferromagnetically (consistent with the reactant in AlkB). The spin density on the oxygen atom of the hydroxyl group in the first intermediate (I1) is -0.86, indicating the existence of a beta electron in its p orbital. This suggests that an alpha electron from the oxygen atom of the hydroxyl group transfers to the p orbital of the oxyl group to form an O-H bond. However, the electron in the p orbital of the oxyl moiety is spin-up. Therefore, this transfer is spin-forbidden and it is not conducive to form the O-H bond, unless the electron configuration of the iron-oxyl moiety changes.

Table 3.4 Comparison of Mulliken charge densities of important atoms involved in the oxidation in TET2, AlkB, ABH2.

Mulliken spin density	Fe	O (Oxyl)	C(1meA)/O(hmC)	H(1meA)/H(hmC)
AlkB reactant	3.26	0.54	0.00	0.00
AlkBH2 reactant	3.20	0.61	0.00	0.00
TET2 reactant	3.25	0.55	0.00	0.00
AlkB I1	4.35	0.26	-0.92	0.005
AlkBH2 I1	4.36	0.26	-0.92	0.004
TET2 I1	4.35	0.25	-0.85	0.01

The spin densities for the first transition state (TS1) shows that high spin iron(III) anti-ferromagnetically coupled to the oxygen. Therefore, the electron in the p orbital of the oxyl moiety is spin-down and can be able to form the O-H bond with the alpha electron of the oxygen atom of the hydroxyl group (5hmC). The change in the electron configuration from the reactant to the first transition state confirms the existence of a crossing point between these two surfaces. These results agree with the results obtained for AlkB. Table 3.4 shows the comparison between spin densities of the wild type in AlkB and ABH2 with TET2. The difference in the spin densities of oxygen in hmC and carbon in 1meA is because hydrogen is abstracted from an electronegative atom in TET2.

As discussed above, the electron configurations of two selected snapshots of T1372E mutant are different due to their different optimized geometry and the orientation of the water molecule in active site. The electron configuration in snapshot number 1 with endothermic reaction energy is $^{IS}Fe(III)\text{-}O_F$. The orientation of water molecule in this snapshot, and consequently, its electron configuration is similar to those in the wild type. However, the oxidation mechanism for this snapshot is thermodynamically unfavorable due to its unstable product. In the snapshot with highly stable product (snapshot number 3), the electron configuration is $^{HS}Fe(III)\text{-}O_{AF}$ and the mechanism is kinetically unfavorable.

3.3.4 ELF/NCI analysis on critical structures

ELF and NCI together can be used to investigate the localization of electrons and the interactions between molecules. ELF can indicate the positions of the electrons in the system and reveal the strong interactions, e.g., atomic centers, lone pairs, and covalent bonds. However, NCI demonstrates weaker, non-covalent interactions such as hydrogen bonding and van der Waals forces. The combination of these two analyses provides a qualitative view of the evolution of the electron structure without the need for complicated orbital analyses. The changes in the basin populations of the bonds

(V($O_{(hmC)}$, H_1), V($O_{(oxo)}$, H_1), V($C_{(hmC)}$, H_2), V($O_{(oxo)}$, H_2) and V($O_{(hmC)}$, $C_{(hmC)}$)) involved with the hydrogen transfers along the pathway for the wild type-catalyzed oxidation are shown in Table 3.5. The basin population of $O_{(hmC)}$-H_1 declines from reactant to I1, while it increases in $O_{(oxo)}$-H_1. This indicates that the H atom is transferred from the hydroxyl group of 5hmC to the ferryl moiety, and $O_{(oxo)}$-H_1 bond is formed. At the TS1, the basin is shared by $O_{(oxo)}$, H_1 and $O_{(hmC)}$ to form a three-center two-electron bond between these atoms. This confirms that the hydrogen atom is abstracted in this step.

Table 3.5 Population change in V($O_{(hmC)}$,H_1), V($O_{(oxo)}$,H_1), V(C(hmC),H2), V($O_{(oxo)}$,H_2) and V($O_{(hmC)}$,$C_{(hmC)}$) basins along the MEP. H_1 and H_2 indicates first and second hydrogen being transferred, respectively. For TS1, the basin is shared by three atoms, indicating hydrogen atom is abstracted from hydroxyl moiety.

	V($O_{(hmC)}$,H_1)	V($O_{(oxo)}$,H_1)	V($C_{(hmC)}$,H_2)	V($O_{(oxo)}$,H_2)	V($O_{(hmC)}$,$C_{(hmC)}$)
Reactant	1.35	0	2.17	0	0.96
TS1	0.5 (V($O_{(hmC)}$,H_1,$O_{(oxo)}$))		2.16	0	0.87
I1	0	1.18	2.21	0	0.86
TS2	0	1.22	2.19	0	0.90
I2	0	1.24	2.17	0	0.92
TS3	0	1.26	2.26	0	0.95
Product	0	1.34	0	1.35	0.97-1.14

By contrast with the first hydrogen abstraction, the second hydrogen is transferred as a proton. At TS3, there is no shared basin between $C_{(hmC)}$, H_2 and $O_{(oxo)}$. This suggests that there is no electron that transfers with the hydrogen atom. In addition, the comparison of basin populations on $C_{(hmC)}$-H_2 and $O_{(oxo)}$-H_2 in I2 and the product confirms that the hydrogen atom is transferred from 5hmC to Fe-hydroxyl moiety. There are two basin populations on $O_{(hmC)}$-$C_{(hmC)}$ in the product, suggesting the formation of double bond for 5fC. This process is also presented in Figure 3.10.

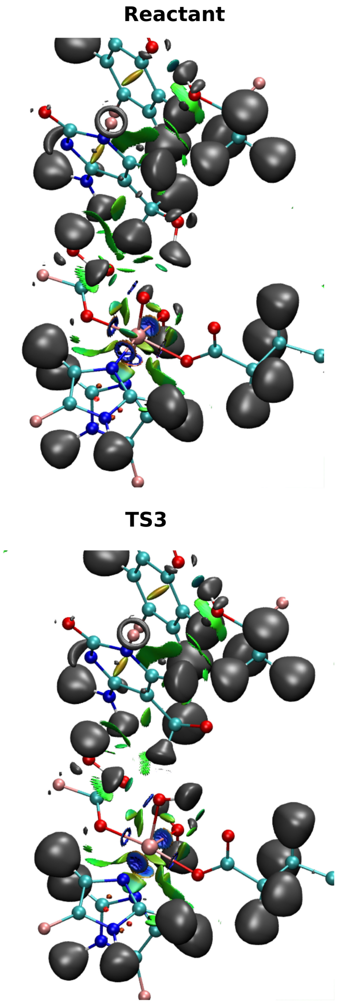

Figure 3.10 ELF and NCI surfaces for the active site of the reactant; TS1; TS3 and product in the wild type TET2-5hmC/fC complex. The isovalue for ELF and NCI is 0.5. For the NCI surfaces, the color scale is chosen so that blue indicates relatively strong attraction, green indicates relatively weak interaction, and red indicates relatively strong repulsion in the region of non-covalent interactions. The MM region is not shown for clarity.

The NCI analysis in Figure 3.10 shows 6 blue surfaces around the iron atom, indicating its octahedral structure. In addition, there is an NCI surface between the oxo ligand and the first hydrogen being transferred. Subsequently, this surface disappears to form a shared ELF basins between $O_{(hmC)}$, H and $O_{(oxo)}$ atoms. As the reaction proceeds, the basin between H and $O_{(oxo)}$ forms. There is also a similar NCI surface between oxo moiety and the second hydrogen being transferred at the beginning of the reaction. This surface vanishes when the first hydrogen is transferred and forms again before transferring the second hydrogen. Then, it disappears again while second hydrogen is transferring. The second hydrogen is transferred without sharing an ELF surface between three atoms, suggesting a proton transfer. When water molecule forms, one blue surface appears between the water and the iron, showing the water is coordinated to the iron. In addition, an ELF surface forms on the oxygen atom of the water molecule as oxygen lone pair. Besides, new water forms hydrogen bond with succinate while its interaction with the “2nd-shell" water molecule decreases.

3.3.5 Energy decomposition analysis on critical structures

To gain a qualitative understanding of the effects of each residue in the MM environment on the QM subsystem, an energy decomposition analysis (EDA) at the residue level has been performed. This analysis is based on the subtraction of the average intermolecular interaction energy (ΔE) between each residue in the MM region and the QM subsystem at all critical structures.[159, 180-183] Figure 3.11 a-c shows the EDA results for TS1, TS2 and TS3 with respect to the reactant. The sign of this energy difference may be used to qualitatively determine the catalytic role of the residues. Positive ΔE indicates that this residue should increase the reaction rate, while the reaction rate should be decreased with residues that have a negative ΔE.

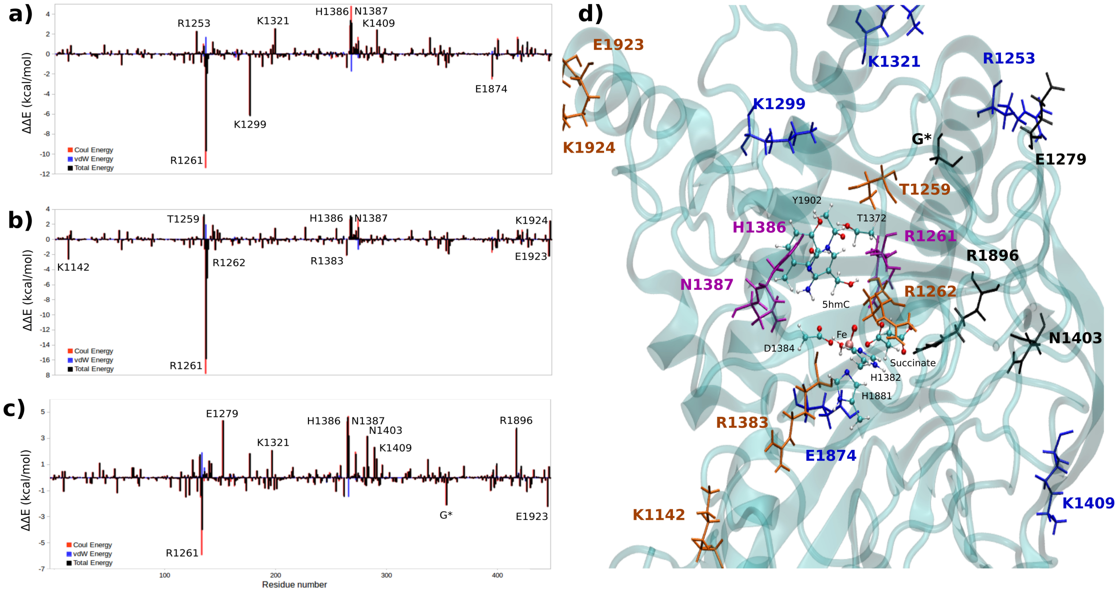

Figure 3.11 Difference of total, Coulomb and vdW energies of a) TS1, b) TS2, c) TS3 and reactant. d) Positions of the residues obtained from (TS1/2/3 and reactant) EDA relative to the QM region. The residues obtained from TS1/2/3 are shown in blue/brown/black. The purple residues are in common for all EDA. G* is the one of the linkers. The QM atoms are shown in ball and stick.

In this study, we consider the residues that have a |ΔΔE| with magnitude larger than 2 kcal/mol (Figure 3.11 and Table 3.6). EDA based on the optimized structures of the TS1 and the reactant reveals eight residues have a |ΔΔE| >2 kcal/mol. These residues include R1261, R1253, K1299, K1321, H1386, N1387, K1409, E1874. The EDA on the TS2 and TS3 indicate the following residues are involved in catalysis: K1142, T1259, R1261, R1262, E1279, K1321, R1381, H1386, N1387, N1403, K1409, R1896, E1923 and E1924. Residues R1261, H1386 and N1387 are common in all three EDA results. The position of these residues in the protein are shown in Figure 3.11 d (see Table 3.6 for ΔΔE values). This result can be used as a guide for experimental mutagenesis.

Table 3.6 Residues with more than 2 kcal/mol difference by EDA. All values are in kcal/mol.

Residue	ΔΔE (TS1-reactant)	ΔΔE (TS2-reactant)	ΔΔE (TS3-reactant)
K1142	-	-2.6	-
R1253	2.3	-	-
T1259	-	3.3	-
R1261	-9.7	-15.9	-4.0
R1262	-	-5.2	-
E1279	-	-	4.4
K1299	-6.2	-	-
K1321	2.5	-	2.1
R1383	-	-2.1	-
H1386	3.4	3.1	4.6
N1387	3.1	2.9	3.2
N1403	-		2.3
K1409	2.4	-	-
E1874	-2.3	-	-
R1896	-	-	3.8
E1923	-	-2.2	-2.2
K1924	-	2.5	-

The sequence alignment for these residues are presented in Figure 3.12. For instance, the mutation of K1299 can significantly change the barrier since it has a large stabilizing impact on the TS1, compared to R1262 which has smaller stability effect on the TS2. Another target for mutagenesis could be the H1386 and N1387 which shows destabilizing effect on all three transition states. The sequence alignment (Figure 3.12) shows that these residues are (partially) conserved in h/m-TET1-3 homologues. Therefore, they could be suitable targets for mutagenesis. Hu et. al. mutated a series of the residues on 5mC substrate,[184] they showed that K1299E/S1303N significantly decreases the TET2 enzymatic activity, while R1262A has a minor effect on the TET2 activity. Their results indicate that R1261G reduces the TET2 activity, and our calculation suggests that R1261 has a large stabilizing effect on all TSs. This high impact of R1261 on the catalytic activity is because of its interaction with α-KG (or NOG in the crystal structure[184]). The EDA results introduce new targets for future research on TET2 mutation.

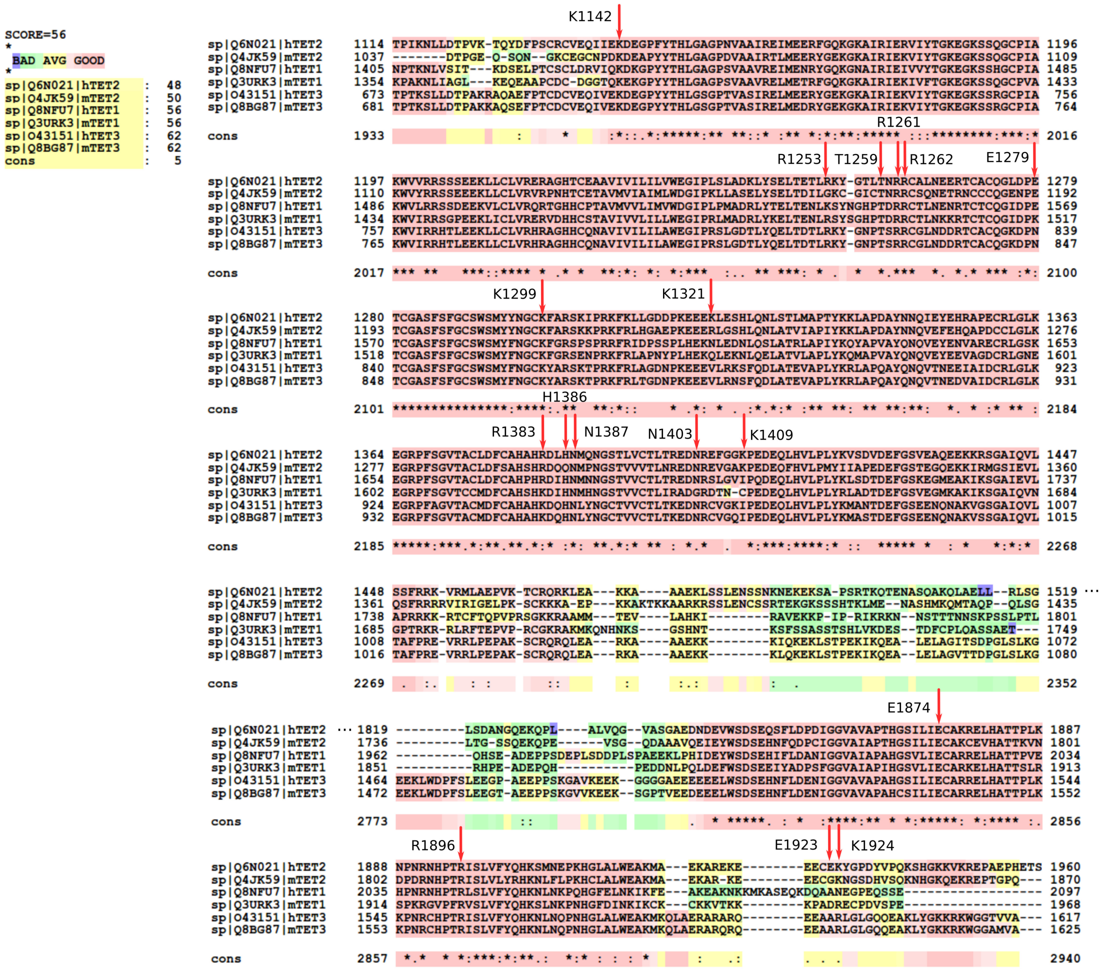

Figure 3.12 Protein sequence alignment of h/m-TET1-3 calculated using T-coffee.[185] The residues with significant change in (de)stabilizing energy ($|\Delta\Delta E| > 2$ kcal/mol) are marked with a red arrow.

3.4 Conclusions

The TET2 mediated oxidation of 5hmC has been investigated by means of QM/MM calculations. Based on the results of our previous molecular dynamics (MD) study on TET2, we performed reaction path calculations on the oxidation of 5hmC to 5fC to understand the source of substrate preference in the wild type and a “5hmC-stalling” mutant (T1372E). The calculated energy barrier of the rate-limiting step in the wild type for the oxidation of 5hmC to 5fC is in good agreement with other Fe/α-KG dependent enzymes. ELF/NCI results show that the first hydrogen transfers as a hydrogen atom, while the second transfer involves a proton. The results also compared with available experimental and computational values for AlkB and ABH2. The comparison of the optimized structure of the reactant, product, and their corresponded electron configurations between the wild type and the T1372E mutant reveals that the orientation of the water molecule in the active site is essential for iterative oxidation. We investigated two possible oxidation paths for T1372E mutant and show that one path is kinetically unfavorable, and the other one is thermodynamically impossible. EDA analysis provides insights on over 18 residues that significantly impact the catalytic step including three residues that have been experimentally shown to impact catalysis. Our calculations provide new targets for mutagenesis studies of human TET2.

CHAPTER 4 COMPUTATIONAL INVESTIGATION OF O_2 DIFFUSION THROUGH AN INTRA-MOLECULAR TUNNEL IN ALKB; INFLUENCE OF POLARIZATION ON O_2 TRANSPORT

4.1 Introduction

E. coli AlkB catalyzes the direct dealkylation of various alkylated bases in damaged DNA. The diffusion of molecular oxygen to the active site in AlkB is an essential step for the oxidative dealkylation activity. Despite detailed studies on the stepwise oxidation mechanism of AlkB, there is no conclusive picture of how O_2 molecules reach the active site of the protein. Yu *et al.*[45] proposed the existence of an intra-molecular tunnel based on their initial crystal structures of AlkB. We have employed computational simulations to investigate possible migration pathways inside AlkB for O_2 molecules. Extensive molecular dynamics (MD) simulations, including explicit ligand sampling and potential of mean force (PMF) calculations, have been performed to provide a microscopic description of the O_2 delivery pathway in AlkB. Analysis of intra-molecular tunnels using the CAVER software indicates two possible pathways for O_2 to diffuse into the AlkB active site. Explicit ligand sampling simulations suggests that only one of these tunnels provides a viable route. The free energy path for an oxygen molecule to travel along each of these tunnels has been determined with AMBER and AMOEBA. Both PMFs indicate passive transport of O_2 from the surface of the protein. However, the inclusion of explicit polarization shows an extremely large barrier for diffusion of the co-substrate out of the active site, compared with the non-polarizable potential. In addition, our results suggest that the mutation of a conserved residue along the tunnel, Y178, has dramatic effects on the dynamics of AlkB and on the transport of O_2 along the tunnel.

4.2 Computational methods

The initial structure for wild type *E. coli* AlkB in complex with DNA was taken from the Protein Data Bank (PDB ID: 2FDG[45]). Hydrogen atoms, counter ions, and TIP3P water molecules were added to the holo structure with the LEaP package[145] from AMBER14.[186] The final system size was 50000 total atoms with 3 counter ions. This system was initially equilibrated in our previous studies on AlkB.[20-21] MD simulations were performed with ff99SB[187] and AMOEBA using the NPT ensemble.[82] Simulations involving the ff99SB potential were performed with the PMEMD.cuda program from AMBER14.[186] AMOEBA simulations were carried out with an in-house, AMOEBA capable, development branch of AMBER based on pmemd termed pmemd.gem. All simulations used a 1 fs step size and a 9 Å cutoff for non-bonded interactions. The non-polarizable simulations were run at 300 K using Langevin dynamics[188] (g = 1 ps^{-1}), with a Berendsen barostat.[189] The parameters for 1-methyladenine (1meA), α-KG and O_2 for both ff99SB and AMOEBA were developed in-house. The iron cation was approximated by Mg(II) parameters based on the precedent established by our previous studies on AlkB[20-21] and TET2.[16] SHAKE[190] was used for all the simulations and the smooth particle mesh Ewald (PME) method[191] was employed to treat long-range Coulomb interactions.

The existence of possible tunnels for O_2 transport in AlkB were determined by analyzing the crystal structure, as well as 250 snapshots out of a 50 ns simulation from the non-polarizable potential using CAVER[192] as implemented in PyMOL.[193] Once the coordinates of the tunnel were obtained, umbrella sampling[194] and WHAM[195-197] were used to calculate the potential of mean force for the transport of oxygen along the tunnel. Two possible channels were obtained from CAVER analysis.[198] Each of these channels was populated by positioning 30-40 oxygen molecules to achieve a spacing of 0.3-0.5 Å between adjacent umbrella windows. A harmonic potential of 10 kcal $mol^{-1}Å^{-2}$ was applied to restrain the motion of the oxygen molecule in each window. The PMF coordinate was set as the distance between the oxygen molecule and T208 (OG1) for the blue tunnel, and α-KG (C2) in the red tunnel. Each window was sampled for 1 ns in explicit solvent (TIP3P water model) at 300 K in NPT ensemble using Langevin thermostat and Berendsen barostat. The free energy associated with the oxygen transport along each tunnel were calculated using 1D-WHAM (one-dimensional weighted histogram analysis method) technique with bootstrapping.[195-197, 199-200] All 1D-WHAM calculations were performed with the WHAM package.[201]

The PMF was also calculated using the multipolar-polarizable potential AMOEBA. The use of higher-order multipoles and/or explicit polarization has been shown to provide accurate description of various system such as water,[77, 202] organic molecules,[203-204] peptides,[205] protein-ligand binding[206] and ion channels.[207] Several methods have been developed to include many-body polarization, such as the fluctuating charge approach,[208-209] the Drude oscillator model[210-211] and atomic induced dipole methods.[136, 212] AMOEBA combines distributed atomic multipoles (up to quadrupole) and (Tholé damped) inducible atomic dipoles, which have been shown to provide an accurate representation for various systems.[76, 78, 82, 161, 204-205, 213-215]

Umbrella sampling/WHAM calculations for the AMOEBA calculations were performed using the amoebabio09 potential[216] using the same windows as for the ff99SB calculations. Various harmonic potentials (5-40 kcal $mol^{-1}Å^{-2}$) were applied to keep the oxygen molecules at the desired positions. Each window was sampled for 100 ps in explicit AMOEBA water at 300 K in the NPT ensemble using the Berendsen thermostat-barostat. Long-range electrostatic interactions were computed employing the smooth particle mesh Ewald method[191, 217-218] with an 8.5 Å cutoff. All PMFs were calculated with the WHAM program from Grossfield *et al.*[201] Bootstrap analysis was done to validate the reproducibility of the data for each free energy calculation. The coordinated water in the primary coordination sphere of iron was deleted in these calculations to avoid the increase in free energies due to close contact between O_2 and water molecules in the active site.

Direct O_2 diffusion was also investigated by running long MD simulations (500 ns) with the ff99SB force field (exclusively) on wild type AlkB and three mutants in the presence of O_2 molecules. To probe for O_2 delivery pathways in AlkB, 10 O_2 molecules, corresponding to 0.03 M [O_2] calculated with respect to the total simulation box volume of 520000 $Å^3$, were added to the equilibrated AlkB model. This O_2 concentration is 23-fold higher than the saturated O_2 concentration in water at ambient condition. This high O_2 concentration was introduced to maximize the sampling of O_2 delivery pathway(s) in the protein within limited time scales (500 ns). Four independent simulation systems were designed for wild type in which O_2 molecules were initially distributed randomly around protein surface in the solution. Each system was visually inspected to ensure that the added O_2 molecules did not overlap with atoms of other molecules, and then was simulated for 500 ns in the NPT ensemble at 300 K using Langevin dynamics (g = 1 ps^{-1}) coupled to a Berendsen barostat. Repeated simulations were performed to improve statistics and to provide a more accurate description of the O_2 delivery pathway.

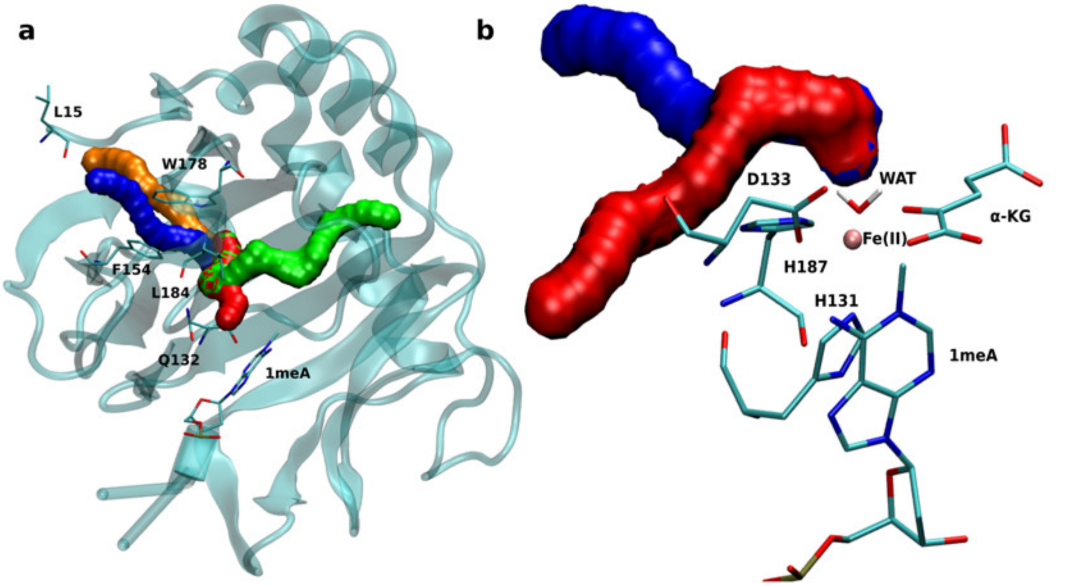

Figure 4.1 a) Tunnels obtained with CAVER for the crystal structure of AlkB. b) AlkB active site with blue and red tunnels.

Three–dimensional (3D) density maps representing the O_2 occupancy profiles were calculated to identify O_2 delivery pathways, using the Volmap tool as implemented in the VMD program.[151] Root mean square deviation (RMSD), root mean square fluctuation (RMSF), hydrogen bond, correlation, histogram and distance analyses were performed using the cpptraj module available in AMBER14 suite.[186]

Energy decomposition analysis (EDA) was used to calculate non-bonded inter-molecular interaction energies (Coulomb and VdW interactions) between selected residues. All EDA calculations were carried out with an in-house FORTRAN90 program.[154-155] The average non-bonded interaction between a particular residue, W178, and every other residue is approximated by $\Delta E_{int} = \langle \Delta E_i \rangle$, where i represents an individual residue, $\langle \Delta E_i \rangle$ represents the non-bonded interaction (Coulomb or VdW) between residue i and the residue of interest, W178, and the broken brackets represent averages over the complete production ensemble obtained from the MD simulations. This analysis has been previously employed for QM/MM and MD simulations to study a number of protein systems.[20-21, 43, 155, 157, 219]

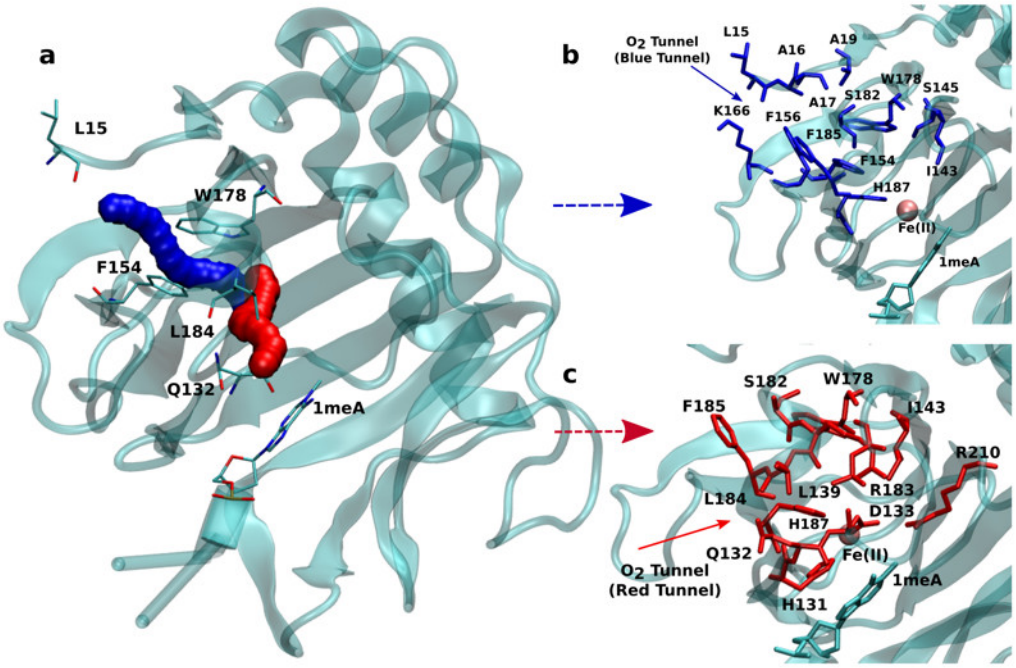

Figure 4.2 a) Two main tunnels obtained with CAVER for the crystal structure of AlkB. Residues that define the blue (b) and red (c) tunnels.

4.3 Results and discussion

This section presents the results obtained from the MD simulations and analysis for the oxygen transport through AlkB and three W178 mutants. Subsection 4.3.1 presents the results of the CAVER analysis for the crystal structure and selected MD snapshots to investigate the prevalence of each tunnel along the trajectory. Long MD simulations on wild type and W178 mutants are discussed in Subsections 4.3.2 and 4.3.3. This is followed by the presentation of the free energy profiles obtained from the umbrella sampling and WHAM calculations in Subsection 4.3.4.

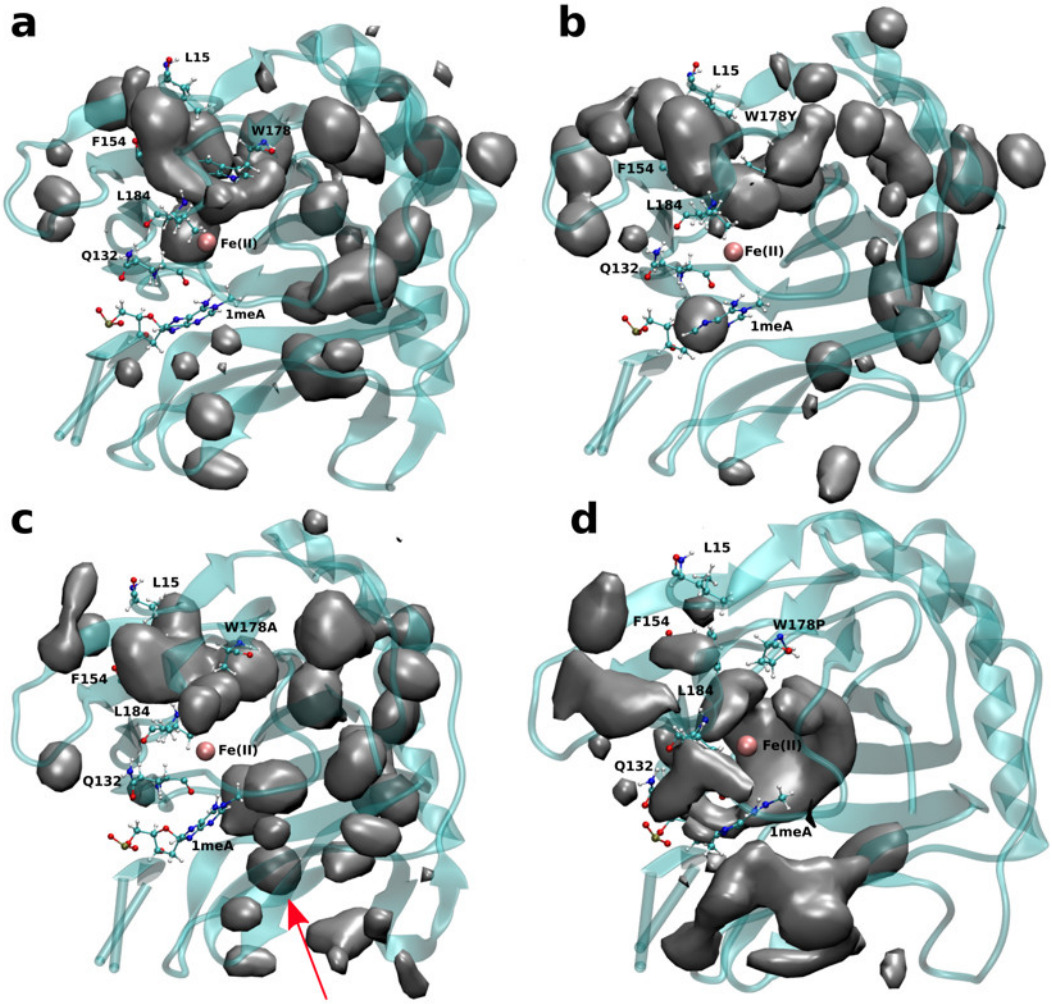

Figure 4.3 3D density man representing the O_2 molecule occupancy in a) WT, b) W178Y, c) W178A, d) W178P. The isovalue for all density maps is 0.006.

4.3.1 Probability of occurrence of different tunnels in wild type AlkB

An initial analysis of the original AlkB crystal structure, 2FDG,[45] using CAVER indicates the existence of four possible tunnels (Figure 4.1 a). All tunnels start from the surface of the protein and reach Fe(II) in the active site. The CAVER analysis further indicates that the blue and red tunnels are the most probable tunnels for O_2 diffusion due to their high throughput, low curvature and short length (Table 4.1, Figure 4.2a). The blue tunnel is defined by residues L15, A16, A17, A19, I143, S145, F154, F156, K166, W178, S182, F185 and H187. The red tunnel which comprises residues H131, Q132, D133, L139, I143, W178, S182, R183, L184, F185, H187, and R210 is consistent with the proposed tunnel by Yu *et al*.[45] Figures 4.2b and 4.2c show the residues that define the red and blue tunnels. Figure 4.1 b shows the position of the blue and red tunnels with respect to the AlkB active site. The coordinated water molecule trans to H131 occupies the site that would be replaced by the oxygen molecule traveling along each tunnel.[220] Given that the initial CAVER analysis is based on a static structure, further analysis was performed to investigate the effects of the protein motion on the predicted tunnels.

Table 4.1 Tunnel properties for crystal structure

Tunnel	Length (Å)	Bottle-neck Radius (Å)	Curvature	Priority-throughputs
Blue	15.8	0.93	1.23	0.54
Red	16.4	0.93	1.47	0.49
Orange	18.1	0.93	1.30	0.40
Green	31.4	0.93	1.97	0.23

To this end, a short MD simulation (50 ns) was carried out on wild type AlkB and 250 random snapshots were extracted and subjected to CAVER analysis. The results for all 250 snapshots indicate that out of these samples 29.4% exhibit the availability of the blue tunnel compared to 28.2% occurrence of the red tunnel (Table 4.2). In addition, the CAVER analysis suggests that the average curvature of the blue tunnel is smaller than that of the red tunnel. Taken together, these results indicate that O_2 molecules should be more easily transported through the blue tunnel.

The CAVER algorithm simplifies the oxygen molecules and protein atoms as hard spheres whose sizes are approximated by van der Waals radii. The algorithm determines the existence of putative tunnels purely based on steric considerations, without considering actual intermolecular interactions. This steric nature of the calculation prevents any insight into the energetics of transport along the tunnel.

Table 4.2 Number of occurrence and properties of each tunnel from 250 snapshots of MD simulations.

Tunnel	Number of Occurrence	Average Length (Å)	Average Bottled-neck Radius (Å)	Average Curvature	Average Priority-throughput
Blue	77 (29.4%)	17.1	1.00	1.32	0.41
Red	74 (28.2%)	15.8	1.01	1.42	0.45
Other	111 (42.4%)	-	-	-	-
Total	262	-	-	-	-

To further test the CAVER prediction, long MD simulations (500 ns) on wild type AlkB in the presence of 10 O_2 molecules randomly positioned in the surface of the protein have been performed to investigate how O_2 molecules travel along these tunnels. The results strikingly confirmed the prediction and show that the O_2 molecules prefer to travel along the blue tunnel rather than other tunnels as described in the next subsection.

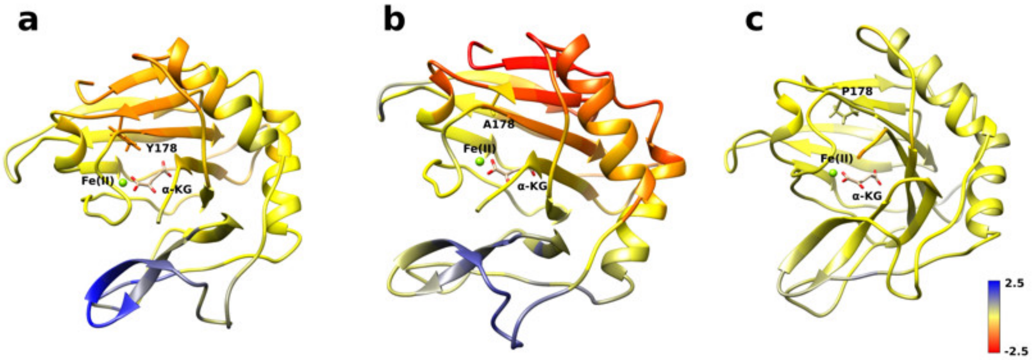

Figure 4.4 RMSF difference between wild type and mutants along the sampled trajectories for (a) W178Y, (b) W178A and (c) W178P. Residues in blue (red) denote longer (smaller) fluctuations in the mutant structures compared to the wild type.

4.3.2 Long MD simulations reveal the main oxygen diffusion pathway in AlkB

Four independent MD simulations on wild type AlkB were performed to investigate the diffusion of O_2 into and out of the active site. Each of the four independent simulations contained 10 O_2 molecules and extended for 0.5 μs. During these simulations, we observe 5 O_2 molecules diffuse into the active site on average. In all cases the diffusion of the ligand occurs exclusively via the blue tunnel pathway.

Table 4.3 Intermolecular energy difference analysis between residue 178 and all protein-DNA residues for the wild type and W178Y/A/P mutants. (b) Difference between ΔE of mutants and wild type. ΔE is the sum of all intermolecular interaction energies for all residues in each system (except water). All energies in kcal/mol.

	a			b		
	ΔE_{Coul}	ΔE_{vdW}	ΔE_{Tot}	$\Delta\Delta E_{Coul}$	$\Delta\Delta E_{vdW}$	$\Delta\Delta E_{Tot}$
WT	-41.2±0.8	-23.7±0.4	-64.9±0.6	0	0	0
W178Y	-43.7	-21.4	-65.1	-48.5	4.4	-44.1
W178A	-41.8	-9.7	-51.5	-28.6	14.5	-14.1
W178P	-32.3	-13.8	-46.1	6.7	8.3	15.0

Figure 4.3a shows the 3D density map representing the occupancy profile for O_2 molecules for the wild type and three W178 mutants (see below for mutant simulations). This density map is an average over all four 0.5 ms trajectories. The calculated density maps show that the cavities occupied by O_2 molecules in the wild type are consistent with the coordinates of the blue tunnel. The distance between the Fe(II) atom and O_2 molecules for each independent simulation are presented in Figure 4.5. Distances smaller than 6 Å between an O_2 molecule and Fe(II) were considered as a complete entrance. In all wild type simulations, we observed that all oxygen molecules that diffuse into the active site spend a very short time close to the Fe(II) atom on average, and then they escape from active site and return to solution.

4.3.3 Mutation of W178 can change the diffusion pathway

Crystal studies on AlkB[45] show that residue W178 is located at the entrance of the binding cavity. Based on its location, Yu *et al*. suggested that Y178 could act as a gate along the tunnel depending on its various side chain conformations. The effect of W178 on O_2 diffusion was probed by performing MD simulations on three AlkB mutants: W178Y, W178A and W178P. These simulations were subjected to various analyses including density maps, RMSD, RMSF and energy decomposition to investigate the effect of the mutation on the structure and dynamics of AlkB.

The occupancy profiles for each mutant are presented in Figures 4.3b-d. The occupancy profile reveals that O_2 molecules in W178Y mutant diffuse through blue tunnel similarly to wild type. The correlation analysis by residue (Figure 4.8e) indicates that W178Y has a pattern more consistent with the wild type, suggesting that the dynamics of this mutant are similar to that of wild type. However, distance analysis (Figure 4.6) indicates that O_2 molecules in the W178Y mutant spend more time in the active site in comparison with wild type (Figure 4.5). This difference in residence time could be due to the difference in size and flexibility of the tyrosine side chain with respect to tryptophan. All systems are observed to be stable throughout the simulation time as per RMSD with respect to the crystal structure (Figure 4.7).

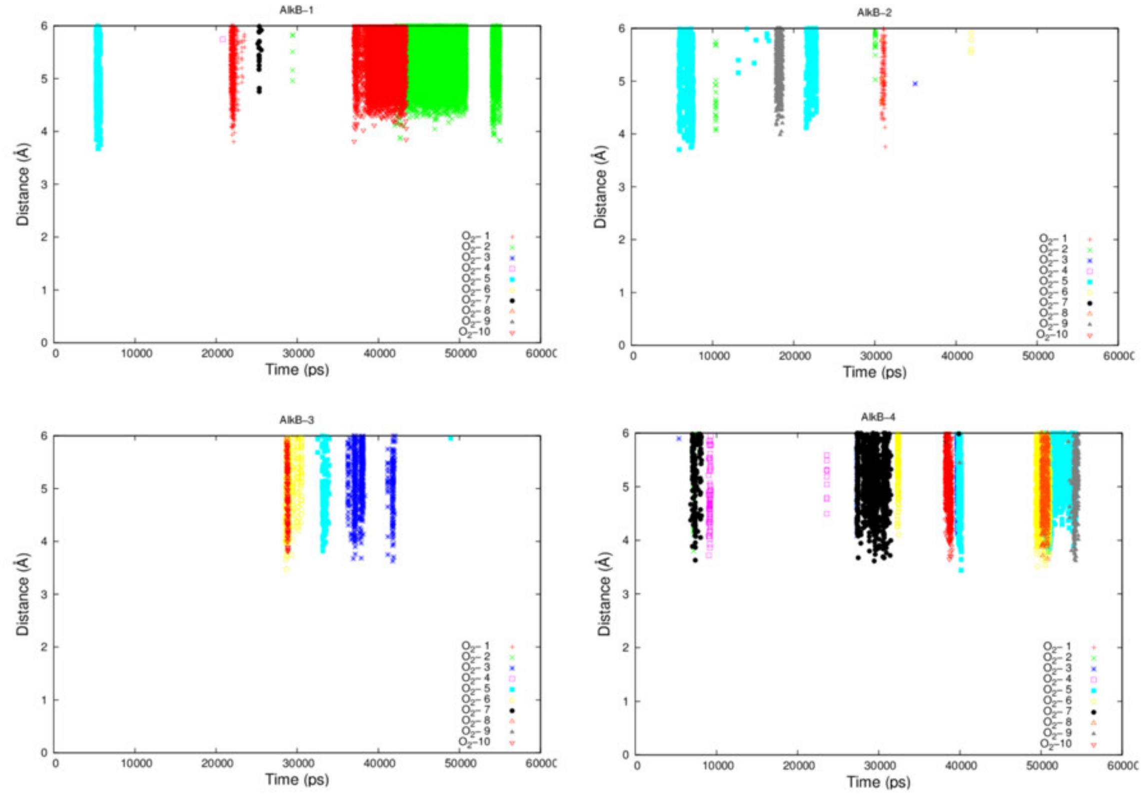

Figure 4.5 Distance between O_2 and Fe(II) in wild type AlkB for 4 independent simulations.

The analysis of the change in fluctuation along the calculated trajectories between the wild type and each mutant shows striking differences. Figure 4.4a suggests that residues around the blue tunnel in the W178Y mutant fluctuate less than the same residues in the wild type. Reduced fluctuation of these residues may explain why O_2 molecules spend more time in the active site compared to the wild type. Energy decomposition analysis (EDA) was carried out to further understand residue-by-residue inter-molecular interactions. The EDA analysis suggests that the mutation of tryptophan to tyrosine changes the stability of this residue in the protein negligibly (Table 4.3a), however, the stability of the whole protein increases significantly (Table 4.3b). Table 4.3 b shows the sum of all interactions (except water molecules interactions) for each mutant in comparison with wild type.

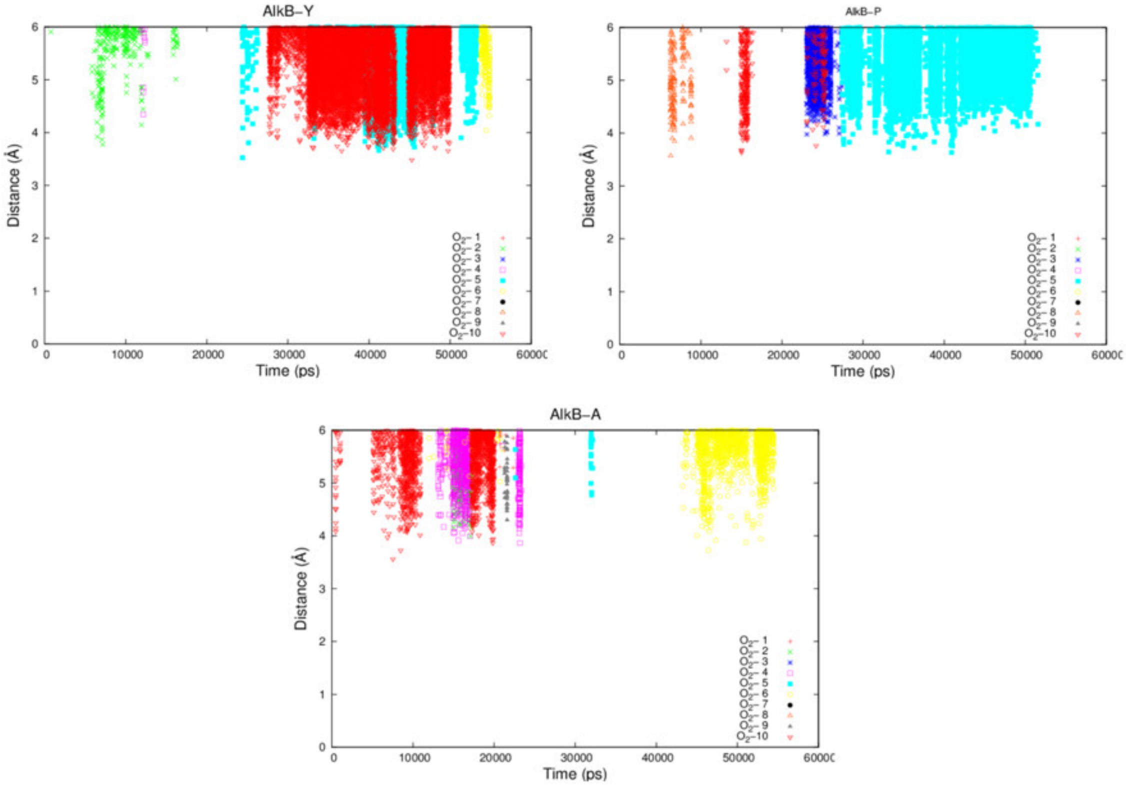

Figure 4.6 Distance between O_2 and Fe(II) for a) W178Y, b) W178P and c) W178A.

Similar to the wild type and W178Y mutant, the O_2 diffusion in the W178P mutant occurs only through the blue tunnel. However, the occupancy profile (Figure 4.3d), distance analysis (Figure 4.6) show that O_2 molecules spend a significant portion of the simulation time around the Fe(II) atom in the active site. These results suggest that the oxygen molecules get trapped in the active site after diffusion. In addition, EDA analysis reveals less stability for residue 178 in this mutant compared to wild type (Table 4.3a). The sum of all intermolecular interactions in this mutant compared to wild type (Table 4.3b) suggests that the mutation of tryptophan to proline destabilizes the protein. The RMSF (Figure 4.4) shows that the residue at the 178 position exhibits larger fluctuations in the P mutant than in the wild type structure.

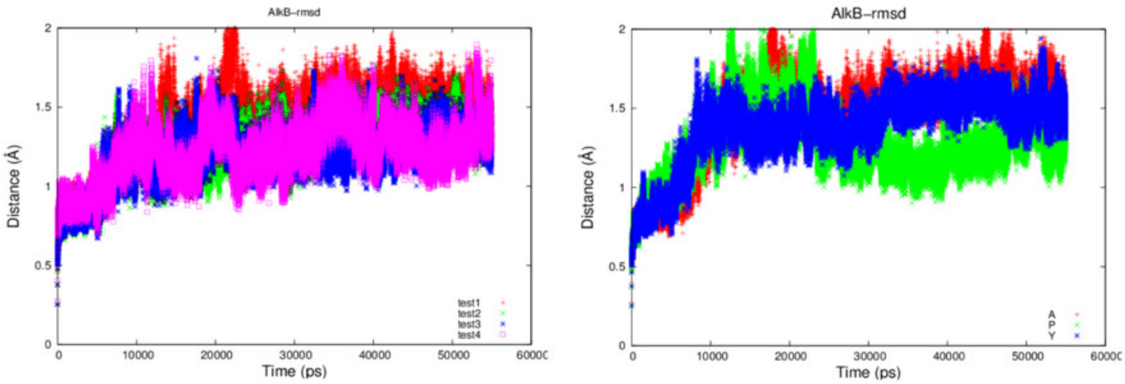

Figure 4.7 Protein backbone RMSD for a) wild type AlkB b) W178A/P/Y mutants.

The main pathway for oxygen diffusion in the W178A mutant is the blue tunnel. However, this mutant reveals a new pathway to transport O_2 molecules from the surface of the protein into the active site (Figure 4.3c). The correlation difference plot (Figure 4.8g) shows that residues 163 to 189 in the W178A mutant and wild type are anti-correlated. In addition, the RMSF difference analysis (Figure 4.4b) shows increased fluctuation for the residues around the new pathway in the W178A mutant compared to the wild type. The increased fluctuation of these residues can help to explain for appearance of a new pathway in the W178A mutant.

In summary, the MD simulations indicate that the main pathway for oxygen diffusion in all mutants is the blue tunnel. However, the access to this tunnel can be facilitated or hindered by the mutation of W178, depending on the size and flexibility of the residue. Although these results support the importance of W178 for oxygen diffusion, experimental studies on this residue are needed to confirm the importance of this site.

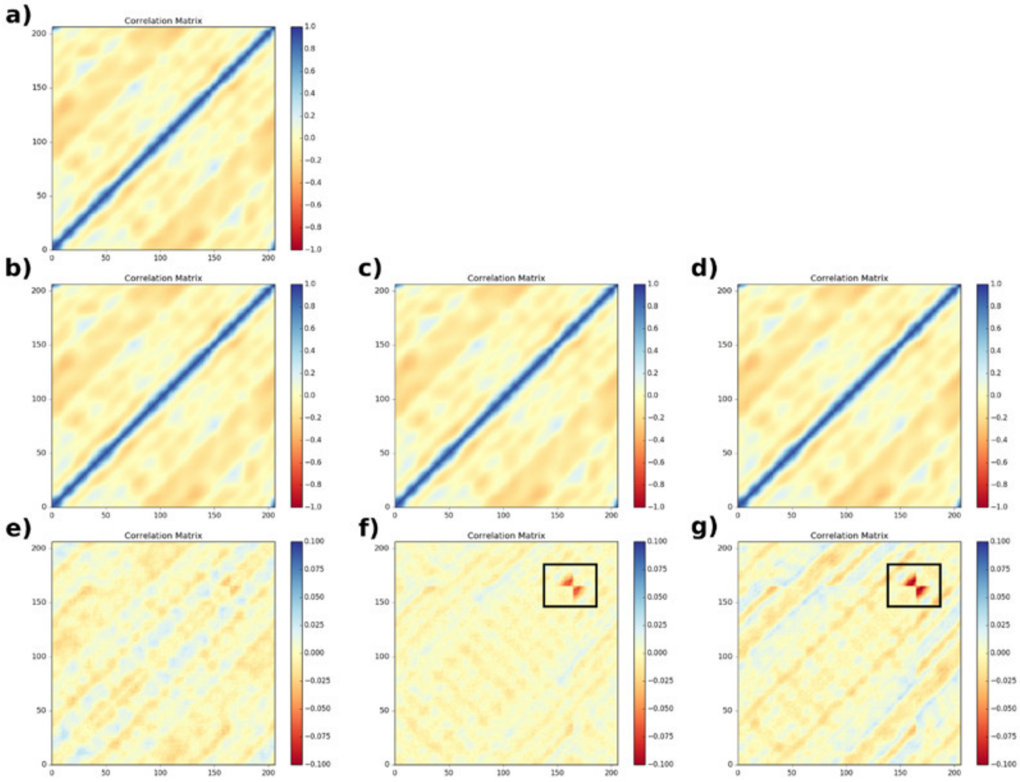

Figure 4.8 The correlation plots for a) WT, b) W178Y, c) W178P and d) W178A. Residue pairs with correlated motions are shown in blue, while anti-correlated motions are shown in red. The correlation difference plots for e) W178Y, f) W178P, and g) W178A compare the mutant correlation plot to that of the WT. Illustrative, regional changes in the single mutants are marked with boxes.

4.3.4 PMF calculation for O_2 pathway to the active site

As described in Subsection 4.3.2, the diffusion MD simulations suggest that the O_2 molecules diffuse to the active site through the blue tunnel exclusively, and remain in the active site for a short time before returning to solution. Therefore, we have calculated the PMFs associated with passive O_2 transport through both tunnels in order to gain further insights for the possible reason for the tunnel preference and causes for the short residence times of O_2 molecules in the active site.

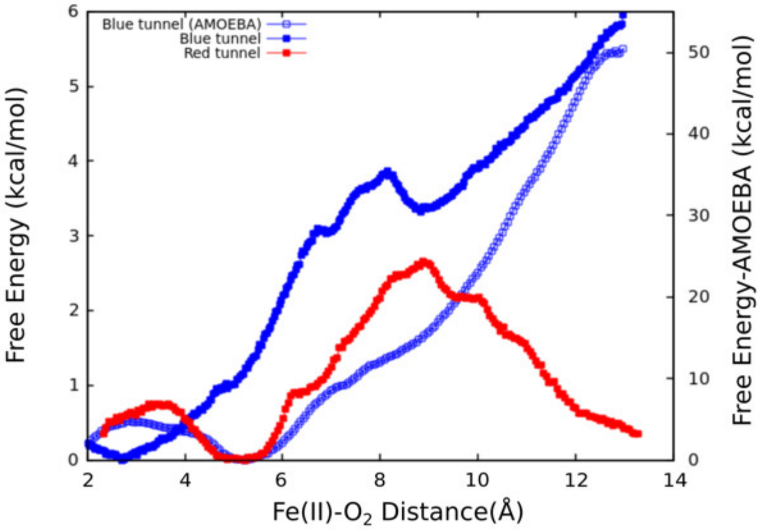

Figure 4.9 Calculated PMFs for the red and blue tunnel using the ff99SB and AMOEBA parameter sets. The y (x) axis in the left (right) shows the free energy calculated with the ff99SB (AMOEBA) parameter set.

The calculated PMFs for the transport of molecular oxygen along the blue and red tunnels using the ff98SB potential are shown in Figure 4.9 (see Figure 4.10 for bootstrap analysis, and Figures 4.11 for histogram analysis). The PMF results suggest that the oxygen molecules do not have to overcome any free energy barriers and the transport into the active site is completely downhill by 6 kcal/mol through the blue tunnel. Conversely,

O_2 that diffuse through the red tunnel need to overcome a barrier of 2.5 kcal/mol to reach the active site.

The barrier observed along the red tunnel likely corresponds to the interaction between the O_2 and various residues along this tunnel. For instance, residues E136, R183, and R210, in the red tunnel form inter-molecular interactions with O_2 molecules for more than 30% of the simulation time. These residues are located around 9-11 Å from the active site and correspond to the region where the O_2 molecules experience the free energy barrier. The interaction between these residues could make the O_2 diffusion more difficult as the ligand molecules need to break these interactions. The calculated barrier in the red tunnel compared to a completely downhill energy path in the blue tunnel provides a possible explanation for the preference of O_2 molecule diffusion through the blue tunnel rather than red tunnel. This prediction is also consistent with the long MD simulations, where O_2 diffusion for wild type AlkB was observed only along the blue tunnel.

Furthermore, the relatively small barrier in the PMF for the blue tunnel for the egress of O_2 from the active site provides a possible explanation as to why oxygen molecules are observed to easily escape from the active site and go back to the solution using the ff99SB potential. However, the short residence time of O_2 molecules in the active site could result in the interruption of stepwise oxidation for this enzyme. This is inconsistent with the available experimental and computational results on the mechanism of AlkB, since it is known that the rate limiting step is the oxidation of the alkyl moiety by the ferryl intermediate.

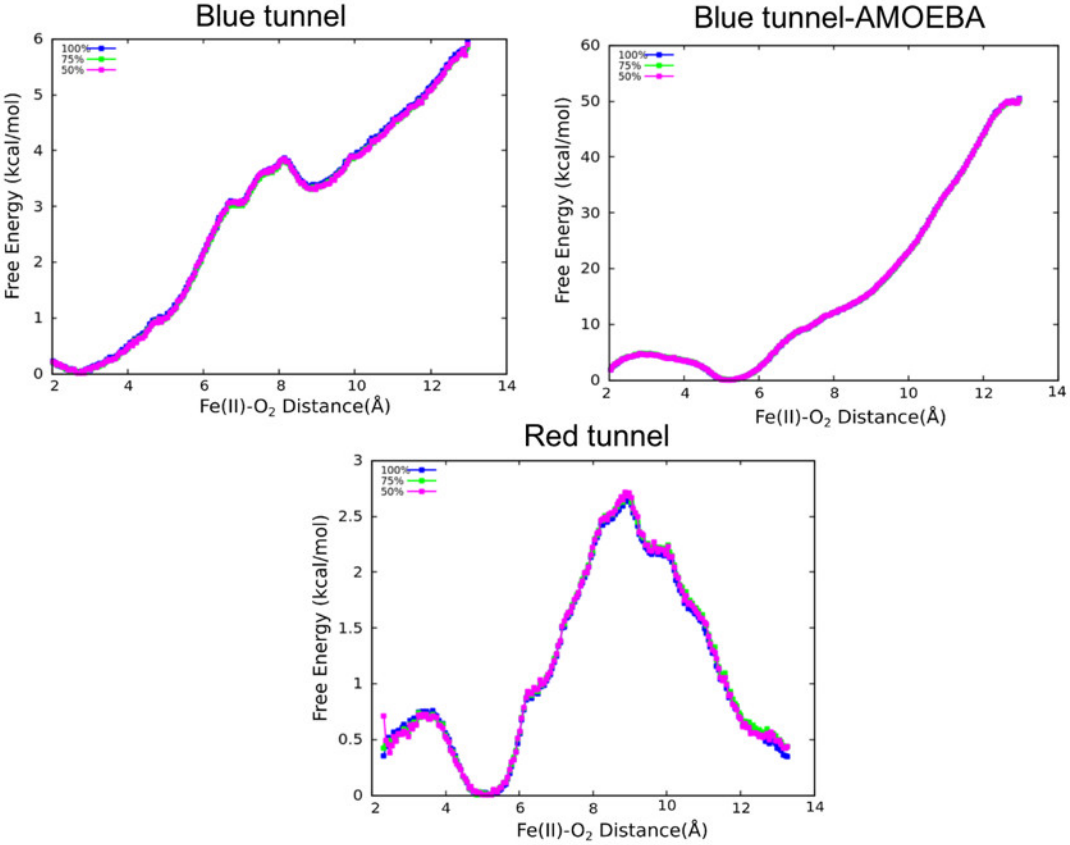

Figure 4.10 Bootstrapping analysis for blue and red tunnel with AMBER and AMOEBA force field.

Therefore, a question arises regarding the accuracy of the PMF based on the non-polarizable potential. In particular, whether this force field can provide a sufficiently accurate description of the inter-molecular interactions given the fact that O_2 is a neutral molecule and therefore most fixed-charge potentials can only represent it by Van der Waals interactions. Moreover, it is known that molecular Oxygen is highly polarizable. Therefore, we performed PMF calculations on the main proposed pathway (blue tunnel) with the multipolar-polarizable AMOEBA potential to examine the effect of polarization on a highly polarizable system.

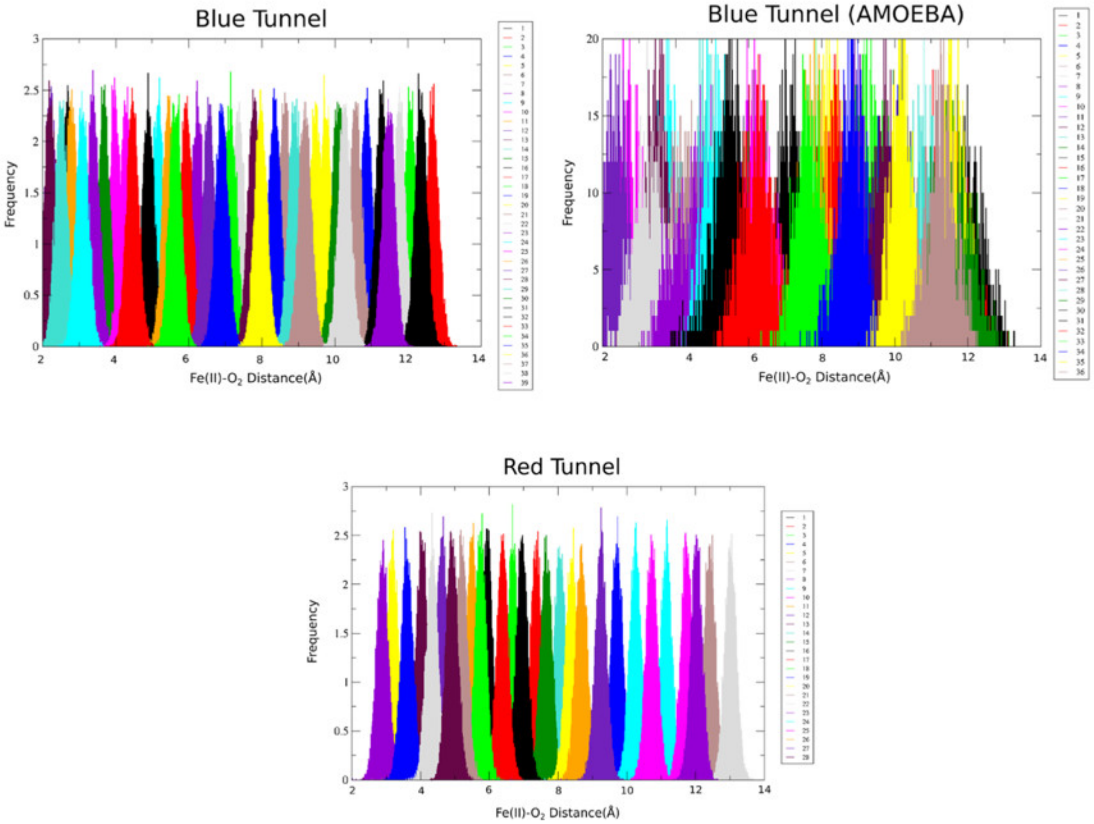

Figure 4.11 Histogram analysis for blue and red tunnel with AMBER and AMOEBA force field.

When explicit polarization is taken into account, the reduction in free energy from the surface of the protein to the active site is calculated to be ≈ 50 kcal/mol (Figure 4.9) (see Figure 4.10 for bootstrap analysis and Figure 4.11 for histogram analysis). Once oxygen gets close to the active site, the molecule needs to overcome a small barrier of around 5 kcal/mol between 2-5 Å from the active site likely due to slightly repulsive interactions with some residues around the Fe cation. In contrast with the results obtained with the ff99SB force field, this very large free energy difference between the surface and the active site indicates that once O_2 molecules diffuse into the active site, it more than likely stays inside the protein and binds to Fe(II) to drive the oxidation forward. This large free energy barrier precludes the egress of molecular oxygen from the active site, and indeed, it provides more time for the O_2 molecule to bind to the Fe(II). Although we are not able to perform ligand diffusion simulations based on long MD using the AMOEBA force field due to computational constraints, our results underscore the importance of polarizability in some biological systems.

4.4 Conclusions

Various computational approaches have been employed to gain insights on the transport of molecular oxygen through intra-molecular tunnels in AlkB. Tunnel analysis based on steric considerations indicates the existence of two possible tunnels in AlkB. Ligand diffusion simulations based on long MD trajectories using a non-polarizable fixed- charge potential (ff99SB) show that only one of these pathways is observed during the simulations on the wild type enzyme. O_2 diffusion in the W178Y/P mutants takes place through the same main tunnel (blue tunnel), while simulations on the W178A mutant reveal a new pathway. The modeling shows that the replacement of W178 with tyrosine has only a modest effect on the O_2 diffusion, while replacement by proline creates a barrier and O_2 can not readily bypass it, therefore, oxygen molecules are trapped in the active site. The calculated PMFs for the wild type are consistent with the diffusion simulations and show that the free energy associated with O_2 transport along the blue tunnel is completely downhill, compared to a barrier of 3 kcal/mol for the red tunnel.

However, the small barrier for O_2 egress from the active site through the blue tunnel is inconsistent with experimental and computational results for the AlkB mechanism.

Complementary PMF simulations with the polarizable AMOEBA potential indicate that the free energy of O_2 transport when electronic polarization is taken into account remains mostly downhill, but the barrier for egress is significantly increased by almost one order of magnitude. Our results provide support for the existence of an intra-molecular O_2 tunnel in AlkB, the role of a key conserved residue (Y178), and the importance of accounting for polarization for O_2 transport simulations.

CHAPTER 5 DEVELOPMENT OF AMOEBA FORCE FIELD FOR 1,3-DIMETHYLIMIDAZOLIUM BASED IONIC LIQUIDS

Portions of the text in this chapter were reprinted or adapted with permission from: *J. Phys. Chem. B., 118, 7156 (2014).* All rights to the work are retained by the authors and any reuse requires permission of the authors.

5.1 Introduction

The development of AMOEBA (a multipolar polarizable force field) for imidazolium based ionic liquids is presented. Our parametrization method follows the AMOEBA procedure and introduces the use of QM intermolecular total interactions as well as QM energy decomposition analysis (EDA) to fit individual interaction energy components. The distributed multipoles for the cation and anions have been derived using both the Gaussian distributed multipole analysis (GDMA) and Gaussian electrostatic model-distributed multipole (GEM-DM) methods.[79] The intermolecular interactions of a 1,3-dimethylimidazolium [dmim$^+$] cation with various anions, including fluoride [F$^-$], chloride [Cl$^-$], nitrate [NO$_3^-$], and tetraflorouborate [BF$_4^-$], were studied using quantum chemistry calculations at the MP2/6-311G(d,p) level of theory. Energy decomposition analysis was performed for each pair using the restricted variational space decomposition approach (RVS) at the HF/6-311G(d,p) level. The new force field was validated by running a series of molecular dynamic (MD) simulations and by analyzing thermodynamic and structural properties of these systems. A number of thermodynamic properties obtained from MD simulations were compared with available experimental data. The ionic liquid structure reproduced using the AMOEBA force field is also compared with the data from neutron diffraction experiment and other MD simulations. Employing GEM-DM force fields resulted in a good agreement on liquid densities (ρ), enthalpies of vaporization (ΔH_{vap}), and diffusion coefficients ($D_{\pm}$) in comparison with conventional force fields.

5.2 Computational methods

In this section, we describe the parameter fitting methodology beginning with the parametrization procedure by means of quantum mechanical (QM) intermolecular interaction and energy decomposition analysis. This is followed by the details of the QM intermolecular interaction calculations employed for the parametrization. Subsection 5.2.3 describes the distributed multipoles calculation. Subsequently, subsections 5.2.4 and 5.2.5 describe the parametrization of intra- and inter-molecular interactions. The optimization procedure of van der Waals parameters is described in subsection 5.2.6.

5.2.1 Parametrization details for AMOEBA using QM EDA data

Our parametrization procedure is modified from the conventional AMOEBA method by introducing the use of QM energy decomposition analysis (EDA) to improve the description of each available individual non-bonded term. We adopt all functional forms of the multipole based AMOEBA force field for the development. The functional form of the AMOEBA potential is given by Eq 5.1.

$$U_{total} = U_{bond} + U_{angle} + U_{torsions} + U_{Coul} + U_{pol} + U_{vdW} \qquad \text{Eq5.1}$$

where the valence functional potentials are described by the potentials due to distortion of bonds (U_{bond}), bends (U_{angle}), torsions ($U_{torsion}$). These functional potentials have been described in detail and are beyond the scope of the current work. In this work, we mainly focus on intermolecular interactions that are described by the last three terms of the potential: U_{Coul}, U_{Pol}, and U_{vdW}.

Permanent multipoles have been fitted for each atomic site (i) based on the QM electron density matrix for each molecule. The fitting procedure is discussed in subsection 5.2.3. The multipole components include point charge (q), dipole (μ), and quadrupole (Q) terms. These components are represented by a polytensor M^T. The potential energy due to the interaction of the permanent multipoles is calculated using Eq 5.2.

$$U_{Coul}(r_{ij}) = M_i^T T_{ij} M_j \qquad \text{Eq 5.2}$$

where T_{ij} is a multipole interaction matrix, M is a polytensor, and r_{ij} is the distance between site i and site j.

The induced polarization is described by placing an inducible atomic polarizable point dipole moment, μ_i, on each interaction site. The induced dipoles are calculated as $\mu_i = \alpha_i E_{ij}$, where α_i is the atomic polarizability and E_i is the external electric field. The polarizability interactions are damped at short range by means of the Tholé scheme[136] to avoid the so-called “polarization catastrophe”.[75] All atomic polarizabilities were adopted unchanged from the AMOEBA force field. A test was performed on the sensitivity of the intermolecular polarization energies due to a variation of the damping parameter value α for the cation-cation and anion-anion pairs. Our results suggest that intermolecular interaction energies are insensitive to the change of the Tholé exponent over the range of α = 0.08-0.45. Therefore, the Tholé damping factor α with a value of 0.39 was used in calculations to be consistent with the AMOEBA force field. In previous studies, the

Tholé parameter was reduced to the value of 0.35 to get a better description of the polarization interactions for the canonical water dimer.[79]

The last term is the potential energy that arises due to van der Waals interactions. These interactions can be described by three functional forms.[221-224] van der Waals interactions are described by the buffered Halgren[224] pairwise potential, as shown by Eq 5.3.

$$U_{vdW}(r_{ij}) = \epsilon_{ij} \left(\frac{1+0.07}{\left(\frac{r_{ij}}{R_{ij}^{0}}\right)+0.07} \right)^{7} \left(\frac{1+0.12}{\left(\frac{r_{ij}}{R_{ij}^{0}}\right)^{7}+0.12} - 2 \right) \qquad \text{Eq 5.3}$$

where ε_{ij} is the potential well, r_{ij} is the separation distance between sites i and j, and R_{ij}^{0} is the minimum energy interaction distance (radius) for sites i and j. This potential provides a better description for interaction of noble gases in comparison with other available potentials.[224] The van der Waals parameters for the unlike atom types are calculated using a combining rule.[82] Validation of the developed force fields is performed by running a series of molecular dynamics simulations on ionic liquids and comparing essential thermodynamic properties with experimental data and other molecular dynamics simulation results. The fitting of van der Waals parameters can be performed if necessary using a well-developed fitting methodology.[140]

5.2.2 Quantum mechanical and monomer AMOEBA calculations

Quantum mechanical calculations were performed for a single [dmim$^+$] cation and $[F^-]$, $[Cl^-]$, $[NO_3^-]$, and $[BF_4^-]$ anions using the Gaussian 09 software package.[225] One-electron, relaxed densities were calculated at the MP2-(full)/6-311G(d,p) level of theory for all ions and ion pairs. Structures of isolated cations and anions were optimized using the same level of theory and basis set. This level of theory was chosen to remain consistent with the AMOEBA force field development methodology.[226] The absence of imaginary frequencies in harmonic vibrational calculations proved that optimized structures correspond to local minima of the energy landscape. There are three [dmim$^+$] conformers to be considered. We find the lowest energy conformer for isolated [dmim$^+$] at the MP2(full)/6-311G(d,p) level by the rotational isomerization of methyl groups. The lowest energy conformer has its methyl hydrogens aligned in the same plane and direction as the hydrogen atom of a middle (C2) carbon on the imidazolium ring (Figure 5.1). Electron-electron correlation effects were also studied at the HF, MP2, and DFT levels of theory.

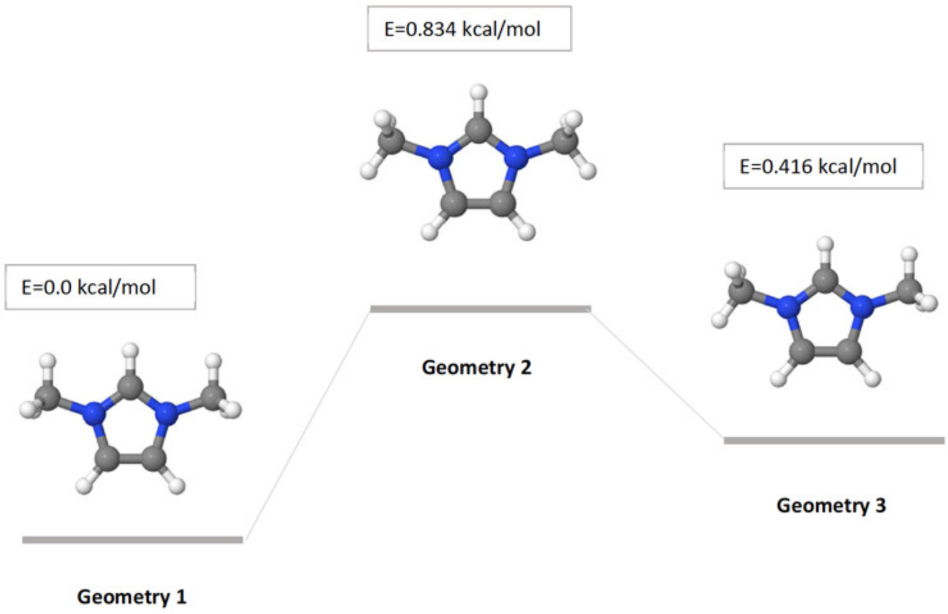

Figure 5.1 [dmim$^+$] relative energies calculated at MP2/6-311G(d,p) level of theory.

We have used two methods for the determination of distributed multipoles. The first involves the use of GEM-DM,[142-143] and the second relies on GDMA.[75, 227] The reason for the use of two different multipole sets is to compare the newly developed GEM-DM multipoles to the GDMA multipoles originally employed in the AMOEBA force field. GEM-DM distributed multipoles were obtained by fitting to one-electron densities from *ab initio* calculations using a single optimized geometry as discussed below. Both analytical and numerical fitting methods are available for the fitting of the QM densities to Hermite Gaussians.[79] The GDMA atomic multipoles were derived using Stone's approach.[75, 227]

Intermolecular interaction energies were calculated next as a function of intermolecular separation distances for all ionic pairs. Total intermolecular energies were calculated using the counterpoise correction to take into account the basis set superposition error (BSSE).[228] Energy decomposition analysis was performed for each pair using the restricted variational space (RVS) decomposition approach[229] at the HF/6-311G(d,p) level of theory as implemented in the GAMESS[230] software package. The Coulomb intermolecular interactions were calculated with an in-house FORTRAN90 program that enables the use of *ab initio* densities for the calculation of the intermolecular Coulomb energies with any one-electron relaxed density matrix for each monomer.

Molecular mechanics (MM) calculations were performed using both AMBER[146] and TINKER[67] simulation packages to ensure a correct reproduction of ion pair optimized geometries, total potential energies, and intermolecular energies from both packages. We have calculated total intermolecular interaction energies and interaction energies by energy components. Excellent agreement between both packages was established in describing intra- and inter-molecular interactions for all ionic pairs.

5.2.3 Distributed multipoles

The fitting methodology to obtain the distributed multipoles from GEM has been previously described.[79, 89] In short, the density fitting method is based on employing auxiliary Gaussian basis functions (ABS) to fit the one-electron relaxed density. There are few available auxiliary basis sets such as A1, A2, and A2DG that can be used in the fitting.[231] We use the A2DG auxiliary basis set to obtain Hermite coefficients for the [dmim$^+$] cation and the A2 auxiliary basis set to obtain Hermite coefficients for anions. A direct mapping has been shown between the elements of the multipole tensor and Hermite Gaussian coefficients hc_{tuv} which enables the direct calculation of distributed Cartesian point multipoles from the fitted Hermite coefficients.[232] In this work, the Hermite coefficient Λ_{tuv} has a highest angular moment of 2 ($t+u+v=2$), which results in distributed multipoles with a highest angular moment of quadrupoles. In GEM, the Hermite spherical Gaussians are normalized to 1, which guarantees that Cartesian multipoles can be calculated as defined by Eq 5.4.

$$hc_{tuv} = hc_{tuv} \int \Lambda_{tuv} \, dr \qquad \text{Eq 5.4}$$

where $\Lambda_{tuv} = (\alpha/\pi)^{3/2} (\partial/\partial x)^t (\partial/\partial y)^u (\partial/\partial z)^v e^{(-\alpha r^2)}$ is the Hermite Gaussian function. Fitting of GDMA multipoles was fitting of GDMA multipoles was accomplished for cation and polyatomic anions according to the well-established GDMA fitting procedure.[75, 227] It should be mentioned that the fitting of the GEM-DM multipoles relies not only on the reproduction of molecular electronic densities but also on the reproduction of intermolecular energies arising from electrostatic interactions of ionic pairs.

In order to enable the rotation of the multipoles, we have to define local coordinate frames for each atom site (i). The local frames are defined following the formalism employed in TINKER.[67] The frames for [dmim$^+$] cation are kept identical for both the GDMA and GEM-DM models, but the definitions of frames for anions vary for the GDMA and GEM-DM models.

Four sets of [dmim$^+$] multipoles have been fit to compare the overall accuracy of the force fields with and without intramolecular polarization effects. All four sets of multipoles are fit to a one-electron density matrix, as discussed in computational section 5.2.1. Intramolecular polarization was not taken into account for the first set while fitting. 1,3- Dimethylimidazolium was considered as a single polarizable group. These multipoles are referred to as GDMA (1G) and GEM-DM (1G) multipoles. The second set of multipoles is fit taking into account intramolecular polarization effects by defining two polarizable segments of [dmim$^+$] following the procedure described elsewhere.[205] The polarization segmentation allows us to fit permanent atomic multipoles that would include an intramolecular polarization contribution from the [dmim$^+$] conformational dependence. The first polarizable group (segment) is defined by the imidazolium ring and a second group is defined by the methyl groups. These groups account for mutual polarization by fitting permanent multipoles upon rotation of methyl groups. The fitted multipoles will be further referred to as GDMA (2G) and GEM-DM (2G) multipoles.

However, no mutual polarization has been applied for polyatomic anions. All anions are defined as a single polarizable group. The fitted multipoles vary only by the fitting methodology used and referred to GDMA and GEM-DM multipoles.

5.2.4 Intramolecular interactions

All intramolecular parameters (bond, bend, torsions, and out-of-plane deformations) were initially taken from the original AMOEBA[78, 205, 216] force field including parameters for 1,3-dimethylimidazolium and anions. Atomic polarizabilities α and van der Waals (ε_{ij} and R^0_{min}) parameters were also taken from the AMOEBA force field. Specifically, intramolecular equilibrium parameters for bonds, bends, torsions, and out-of-plane deformations of 1,3-dimethylimidazolium were adopted from the 4-ethylimidazole compound. However, the AMOEBA force field was augmented with (C2−N−Cm−Hm) torsional parameters for the [dmim$^+$] cation. A potential energy surface was generated using a 10° scan step fit to the MP2/6-311G(d,p) level of theory to get the best description of the conformational energy for this model compound. Torsional profiles are given in Figure 5.2 a.

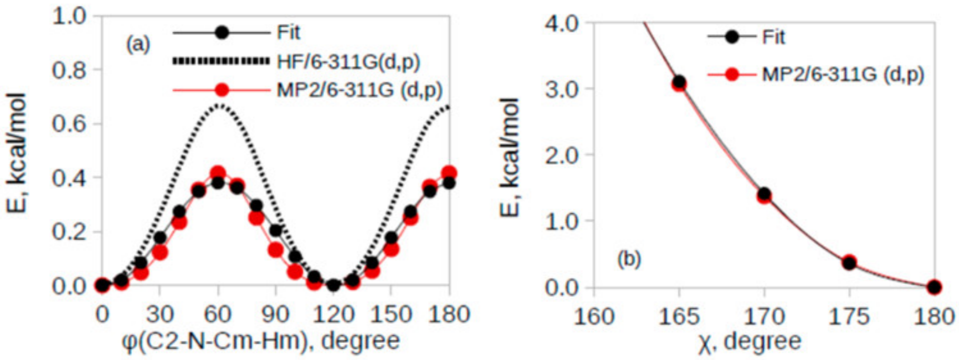

Figure 5.2 Conformational energy change for the (a) C2-N-C-H torsion, (b) out-of-plane deformation for NO3- anion.

Parameters for the (C2−N−Cm−Hm) torsion have been previously fit to the MP2/6-31+G(d) level by Liu *et al*. for the AMBER force field.[233] The torsional barrier is about 0.4 kcal/mol higher than the one obtained in our current calculations due to the smaller basis set used. However, we obtain a smaller rotational barrier using the HF/6-311G(d,p) level than the one reported by Liu *et al*. Nevertheless, we were unable to get a correct dimer optimized geometry in the gas phase with a torsional energy barrier of 0.9 kcal/mol. In contrast, Cadena and Maginn used C2−N−Cm−Hm with a torsional barrier of 0.195 kcal/mol.[54] This barrier is too small in comparison with *ab initio* calculations.

We adopted equilibrium constants (k_0), bond lengths (r_0), and angles (θ_0) for [NO_3^-] from the parameters reported by Cadena and Maginn.[234] Van der Waals parameters were taken from amine nitrogen and amide oxygen as developed for the AMOEBA force field.[216] The optimized geometry of the [NO_3^-] anion is planar, which corresponds to the lowest energy state.

Therefore, we fit the out-of-plane deformation constant (K_{out}) to ensure that the [NO_3^-] molecular model stays in the planar configuration; see Figure 5.2b. A good description of out-of-plane energies ($U_{tot} = K_{out}\chi^2$) was obtained as compared with MP2/6-311G(d,p) calculations. Normal mode vibrational frequency analysis was performed next. The average unsigned difference was 119 cm^{-1}. We were not able to get a better agreement with QM calculations.

Intramolecular parameters for [BF_4^-] were taken from the AMOEBA[78, 205, 216] force field. Normal mode vibrational analysis was also performed for this model compound. Good agreement in vibrational frequencies with available experimental calculations was attained using AMOEBA parameters for the B-F bond and the F-B-F angle. The difference in asymmetric stretch was 192 cm^{-1}. Gas phase normal mode vibrations for [NO_3^-] and [BF_4^-] were carried out using the VALENCE tool available in TINKER.

5.2.5 Intermolecular interactions

In order to calculate intermolecular interaction energies for ionic pairs, we use the same methodology implemented for non-ionic compounds.[79, 235] We systematically change the distance between cation center of mass and anion center of mass as shown in Figure 5.3 along the vector that is perpendicular to the plane of the [$dmim^+$] ring. In the case of [NO_3^-] and [BF_4^-], we use the same pathway as that for [F^-] and [Cl^-] varying the distance between the middle nitrogen and boron as shown. The total interaction energy in the gas phase was calculated at the MP2(full)/6-311G(d,p) level of theory with a step size of 0.2 Å using the optimized monomer geometries.

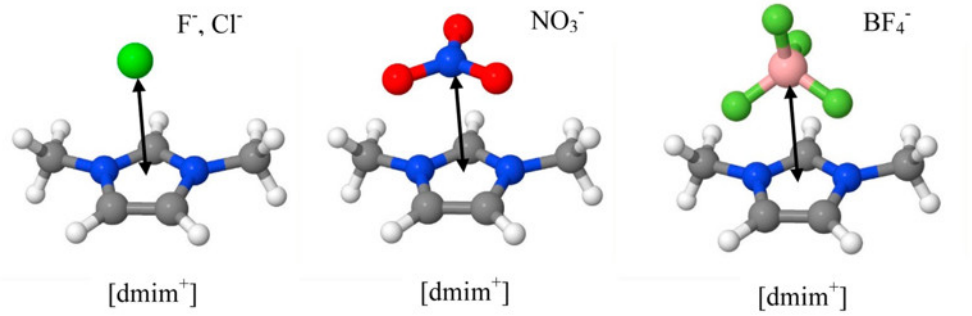

Figure 5.3 Schematic representations of ionic dimers, [$dmim^+$][F^-] and [$dmim^+$][Cl^-] (left), [$dmim^+$][NO_3^-] (middle), and [$dmim^+$][BF_4^-] (right). Arrows indicate the directions along which the anions were moved.

The total intermolecular interaction energies were decomposed using the RVS decomposition analysis.[229] Energy decomposition analysis can be performed only at the HF level of theory. Therefore, we plotted intermolecular potentials from calculations using both MP2 and HF levels for comparison. The resulting intermolecular potential along with the corresponding energy components are given in Figure 5.4 for the [dmim$^+$][Cl$^-$] pair. Intermolecular potentials for other compounds are given in Figure 5.5-5.11. A good agreement on total intermolecular interactions is established using both the MP2 and HF levels. Intermolecular energies are systematically underestimated (~1 kcal/mol) using the HF level for all pairs but [dmim+][F−] where the minimum energy is overestimated by ~7 kcal/mol. We found the minimum interaction distance is at 3 Å for the [dmim$^+$][Cl$^-$] pair with a corresponding intermolecular energy of -82.81 kcal/mol. The intermolecular interaction distance is at 2.4 Å for the [dmim$^+$][F$^-$] pair with a minimum energy of -94.97 kcal/mol, which is more favorable than that for [dmim$^+$][Cl$^-$]. An interaction distance and a minimum energy for [dmim$^+$][NO3$^-$] are comparable with those for [dmim$^+$][Cl$^-$]. The minimum interaction distance is 3.3 Å for [dmim$^+$][BF4$^-$] with an energy of −79.95 kcal/mol, which is less favorable than energies for other compounds.

These calculations indicate that intermolecular interactions become more favorable as the ion size decreases, and the interaction distances become shorter as the size of the anions becomes smaller as expected. The order for the intermolecular interaction energies can be placed as $E_{F^-}^{int} < E_{Cl^-}^{int} \approx E_{NO_3^-}^{int} < E_{BF_4^-}^{int}$ .On the basis of these results, we expect to obtain a similar tendency on enthalpies of vaporization ΔH_{vap} for these model compounds. The highest energy of vaporization is expected for the [dmim$^+$][F$^-$] pair, and the lowest energy is expected for the [dmim$^+$][BF$_4^-$] pair.

Intermolecular energies are systematically overestimated employing both sets of force fields; see Figure 5.4 a. The intermolecular energy minimum is overestimated by ≈ 9 kcal/ mol using one group multipoles (1G) and by ≈ 6 kcal/mol using two group multipoles (2G) for the [dmim$^+$][Cl$^-$] ionic pair. Better agreement in the minimum energies is obtained for all other ionic compounds. Next, we perform energy decomposition analysis and plot intermolecular energies due to permanent multipoles U_{Coul}, induced polarization U_{Pol}, and van der Waals interactions U_{vdW} using both sets of force fields and compare those energies with the ones from the reference *ab initio* calculations.

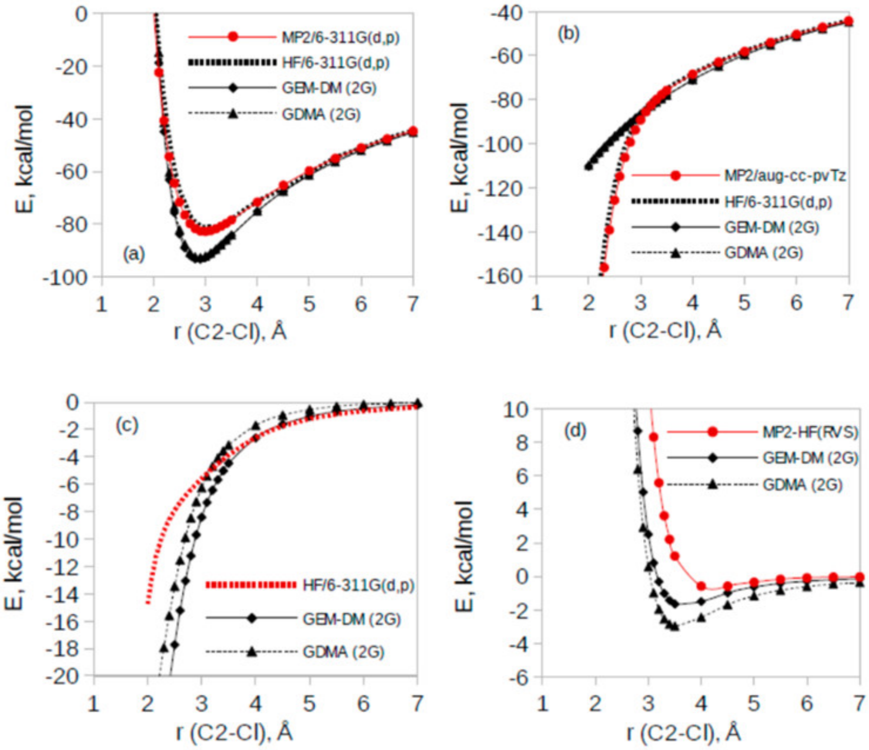

Figure 5.4 (a) Total binding energy, (b) Coulomb energy, (c) polarization energy, (d) van der Waals energy for [dmim$^+$][Cl$^-$] employing 2 polarizable groups.

Intermolecular Coulomb interactions are shown in Figure 5.4b. Intermolecular electrostatic interactions are well described by both sets of permanent multipoles at medium and long range. The multipoles with intramolecular induced polarization (2G sets) slightly overestimate the electrostatic interactions. A better description of intermolecular electrostatic interactions is attained using one group based multipoles as compared to the reference energies. We can see a greater deviation of electrostatic energies at short intermolecular distances.

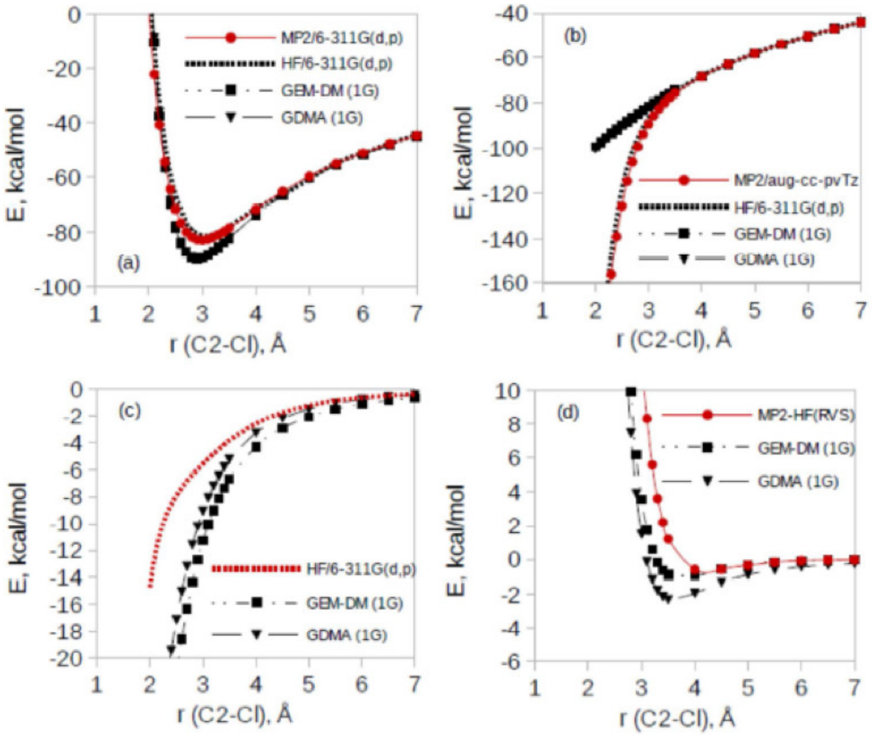

Figure 5.5 (a) Total binding energy, (b) Coulomb energy, (c) polarization energy, (d) van der Waals energy for [dmim$^+$][Cl$^-$] employing 1 polarizable group.

These energy deviations arise since the current version of these force fields does not account for the charge penetration effects. These effects can be included by means of damping functions.[86-87] The largest deviation between the AMOEBA force fields and the *ab initio* reference comes from the energies due to the polarization interactions U_{Pol}; see Figure 5.4c. GDMA (2G) tends to underestimate the polarization energies, while the forced fields based on only one polarization group tend to overestimate intermolecular energies for all ionic pairs as compared to the energies from HF/6-311G(d,p) calculations.

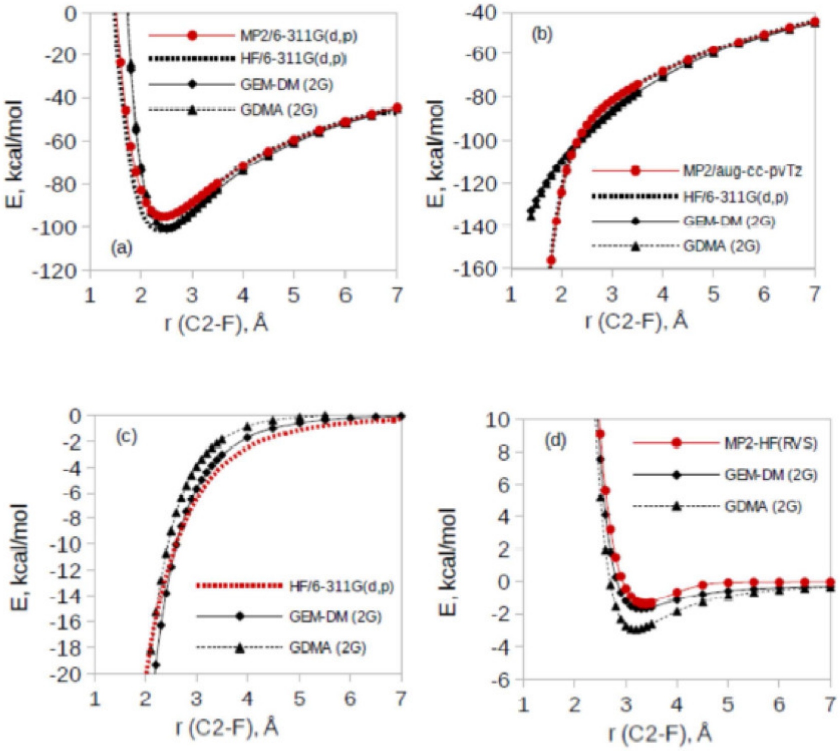

Figure 5.6 (a) Total intermolecular energies (b) Coulomb energies (c) polarization energies and (d) van der Waals energies for [dmim$^+$][F$^-$] employing 2 polarizable groups.

The maximum energy difference between GDMA and GEMDM based force fields is ≈ 1.2 kcal/mol. The potential energy $U_{Coul(MP2)}$ due to electrostatic interactions was calculated using an in-house FORTRAN90 program. The polarization interaction, $U_{Pol(HF)}$, was calculated using the RVS method. Intermolecular energies due to van der Waals interactions are given in Figure 5.4d. van der Waals parameters were taken directly from the AMOEBA force field. Reference van der Waals energies were calculated as the difference between total intermolecular energies at the MP2 level and the Coulomb energies due to permanent multipoles, and the energies due to polarization interactions as given by Eq 5.5.

$$U_{vdW} = U_{Total(MP2)} - U_{Coul(MP2)} - U_{Pol(HF)} \qquad \text{Eq 5.5}$$

A comparison of van der Waals energies using GDMA and GEM-DM reveals that both potentials are more attractive as compared to *ab initio* calculations. The van der Waals interactions are shifted to shorter interatomic distances for all ionic pairs except for the [dmim$^+$][F$^-$] pair; see Figure 5.4 to 5.11. A better agreement of van der Waals interactions is reached for [dmim$^+$][F$^-$] using the GEM-DM (2G) force field. However, the GDMA based force fields overestimate the van der Waals interactions at short and long interaction distances while the GEM-DM based force fields show good agreement at long range.

As noted above, the total intermolecular interactions calculated with both GEM-DM and GDMA based force fields show good agreement with the QM reference. This is due to a cancelation of errors between the overestimation of the van der Waals and polarization terms and the underestimation of the Coulomb term at short range.

One way to test the accuracy of the intermolecular interactions is by calculating the optimized geometries for the gas phase dimers. We performed geometry optimizations for all ionic pairs in the gas phase and compared geometries and corresponding energies with the ones obtained from *ab initio* calculations at the MP2/6-311G(d,p) level of theory. Almost all optimized geometries are well reproduced by both sets of force fields with the exception of the [dmim$^+$][F$^-$] pair, which optimizes to a different geometry as compared with *ab initio* results.

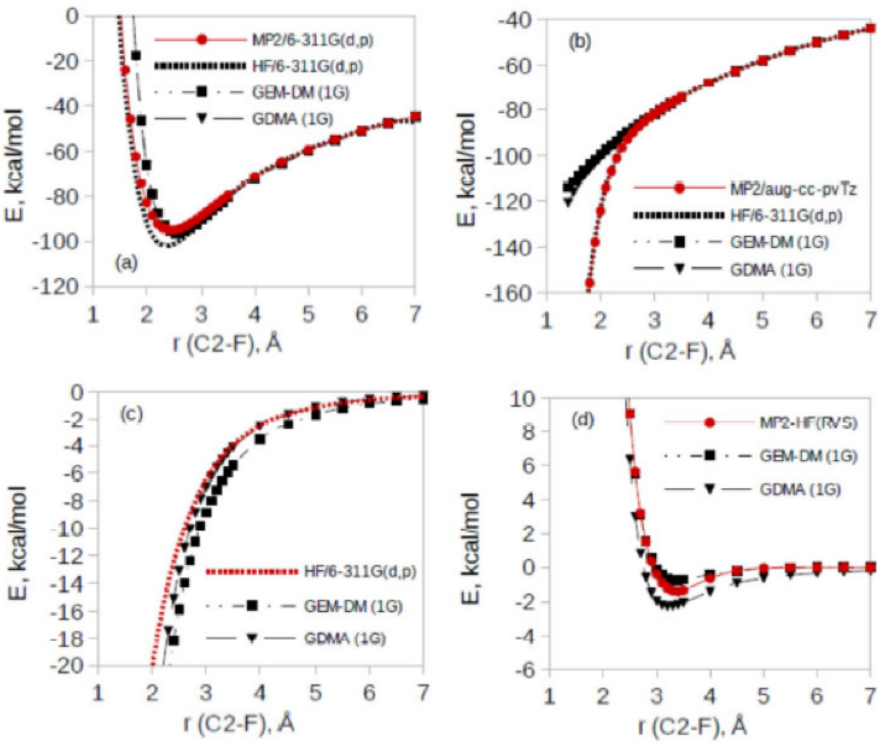

Figure 5.7 (a) Total intermolecular energies (b) Coulomb energies (c) polarization energies and (d) van der Waals energies for [dmim$^+$][F$^-$] employing 1 polarizable group.

In general, the GDMA based force fields have a tendency to underestimate the intermolecular interactions of the gas phase dimers, while the GEM-DM based force fields show the opposite trend. These optimized geometries also show that the calculated parameters reproduce not only the potential energy surface (PES) for which they were parametrized but also those for other structures along the PES.

5.2.6 Van der Waals parameter fitting

We have previously developed a method for the efficient optimization of van der Waals parameters.[140] This method allows the combination of QM and experimental data for the optimization of van der Waals parameters by means of an active-space optimization approach with quadratic convergence. In the case of the ion pairs used in the current study, the only experimental data available is liquid density for [dmim$^+$][Cl$^-$].

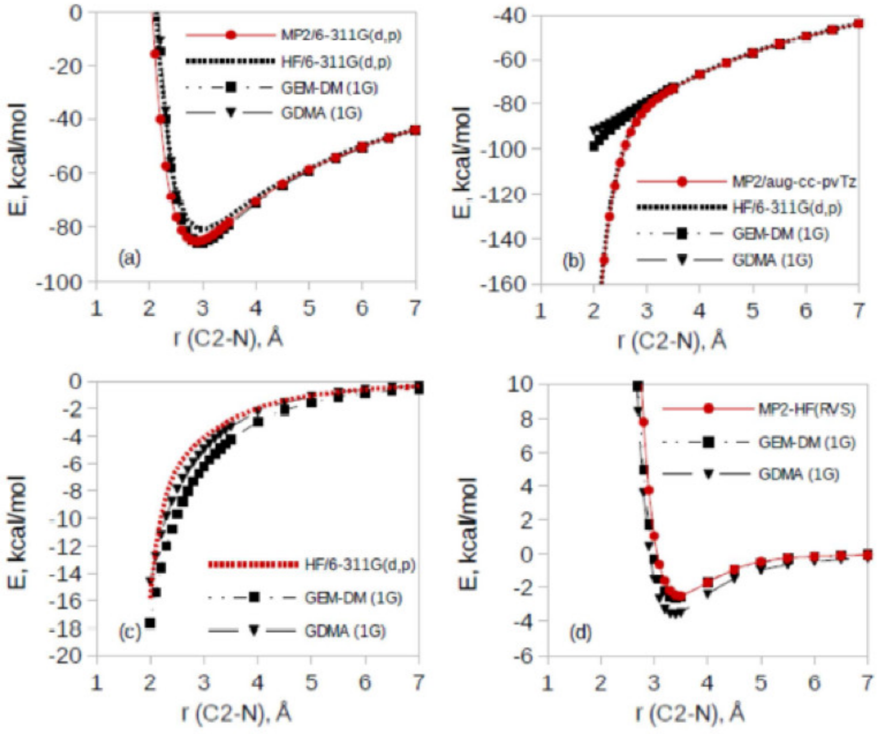

Figure 5.8 (a) Total binding energy, (b) Coulomb energy, (c) polarization energy, (d) van der Waals energy for [dmim$^+$][NO$_3^-$] employing 1 polarizable group.

Therefore, in this particular instance, we were unable to employ the active space optimization approach due to the lack of experimental data in the literature. Initially, we have used an original set of van der Waals parameters for [dmim$^+$] and [Cl$^-$] and these parameters were optimized to match the available experimental data.

After scaling the interatomic radii, the liquid density at 425 K shows a marked improvement with respect to the experimental result (1.127 g/cm^3). Liquids become denser at all other simulated temperatures. As can be seen in Figure 5.12, the energies due to van der Waals interactions are shifted to the shorter intermolecular distances, making the force fields even more attractive. This shift to lower energies may be due to the fact that the parameters are fitted to an intermediate level of theory (MP2/6-311G(d,p)). Recent force fields for different molecules fitted only to QM data at a higher level of theory have shown very good agreement with the experiment.[132, 236-237] In a future contribution, we plan to employ a higher level of theory and more orientations for the ionic dimers to fit a larger set of ionic pairs.

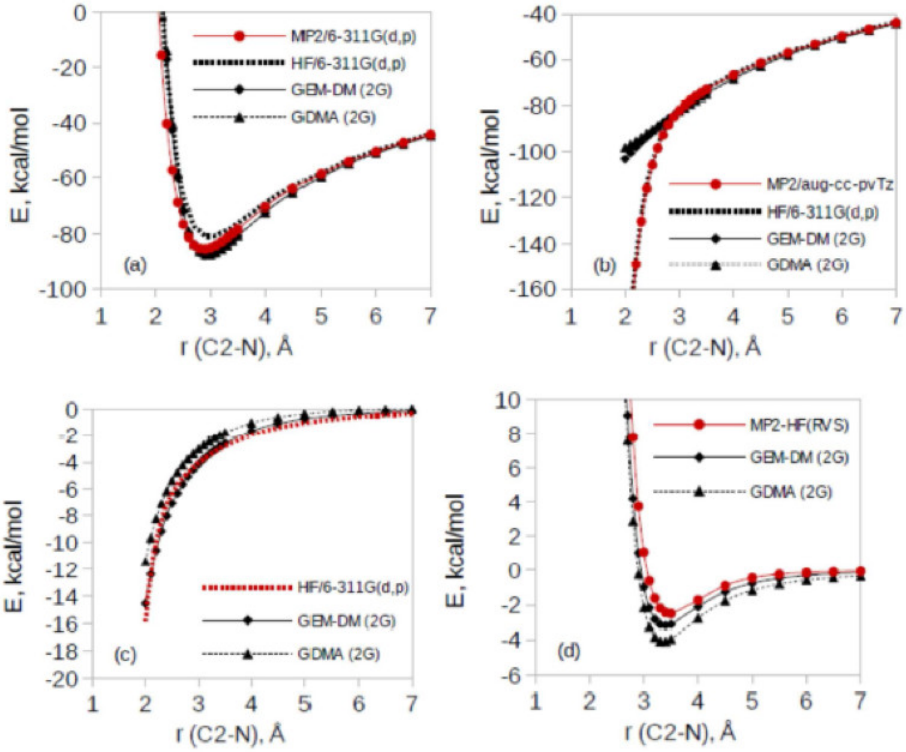

Figure 5.9 (a) Total binding energy, (b) Coulomb energy, (c) polarization energy, (d) van der Waals energy for [dmim$^+$][NO$_3^-$] employing 2 polarizable groups.

5.2.7 Molecular dynamics simulation methodology

Molecular dynamics (MD) simulations were performed for each ionic liquid pair using the AMBER[78, 238] simulation package. The AMOEBA force field was used to carry out MD simulations using a cubic simulation cell with applied periodic boundary conditions. The ionic liquid systems were set up on a periodic SC lattice that included 216 ionic pairs (3672-4536 atoms depending upon the system).

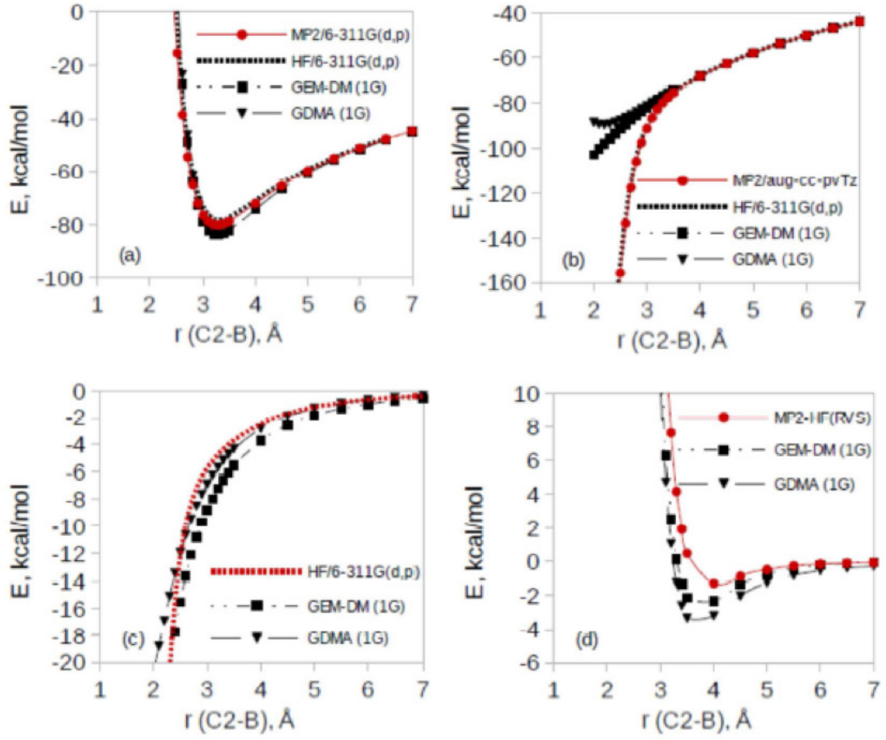

Figure 5.10 (a) Total binding energy, (b) Coulomb energy, (c) polarization energy, (d) van der Waals energy for [dmim$^+$][BF$_4^-$] employing 1 polarizable group.

Energy minimization was performed using the conjugate gradient minimization algorithm to reduce energetic strains. Simulations were carried out in the NVT ensemble to heat up the system to 600 K. The cutoff radius was 8.5 Å for non-bonded and electrostatic interactions. Isobaric-isothermal simulations were performed next until steady state conditions were reached with an integration time step of 1 fs. The Beeman integration algorithm[239-240] was employed for integrating equations of motion. Long-range electrostatic interactions were computed employing the smooth particle mesh

Ewald method[149, 241-242] with an 8 Å direct cutoff. The Berendsen thermostat and barostat[189] were used to control temperature and pressure with a relaxation time of 2.0 ps. Sampling trajectories were generated for 8 ns.

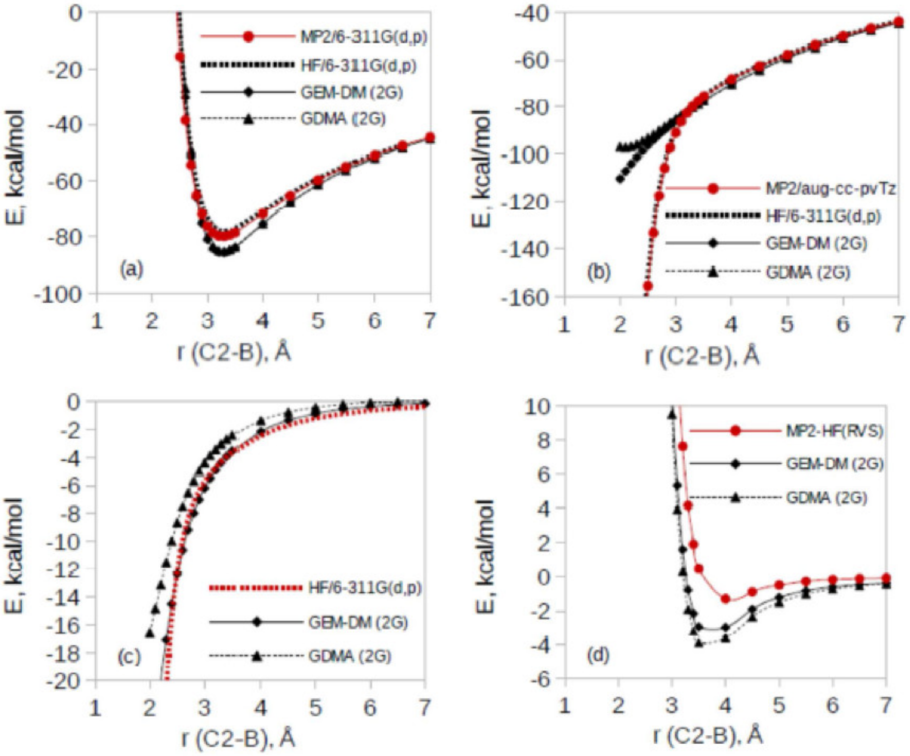

Figure 5.11 (a) Total binding energy, (b) Coulomb energy, (c) polarization energy, (d) van der Waals energy for [dmim$^+$][BF$_4^-$] employing 2 polarizable groups.

5.3 Simulation results

Molecular dynamics simulations were performed to demonstrate the accuracy of the newly developed AMOEBA force fields. Specifically, after implementing GDMA and GEM-DM multipoles and associated parameters, we obtain a good agreement (1% deviation) with the experimental result on liquid densities, ρ, after optimization of the van der Waals parameters for [dmim$^+$][Cl$^-$]. We also calculated enthalpies of vaporization, ΔH_{vap}, and diffusion coefficients, $D_{\pm}$, as a function of temperature and compared the results with available MD simulation data from the literature (See Table 5.1-5.3, and Figure 5.13-5.14).

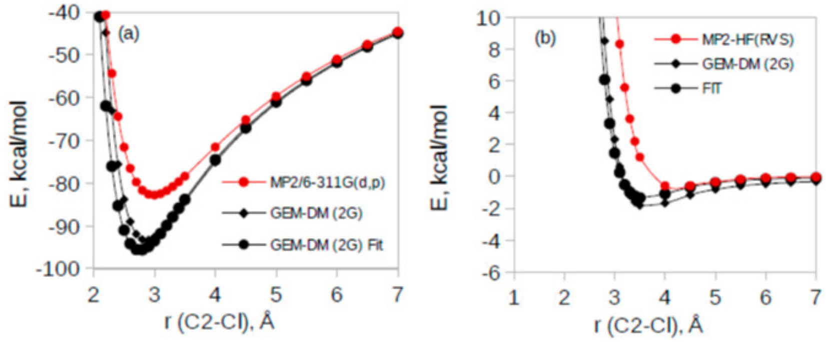

Figure 5.12 Total intermolecular energies (a) and van der Waals energies (b) before and after the fitting of van der Waals parameters for [dmim$^+$][Cl$^-$] employing the GEM-DM (2G) force field.

5.3.1 Ionic liquid volumes and density

Volumes of ionic pairs are estimated at T = 425 K and are based on an average density from MD simulations. The volumes of [dmim$^+$][F$^-$] and [dmim$^+$][Cl$^-$] ion pairs were predicted to be 177.3 and 203.7 Å^3, respectively. Larger volumes were predicted for [dmim$^+$][NO$_3^-$] and [dmim$^+$][BF$_4^-$] as 223.5 and 241.5 Å^3, respectively; see Table 5.1. These results are well correlated with the anion radii where [F$^-$] is the smallest and tetrafluoroborate [BF$_4^-$] is the largest among four studied anions. The volume of [BF$_4^-$] is 18 Å^3 larger than the volume of [NO$_3^-$] and 37.8 Å^3 larger than that of [Cl$^-$].

Comparison of the MD simulation results with the experiment for liquid densities of [dmim$^+$][Cl$^-$] is given in Figure 5.13. The only available experimental data for the ionic liquid studied herein is for the liquid density of [dmim$^+$][Cl$^-$] as mentioned above. Using the original van der Waals parameters, liquid densities of [dmim$^+$][Cl$^-$] are 4% less than the experimental value at 425 K (see Figure 5.13a). However, after the van der Waals parameters are scaled, the GEM-DM (2G) force field shows very good agreement with the experiment with the error below 1%, as shown in Figure 5.13b and Table 5.1.

Table 5.1 Liquid volume and densities at T = 425 K Liquid densities. The density value in parentheses corresponds to data after scaling of van der Waals parameters.

	V (Å^3)		ρ(g/cm^3)		
ILs	GDMA(2G)	GEM-DM(2G)	GDMA(2G)	GEM-DM(2G)	exp
[dmim$^+$][F$^-$]	174.7	177.3	1.105	1.087	
[dmim$^+$][Cl$^-$]	203.8	203.7	1.082	1.118 (1.127)	1.123
[dmim$^+$][NO$_3^-$]	224.0	223.5	1.179	1.181	
[dmim$^+$][BF$_4^-$]	-	241.6	1.370	1.265	

Simulation results for other ionic liquids are in good agreement with the results from other MD simulations [243] for ionic liquids with the greater molecular volume cations. For instance, a liquid density of 1.198 g/cm^3 was predicted from simulations for [dmim$^+$][BF$_4^-$], while a density of 1.206 g/cm^3 was predicted from experiment for [emim$^+$][BF$_4^-$] at 393 K.

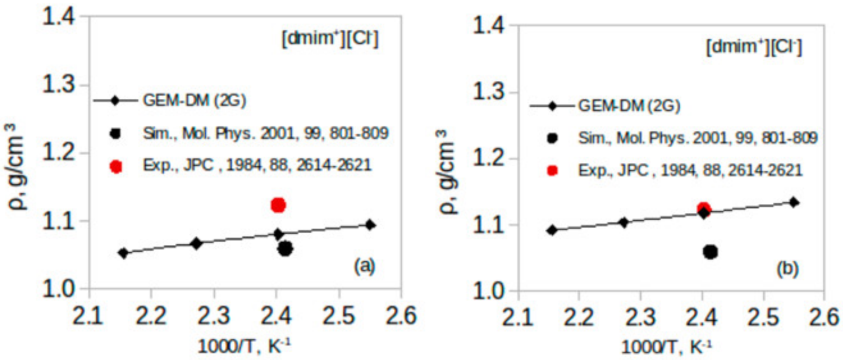

Figure 5.13 Liquid densities for [dmim$^+$][Cl$^-$] employing GDMA and GEM-DM force fields (a) before and (b) after the fitting of van der Waals interactions.

We predict the liquid density for [dmim$^+$][BF$_4^-$] at 400 K to be 1.284 g/cm^3. The resulting difference of 6% can be ascribed to the larger volume of the [emim$^+$] cation.

The liquid density of [bmim$^+$][NO$_3^-$] was reported to be 1.092 kcal/mol at 393 K.[74] MD simulations showed the liquid density to be 1.198 kcal/mol for [dmim$^+$][NO$_3^-$] at 400 K consistent with the change in molecular volume of the cation as above. Some variance in liquid densities is observed between our various multipolar force fields based on the different multipoles or number of polarization groups (see Table 5.1). The maximum deviation can be as large as ≈ 10%. The density calculated in [dmim$^+$][BF$_4^-$] with GDMA and GEM-DM shows a significant difference. The Figure 5.19 shows two snapshots from MD simulation for this system using GDMA and GEM-DM that can explain the difference in densities.

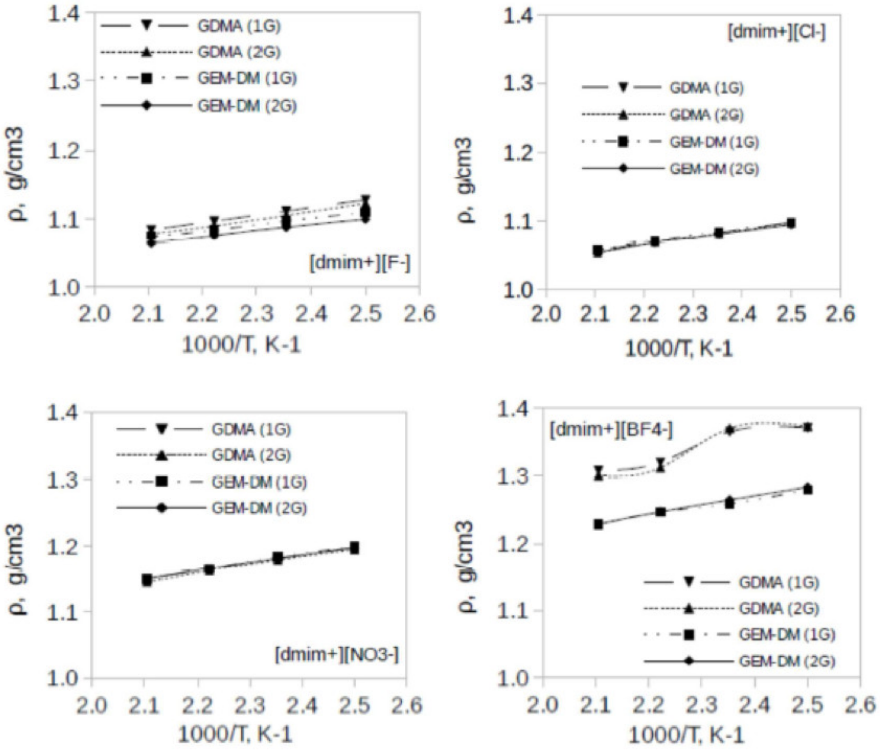

Figure 5.14 MD simulation results for liquid densities using GDMA and GEM-DM force fields at various temperatures.

5.3.2 Intermolecular structure

Liquid structures of ionic liquids are usually compared to X-ray[98] scattering or neutron[244] diffraction experimental data. In particular, interionic correlations are well described by radial distribution functions and can be compared with the structural factor S(Q) obtained from neutron diffraction.[244] Radial distribution functions (RDFs) were calculated for the four ionic liquids at T = 475 K. A summary of these structures for [dmim$^+$][Cl$^-$] is shown in Figure 5.15. These figures include cation-cation (C2-C2), cation- anion (C2-Cl$^-$), and anion-anion (Cl$^-$-Cl$^-$) interatomic correlations.

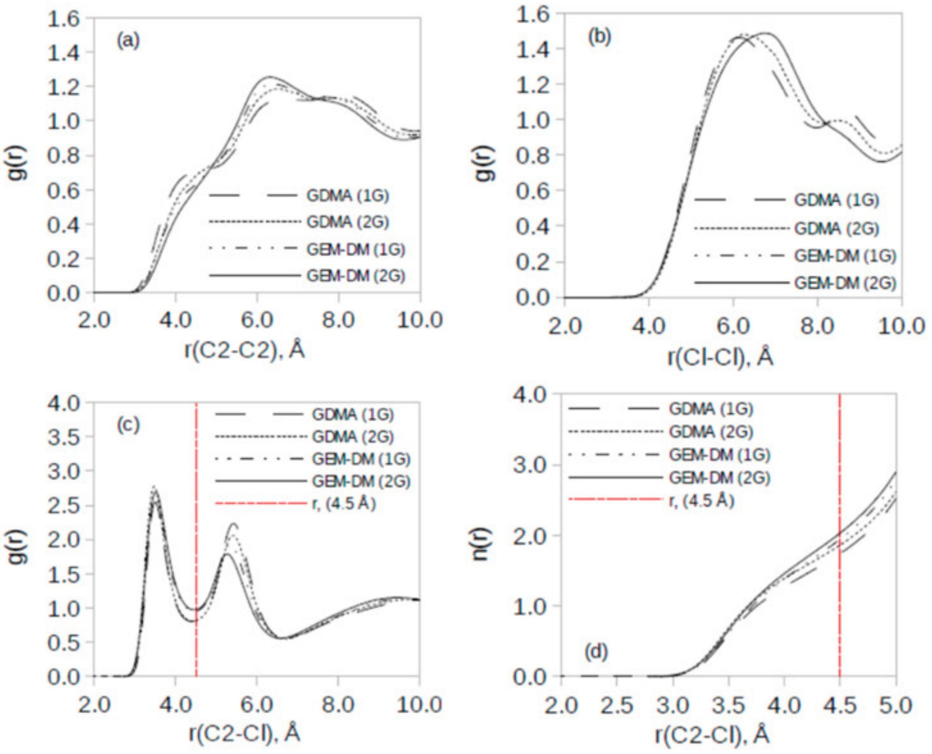

Figure 5.15 Radial distribution function for (a) cation-cation (b), anion-anion (c), and cation-anion for [dmim$^+$][Cl$^-$] employing GDMA and GEM-DM multipole force fields. (d) The coordination number n(r) is given for the cation-anions coordination shell.

Qualitatively similar results were obtained for all ionic liquids studied (Figure 5.16-5.18). A typical correlation is observed for the ionic liquid structures as compared to the RDFs from other molecular dynamics simulations.[54-55, 60] The first peak of the coordination shell is shifted to longer distances as the ionic pairs change from [dmim$^+$][F$^-$] to [dmim$^+$][BF$_4$$^-$]. These results are in good agreement with the results from gas phase optimized geometries. As the size of the anions increases, the intermolecular equilibrium distance increases for the dimer optimized geometries.

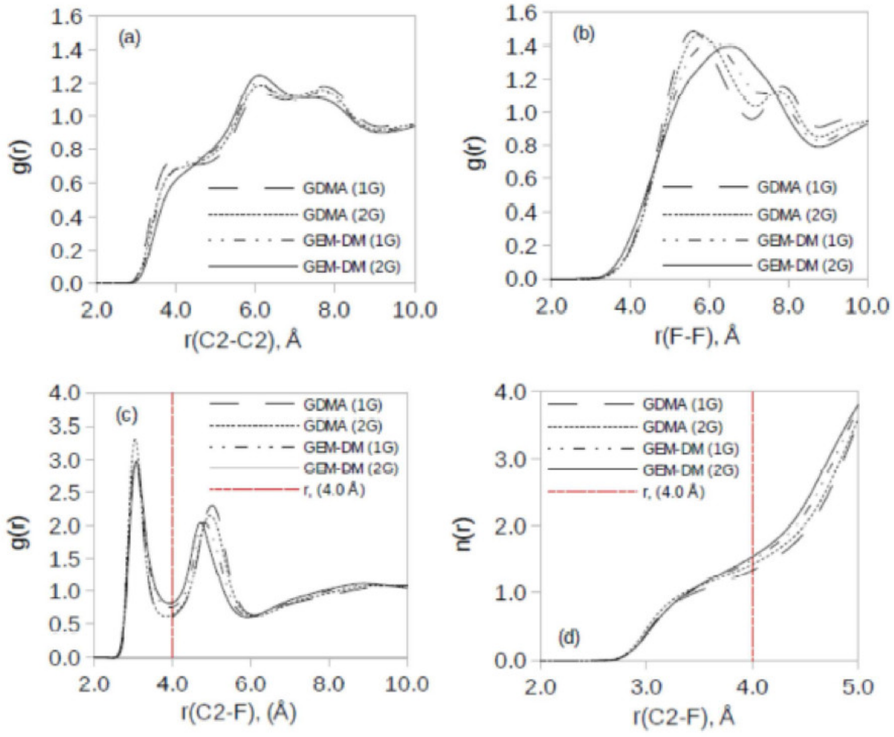

Figure 5.16 Radial distribution function for (a) cation-cation (b), anion-anion (c), and cation-anion for [dmim$^+$][F$^-$] employing GDMA and GEM-DM multipole force fields. (d) The coordination number n(r) is given for the cation-anions coordination shell.

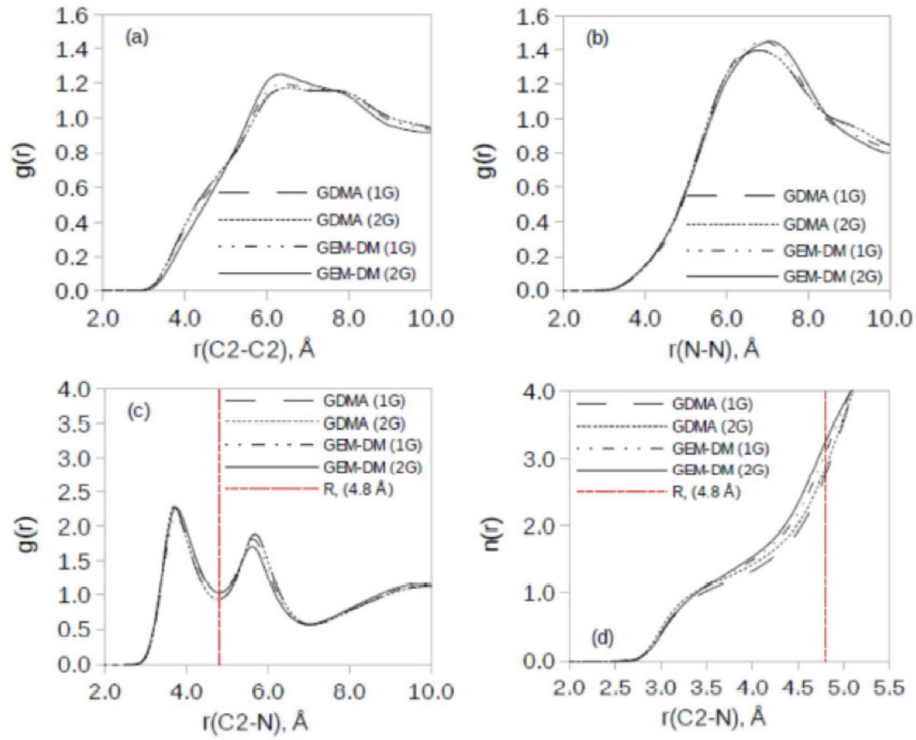

Figure 5.17 Radial distribution function for (a) cation-cation (b), anion-anion (c), and cation-anion for [dmim$^+$][NO$_3^-$] employing GDMA and GEM-DM multipole force fields. (d) The coordination number n(r) is given for the cation-anions coordination shell.

There is not a significant difference in C2-C2 radial distribution functions that were calculated for two sets of the force fields. The GEM-DM (2G) force field resulted in less structured correlations, as indicated by a smooth interatomic curve in comparison with RDF curves using other force fields (see Figure 5.15a). Other force fields show a correlation peak at 4.0 Å. Similar structures were obtained from *ab initio* molecular dynamics simulations.[60] The first (C2-C2) correlation peak can be found at a distance of 6.2 Å. The r_g (C2-C2) radial distributions are in good agreement with the structural model that was fit to the data from neutron diffraction experiments[244] and other molecular dynamics simulation results.[60]

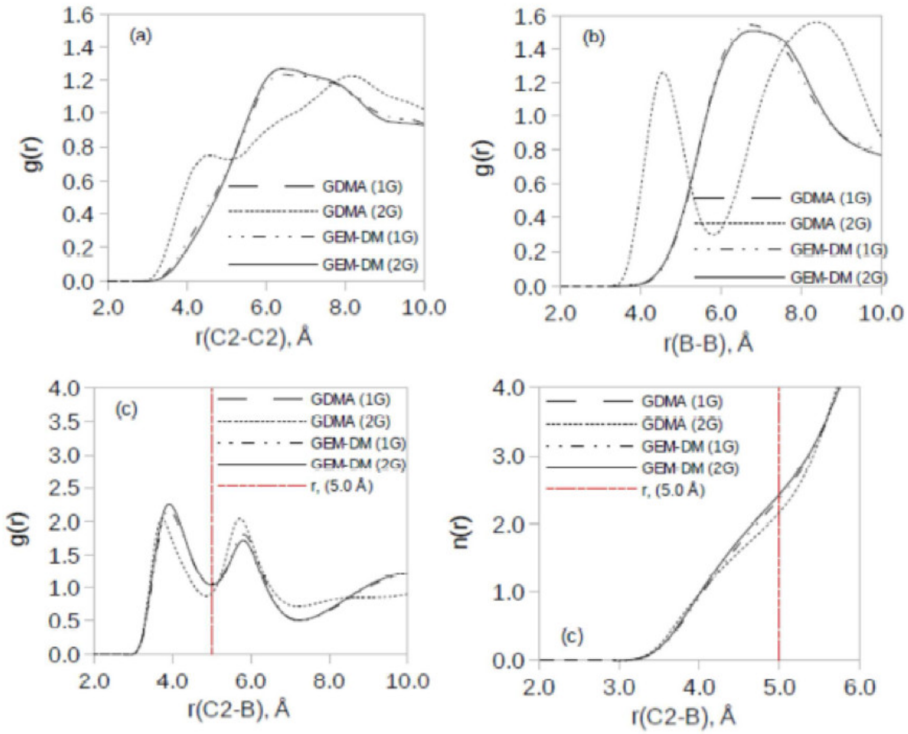

Figure 5.18 Radial distribution function for (a) cation-cation (b), anion-anion (c), and cation-anion for [dmim$^+$][BF$_4^-$] employing GDMA and GEM-DM multipole force fields. (d) The coordination number n(r) is given for the cation-anions coordination shell.

Despite the difference in the description of charge densities for partial point charges vs atomic multipoles, the structural differences are negligible in these particular correlations. Similar tendencies are shown for the (Cl$^-$−Cl$^-$) anion−anion interatomic correlations. The next structure is the cation-anion (C2−Cl$^-$) radial distribution; see Figure 5.15c. Sharper peaks are obtained using the GDMA based force fields, again indicating more structured correlations in comparison with the GEM-DM (2G) force field. The first peak of the radial distribution is at a distance of 3.5 Å. These results are well correlated to gas phase optimized geometries for the dimers of ionic liquids. The first peaks for the [dmim$^+$][F$^-$] and [dmim$^+$][BF$_4^-$] are at interatomic distances of 3.0 and

3.9 Å, respectively. Qualitatively similar two peak distributions were obtained from molecular dynamics simulations of [bmim$^+$][PF$_6^-$] at 298 K.[95] The results indicated that the first peak is located at 4.0 Å interatomic distance. We also calculate the coordination number n(r) by integrating the C2-Cl$^-$ radial distribution function. The coordination shell is defined by the dotted red line. It can be seen that the [dmim$^+$] cation is coordinated by two anions on average. If we define the coordination shell at the distance of second minimum, then we can see that the [dmim$^+$] cation will be coordinated by six to seven anions depending upon the force field and anion types.

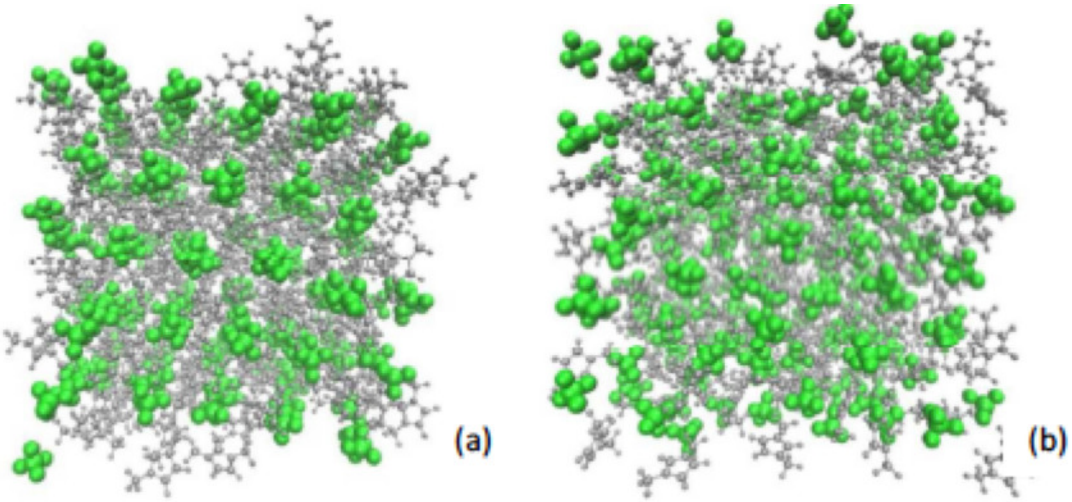

Figure 5.19 Two snapshots from MD simulations for liquid structure of [dmim$^+$][BF$_4^-$] at T=475 K using (a) GDMA and (b) GEM-DM force fields. Green molecules are [BF$_4^-$] and gray ones are [dmim$^+$].

5.3.3 Enthalpy of vaporization

We calculated the heat of vaporization to estimate the strength of intermolecular interactions for ionic pairs. The enthalpy of vaporization is calculated as the energy required to bring an ionic pair from the liquid to the gas phase. Gas phase simulations were carried out using stochastic molecular dynamics simulations of a single ionic pair at 425 K. Making an assumption of an ideal gas behavior, the enthalpy of vaporization was estimated using the following relation.

$$\Delta H_{vap} = \Delta E + \Delta PV = \Delta E + nRT = \langle U_{gas} \rangle - \frac{\langle U_{tot}^{liq} \rangle}{N} + RT \qquad \text{Eq 5.6}$$

where ΔE is the difference between potential energies from gas (−106.8 kcal/mol) and liquid phases (−141.9 kcal/mol) and ΔPV is the volume difference at constant pressure. Enthalpies of vaporization are listed in Table 5.2. Higher energies of vaporization correspond to the ionic liquids with anions of smaller radii, indicating stronger intermolecular interactions.

Table 5.2 Enthalpy of vaporization ΔH_{vap} at T =425 K. The enthalpy value in parentheses corresponds to data after scaling of van der Waals parameters. [a] Reference [233] T =423 K.

	ΔH_{vap} (kj/mol)				
ILs	GDMA(1G)	GDMA(2G)	GEM-DM(1G)	GEM-DM(2G)	simulation
$[dmim^+][F^-]$	175.7	172.2	168.2	168.5	
$[dmim^+][Cl^-]$	159.4	157.0	154.6	145.3 (175.0)	187.1[a]
$[dmim^+][NO_3^-]$	169.4	167.1	164.4	162.4	
$[dmim^+][BF_4^-]$			147.3	139.8	

These results are consistent with the results for other pairwise potentials. These potentials predict more favorable energies for $[dmim^+][F^-]$ by 13 kJ/mol than for $[dmim^+][Cl^-]$ or $[dmim^+][NO_3^-]$ and by 15 kJ/mol in comparison with $[dmim^+][BF_4^-]$.

The GDMA based force fields resulted in higher energies of vaporization compared to the GEM-DM based force fields. These results can be referred to the difference in describing van der Waals interaction energies. The GEM-DM based force field has a better agreement on van der Waals interactions with the quantum chemistry results, while the GDMA force field overestimates van der Waals interactions by ≈ 2 kcal/mol.

The calculated enthalpy of vaporization for [dmim$^+$][Cl$^-$] ionic liquid with GEM-DM (2G) is 145.3 kJ/mol, which is significantly lower than the result of 187.1 kJ/mol at 423 K from simulations using point charges. After the scaling of van der Waals parameters, the calculated enthalpy of vaporization is 175.0 kJ/mol, which is 29.7 kJ/mol more favorable. This error might arise due (at least in part) to intermolecular energy deviations in electrostatic and polarization interactions.

Table 5.3 Intermolecular van der Waals parameters for N-Cl$^-$, C-Cl$^-$ and H-Cl$^-$ before and after the fitting.

		Before fitting		After fitting	
Atom type pair	Atom class pair	R_{min} (Å)	ε (kcal/mol)	R_{min} (Å)	ε (kcal/mol)
N-Cl$^-$	62-12	3.942	0.174	3.842	0.174
C-Cl$^-$	63-12	3.970	0.169	3.870	0.169
H-Cl$^-$	64-12	3.740	0.599	3.640	0.599

However, the empirical adjustment of van der Waals parameters can also have a significant influence on the strength of intermolecular interactions (see Table 5.3). An increase in enthalpy of vaporization was predicted to be ≈ 4.0 kJ/mol per one CH_2 group from simulations of imidazolium based ionic liquids. Extrapolation of simulation results for [emim$^+$] and [bmim$^+$] with [BF_4^-] at 298 K would result in an enthalpy of vaporization of ≈ 130 kJ/mol for [dmim$^+$][BF_4^-]. This value would be too low. Nevertheless, [dmim$^+$][BF_4^-] ionic liquid is more likely to be in a crystalline state at 298

K. We also calculated the enthalpy of vaporization at 400 K for [dmim$^+$][BF$_4^-$] obtaining 148 kJ/mol employing the GEM-DM force field.

5.3.4 Ion self-diffusion coefficients

The self-diffusion coefficient is another liquid property that reflects the quality of the force fields implemented in molecular dynamics simulations. The mobility of ionic species depends on many parameters including the geometrical structure, ion size, charge delocalization, and strength of intermolecular interactions.[56, 59] A minor effect has been observed on the ion self-diffusion coefficient due to a change in a conformational barrier (C2−N−C−C).[56] Faster ionic diffusion is also achieved using polarizable force fields in comparison with nonpolarizable ones.

Inclusion of many body interactions speeds up the ion diffusion. Here, we study the influence of multipolar force fields on self-diffusion coefficients of ionic species. We compare GDMA and GEM-DM based force fields with available simulation and experimental data. The ion self-diffusion coefficients were calculated using the Einstein relation.[245]

$$D_{\pm} = \lim_{t\to\infty} \frac{\langle MSD\ (t)_{\pm}\rangle}{6t} \qquad \text{Eq 5.7}$$

where $MSD_{\pm}(t)$ is the mean square displacement of the molecule center of mass, t is the time, and ⟨ ⟩ defines an ensemble average. During the molecular dynamics simulations, the ions moved around ≈ 1000 and ≈ 4000 Å^2 over the production runs corresponding to 400 and 475 K, respectively.

This indicates that the diffusion regime has been reached, since the mean square displacement of an ion is clearly much larger than two ionic radii of gyration ($R_g \approx$ 2.79

Å). We fit the linear equation (m*t+c) to the mean square displacement (MSD(t)) in the diffusive regime from the production trajectories. Production trajectories were generated for over 8 ns. Calculated self-diffusion coefficients for the [dmim$^+$] cation and various anions are given in Figure 5.20. The trends in the self-diffusion coefficients are well correlated with the enthalpies of vaporization.

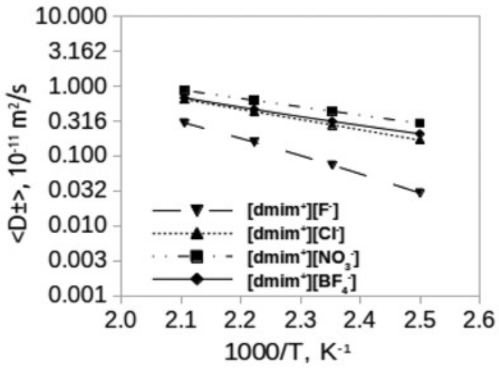

Figure 5.20 Comparison of self-diffusion coefficients $< D_{\pm} >$ using the GEM-DM (2G) force field. Average self-diffusion coefficients are calculated for [dmim$^+$][F$^-$], [dmim$^+$][Cl$^-$], [dmim$^+$][NO$_3^-$], and [dmim$^+$][BF$_4^-$].

Borodin *et al.* have shown a correlation between enthalpy of vaporization and diffusion coefficients for a number of ionic liquids.[59] We get similar relations, as faster diffusion rates are obtained for [dmim$^+$][NO$_3^-$] with lower enthalpy of vaporization in comparison with the [dmim$^+$][F$^-$] ionic liquid.

However, we obtained lower diffusion rates for [dmim$^+$][BF$_4^-$] in comparison with [dmim$^+$][NO$_3^-$] despite having a lower enthalpy of vaporization. These results can be attributed to the geometry of [BF$_4^-$]. Previous studies have shown that the geometries of ions do have an impact on diffusion rates due to their weaker or stronger ordering.[56]

The difference in diffusion rates between cations and anions is not significant. Diffusion rates reported by Borodin *et al.* for [emim$^+$][BF$_4^-$] using a polarizable force field [74] are somewhat similar to [dmim$^+$][BF$_4^-$] using multipolar force fields. Therefore, the anisotropy of electrostatic interactions did not have a significant effect on the overall diffusion rates becoming comparable with the rates that result from employing partial charge based polarizable force fields.

5.4 Conclusions

We have developed AMOEBA polarizable force fields, that employ higher order multipoles to describe electrostatic interactions, for imidazolium based ionic liquids. The multipoles obtained by both the GDMA and the GEM-DM methods well describe intermolecular pair electrostatic interactions as compared to the energies from EDA analysis at medium and long range. A good description of the gas phase dimer geometries is also obtained without any empirical adjustments of the multipoles, atomic polarizabilities, or van der Waals parameters for all ionic liquids but [dmim$^+$][Cl$^-$]. The use of GDMA and GEM-DM multipoles in the newly developed AMOEBA force fields has shown initially a 4% deviation from experimental liquid densities ρ for [dmim$^+$][Cl$^-$]. The corresponding enthalpy of vaporization ΔH_{vap} is 145.3 kJ/ mol. A limited number of experimental data is available for [dmim$^+$][Cl$^-$]. Therefore, empirical adjustments of the van der Waals parameters have been carried out only to experimental (ρ = 1.127 g/cm3) for [dmim$^+$][Cl$^-$] and an enthalpy of vaporization ΔH_{vap} of 175 kJ/mol after the optimization of the van der Waals parameters for this ion pair. Overall results indicate that the implementation of the GEM-DM (2G) force field has resulted in a good agreement in intermolecular energies as compared to QM calculations for all ionic liquids except for $[\text{dmim}^+][\text{Cl}^-]$. Thermodynamics and transport properties of the studied ionic liquids are in reasonable agreement with available data making the GEM-DM (2G) force field a good model to simulate 1,3-dimethylimidazolium based ionic liquids.

CHAPTER 6 COMPUTATIONAL CHARACTERIZATION OF PYRROLIDINIUM BASED IONIC LIQUID FOR ELECTROLYTE APPLICATION

6.1 Introduction

The development of Li ion batteries for energy storage has received significant attention. The synthesis and characterization of electrolytes in these batteries is an important component of this development. Ionic liquids (ILs) have been proposed as possible alternatives for electrolytes in these devices. The accurate determination of basic thermophysical properties of these solvents are vital for determining the applicability of these solvents as electrolytes. Here, we present the computational characterization of thermodynamic and transport properties of a pyrrolidinium based ionic liquid as a first step to investigate the possible applicability of this class of ILs for Li ion batteries.

6.2 Computational methods

6.2.1 Parameter determination

The accuracy of the parametrization of a force field may be verified by the calculation of three sets of molecular data: (i) molecular geometry and intramolecular energies, (ii) intermolecular potential and (iii) bulk thermodynamic and structural properties. Most classical force fields are comprised of bonded and non-bonded parameters. In our model, the bonded terms are taken directly from AMOEBA without any adjustment,[78, 216] whereas the non-bonded terms, described by Eq. 5.1, have been fitted to reproduce Coulomb, polarization and the Van der Waals energies based on QM energy decomposition analysis (EDA).

We first calculated the total inter-molecular, Coulomb, polarization and Van der Waals energies using quantum mechanics as a reference for fitting our parameters. Total inter molecular energies were calculated using the counterpoise correction to take into account the basis set superposition error (BSSE)[246] at the MP2/6-311G(d) level for dimers. The inter molecular polarization interactions for each pair were calculated using the restricted variational space (RVS) decomposition approach[247] at the HF/6-311G(d) level of theory as implemented in GAMESS.[248-249] An in-house FORTRAN90 program that uses ab initio monomer electron densities was employed to calculate the Coulomb inter molecular energies for each dimer. Finally, the Van der Waals energies were obtained by subtracting the Coulomb and polarization energy from total intermolecular energy for each dimer, as described in Eq. 5.5.

Briefly, the non-bonded parameters in the AMOEBA potential include polarization, Van der Waals and Coulomb terms. The polarization energy is calculated by inducible atomic dipoles on each interaction (atomic) site. Here, each induced dipole is obtained by $\mu_{i,\alpha}^{ind} = \alpha_i E_i$ where α_i is the atomic polarizability and E_i is the external electric field generated by both permanent multipoles and induced dipoles. The Tholé damping function [75, 136] is employed to avoid the "polarization catastrophe" at short range.

In this study, the intra molecular polarization is also taken into account for spirocyclic pyrrolidinium.[216] Since spirocyclic pyrrolidinium is not a planar molecule, the changes in each ring of this molecule could change the electron density distribution of the whole molecule. Hence, two sets of parameters have been developed: 1) One with no intra molecular polarization for [sPyr$^+$], i.e. 1 polarization group, termed 1G; and 2) One with intra molecular polarization in [sPyr$^+$] where each ring, and the central nitrogen atom were polarized separately, i.e., three polarizable groups (3G). The Van der Waals potential energy is described by the buffered Halgren pairwise potential (see Eq. 5.3).[82]

Finally, distributed atomic multipoles are used to describe the electrostatic (Coulomb) interactions. In this case, following our previous work (Chapter 5), the distributed multipoles obtained from the fitting of quantum mechanical (QM) electronic density via the GEM procedure are employed. The Hermite coefficients and associated distributed multipoles (GEM-DM) up to quadrupoles on each atom for each monomer (in the optimized geometry) were obtained using relaxed electronic densities for the monomers calculated at the MP2/6-311G(d) level in the GEM fit program.[79]

The accuracy of the fitted GEM density is tested by comparing the intermolecular Coulomb interaction calculated with GEM, with its QM counterpart from RVS.[248] In our previous work,[76] the GEM densities were fitted as a function of the change in a 1D scan along the distance between the dimers. This leads to less accurate sampling for all the possible interactions in the bulk. Hence, in this study, our fitting methodology has been modified, and instead, for this work the GEM densities have been fitted using a series of random dimers.

The optimization of the GEM densities and associated distributed multipoles for sPyr$^+$ and BF_4^- have been optimized separately by calculating the intermolecular Coulomb interaction of each ion with a single water molecule. The GEM densities and distributed multipoles for the water fragment have been previously fitted and reported in our previous paper.[77] Once the densities and multipoles for the ions were optimized, the intermolecular interaction between the cation and anion was used to validate the fit.

One additional advantage of performing the RVS calculations for the random dimers is that this provides a QM reference for each non-bonded term in the intermolecular interaction. Thus, the polarization and Van der Waals terms are also optimized by comparing to their RVS counterparts as described above.[77, 89] Once all parameters for the new molecules have been fitted by comparing to QM data, the total intermolecular energies for the [sPyr$^+$][BF$_4^-$] dimers are compared with counterpoise corrected total intermolecular interactions obtained with the Gaussian09 package.[162-163] The inter molecular, Coulomb, polarization, Van der Waals and total energies using the optimized parameters for a series of random dimers of [sPyr$^+$][BF$_4^-$] were obtained with the analyze module available in TINKER.[163]

6.2.2 MD simulation details

MD simulations were carried out in the AMBER12 simulation package,[186] using the AMOEBA/GEM-DM force field. The MD simulations were performed using an orthorhombic simulation cell with periodic boundary conditions. The calculated system included 216 dimers. For the Li-doped system 22 spirocyclic pyrrolidinium cations were replaced with Li$^+$ ions to make a mixture of 10% Li ion doped into the IL. All systems were minimized with 1000 conjugate steps, and heated up to 600 K to make sure they are in liquid phase. Subsequently, all systems were cooled down to the desired temperatures. The productions were carried out in the NPT ensemble with an integration time step of 1 fs. Long-range electrostatic effects were computed employing the smooth particle mesh

Ewald method[217-218] with an 8 Å direct cutoff. Sampling trajectories for neat IL were generated for 10 ns, and for Li salt doped mixture for 20 ns.

Calculated properties from the generated ensembles include density (ρ), heat of vaporization (H_{vap}), diffusion coefficient (D), and radial distribution functions (RDFs). The density and volume data are obtained directly from MD outputs. The heat of vaporization, which is an estimation of the inter-molecular interaction strength of ionic pairs, is calculated using Eq. 5.6. where ΔE is the difference between potential energy in gas and liquid phase. The enthalpy of vaporization is calculated as the energy required to take an ionic pair from the liquid to the gas phase. Gas phase simulations were carried out using stochastic molecular dynamics simulations of a single ionic pair at each temperature using the dynamic module in TINKER.[163]

The self-diffusion coefficient is another liquid property that reflects the quality of the force fields implemented in MD simulations. The mobility of ionic species depends on many parameters including the geometric structure, ion size, charge delocalization, and strength of intermolecular interactions. The ion self-diffusion coefficients were calculated using the Einstein relation (Eq. 5.7).

Liquid structures of ionic liquids are usually compared to X-ray scattering[11] or neutron diffraction experimental data.[244] In particular, inter-ionic correlations are well described by radial distribution functions and can be compared with the structural factor S(Q) obtained from neutron diffraction. Radial distribution functions (RDF) were calculated at different temperatures for Cation-Cation, Anion-Anion and Cation-Anion pairs.

6.3 Results and discussion

6.3.1 Parameter fitting

In order to fit the multipoles for each molecular fragment, the same methodology implemented for nonionic compounds was used as described previously.[77] We randomly generated thirty dimers of cation+water and twenty-five dimers of anion+water, and calculated the Coulomb energies for each dimer. The Coulomb energies are calculated with GEM, GEMDM and QM (from the relaxed QM densities of the monomers).

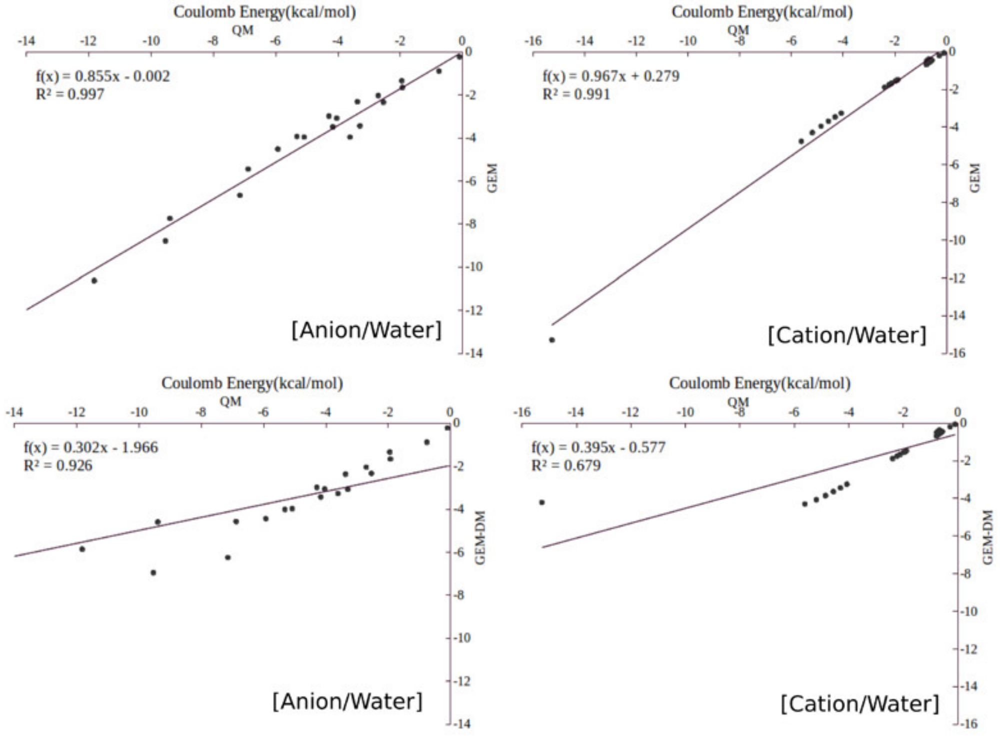

Figure 6.1 Coulomb energies calculated for cation/anion+water dimer using GEM, GEM-DM and QM.

Figure 6.1 shows the comparison of Coulomb energies calculated for spirocyclic pyrrolidinum+water and tetrafluoroborate+water using GEM and GEM-DM with full QM. As expected, the Coulomb energies calculated with the full GEM densities are well correlated with QM energies. Conversely, this correlation decreases when multipoles (GEM-DM) are employed due to the penetration error.

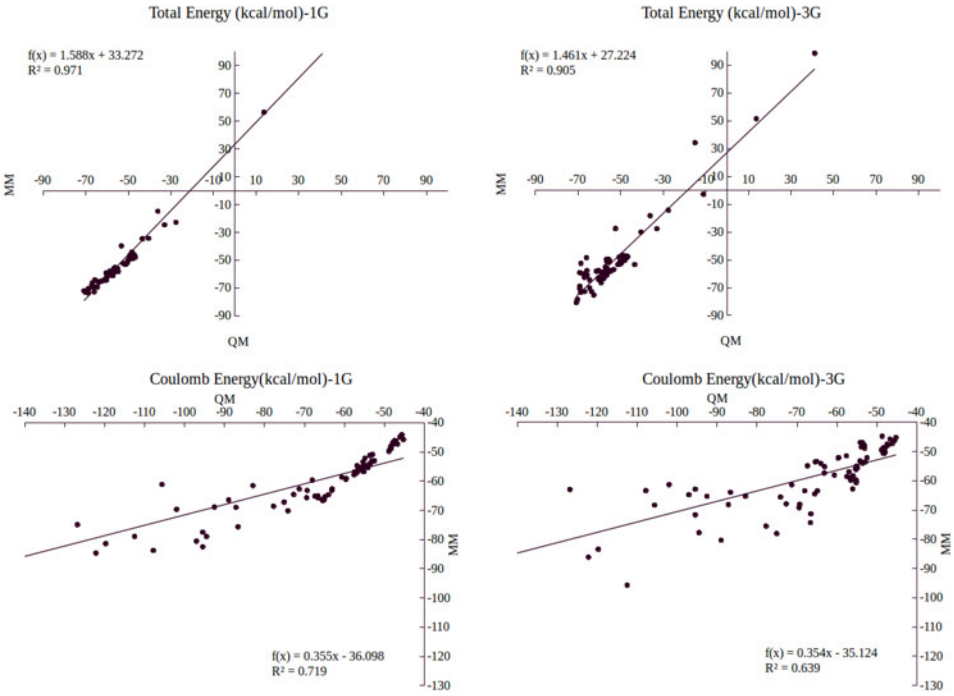

Figure 6.2 Total and Coulomb energies calculated for $[sPyr^{+}][BF_4^{-}]$ dimers in MM and QM.

Once the multipoles for the ions had been obtained, intermolecular interaction energies were determined for seventy seven random dimers of $[sPyr^{+}][BF_4^{-}]$ using QM EDA as a reference and the new force field using both sets of parameters (with (3G) and without (1G) intra-molecular polarization for $[sPyr^{+}]$. The total interaction energy in the gas phase was calculated at the MP2(full)/6-311G(d) level of theory. The total inter molecular interaction energies were decomposed using the RVS[247] decomposition analysis. The Figures 6.2 and 6.3 shows the decomposed energies for [sPyr$^+$][BF$_4^-$].

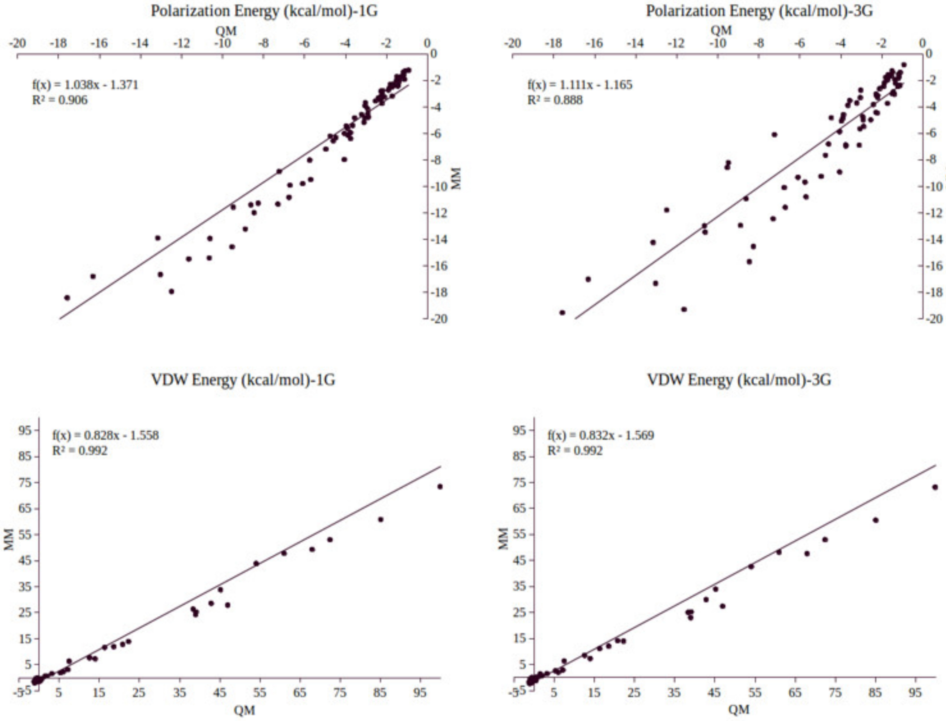

Figure 6.3 Polarization and vdW energies calculated for [sPyr$^+$][BF$_4^-$] dimers in MM and QM.

As can be observed, both sets of parameters provide good agreement for total intermolecular interactions. The correlation between QM and MM energies for both sets of force fields are very similar, therefore, MD simulations were performed with both sets of force fields to find out which one describes the behavior of the system better and in more agreement with experimental data.

6.3.2 Density

The density of neat [sPyr$^+$][BF$_4^-$] was calculated for a range of temperatures between 300-500 K as shown in Figure 6.4. Only a small change in density between 400-450 K is observed, however, the density plot shows a significant change in density between 450-453 K. This change seems to correspond to a phase change of the system. Interestingly, subsequent DSC results show that the melting point for [sPyr$^+$][BF$_4^-$] is within this temperature range as described below.

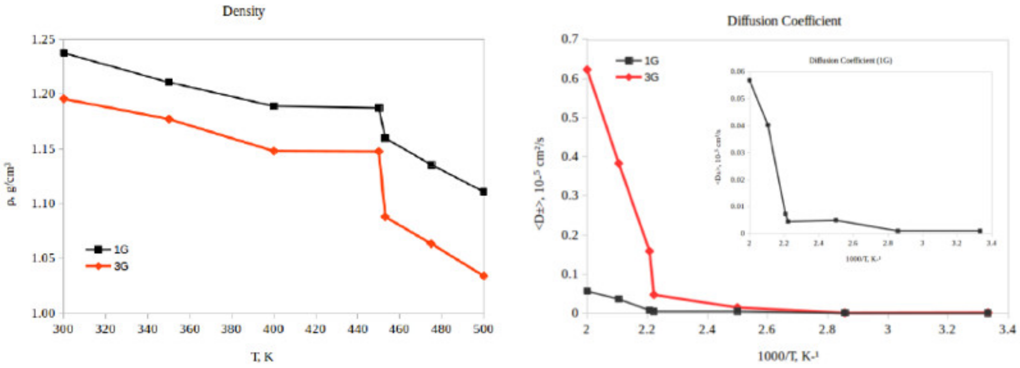

Figure 6.4 Densities and diffusion coefficients calculated for [sPyr$^+$][BF$_4^-$] with 1 and 3 polarizable groups for 300 to 500 K.

The density and volume using the parameters with 3 polarizable groups at 450 K is 1.15 kg/m^3 and 308.12 Å^3, respectively, compared with 1.09 kg/m^3 and 325.06 Å^3 at 453 K. The density decreased by 6% from 450 K to 453 K. The densities at 450 K and 453 K obtained with the other parameter set (1G) are 1.19 kg/m^3 and 1.16 kg/m^3, which shows the density with no intra molecular polarization decreased by 2:8%. Thus, the change in the density is larger when using 3 polarizable groups in comparison with 1 polarizable group, with both parameter sets suggesting a phase transition in this temperature range.

To further investigate the applicability of these newly designed electrolytes in Li ion batteries, a system was created where 10% Li ions were doped in the ionic liquid simulation box to observe the changes in thermodynamic properties of our system.

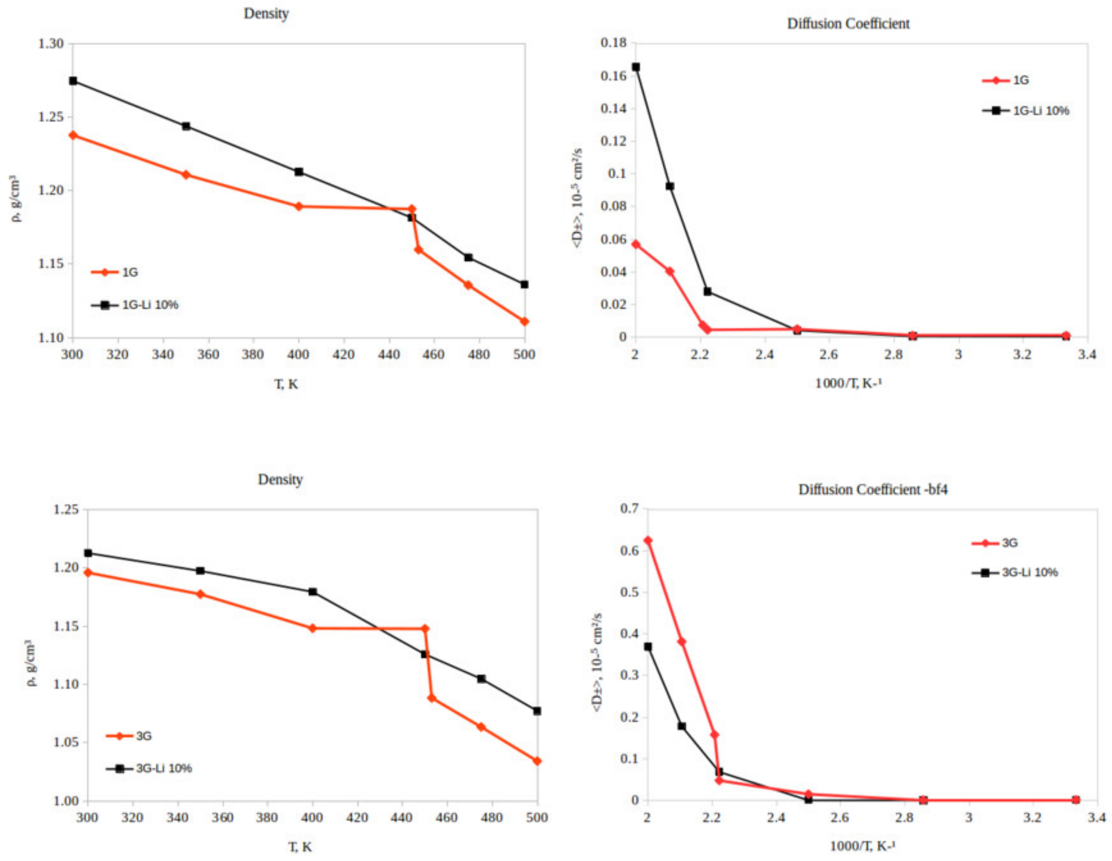

Figure 6.5 Densities and diffusion coefficients calculated for neat IL and [10% Li^+][BF_4^-] -[90% $sPyr^+$][BF_4^-] with 1 and 3 polarizable groups for 300 to 500 K.

Overall, the density of the mixture is 2-4% higher than the density in the neat IL for all the temperature except at 450 K (or at 430-450K) using both sets of force fields (Figure 6.5). In addition, a significant change in density (5.4% and 3.1% for 3G and 1G, respectively) between 400-450 is observed for the mixture. This change may correspond to the phase change of the mixture at a slightly lower temperature in comparison with neat IL. The comparison of the calculated densities with 1 and 3 polarizable groups for the Li-doped [sPyr$^+$][BF$_4^-$] IL is presented in Figure 6.6. The density values calculated without intra-molecular polarization (1G) is larger than those values using 3 polarizable groups at all temperatures.

6.3.3 Enthalpy of vaporization

The enthalpy of vaporization is calculated as an estimation of the strength of inter-molecular interactions for ionic liquid. The heat of vaporization energies were calculated at T=300-500 K with both sets of force fields. Table 6.1 shows the enthalpies of vaporization at different temperatures.

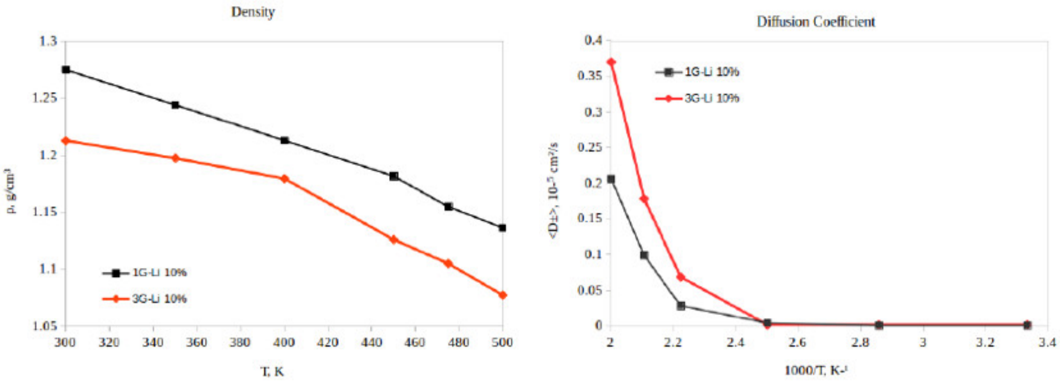

Figure 6.6 Densities and diffusion coefficients calculated for [10% Li$^+$][BF$_4^-$]-[90% sPyr$^+$][BF$_4^-$] with 1 and 3 polarizable groups for 300 to 500 K.

Gas phase energies presented in Table 1 are calculated for a single dimer using dynamic module in TINKER and bulk energies are obtained directly from MD simulations in different temperatures. As seen in Table 6.1, the heat of vaporization values obtained with the 3G parameter set are bigger than the values obtained with the 1G force field. These results suggest that taking the intra-molecular polarization into the account results in stronger inter-ionic interactions in comparison with the model without intra molecular polarization.

Table 6.1 Heat of vaporization for [sPyr$^+$][BF$_4^-$] in kcal/mol.

T	AMOEBA/GEM-DM (1G)			AMOEBA/GEM-DM (2G)		
	gas	bulk	ΔH_{vap}	gas	bulk	ΔH_{vap}
300	-69.06	-104.21	-67.94	-110.4	-180.58	-108.61
350	-69.94	-99.14	-63.78	-106.53	-176.71	-105.01
400	-60.42	-93.62	-59.19	-101.42	-172.14	-99.83
450	-56.91	-93.63	-55.58	-97.18	-167.61	-95.51
453	-55.91	-87.65	-54.61	-97.16	-166.13	-95.49
475	-54.91	-84.94	-53.57	-95.57	-163.62	-93.87
500	-52.22	-81.98	-50.84	-92.84	-160.74	-91.10

6.3.4 Diffusion coefficients

Self-diffusion coefficients for [sPyr$^+$][BF$_4^-$] using AMOEBA/GEM-DM at T=300-500 K are shown in Figure 6.4 for the 3G set. The jump between 450 K and 453 K provides further support for the phase change at this temperature. The average diffusion coefficients for anion and cations (D) at 450 K is 0.07×10^{-5} cm^{2}s^{-1}, and at 453 K is 0.16×10^{-5} cm^{2}s^{-1}. Faster ionic diffusion is also achieved using the 3G potential in comparison with the non intra-polarizable parameter set (Figure 6.4). This suggests that the inclusion of a better description of many body interactions speeds up the ion diffusion for this system.

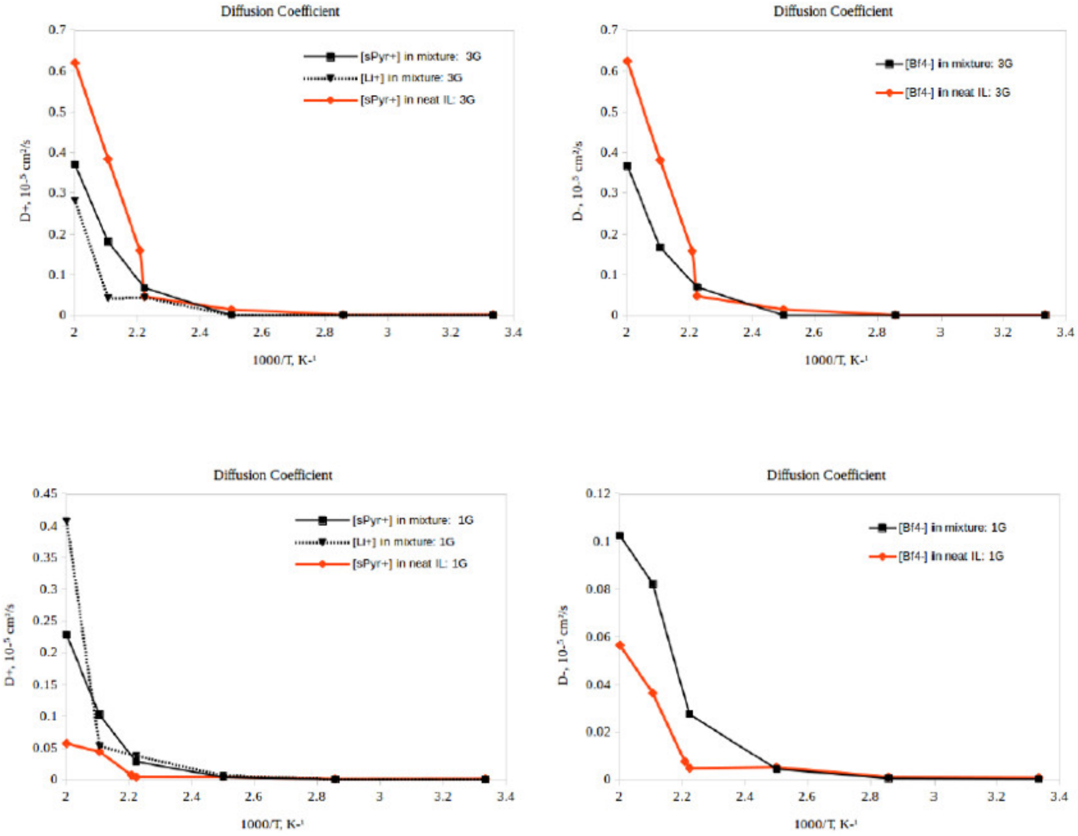

Figure 6.7 Diffusion coefficients for cations and anions for neat IL and [10% Li$^+$][BF$_4^-$]-[90% sPyr$^+$][BF$_4^-$] with one and three polarizable groups.

The average diffusion coefficients of [10% Li$^+$][BF$_4^-$] 90%sPyr$^+$][BF$_4^-$] for the Li-doped mixture at T=300-500 K using 3 polarizable groups are presented in Figure 6.5 and 6.7. The significant change observed at 450 K for the neat IL shifts to lower temperature in the mixture. In addition, the ions diffusion in the mixture is very similar to the neat IL at lower temperatures. However, ions diffuse slower in the Li-doped mixture in comparison with neat IL at higher temperatures (except around 450 K), due to stronger inter ionic interactions between the Li cations and the anions. Conversely, the diffusion coefficients at 430-450 K are slightly larger than their neat IL counterparts at the same temperature (2.2 % higher at 450 K), which is in good agreement with the lower density of mixture at these temperatures.

The diffusion coefficients calculated using the 1G parameter set for the mixture and its comparison to neat IL are presented in the Figure 6.5 and 6.7. Conversely, the ions diffuse faster in the 10% Li^+/IL mixture than in the neat IL when the 1G parameter set is employed. This may be due to the reduced accuracy in the description of the many-body effects.

6.3.5 Radial distribution function

The RDF for the Anion-Anion (B-B), Anion-Cation (B-N) and Cation-Cation (N-N) pairs for the neat IL in all the tested temperatures (300-500 K) using the 3G parameter set are depicted in Figure 6.8. A small peak at 4 Å is observed for the B-B RDF at 300 K and 350 K, which is correlated to a close distance between [BF_4^-] ions at these temperatures.

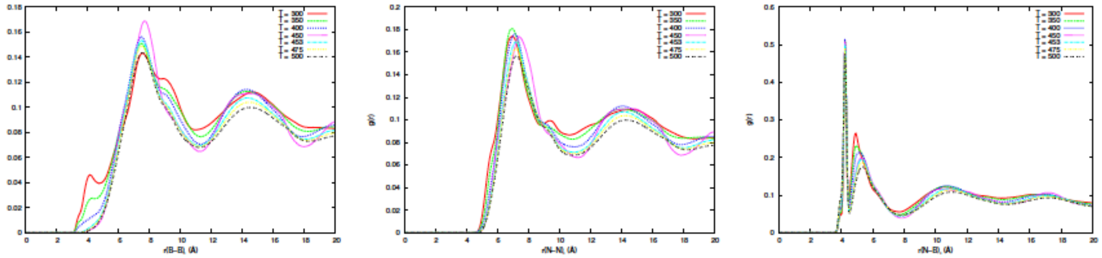

Figure 6.8 Radial distribution function for [$sPyr^+$][BF_4^-] with three polarizable groups.

This peak is not observed at higher temperatures due to the increase in ion distances and the phase transition of the IL. The second shell (at 8 Å) of the Anion-Anion RDF shows that the largest peak corresponds to 450 K, which decreases significantly as the temperature is increased to 453 K. The RDF plots for the neat IL obtained without intra-molecular polarization (1G) are presented in Figure 6.9.

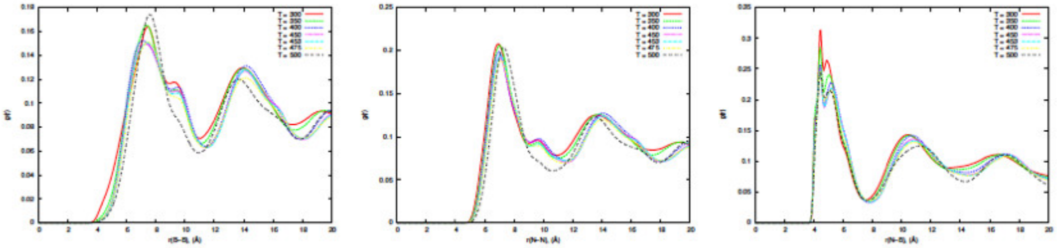

Figure 6.9 Radial distribution function for [sPyr$^+$][BF$_4^-$] with one polarizable groups.

The overlay of the RDFs calculated with 1 and 3 polarizable groups for the mixture are presented in Figure 6.10 and 6.11. The RDF for Anion-Anion using 3 polarizable groups shows that the largest peak in the first and second shells correspond to 300 K and 400 K, respectively. It may correspond to the phase change of Li-doped mixture at lower temperature in comparison with the neat IL. In addition, the first peak of this plot (4 Å) is eliminated by increasing the temperature using both force fields, however, this peak is smaller using 1 polarizable group.

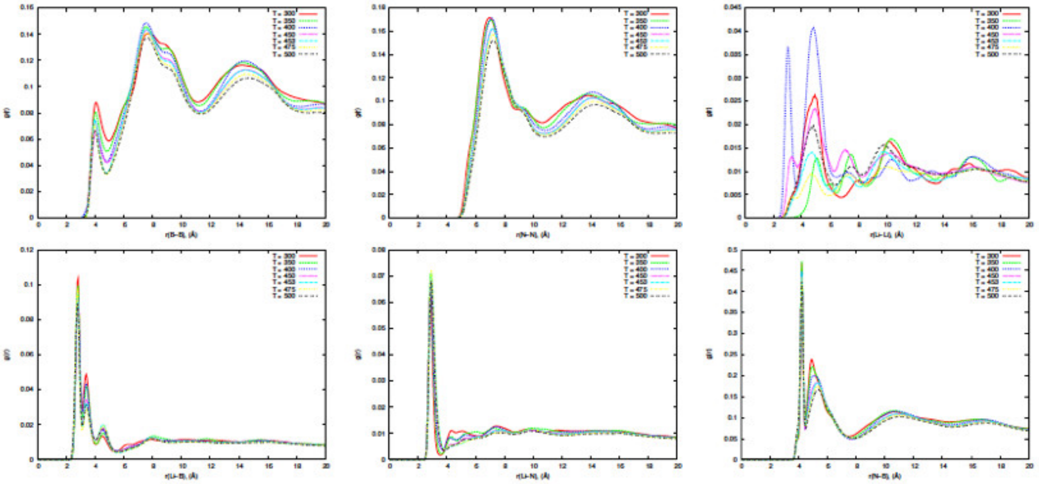

Figure 6.10 Radial distribution function for [10% Li$^+$][BF$_4^-$]-[90% sPyr$^+$][BF$_4^-$] with three polarizable groups.

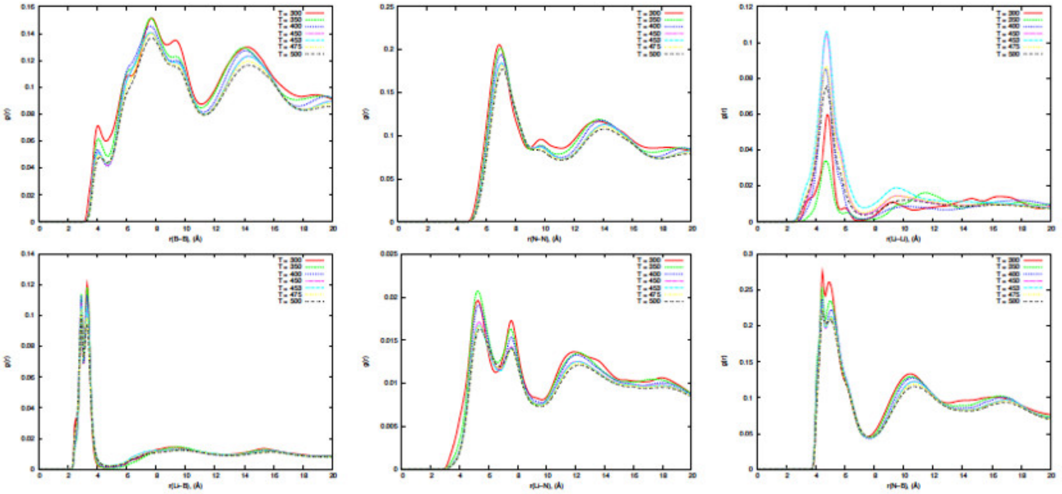

Figure 6.11 Radial distribution function for [10% Li$^+$][BF$_4^-$]-[90% sPyr$^+$][BF$_4^-$] with one polarizable groups.

The Li-Li RDF shows two split peaks between 3-4 Å at 400 and 450 K, which correspond to the existence of Li-BF$_4$ clusters at these temperatures using 3 polarizable groups. The RDF using 1 polarizable group for Li-Li ions, on the other hand, doesn't show the split peaks. Instead, there is a shoulder in the first shell at 300 K, which vanishes at higher temperatures. Overall, the calculated RDFs for the mixture using 1 polarizable group are more similar to the neat IL using the 1G parameter set. Further studies are under way to investigate the Li ion behavior in this mixture at different temperatures.

6.3.6 Summary of experimental results

Following the interesting results obtained from the computational simulations regarding the observed phase transition around 400-450 K, differential scanning calorimetry was performed on the synthesized [sPyr$^+$][BF$_4^-$] by our experimental collaborators at UC-Merced. Figure 6.12 shows a typical DSC run from 360 K to 480 K for the BF_4 compound. The DSC for the BF_4 compound shows that the melting temperature is about 448 K, and an enthalpy of fusion of 181 J/g, and a crystallization onset of 446 K and an enthalpy of crystallization of 350 J/g. However, examination of the DSC for the BF_4 compound reveals extra peaks, suggesting the presence of impurities in the compound making the enthalpy measurements unreliable.

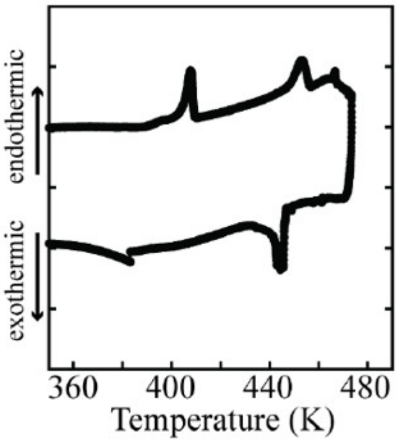

Figure 6.12 Experimental DSC for $[sPyr^+][BF_4^-]$ at T = 360-480 K.

6.4 Conclusions

The computational characterization of spirocyclic pyrrolidinium tetraflorouborate has been explored as an initial step toward its evaluation as an electrolyte in Li-ion batteries. AMOEBA parameters have been developed based on QM reference data and employed to calculate several thermodynamic properties including ρ, ΔH_{vap}, and D for a range of temperatures. Our results suggest a phase change around 450 K, which was subsequently confirmed by experimental DSC characterization of the $[sPyr^+][BF_4^-]$.

The obtained thermogram shows that the melting point is around 448 K. Further computational studies on the Li-doped mixture predict that the phase change temperature shifts to lower value in the mixture. Our results for the mixture shows higher density and slower diffusion than neat IL at all temperature except the temperatures between 400 to 450 K.

CHAPTER 7 DEVELOPMENT OF AN AMOEBA WATER MODEL USING GEM DISTRIBUTED MULTIPOLES

Portions of the text in this chapter were reprinted or adapted with permission from: *Theor. Chem. Acc., 134, 101:1 (2015)*. All rights to the work are retained by the authors and any reuse requires permission of the authors.

7.1 Introduction

Distributed multipoles obtained from the Gaussian electrostatic model (GEM) have been previously shown to be amenable for use in the AMOEBA force field.[79] GEM distributed multipoles (GEM-DM) were determined for several systems including water. This previous AMOEBA water model with GEM-DM included only monopoles on the hydrogens and multipoles up to quadrupoles on the oxygen. This model showed good agreement with experiment for several properties at room temperature, but not at higher temperatures. In this contribution, we present the development of an AMOEBA water model using GEM-DM with distributed multipoles for each atomic site up to the quadrupole level. Quantum mechanical energy decomposition analysis has been employed to compare each term of the force field for parametrization. The inclusion of higher-order multipoles on hydrogen atoms is shown to provide better agreement with experiment on a number of properties including liquid density (ρ), enthalpy of vaporization (ΔH_{vap}), heat capacity (C_p), and self-diffusion coefficient (D) for a range of temperatures.

7.2 Computational methods

The parametrization details employed in the present study have been described in detail in previous papers.[76, 79] A brief explanation of the parameter fitting methodology is presented in Section 7.2.1, followed by a description of the MD simulation setup in Section 7.2.2.

7.2.1 Parametrization details

AMOEBA potential including five bonded and three non-bonded terms is shown in Eq. 5.1.[82] As explained previously,[76, 79] the bonded terms are taken directly from AMOEBA without any adjustment.[78, 205, 216] In the case of the non-bonded terms, described by the last three terms of the potential (see Eq. 5.1): U_{Coul} , U_{Pol} , and U_{vdw} , all parameters have been fitted to reproduce results based on QM energy decomposition analysis (EDA) for the Coulomb and Polarization terms. For the van der Waals term, the initial parameters were fitted to reproduce the gas-phase EDA results and subsequently refined against two experimental data points. Total intermolecular energies were calculated using the counterpoise correction to take into account the basis set superposition error (BSSE)[228] at the MP2/aug-ccpVTZ level as a function of intermolecular separation for the canonical water dimer. The intermolecular polarization interactions for each pair were calculated using the restricted variational space (RVS) decomposition approach[250] at the HF/aug-cc-pVTZ level of theory as implemented in GAMESS.[230] An in-house FORTRAN90 program that uses *ab initio* monomer densities was employed to calculate the Coulomb intermolecular energies for each dimer. Finally, the van der Waals energies were obtained by subtracting the Coulomb and polarization energies from total intermolecular energy for each dimer, as described in Eq. 5.5.

GEM distributed multipoles (GEM-DM) up to quadrupoles on each atom were obtained using the GEM-fit program.[79] The Hermite coefficients and associated distributed multipoles were calculated by fitting the relaxed one-electron density from a single water molecule (in the AMOEBA optimized geometry) calculated at the MP2/aug-cc-pVTZ level and fitted using the A2[231] auxiliary basis set. These distributed multipoles are used to determine the intermolecular Coulomb interactions by Eq. 5.2.

Table 7.1 Comparison of multipoles using A1 and A2 ABSs for water

ABS	Atoms	Multipoles			Atoms	Multipoles		
A1	Oxygen	-0.92473			Hydrogen	0.46237		
		0.00192	-	-		0.00000	0.00000	0.00000
		0.49614				0.00000		
		0.00067	-			0.00000	0.00000	
		0.000182	0.00041	-		0.00000	0.00000	0.00000
A2	Oxygen	-0.41444			Hydrogen	0.20722		
		0.00000	0.0000	0.01503		-0.00600	0.0000	0.21386
		0.61476				0.05872		
		0.00000	-			0.00000	0.04492	
		0.00000	0.0000	-		0.00216	0.00000	-

In our previous work,[79] multipole moments were calculated using the A1 ABS, which contains only s-type functions on the hydrogen atoms. This results in only monopole components available on the hydrogen atoms. As described above, the results obtained with this model showed good agreement with QM reference data for intermolecular interactions and with experimental values for thermodynamic properties at 298 K. However, this model was not able to describe the behavior of system at elevated temperatures. In the current water model, we have employed the same reference electronic density as in our original model, except that the A2 auxiliary fitting basis set was employed. This ABS includes higher-order angular momentum functions (l = 2) on the hydrogen atoms.[79] The old and new multipoles are compared in Table 7.1.

The polarization term is described by inducible atomic dipoles on each interaction (atomic) site. Here, each induced dipole is obtained by $\mu_{i,\alpha}^{ind} = \alpha_i \, E_{i,\alpha}$, where α is the atomic polarizability, and $E_{i,\alpha}$ is the external electric field generated by both permanent multipoles and induced dipoles. The Tholé damping function[75, 136] is employed to avoid the "polarization catastrophe" at short range. In our previous study, the Tholé parameter was reduced to 0.35 to get a better description of the polarization interactions for the canonical water dimer.[79] In the current work, the damping factor was set to 0.39 (consistent with AMOEBA03) and resulted in good agreement for the polarization contribution (see below). The last term is the van der Waals potential energy, which is described by the buffered Halgren[224] pairwise potential (see Eq. 5.3)

The van der Waals parameters for molecules are calculated using the Lorentz–Berthelot combining rule.[82]

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_{ii} + \sigma_{jj}\right) \qquad \text{Eq 7.1}$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \qquad \text{Eq 7.2}$$

Validation of the new parameters was performed by running a series of molecular dynamics simulations for liquid and gas phases and comparing the resulting calculated properties with experimental data and other molecular dynamics simulation results.

7.2.2 MD simulations

MD simulations were carried out using the AMOEBA/ GEM-DM force field in the AMBER12[238] simulation package. The MD simulations were performed using a cubic simulation cell with periodic boundary conditions. The calculated system included 216 water molecules (648 atoms). Simulations were carried out in the NPT ensemble with an integration time step of 1 fs. Long-range electrostatic effects were computed employing the smooth particle mesh Ewald method[149, 218, 241] with an 8 Å direct cutoff. Sampling trajectories were generated for 4 ns.

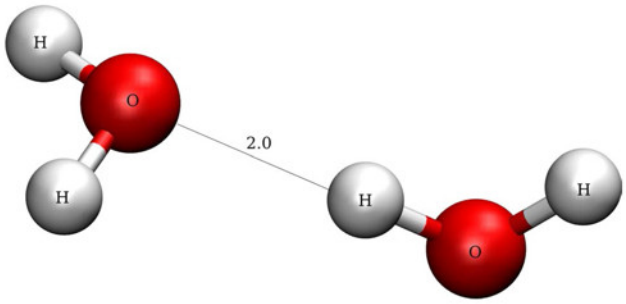

Figure 7.1 Distance between the center of mass of the oxygen atom in one monomer with respect to the hydrogen atom in another monomer in Å

7.3 Results and discussion

This section presents the parametrization procedure, and results are obtained for the new AMOEBA/GEM-DM water model. Section 7.3.1 presents the results for the parametrization calculations for the intermolecular parameters. This is followed by the results for the parametrization of the van der Waals term in Section 7.3.1.2, and the validation of the water model by calculating various thermodynamic and transport properties as discussed in Section 7.3.2.

7.3.1 Force field parametrization

7.3.1.1 Intermolecular interactions

A series of water dimers was generated by systematically varying the distance (increments of 0.1 Å) between the center of mass of the oxygen atom in one monomer with respect to the hydrogen atom in another monomer to calculate intermolecular interaction energies for the canonical water pair (Figure 7.1). The total interaction energy for these dimers in the gas phase was calculated at the MP2(full)/aug-ccpVTZ level of theory using BSSE in Gaussian09. For these dimers, the minimum interaction distance is located at 2 Å with a corresponding intermolecular interaction energy of -4.71 kcal/mol compared with -4.69 using AMOEBA/GEM-DM. The complete intermolecular potential energy surface is shown in Figure 7.2.

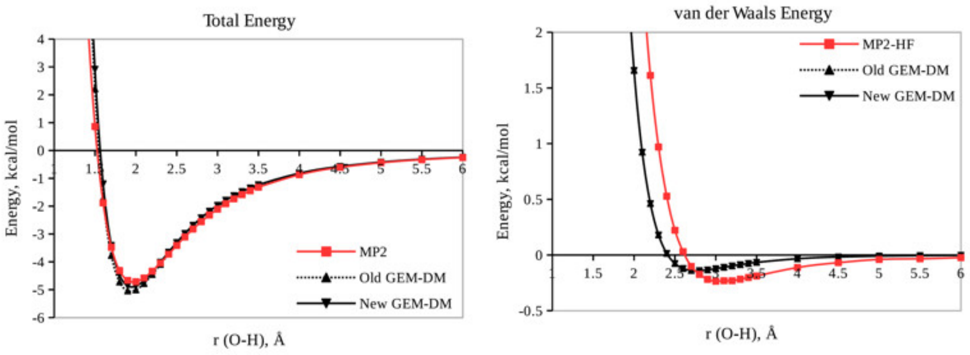

Figure 7.2 Total interaction energy and vdW interactions for water dimers

Next, we performed EDA to compare the intermolecular energies due to Coulomb, polarization, and Van der Waals interactions as a function of separation distance with the reference *ab initio* calculations. Intermolecular Coulomb interactions and polarization interactions are shown in Figure 7.3. Intermolecular electrostatic interactions are well described at medium and long range, and a deviation at short intermolecular distances is observed, which is due to penetration errors. The van der Waals parameters were taken directly from the AMOEBA force field (Figure 7.2). Reference van der Waals energies were calculated as the difference between total intermolecular energies at the MP2 level and the Coulomb energies due to permanent multipoles, and the energies due to the polarization interaction.

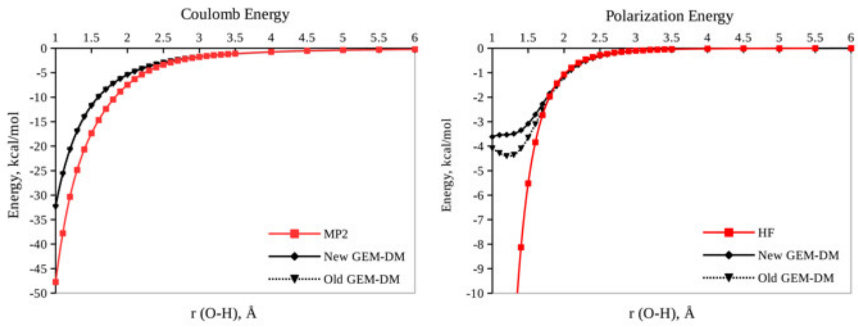

Figure 7.3 Intermolecular Coulomb interactions and polarization interactions for water dimers

7.3.1.2 Van der Waals parameter fitting

As explained above, we used the original AMOEBA van der Waals parameters as a starting point for the parametrization. Our simulation results showed that the liquid density and heat of vaporization at 298 K were overestimated (ρ = 1.062 g cm^{-3}, ΔH_{vap} = 11.22 kcal mol^{-1}) compared to the experimental data (ρ = 0.997 g cm^{-3}, ΔH_{vap} = 10.51 kcal mol^{-1}).[114] Therefore, we optimized these parameters initially to fit to the QM data in Figure 7.2 and subsequently refined them to match the experimental density and heat of vaporization at this single temperature.

Table 7.2 Van der Waals parameter, density, and heat of vaporization (a) before and (b) after scaling the vdW parameters.

Atoms	R	ϵ	ρ(g/cm^3)	ΔH_{vap} (kcal/mol)
a				
O	3.405	0.1100	1.062	11.22
H	2.655	0.0135		
b				
O	3.480	0.1077	1.005	10.53
H	2.680	0.0113		

The fitted van der Waals parameters are shown in Table 7.2. The vdW and total intermolecular energy for a series of water dimers show excellent agreement after fitting the vdW parameters (Figure 7.4). After scaling of the van der Waals parameters, MD simulations were carried out to calculate other thermodynamic and transport properties to validate these new parameters.

7.3.2 Simulation results

Molecular dynamics simulations were performed to validate the accuracy of the newly developed AMOEBA parameters. This section presents the results obtained for the calculation of a series of thermodynamic and structural properties including liquid densities, heat of vaporization, heat capacity, diffusion coefficients, and liquid structure (from radial distribution functions).

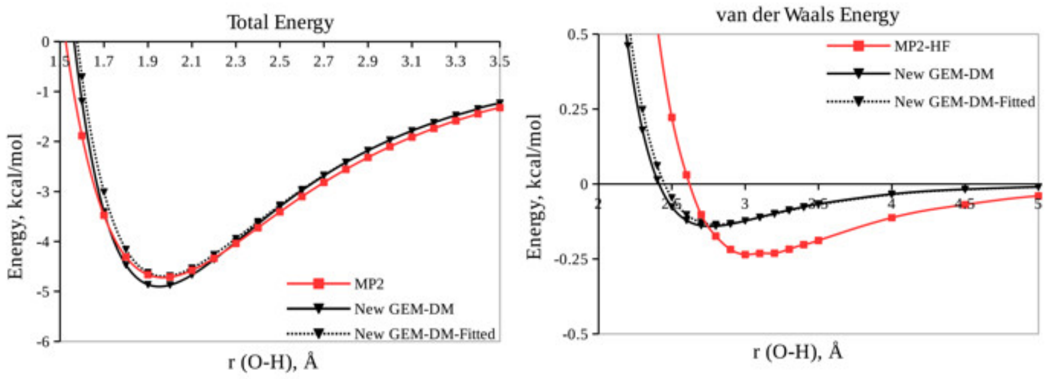

Figure 7.4 Total interaction energy and vdW interactions for water dimers

7.3.2.1 Liquid volume and density

The calculated volume is 29.76 Å^3 at T = 298 K based on the average density from the MD simulations. Using the original van der Waals parameters, the liquid density is 6.5 % higher than the experimental value at 298 K. In contrast, after the van der Waals parameters were scaled, the liquid density decreases to 1.005 g cm^{-3} which is in good agreement with the experimental result, with an error of 0.8 % as shown in Table 7.3. Liquid densities calculated with AMOEBA14 and AMOEBA03 at 298.15 K are 0.998 and 1.000 g cm^{-3} respectively.[141] Comparison of the MD simulation results with experiment for liquid densities in the range of 250-370 K is shown in Figure 7.5. Both MD simulation results and experiment show a maximum density at 280 K with a value of 1.007 and 1.000 g cm^{-3}, respectively (Figure 7.5). The mean absolute deviation (MAD) from experiment for liquid densities calculated across all temperatures (250-370 K) is

0.006 g cm^{-3} (Table 7.3), which is slightly higher than the MAD value obtained with AMOEBA14 (0.001 g cm^{-3}).[141]

Table 7.3 Mean absolute deviation (MAD) from experimental liquid-phase properties calculated by AMOEBA/GEM-DM and AMOEBA14.

Property	AMOEBA/GEM-DM	AMOEBA	Units
ρ	0.006	0.001	g cm^{-3}
ΔH_{vap}	-0.005	0.103	Kcal mol^{-1}
C_p	1.54	2.28	Cal mol^{-1} K^{-1}

7.3.2.2 Enthalpy of vaporization

The enthalpy of vaporization is calculated as an estimation of the strength of intermolecular interactions for water molecules. We used the Eq 5.6 with the assumption of ideal gas behavior to calculate the heat of vaporization. The gas-phase energy for a single water molecule at 298 K is calculated using stochastic molecular dynamics simulations (E_{gas} = 0.91 kcal mol^{-1}). The calculated enthalpy of vaporization for water with GEM-DM is 10.53 kcal mol^{-1}, which is 2 % higher than the experiment value of 10.51 kcal mol^{-1} at 298 K. By comparison, the results obtained with iAMOEBA (10.94 kcal mol^{-1}) and AMOEBA14 (10.63 kcal mol^{-1}) are slightly higher than the experimental result.[137, 141] The heat of vaporization was also calculated outside of ambient conditions and compared to experimental data (Figure 7.5). The MAD from experiment for heat of vaporization is 0.005 kcal mol^{-1} (Table 7.3).

7.3.2.3 Heat capacity

The constant-pressure-specific heat capacity is $c_p = \left(\frac{\delta H}{\delta T}\right)_p$. By using $\langle H\rangle = \langle E_{tot}\rangle + \langle PV\rangle$, in which PV is independent of temperature for liquid systems, the heat capacity at constant pressure can be calculated. A correction of 6 kcal $mol^{-1}K^{-1}$ has been suggested to be deducted from the heat capacity, which is obtained from energy fluctuations of a classical flexible model.[251] This correction is employed to compensate for the error due to the intra- and intermolecular vibrations compared to a quantum oscillator model.[252] We subtracted the quantum corrections for the high frequency vibrational modes at different temperatures[137] to calculate the heat capacity in this study. We calculated the heat capacity via the differentiation of total energy with respect to temperature. The heat capacity at 298 K and 1 atm for water is computed to be 18.75 cal $mol^{-1}K^{-1}$ which is 0.75 cal $mol^{-1}K^{-1}$ greater than the experimental value of 18.00 cal $mol^{-1}K^{-1}$. The values for heat capacity calculated with AMOEBA14 and AMOEBA03 are 20.48 and 22.36 cal $mol^{-1}K^{-1}$, respectively.[141] In the AMOEBA14 model, the heat capacity has also been used as data for the parametrization in ForceBalance.[141] In contrast, in this study, we only fitted the parameters to reproduce the experimental heat of vaporization and density at 298 K. Figure 7.5 shows the experimental and calculated heat capacity in the range of 250-370 K. The MAD from experiment for heat capacity calculated across all temperatures is 1.54 cal $mol^{-1}K^{-1}$ using AMOEBA/GEM-DM, which is lower than the MAD value using AMOEBA14 (2.28 cal $mol^{-1}K^{-1}$) (Table 7.3).

7.3.2.4 Self-diffusion coefficients

Another liquid property that can be used to validate the quality of the force fields is the self-diffusion coefficient. The self-diffusion coefficients were calculated using the Einstein relation (equation 5.7).[245] The mean square displacement (MSD(t)) in the diffusive regime was calculated from the production trajectories (the last 1.5 ns).

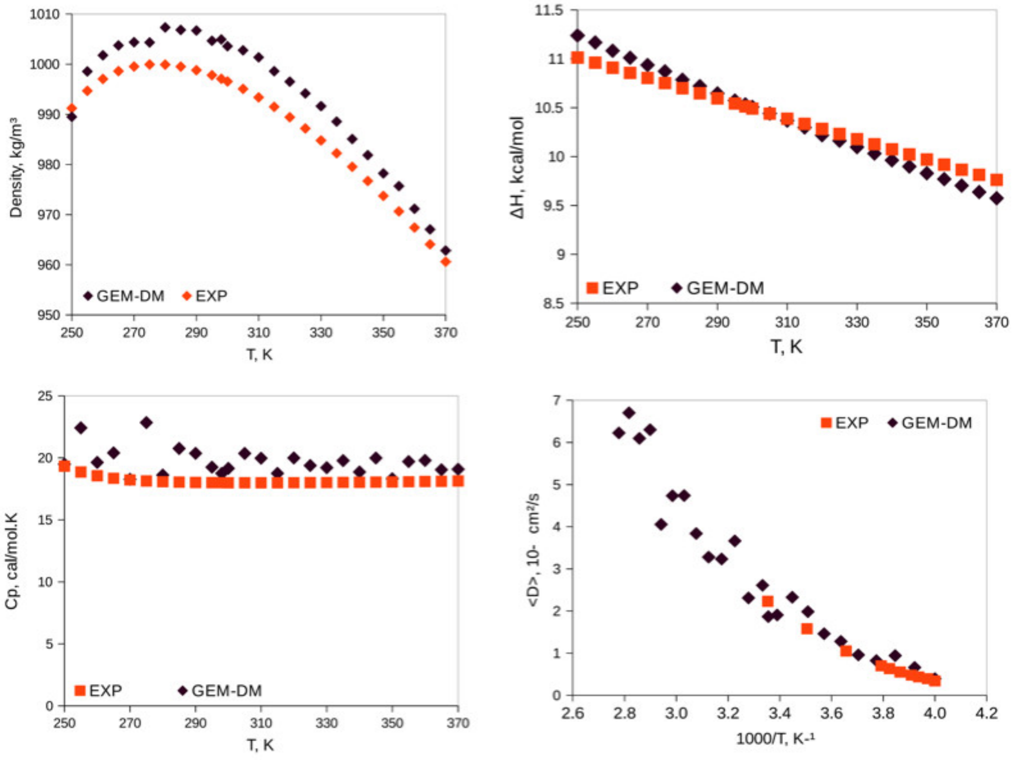

Figure 7.5 Comparison of simulation result for water density, heat of vaporization, heat capacity and diffusion coefficient with experiment between 250 and 370 K

Self-diffusion coefficients for the water molecules using AMOEBA/GEM-DM at 298 K are 1.91×10^{-5} cm^2s^{-1}, which is slightly lower than experimental data (2.29×10^{-5} cm^2s^{-1}). By comparison, the diffusion coefficients calculated with AMOEBA and iAMOEBA are 2×10^{-5} cm^2s^{-1} and 2.54×10^{-5} cm^2s^{-1} [78, 138]. The calculated diffusion coefficients for temperatures between 250 and 370 K are shown and compared to available experimental data in Figure 7.5. The trend in the self-diffusion coefficients is well correlated with the enthalpies of vaporization. At higher temperature, the present model predicts faster diffusion rate and lower heat of vaporization.

7.3.2.5 Intermolecular structure

The liquid structure of water is usually described by radial distribution functions (RDFs). These data can be compared to experimental neutron diffraction data.[106] The calculated oxygen-oxygen, oxygen–hydrogen, and hydrogen-hydrogen RDFs and experimental data are depicted in Figure 7.6. AMOEBA/GEM-DM model is in good agreement with experimental data for all RDFs. The first peak for the oxygen-oxygen RDF is located at the same distance as experiment (2.8 Å) with a height of 3.12, which is slightly higher than experimental value 2.73. The second peak, as well as the first and second troughs of the oxygen-oxygen RDF, is in excellent agreement with experiment. RDFs for oxygen-hydrogen and hydrogen-hydrogen show similar level of agreement.

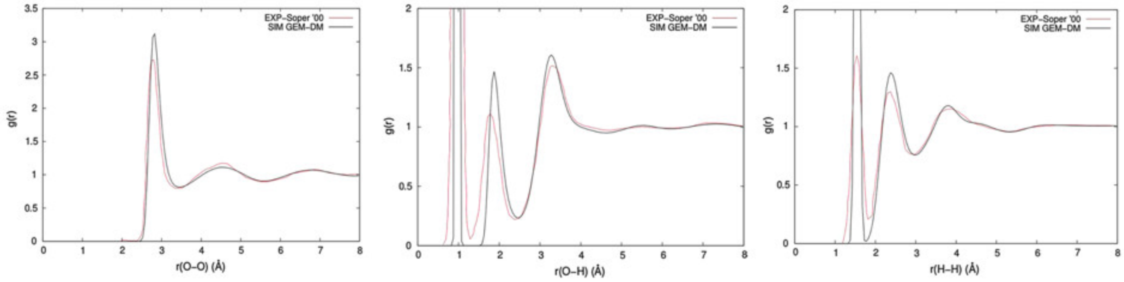

Figure 7.6 Radial Distribution function for water

7.4 Conclusions

We have presented a new AMOEBA polarizable model for water that uses GEM distributed multipoles with higher order multipoles on the oxygen and hydrogen atoms. The new parameters improved accuracy for several properties, including ρ, ΔH_{vap}, C_p and D for a range of temperatures compared to our previous parameters. This improvement is mainly due to the inclusion of dipoles and quadrupoles on the hydrogen atoms, which provides a better description of polarization and electrostatics. Although we only adjusted the van der Waals parameters to reproduce the experimental density and heat of vaporization at 298 K, the model shows very good agreement for a series of thermodynamic properties including density, heat of vaporization, heat capacity, and diffusion coefficient for a range of temperatures. Mean absolute deviations calculated for density, heat of vaporization, and heat capacity confirm the excellent agreement between experimental data and simulation results in the range of 250–370 K.

CHAPTER 8 CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

In this work, we performed MD and QM/MM studies on Fe(II)-α-KG dependent enzymes (TET2 and AlkB), and developed multipolar/polarizable force field for imidazolium- and pyrrolidinium-based ionic liquids and water. A brief introduction of all studies systems in this thesis was presented in Chapter one. In Chapter two, we extensively studied human TET2 to characterize the source of the substrate preference in this enzyme. To this end, we performed MD simulations on wild type and 10 different mutants and we uncovered that the scaffold stablished by T1372 and Y1902 is essential for iterative oxidation steps. We showed that the mutation of one of these hydrogen bonding partners can disrupt the alignment of 5hmC in the active site, resulting in stalling oxidation after the formation of 5hmC.

In Chapter three, we performed *ab initio* QM/MM studies on the active site of wild type TET2 and one of its mutant for further understanding of the mechanism for 5hmC to 5fC oxidation. We showed that the first step which is the hydrogen abstraction from hydroxyl group of 5hmC is the rate-limiting step and the second hydrogen is transferred via third transition state as a proton. The large barrier for T1372E mutant confirms our prediction from MD simulations that this mutant prevents oxidation after the formation of 5hmC. We showed that this mechanism is kinetically unfavorable due to the orientation of hydroxyl group of 5hmC. These results are also in very good agreement with the results obtained for other enzymes of this family such as AlkB.

In Chapter 4, we applied two different MD approaches to study the diffusion of O_2 molecules into AlkB active site. Here, we performed explicit ligand sampling and PMF calculations to provide a microscopic description of the O_2 delivery pathway in AlkB. We initially showed that there are two possible tunnels for oxygen diffusion based on the static structure of the enzyme using CAVER analysis. Explicit ligand sampling results show that oxygen molecules prefer one of these channels to access to the active site. The PMF calculations also confirmed the results obtained with explicit ligand sampling approach. We also used AMOEBA potential for PMF calculation to investigate the effect of multipolar-polarizable force filed on the systems that are highly polarizable. We also mutated one single residue (W178) which is experimentally shown to have an important role in opening and closing the channel gate. We showed that mutation of this single residue can hinder the access of oxygen to the active site or open new pathways for oxygen delivery.

Chapter five focused on the importance of including the multipoles and polarizability in force fields to have an accurate description of intermolecular interactions in condensed systems. Here, we developed new AMOEBA/GEM-DM force field for imidazolium-based ionic liquids. We fitted all non-bonded parameters to QM and then performed MD simulations to calculate a series of properties. Our results show very good agreement with available experimental data.

In Chapter six, we employed the same approach to develop a multipolar-polarizable force field to model a pyrrolidinium-based ionic liquid as an electrolyte in Li ion batteries. Here, we performed MD simulations on neat and Li-doped ionic liquids. We modelled our systems with and without intramolecular polarization and showed that it is important to include the intramolecular polarization to have an accurate description of non-bonded interactions.

We also developed a new AMOEBA/GEM-DM force field for water. We included the higher order multipoles (up to quadrupoles) on both oxygen and hydrogen atoms and calculated a series of properties for water at T=250-370 K and compared with experimental data. Our model shows very good agreement with experimental data.

8.2 Future work

This thesis has demonstrated various computational methods and their application in biomolecules and condensed systems studies. We believe this work will encourage more scientists to employ computational tools to aid in the understanding of complicated systems.

Although this work provides details in structure and mechanism of Fe(II)-α-KG enzymes, there are still number of challenges that need to be addressed. We studied the oxidation of 5hmC substrate to 5fC in TET2 with MD and QM/MM approach, however, we didn’t investigate the next iterative oxidation step. One can work on the oxidation of 5fC to 5caC and 5mC to 5hmC. The oxidation of 5mC to 5hmC by TET can be compared with the oxidation of 1meA to understand why 1meA doesn’t go through the iterative oxidation steps. Besides, the DNA base flipping effect on the oxidation is another challenge that we found in some of the mutants could be another potential source for their unique behavior. In addition, the mutation of the residues we introduced to be effective on active site can be addressed by both computational and experimental research.

Moreover, this work shows the importance of employing multipolar-polarizable force fields for both biomolecular systems and ionic liquids. Here, the effect of AMOEBA force field on the diffusion of oxygen molecule in AlkB is presented. This approach can be used for other polarizable biomolecules to gain an accurate description of their behavior.

The AMOEBA/GEM force field presented here can describe the properties of ionic liquids and water very well. However, there are still some shortcomings due to penetration effect that one can address them by employing a continuous description of the charge density, for instance, Gaussian Electrostatic Model.

REFERENCES

1. Tahiliani, M., Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **2009,** *324*, 930-935.

2. Kriaucionis, S.; Heintz, N., The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **2009,** *324*, 929-930.

3. Ito, S., Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **2010,** *466*, 1129-1133.

4. Pfaffeneder, T., The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Edn Engl.* **2011,** *50*, 7008-7012.

5. He, Y. F., Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **2011,** *333*, 1303-1307.

6. Ito, S., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011,** *333*, 1300-1303.

7. Kohli, R. M.; Zhang, Y., TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **2013,** *502* (7472), 472--479.

8. Bachman, M., 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* **2014,** *6*, 1049-1055.

9. Wu, H.; Zhang, Y., Charting oxidized methylcytosines at base resolution. *Nat. Struct. Mol. Biol.* **2015,** *22*, 656-661.

10. Liu, M. Y.; DeNizio, J. E.; Schutsky, E. K.; Kohli, R. M., The expanding scope and impact of epigenetic cytosine modifications. *Curr. Opin. Chem. Biol.* **2016,** *33*, 67-73.

11. Zheng, G.; Fu, Y.; He, C., Nucleic acid oxidation in DNA damage repair and epigenetics. *Chem. Rev.* **2014,** *114*, 4602-4620.

12. Hu, L., Structural insight into substrate preference for TET-mediated oxidation. *Nature* **2015,** *527*, 118-122.

13. Hu, L., Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* **2013,** *155*, 1545-1555.

14. Lu, J., A computational investigation on the substrate preference of ten-eleven-translocation 2 (TET2). *Phys. Chem. Chem. Phys.* **2016,** *18*, 4728-4738.

15. Crawford, D. J., Tet2 catalyzes stepwise 5-methylcytosine oxidation by an iterative and de novo mechanism. *J. Am. Chem. Soc.* **2016,** *138*, 730-733.

16. Liu, M. Y.; Torabifard, H.; Crawford, D. J.; DeNizio, J. E.; Cao, X. J.; Garcia, B. A.; Cisneros, G. A.; Kohli, R. M., Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine. *Nat Chem Biol* **2016**.

17. Drablos, F.; Feyzi, E.; Aas, P. A.; Vaagbo, C. B.; Kavli, B.; Bratlie, M. S.; Pena-Diaz, J.; Otterlei, M.; Slupphaug, G.; Krokan, H. E., Alkylation damage in DNA and RNA--repair mechanisms and medical significance. *DNA Repair (Amst)* **2004,** *3* (11), 1389-407.

18. Fu, D.; Calvo, J. A.; Samson, L. D., Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat Rev Cancer* **2012,** *12* (2), 104-20.

19. Hausinger, R. P., Fe(II)/α--ketoglutarate--dependent hydroxylases and related enzymes. *Crit. Rev. Biochem and Mol. Bio.* **2004,** *39*, 21-68.

20. Fang, D.; Lord, R. L.; Cisneros, G. A., Ab Initio QM/MM Calculations Show an Intersystem Crossing in the Hydrogen Abstraction Step in Dealkylation Catalyzed by AlkB. *J. Phys. Chem. B* **2013,** *117* (21), 6410-6420.

21. Fang, D.; Cisneros, G. A., Alternative pathway for the reaction catalyzed by DNA dealkylase AlkB from ab initio QM/MM calculations. *J. Chem. Theory Comput.* **2014,** *10*, 5136-5148.

22. Wang, B.; Usharani, D.; Li, C.; Shaik, S., Theory uncovers an unusual mechanism of DNA repair of a lesioned adenine by AlkB enzymes. *J Am Chem Soc* **2014,** *136* (39), 13895-901.

23. Schenk, G.; Pau, M. Y. M.; Solomon, E. I., Comparison between the geometric and electronic structures and reactivities of \FeNO\$^7$ and \FeO$_2$\$^8$ complexes: A density functional theory study. *J. Am. Chem. Soc.* **2004,** *126*, 505-515.

24. Sinnecker, S.; Svensen, N.; Barr, E. W.; Ye, S.; Bollinger, J. M., Jr.; Neese, F.; Krebs, C., Spectroscopic and computational evaluation of the structure of the high-spin Fe(IV)-oxo intermediates in taurine: alpha-ketoglutarate dioxygenase from Escherichia coli and its His99Ala ligand variant. *J Am Chem Soc* **2007,** *129* (19), 6168-79.

25. Grzyska, P. K.; Appelman, E. H.; Hausinger, R. P.; Proshlyakov, D. A., Insight into the mechanism of an iron dioxygenase by resolution of steps following the FeIV=HO species. *Proc Natl Acad Sci U S A* **2010,** *107* (9), 3982-7.

26. Eichhorn, E.; van der Ploeg, J. R.; Kertesz, M. A.; Leisinger, T., Characterization of alpha-ketoglutarate-dependent taurine dioxygenase from Escherichia coli. *J Biol Chem* **1997,** *272* (37), 23031-6.

27. Price, J. C.; Barr, E. W.; Tirupati, B.; Bollinger, J. M., Jr.; Krebs, C., The first direct characterization of a high-valent iron intermediate in the reaction of an alpha-ketoglutarate-dependent dioxygenase: a high-spin FeIV complex in taurine/alpha-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **2003,** *42* (24), 7497-508.

28. Godfrey, E.; Porro, C. S.; de Visser, S. P., Comparative quantum mechanics/molecular mechanics (QM/MM) and density functional theory calculations on the oxo-iron species of taurine/alpha-ketoglutarate dioxygenase. *J Phys Chem A* **2008,** *112* (11), 2464-8.

29. de Visser, S. P., What factors influence the ratio of C-H hydroxylation versus C=C epoxidation by a nonheme cytochrome P450 biomimetic? *J Am Chem Soc* **2006,** *128* (49), 15809-18.

30. Usharani, D.; Janardanan, D.; Shaik, S., Does the TauD enzyme always hydroxylate alkanes, while an analogous synthetic non-heme reagent always desaturates them? *J Am Chem Soc* **2011,** *133* (2), 176-9.

31. Ye, S.; Price, J. C.; Barr, E. W.; Green, M. T.; Bollinger, J. M., Jr.; Krebs, C.; Neese, F., Cryoreduction of the NO-adduct of taurine:alpha-ketoglutarate dioxygenase (TauD) yields an elusive {FeNO}(8) species. *J Am Chem Soc* **2010,** *132* (13), 4739-51.

32. Shadrina, M. S.; English, A. M.; Peslherbe, G. H., Effective simulations of gas diffusion through kinetically accessible tunnels in multisubunit proteins: O2 pathways and escape routes in T-state deoxyhemoglobin. *J Am Chem Soc* **2012,** *134* (27), 11177-84.

33. Wang, P. H.; Bruschi, M.; De Gioia, L.; Blumberger, J., Uncovering a dynamically formed substrate access tunnel in carbon monoxide dehydrogenase/acetyl-CoA synthase. *J Am Chem Soc* **2013,** *135* (25), 9493-502.

34. Di Russo, N. V.; Condurso, H. L.; Li, K.; Bruner, S. D.; Roitberg, A. E., Oxygen diffusion pathways in a cofactor-independent dioxygenase. *Chem Sci* **2015,** *6* (11), 6341-6348.

35. Elber, R., Ligand diffusion in globins: simulations versus experiment. *Curr Opin Struct Biol* **2010,** *20* (2), 162-7.

36. Pesce, A.; Nardini, M.; Dewilde, S.; Capece, L.; Marti, M. A.; Congia, S.; Salter, M. D.; Blouin, G. C.; Estrin, D. A.; Ascenzi, P.; Moens, L.; Bolognesi, M.; Olson, J. S., Ligand migration in the apolar tunnel of Cerebratulus lacteus mini-hemoglobin. *J Biol Chem* **2011,** *286* (7), 5347-58.

37. Colloc'h, N.; Gabison, L.; Monard, G.; Altarsha, M.; Chiadmi, M.; Marassio, G.; Sopkova-de Oliveira Santos, J.; El Hajji, M.; Castro, B.; Abraini, J. H.; Prange, T., Oxygen pressurized X-ray crystallography: probing the dioxygen binding site in cofactorless urate oxidase and implications for its catalytic mechanism. *Biophys J* **2008,** *95* (5), 2415-22.

38. Tomita, A.; Kreutzer, U.; Adachi, S.; Koshihara, S. Y.; Jue, T., 'It's hollow': the function of pores within myoglobin. *J Exp Biol* **2010,** *213* (Pt 16), 2748-54.

39. Ruscio, J. Z.; Kumar, D.; Shukla, M.; Prisant, M. G.; Murali, T. M.; Onufriev, A. V., Atomic level computational identification of ligand migration pathways between solvent and binding site in myoglobin. *Proc Natl Acad Sci U S A* **2008,** *105* (27), 9204-9.

40. Bocahut, A.; Bernad, S.; Sebban, P.; Sacquin-Mora, S., Relating the diffusion of small ligands in human neuroglobin to its structural and mechanical properties. *J Phys Chem B* **2009,** *113* (50), 16257-67.

41. Orlowski, S.; Nowak, W., Locally enhanced sampling molecular dynamics study of the dioxygen transport in human cytoglobin. *J Mol Model* **2007,** *13* (6-7), 715-23.

42. Javanainen, M.; Vattulainen, I.; Monticelli, L., On Atomistic Models for Molecular Oxygen. *J Phys Chem B* **2017,** *121* (3), 518-528.

43. Dewage, S. W.; Cisneros, G. A., Computational analysis of ammonia transfer along two intramolecular tunnels in Staphylococcus aureus glutamine-dependent amidotransferase (GatCAB). *J. Phys. Chem. B* **2015,** *119*, 3669-3677.

44. Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S., Accelerating molecular dynamic simulation on graphics processing units. *J Comput Chem* **2009,** *30* (6), 864-72.

45. Yu, B.; Edstrom, W. C.; Benach, J.; Hamuro, Y.; Weber, P. C.; Gibney, B. R.; Hunt, J. F., Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* **2006,** *439*, 879-884.

46. Yu, B.; Hunt, J. F., Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. *Proc Natl Acad Sci U S A* **2009,** *106* (34), 14315-20.

47. Wilkes, J. S.; Levisky, J. A.; Wilson, R. A.; Hussey, C. L., Dialkylimidazolium chloroaluminate melts: a new class of room-temperature ionic liquids for electrochemistry, spectroscopy and synthesis. *Inorg Chem* **1982,** *21* (3), 1263-1264.

48. Armand, M.; Endres, F.; MacFarlane, D. R.; Ohno, H.; Scrosati, B., Ionic-liquid materials for the electrochemical challenges of the future. *Nat. Mater.* **2009,** *8* (8), 621.

49. Garcia, B.; Lavallée, S.; Perron, G.; Michot, C.; Armand, M., Room Temperature Molten Salts as Lithium Battery Electrolyte. *Electrochim. Acta* **2004,** *49* (26), 4583.

50. Olivier-Bourbigou, H.; Magna, L.; Morvan, D., Ionic Liquids and Catalysis: Recent Progress From Knowledge to Applications. *Appl. Catal., A* **2010,** *373* (1–2), 1.

51. Zhang, S. J.; Sun, N.; He, X. Z.; Lu, X. M.; Zhang, X. P., Physical properties of ionic liquids: Database and evaluation. *J. Phys. Chem. Ref. Data* **2006,** *35* (4), 1475.

52. Lopes, J. N. C.; Deschamps, J.; Padua, A. A. H., Modeling ionic liquids using a systematic all-atom force field. *J. Phys. Chem. B* **2004,** *108* (6), 2038.

53. Lopes, J. N. C.; Padua, A. A. H., Molecular force field for ionic liquids composed of triflate or bistriflylimide anions. *J. Phys. Chem. B* **2004,** *108* (43), 16893.

54. Cadena, C.; Maginn, E. J., Molecular simulation study of some thermophysical and transport properties of triazolium-based ionic liquids. *J. Phys. Chem. B* **2006,** *110* (36), 18026.

55. Morrow, T. I.; Maginn, E. J., Molecular Dynamics Study of the Ionic Liquid 1-n-Butyl-3-methylimidazolium Hexafluorophosphate. *J. Phys. Chem. B* **2002,** *106*, 12807.

56. Hooper, J. B.; Starovoytov, O. N.; Borodin, O.; Bedrov, D.; Smith, G. D., Molecular dynamics simulation studies of the influence of imidazolium structure on the properties of imidazolium/azide ionic liquids. *J. Chem. Phys.* **2012,** *136* (19), 194506.

57. Borodin, O.; Smith, G. D., Structure and Dynamics of N--Methyl--N--propylpyrrolidinium Bis(trifluoromethanesulfonyl)imide Ionic Liquid from Molecular Dynamics Simulations. *J. Phys. Chem. B* **2006,** *110*, 14426-14435.

58. Borodin, O.; Smith, G. D.; Geiculescu, O.; Creager, S. E.; Hallac, B.; DesMarteau, D., Li+ transport in lithium sulfonylimide-oligo(ethylene oxide) ionic liquids and oligo(ethylene oxide) doped with LiTFSI. *J. Phys. Chem. B* **2006,** *110* (47), 24266.

59. Borodin, O., Relation between heat of vaporization, ion transport, molar volume, and cation-anion binding energy for ionic liquids. *J. Phys. Chem. B* **2009,** *113* (36), 12353.

60. Bühl, M.; Chaumont, A.; Schurhammer, R.; Wipff, G., Ab initio molecular dynamics of liquid 1,3-dimethylimidazolium chloride. *J. Phys. Chem. B* **2005,** *109* (39), 18591.

61. deAndrade, J.; B\"oes, E. S.; Stassen, H., Computational Study of Room Temperature Molten Salts Composed by 1--Alkyl--3--methylimidazolium Cations---Force--Field Proposal and Validation. *J. Phys. Chem. B* **2002,** *106*, 13344-13351.

62. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005,** *26* (16), 1668.

63. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995,** *91* (1–3), 43.

64. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J., GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005,** *26* (16), 1701.

65. Plimpton, S., Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995,** *117*, 1.

66. Todorov, I. T.; Smith, W., DL_POLY_3: the CCP5 national UK code for molecular-dynamics simulations. *Philos. Trans.: Math., Phys. Eng. Sci.* **2004,** *362* (1822), 1835.

67. Ponder, J. W., *TINKER, Software Tools for Molecular Design, Version 4.2: the most updated version for the TINKER program can be obtained from J.W. Ponder's WWW site at* http://dasher.wustl.edu/tinker. Washington University: 2004.

68. Schuler, L. D.; Daura, X.; Van Gunsteren, W. F., An Improved GROMOS96 Force Field for Aliphatic Hydrocarbons in the Condensed Phase. *J. Comput. Chem.* **2001,** *22* (11), 1205.

69. McDonald, N. A.; Jorgensen, W. L., Development of an All-Atom Force Field for Heterocycles. Properties of Liquid Pyrrole, Furan, Diazoles, and Oxazoles. *J. Phys. Chem. B* **1998,** *102* (41), 8049.

70. Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996,** *118*, 11225.

71. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for The Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995,** *117*, 5179-5197.

72. Borodin, O.; Smith, G. D., Development of many-body polarizable force fields for Li-battery applications: 2. LiTFSI-doped Oligoether, polyether, and carbonate-based electrolytes. *J. Phys. Chem. B* **2006,** *110* (12), 6293.

73. Borodin, O.; Smith, G. D., Development of many-body polarizable force fields for Li-battery components: 1. Ether, alkane, and carbonate-based solvents. *J. Phys. Chem. B* **2006,** *110* (12), 6279.

74. Borodin, O., Polarizable Force Field Development and Molecular Dynamics Simulations of Ionic Liquids. *J. Phys. Chem. B* **2009,** *113*, 11463-11478.

75. Stone, A. J., Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.* **1981,** *83*, 233-239.

76. Starovoytov, O. N.; Torabifard, H.; Cisneros, G. A., Development of AMOEBA force field for 1,3-dimethylimidazolium based ionic liquids. *J Phys Chem B* **2014,** *118* (25), 7156-66.

77. Torabifard, H.; Starovoytov, O. N.; Ren, P.; Cisneros, G. A., Development of an AMOEBA water model using GEM distributed multipoles. *Theor Chem Acc* **2015,** *134* (8), 101.

78. Ren, P.; Ponder, J. W., Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003,** *107*, 5933-5947.

79. Cisneros, G. A., Application of Gaussian Electrostatic Model (GEM) Distributed Multipoles in the AMOEBA Force Field. *J. Chem. Theory Comput* **2012,** *8*, 5072.

80. Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P., Anisotropic, polarizable molecular mechanics studies of inter--, intra--molecular interactios, and ligand--macromolecule complexes. A bottom--up strategy. *J. Chem. Theo. Comp.* **2007,** *3*, 1960-1986.

81. Hermida-Ramon, J. M.; Brdarski, S.; Karlstrom, G.; Berg, U., Inter- and intramolecular potential for the N-formylglycinamide-water system. A comparison between theoretical modeling and empirical force fields. *J. Comput. Chem.* **2003,** *24* (2), 161.

82. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N.; Johnson, M.

E.; Head-Gordon, T., Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010,** *114* (8), 2549.

83. Kosov, D. S.; Popelier, P. L. A., Atomic partitioning of molecular electrostatic potentials. *The Journal of Physical Chemistry A* **2000,** *104* (31), 7339--7345.

84. Popelier, P. L. A.; Joubert, L.; Kosov, D. S., Convergence of the Electrostatic Interaction Based on Topological Atoms. *J. Phys. Chem. A* **2001,** *105* (35), 8254.

85. Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J., Evaluation of charge penetration between distributed multipolar expansions. *J. Chem. Phys.* **2000,** *112*, 7300-7306.

86. Cisneros, G. A.; Tholander, S. N.; Parisel, O.; Darden, T. A.; Elking, D.; Perera, L.; Piquemal, J. P., Simple Formulas for Improved Point-Charge Electrostatics in Classical Force Fields and Hybrid Quantum Mechanical/Molecular Mechanical Embedding. *Int. J. Quantum Chem.* **2008,** *108* (11), 1905.

87. Piquemal, J.-P.; Cisneros, G. A. e.; Reinhardt, P.; Gresh, N.; Darden, T. A., Towards a force field based on density fitting. *J. Chem. Phys.* **2006,** *124*, 104101.

88. Piquemal, J.-P.; Gresh, N.; Giessner--Prettre, C., Improved formulas for the calculation of the electrostatic contribution to the intermolecular interaction energy from multipolar expansion of the electronic distribution. *J. Phys. Chem. A* **2003,** *107*, 10353-10359.

89. Duke, R. E.; Starovoytov, O. N.; Piquemal, J.-P.; Cisneros, G. A., GEM*: A molecular electronic density--based force field for molecular dynamics simulations. *J. Chem. Theo. Comp.* **2014,** *10*, 1361-1365.

90. Armand, M.; Tarascon, J.-M., Building better batteries. *Nature* **2008,** *451* (7179), 652--657.

91. Markevich, E.; Baranchugov, V.; Aurbach, D., On the possibility of using ionic liquids as electrolyte solutions for rechargeable 5V Li ion batteries. *Electrochemistry communications* **2006,** *8* (8), 1331--1334.

92. Zhao, L.; Hu, Y.-S.; Li, H.; Wang, Z.; Chen, L., Porous Li4Ti5O12 Coated with N-Doped Carbon from Ionic Liquids for Li-Ion Batteries. *Advanced Materials* **2011,** *23* (11), 1385--1388.

93. Canongia Lopes, J. N.; Deschamps, J.; Pádua, A. A., Modeling ionic liquids using a systematic all-atom force field. *J. Phys. Chem. B* **2004,** *108* (6), 2038--2047.

94. Canongia Lopes, J. N.; Pádua, A. A., Molecular force field for ionic liquids composed of triflate or bistriflylimide anions. *J. Phys. Chem. B* **2004,** *108* (43), 16893--16898.

95. Morrow, T. I.; Maginn, E. J., Molecular Dynamics Study of the Ionic Liquid 1--n--Butyl--3--methylimidazolium Hexafluorophosphate. *J. Phys. Chem. B* **2002,** *106*, 12807-12813.

96. Borodin, O.; Smith, G. D.; Henderson, W., Li\$^+\$ Cation Environment, Transport, and Mechanical Properties of the LiTFSI Doped N-Methyl-N-alkylpyrrolidinium\$^+\$TFSI\$^-\$ Ionic Liquids. *J. Phys. Chem. B* **2006,** *110*, 16879-16886.

97. Buhl, M.; Chaumont, A.; Schurhammer, R.; Wipff, G., Ab Initio Molecular Dynamics of Liquid 1,3-Dimethylimidazolium Chloride. *J. Phys. Chem. B* **2005,** *109* (39), 18591-18599.

98. de Andrade, J.; Boes, E. S.; Stassen, H., Computational study of room temperature molten salts composed by 1-alkyl-3-methylimidazolium cations-force-field proposal and validation. *J. Phys. Chem. B* **2002,** *106* (51), 13344.

99. Son, C. Y.; McDaniel, J. G.; Schmidt, J. R.; Cui, Q.; Yethiraj, A., First Principles United Atom Force Field for the Ionic Liquid [BMIM][BF4]: An Alternative to Charge Scaling. *The Journal of Physical Chemistry B* **2016.**

100. Popelier, P.; Joubert, L.; Kosov, D., Convergence of the electrostatic interaction based on topological atoms. *J. Phys. Chem. A* **2001,** *105* (35), 8254--8261.

101. Corongiu, G.; Clementi, E., Erratum: Liquid water with an abinitio potential: X-ray and neutron scattering from 238 to 368 K [J. Chem. Phys. 97, 2030 (1992)]. *J. Chem. Phys.* **1992,** *97* (11), 8818--8818.

102. Narten, A. H.; Levy, H. A., Liquid Water: Molecular Correlation Functions from X-Ray Diffraction. *J. Chem. Phys.* **1971,** *55* (5), 2263--2269.

103. Neilson, G. W.; Enderby, J. E., *Water and aqueous solutions: proceedings of the Thirty-seventh Symposium of the Colston Research Society, held in the University of Bristol in April 1985*. Taylor \& Francis: 1986; Vol. 37.

104. Neilson, G. W.; Enderby, J. E., Aqueous solutions and neutron scattering. *J. Phys. Chem.* **1996,** *100* (4), 1317--1322.

105. Sellberg, J. A.; Huang, C.; McQueen, T. A.; Loh, N. D.; Laksmono, H.; Schlesinger, D.; Sierra, R. G.; Nordlund, D.; Hampton, C. Y.; Starodub, D.; others, Ultrafast X-ray probing of water structure below the homogeneous ice nucleation temperature. *Nature* **2014,** *510* (7505), 381--384.

106. Soper, A. K., The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chem. Phys.* **2000,** *258* (2-3), 121-137.

107. Sorenson, J. M.; Hura, G.; Glaeser, R. M.; Head-Gordon, T., What can x-ray scattering tell us about the radial distribution functions of water? *J. Chem. Phys.* **2000,** *113* (20), 9149--9161.

108. Fuentes-Azcatl, R.; Alejandre, J., Non-Polarizable Force Field of Water Based on the Dielectric Constant: TIP4P/ε. *J. Phys. Chem. B* **2014,** *118* (5), 1263--1272.

109. Mark, P.; Nilsson, L., Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* **2001,** *105* (43), 9954--9960.

110. Paesani, F.; Voth, G. A., The Properties of Water: Insights from Quantum Simulations†. *J. Phys. Chem. B* **2009,** *113* (17), 5702--5719.

111. Guillot, B., A reappraisal of what we have learnt during three decades of computer simulations on water. *Journal of Molecular Liquids* **2002,** *101* (1), 219--260.

112. Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragones, J. L., What ice can teach us about water interactions: a critical comparison of the performance of different water models. *Faraday discussions* **2009,** *141*, 251--276.

113. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926--935.

114. Vega, C.; Abascal, J. L. F., Simulating water with rigid non-polarizable models: a general perspective. *Physical Chemistry Chemical Physics* **2011,** *13* (44), 19663--19688.

115. Kiss, P. T.; Bertsyk, P.; Baranyai, A., Testing recent charge-on-spring type polarizable water models. I. Melting temperature and ice properties. *J. Chem. Phys.* **2012,** *137* (19), 194102.

116. Vega, C.; Sanz, E.; Abascal, J. L. F., The melting temperature of the most common models of water. *J. Chem. Phys.* **2005,** *122* (11), 114507--114507.

117. Barnes, P.; Finney, J. L.; Nicholas, J. D.; Quinn, J. E., Cooperative effects in simulated water. *Nature* **1979,** *282*, 459--464.

118. Stone, A. J., *The theory of intermolecular forces*. Oxford University Press: 2000.

119. Kairys, V.; Jensen, J. H., Evaluation of the charge penetration energy between non-orthogonal molecular orbitals using the Spherical Gaussian Overlap approximation. *Chem. Phys. Lett.* **1999,** *315* (1-2), 140-144.

120. Stone, A. J., Electrostatic Damping Functions and the Penetration Energy. *J. Phys. Chem. A* **2011,** *115* (25), 7017-7027.

121. Wang, B.; Truhlar, D. G., Including Charge Penetration Effects in Molecular Modeling. *J. Chem. Theo. Comp.* **2010,** *6* (11), 3330-3342.

122. Halgren, T. A.; Damm, W., Polarizable force fields. *Current opinion in structural biology* **2001,** *11* (2), 236--242.

123. Rick, S. W.; Stuart, S. J., Potentials and algorithms for incorporating polarizability in computer simulations. *Reviews in computational chemistry* **2002,** *18*, 89--146.

124. Caldwell, J. W.; Kollman, P. A., Structure and properties of neat liquids using nonadditive molecular dynamics: water, methanol, and N-methylacetamide. *The Journal of Physical Chemistry* **1995,** *99* (16), 6208--6219.

125. Caldwell, J.; Dang, L. X.; Kollman, P. A., Implementation of nonadditive intermolecular potentials by use of molecular dynamics: development of a water-water potential and water-ion cluster interactions. *J. Am. Chem. Soc.* **1990,** *112*, 9144-9147.

126. Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D., An effective fragment method for modeling solvent effects in quantum mechanical calculations. *J. Chem. Phys.* **1996,** *105*, 1968-1986.

127. Xie, W.; Gao, J., Design of a next generation force field: the X-POL potential. *Journal of chemical theory and computation* **2007,** *3* (6), 1890--1900.

128. Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J., X-Pol potential: an electronic structure-based force field for molecular dynamics simulation of a solvated protein in water. *Journal of chemical theory and computation* **2009,** *5* (3), 459--467.

129. Hermida-Ramón, J. M.; Brdarski, S.; Karlström, G.; Berg, U., setminusInter-and intramolecular potential for the N-formylglycinamide-water system. A comparison between theoretical modeling and empirical force fields. *J. Comput. Chem.* **2003,** *24* (2), 161--176.

130. Popelier, P. L. A.; Kosov, D. S., Atom--atom partitioning of molecular coulomb energy. *J. Chem. Phys.* **2001,** *114*, 6539-6547.

131. Popelier, P. L. A.; Hall, P., *Atoms in molecules: an introduction*. Prentice Hall London: 2000.

132. Babin, V.; Leforestier, C.; Paesani, F., Development of a "First Principles" Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. *Journal of Chemical Theory and Computation* **2013,** *9* (12), 5395--5403.

133. Babin, V.; Medders, G. R.; Paesani, F., Development of a "first principles" water potential with flexible monomers. II: trimer potential energy surface, third virial coefficient, and small clusters. *Journal of Chemical Theory and Computation* **2014,** *10* (4), 1599--1607.

134. Medders, G. R.; Babin, V.; Paesani, F., Development of a "First-Principles" Water Potential with Flexible Monomers. III. Liquid Phase Properties. *Journal of Chemical Theory and Computation* **2014,** *10* (8), 2906-2910.

135. Ren, P.; Ponder, J. W., Temperature and pressure dependence of the AMOEBA water model. *J. Phys. Chem. B* **2004,** *108* (35), 13427--13437.

136. Thole, B. T., Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981,** *59*, 341-350.

137. Wang, L.-P.; Head-Gordon, T. L.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Mart\'i-nez, T. J.; Pande, V. S., Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B* **2013**.

138. Wang, L. P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S., Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B* **2013,** *117*, 9956.

139. Wang, L. P., ForceBalance: Systematic force field optimization. 2013.

140. Burger, S. K.; Cisneros, G. A., Efficient optimization of van der Waals parameters from bulk properties. *J. Comput. Chem.* **2013,** *34* (27), 2313--2319.

141. Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W., Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *The Journal of Physical Chemistry B* **2015**.

142. Cisneros, G. A. e.; Piquemal, J.-P.; Darden, T. A., Intermolecular electrostatic energies using density fitting. *J. Chem. Phys.* **2005,** *123*, 044109.

143. Cisneros, G. A. e.; Piquemal, J.-P.; Darden, T. A., Generalization of the Gaussian Electrostatic Model: extension to arbitrary angular momentum, distributed multipoles and computational speedup with reciprocal space methods. *J. Chem. Phys.* **2006,** *125*, 184101.

144. Chen, V. B., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66*, 12-21.

145. Schafmeister, C. E. A. F.; Ross, W. S.; Romanovski, V., The Leap Module of AMBER. 1995.

146. Case, D. A.; T. E. Cheatham, I.; Darden, T. A.; Gohlke, H.; Luo, R.; K. M. Merz, J.; Onufirev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005,** *26*, 1668-1688.

147. Dolinsky, T. J., PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007,** *35*, W522-W525.

148. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A., PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004,** *32*, W665-W667.

149. Essmann, U., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103*, 8577-8593.

150. Roe, D. R.; Cheatham Iii, T. E., PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013,** *9*, 3084-3095.

151. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graph.* **1996,** *14*, 33-38.

152. Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W., Revealing Noncovalent Interactions. *J. Am. Chem. Soc.* **2010,** *132* (18), 6498-6506.

153. Contreras-Garcia, J., NCIPLOT: a program for plotting non-covalent interaction regions. *J. Chem. Theory Comput.* **2011,** *7*, 625-632.

154. Graham, S. E.; Syeda, F.; Cisneros, G. A., Computational prediction of residues involved in fidelity checking for DNA synthesis in DNA polymerase I. *Biochemistry* **2012,** *51*, 2569-2578.

155. Elias, A. A.; Cisneros, G. A., Computational study of putative residues involved in DNA synthesis fidelity checking in Thermus aquaticus DNA polymerase I. *Adv. Protein Chem. Struct. Biol.* **2014,** *96*, 39-75.

156. Cui, Q.; Karplus, M.; Valerie, D., *``Catalysis and specificity in enzymes: A study of triosephosphate isomerase and comparison with methyl glyoxal synthase'' in Advances in protein chemistry*. Advanced Protein Chemistry: 2003; Vol. 66, p 315-372.

157. Marti, S.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J., Preorganization and reorganization as related factors in enzyme catalysis: The chorismate mutase case. *Chem. Eur. J.* **2003,** *9*, 984-991.

158. Senn, H. M.; O'Hagan, D.; Thiel, W., Insight into Enzymatic C--F Bond Formation from QM and QM/MM Calculations. *J. Am. Chem. Soc.* **2005,** *127*, 13643-13655.

159. Cisneros, G. A.; Perera, L.; Schaaper, R. M.; Pedersen, L. C.; London, R. E.; Pedersen, L. G.; Darden, T. A., Reaction Mechanism of the ϵ Subunit of \em E. coli DNA

Polymerase III: Insights into Active Site Metal Coordination and Catalytically Significant Residues. *J. Am. Chem. Soc.* **2009,** *131*, 1550-1556.

160. Uebele, V. N.; Nuss, C. E.; Fox, S. V.; Garson, S. L.; Cristescu, R.; Doran, S. M.; Kraus, R. L.; Santarelli, V. P.; Li, Y.; Barrow, J. C.; Yang, Z. Q.; Schlegel, K. A.; Rittle, K. E.; Reger, T. S.; Bednar, R. A.; Lemaire, W.; Mullen, F. A.; Ballard, J. E.; Tang, C.; Dai, G.; McManus, O. B.; Koblan, K. S.; Renger, J. J., Positive allosteric interaction of structurally diverse T-type calcium channel antagonists. *Cell Biochem Biophys* **2009,** *55* (2), 81-93.

161. Kratz, E. G.; Walker, A. R.; Lagardere, L.; Lipparini, F.; Piquemal, J. P.; Andres Cisneros, G., LICHEM: A QM/MM program for simulations with multipolar and polarizable force fields. *J Comput Chem* **2016,** *37* (11), 1019-29.

162. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., *Gaussian 09*. 2009.

163. Ponder, J. W., TINKER, Software Tools for Molecular Design, version 7.0. 2015.

164. Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A., Conversion of 5-methylcytosine to 5-

hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **2009,** *324* (5929), 930-5.

165. Ye, S.; Neese, F., Nonheme oxo-iron(IV) intermediates form an oxyl radical upon approaching the C-H bond activation transition state. *Proc Natl Acad Sci U S A* **2011,** *108* (4), 1228-33.

166. Chai, J. D.; Head-Gordon, M., Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys Chem Chem Phys* **2008,** *10* (44), 6615-20.

167. Chai, J. D.; Head-Gordon, M., Systematic optimization of long-range corrected hybrid density functionals. *J Chem Phys* **2008,** *128* (8), 084106.

168. Lonsdale, R.; Harvey, J. N.; Mulholland, A. J., Compound I reactivity defines alkene oxidation selectivity in cytochrome P450cam. *J Phys Chem B* **2010,** *114* (2), 1156-62.

169. Fang, D.; Chaudret, R.; Piquemal, J.-P.; Cisneros, G. A. e., Toward a Deeper Understanding of Enzyme Reactions Using the Coupled ELF/NCI Analysis: Application to DNA Repair Enzymes. *J. Chem. Theo. Comp.* **2013,** *9* (5), 2156-2160.

170. Liu, H.; Lu, Z.; Cisneros, G. A. e.; Yang, W., Parallel iterative reaction path optimization in \em ab initio quantum mechanical/molecular mechanical modeling of enzyme reactions. *J. Chem. Phys.* **2004,** *121*, 697-706.

171. Zhang, Y.; Liu, H.; Yang, W., Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.* **2000,** *112*, 3483-3491.

172. Parks, J. M.; Hu, H.; Cohen, A. J.; Yang, W., A pseudobond parametrization for improved electrostatics in quantum mechanical/molecular mechanical simulations of enzymes. *J. Chem. Phys.* **2008,** *129*, 154106.

173. Henkelman, G.; Jonsson, H., Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **2000,** *113*, 9978-9985.

174. Jonsson, H.; Mills, G.; Jacobsen, K. W., *``Nudged Elastic Band Method'', in Classical and quantum dynamics in condensed phase simulations*. World Scientific: 1998; p 387-404.

175. Becke, A. D.; Edgecombe, K. E., A simple measure of electron localization in atomic and molecular systems. *J. Chem. Phys.* **1990,** *92*, 5397-5403.

176. Piquemal, J.-P.; Pilme, J.; Parisel, O.; Gerard, H.; Fourre, I.; Berges, J.; Gourlaouen, C.; Lande, A. D. L.; Severen, M. C. V.; Silvi, B., What can be learnt on biologically relevant systems from the topological analysis of the electron localization function? *Int. J. Quantum Chem.* **2008,** *108*, 1951-1969.

177. Silvi, B.; Savin, A., Classification of chemical bonds based on topological analysis of electron localization functions. *Nature* **1994,** *371*, 683-686.

178. Noury, S. e.; Krokidis, X. e.; Fuster, F.; Silvi, B., Computational tools for the electron localization function topological analysis. *Comp. \& Chem.* **1999,** *23*, 597-604.

179. Koivisto, P.; Duncan, T.; Lindahl, T.; Sedgwick, B., Minimal methylated substrate and extended substrate range of Escherichia coli AlkB protein, a 1-methyladenine-DNA dioxygenase. *J Biol Chem* **2003,** *278* (45), 44348-54.

180. Cisneros, G. A.; Liu, H.; Zhang, Y.; Yang, W., Ab--initio QM/MM study shows there is no general acid in the reaction catalyzed by 4--Oxalocrotonate Tautomerase. *J. Am. Chem. Soc.* **2003,** *125*, 10348-10393.

181. Cisneros, G. A. e.; Wang, M.; Silinski, P.; Fitzgerald, M. C.; Yang, W., The protein backbone makes important contributions to 4-Oxalocrotonate Tautomerase enzyme catalysis: Understanding from theory and experiment. *Biochemistry* **2004,** *43*, 6885-6892.

182. Li, G.; Cui, Q., What is so special about Arg 55 in the catalysis of cyclophilin A? insights from hybrid QM/MM simulations. *J Am Chem Soc* **2003,** *125* (49), 15028-38.

183. Cisneros, G. A.; Perera, L.; Garc\ia--D\iaz, M.; Bebenek, K.; Kunkel, T. A.; Pedersen, L. G., Catalytic mechanism of human DNA polymerase λ with Mg$^2+$ and Mn$^2+$ from ab initio quantum mechanical/molecular mechanical studies. *DNA Rep.* **2008,** *7*, 1824-1834.

184. Hu, L.; Li, Z.; Cheng, J.; Rao, Q.; Gong, W.; Liu, M.; Shi, Y. G.; Zhu, J.; Wang, P.; Xu, Y., Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* **2013,** *155* (7), 1545-55.

185. Notredame, C.; Higgins, D. G.; Heringa, J., T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **2000,** *302* (1), 205-17.

186. Case, D.; Darden, T.; Cheatham III, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K.; others, AMBER 14. *University of California, San Francisco* **2014,** *1* (2), 3.

187. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006,** *65* (3), 712-25.

188. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W., Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **1992,** *32* (5), 523-35.

189. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984,** *81* (8), 3684.

190. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., FEP. *J. Comp. Phys.* **1977,** *23*, 327.

191. Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T. A.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103*, 8577-8593.

192. Pavelka, A.; Sebestova, E.; Kozlikova, B.; Brezovsky, J.; Sochor, J.; Damborsky, J., CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules. *IEEE/ACM Trans Comput Biol Bioinform* **2016,** *13* (3), 505-17.

193. Schrödinger, L., The PyMOL Molecular Graphics System. **2010**.

194. G. M. Torrie and J. P. Valleau, *J. Comput. Phys.,* **1977,** *23*, 187-199.

195. Banci, L.; Schroder, S.; Kollman, P. A., Molecular dynamics characterization of the active cavity of carboxypeptidase A and some of its inhibitor adducts. *Proteins* **1992,** *13* (4), 288-305.

196. Kollmann, C., [Results of a study of quality control of diagnostic ultrasound equipment]. *Ultraschall Med* **1995,** *16* (5), 206-9.

197. Carteaux, J. P.; Gast, A.; Tschopp, T. B.; Roux, S., Activated clotting time as an appropriate test to compare heparin and direct thrombin inhibitors such as hirudin or Ro 46-6240 in experimental arterial thrombosis. *Circulation* **1995,** *91* (5), 1568-74.

198. Petrek, M.; Otyepka, M.; Banas, P.; Kosinova, P.; Koca, J.; Damborsky, J., CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics* **2006,** *7*, 316.

199. Efron, B., Breakthroughs in Statistics. *Springer* **1992**, 569-593.

200. Efron, B., The jackknife, the bootstrap and other resampling plans. *SIAM* **1982**.

201. A. Grossfield, WHAM: the weighted histogram analysis method. **2012,** *2*, 06.

202. Ren, Y.; Tsui, H. T.; Poon, R. T.; Ng, I. O.; Li, Z.; Chen, Y.; Jiang, G.; Lau, C.; Yu, W. C.; Bacher, M.; Fan, S. T., Macrophage migration inhibitory factor: roles in regulating tumor cell

migration and expression of angiogenic factors in hepatocellular carcinoma. *Int J Cancer* **2003,** *107* (1), 22-9.

203. G. A. Kaminski, H. A. S., B. J. Berne and R. A. Friesner, *J. Phys. Chem. A* **2004,** *108*, 621-627.

204. Shi, Y.; Wu, C.; Ponder, J. W.; Ren, P., Multipole electrostatics in hydration free energy calculations. *J Comput Chem* **2011,** *32* (5), 967-77.

205. Ren, P.; Ponder, J. W., Consistent treatment of inter-and intramolecular polarization in molecular mechanics calculations. *Journal of computational chemistry* **2002,** *23* (16), 1497--1506.

206. Jiao, D.; Zhang, J.; Duke, R. E.; Li, G.; Schnieders, M. J.; Ren, P., Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential. *J Comput Chem* **2009,** *30* (11), 1701-11.

207. Patel, S.; Davis, J. E.; Bauer, B. A., Exploring ion permeation energetics in gramicidin A using polarizable charge equilibration force fields. *J Am Chem Soc* **2009,** *131* (39), 13890-1.

208. Ando, K., *J. Chem. Phys.* **2001,** *115*, 5228-5237.

209. III, S. P. a. C. L. B., *Mol. Simul.* **2006,** *32*, 231-249.

210. Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D., Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. *J Chem Theory Comput* **2005,** *1* (1), 153-68.

211. Lamoureux, G.; Alexander D. MacKerell, Jr.; Roux, B., A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003,** *119*, 5185-5197.

212. J. Applequist, J. R. C. a. K.-K. F., *J. Am. Chem. Soc* **1972,** *94*, 2952-2960.

213. Laury, M. L.; Wang, L. P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W., Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J Phys Chem B* **2015,** *119* (29), 9423-37.

214. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P., The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins. *J Chem Theory Comput* **2013,** *9* (9), 4046-4063.

215. Tu, Y. J.; Allen, M. J.; Cisneros, G. A., Simulations of the water exchange dynamics of lanthanide ions in 1-ethyl-3-methylimidazolium ethyl sulfate ([EMIm][EtSO4]) and water. *Phys Chem Chem Phys* **2016,** *18* (44), 30323-30333.

216. Ren, P.; Wu, C.; Ponder, J. W., Polarizable atomic multipole-based molecular mechanics for organic molecules. *Journal of chemical theory and computation* **2011,** *7* (10), 3143--3161.

217. Darden, T. A.; York, D.; Pedersen, L. G., Particle mesh Ewald: an Nlog(N) method for Ewald sums. *J. Chem. Phys.* **1993,** *98*, 10089-10092.

218. Nam, K.; Gao, J.; York, D. M., An efficient linear--scaling Ewald method for long--range electrostatic interactions in combined QM/MM calculations. *J. Chem. Theo. Comp.* **2005,** *1*, 2-13.

219. Cui, Q.; Karplus, M., Catalysis and specificity in enzymes: a study of triosephosphate isomerase and comparison with methyl glyoxal synthase. *Adv. Protein Chem.* **2003,** *66*, 315-372.

220. Quesne, M. G.; Latifi, R.; Gonzalez-Ovalle, L. E.; Kumar, D.; de Visser, S. P., Quantum mechanics/molecular mechanics study on the oxygen binding and substrate hydroxylation step in AlkB repair enzymes. *Chemistry* **2014,** *20* (2), 435-46.

221. Lennard-Jones, J. E., Cohesion. *Proc. Phys. Soc., London* **1931,** *43* (5), 461.

222. Lennard-Jones, J. E.; S, F. R.; Devonshire, A. F., Critical Phenomena in Gases - I. *Proc. R. Soc. London, Ser. A* **1937,** *163*, 53.

223. Buckingham, R. A., The Classical Equation of State of Gaseous Helium, Neon, and Argon. *Proc. R. Soc. London, Ser. A* **1938,** *168*, 264.

224. Halgren, T. A., The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *Journal of the American Chemical Society* **1992,** *114* (20), 7827--7843.

225. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Jr; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; and, D. J. F., *Gaussian 09, Revision A.02*. Gaussian, Inc.: 2010.

226. Wu, J. C.; Chattree, G.; Ren, P., Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012,** *131* (3), 1138.

227. Stone, A. J., Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theo. Comp.* **2005,** *1*, 1128-1132.

228. Stevens, W. J.; Fink, W. H., Frozen fragment reduced variational space analysis of hydrogen bonding interactions. Applications to the water dimer. *Chem. Phys. Lett.* **1987,** *139*, 15-22.

229. Kitaura, K.; Morokuma, K., A new energy decomposition scheme for molecular interactions within the hartree-Fock approximation. *Int. J. Quantum Chem.* **1976,** *10*, 325-340.

230. Boys, S. F.; Bernardi, F., The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970,** *19* (4), 553.

231. Andzelm, J.; Wimmer, E., Density functional gaussian--type--orbital approach to molecular geometries, vibrations and reaction energies. *J. Chem. Phys.* **1992,** *96*, 1280-1303.

232. Challacombe, M.; Schwgler, E.; Almlof, J., *Modern developments in Hartree-Fock theory: Fast methods for computing the Coulomb matrix*. 1996.

233. Liu, Z. P.; Huang, S. P.; Wang, W. C., A refined force field for molecular simulation of imidazolium-based ionic liquids. *J. Phys. Chem. B* **2004,** *108* (34), 12978.

234. Arduengo, A. J.; Dias, H. V. R.; Harlow, R. L.; Kline, M., Electronic Stabilization of Nucleophilic Carbenes. *J. Am. Chem. Soc.* **1992,** *114* (14), 5530.

235. Starovoytov, O. N.; Borodin, O.; Bedrov, D.; Smith, G. D., Development of a Polarizable Force Field for Molecular Dynamics Simulations of Poly (Ethylene Oxide) in Aqueous Solution. *J. Chem. Theory Comput.* **2011,** *7* (6), 1902.

236. Donchev, A. G.; Galkin, N. G.; Pereyaslavets, L. B.; Tarasov, V. I., Quantum mechanical polarizable force field (QMPFF3): Refinement and validation of the dispersion interaction for aromatic carbon. *J. Chem. Phys.* **2006,** *125*, 244107.

237. Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; vander Avoird, A., Predictions of the Properties of Water from First Principles. *Science* **2007,** *315*, 1249-1252.

238. Case, D. A.; Darden, T. A.; Cheatham Iii, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; others, AMBER 12. *University of California, San Francisco* **2012,** *1* (2), 3.

239. Schofield, P., Computer simulation studies of the liquid state. *Comput. Phys. Commun.* **1973,** *5* (1), 17.

240. Beeman, D., Some multistep methods for use in molecular dynamics calculations. *Journal of Computational Physics* **1976,** *20* (2), 130--139.

241. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98* (12), 10089.

242. Darden, T.; Perera, L.; Li, L.; Pedersen, L., New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **1999,** *7* (3), R55.

243. Hanke, C. G.; Price, S. L.; Lynden-Bell, R. M., Intermolecular potentials for simulations of liquid imidazolium salts. *Mol. Phys.* **2001,** *99*, 801-809.

244. Hardacre, C.; Holbrey, J. D.; McMath, S. E. J.; Bowron, D. T.; Soper, A. K., Structure of molten 1,3-dimethylimidazolium chloride using neutron diffraction. *J. Chem. Phys.* **2003,** *118* (1), 273-278.

245. Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*. 1987.

246. Jansen, H.; Ros, P., Non-empirical molecular orbital calculations on the protonation of carbon monoxide. *Chemical Physics Letters* **1969,** *3* (3), 140--143.

247. Stevens, W. J.; Fink, W. H., Frozen fragment reduced variational space analysis of hydrogen bonding interactions. Application to the water dimer. *Chem. Phys. Lett.* **1987,** *139* (1), 15--22.

248. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Mongomery, J. A., General Atomic and Molecular Electronic Structure System. *J. Comp. Chem.* **1993,** *14*, 1347-1363.

249. Gordon, M. S.; Schmidt, M. S., Advances in electronic structure theory: GAMESS a decade later. Amsterdam.

250. Kitaura, K.; Morokuma, K., New Energy Decomposition Scheme for Molecular-Interactions within Hartree-Fock Approximation. *Int. J. Quantum Chem.* **1976,** *10*, 325.

251. Levitt, M.; Hirshberg, M.; Sharon, R.; Laidig, K. E.; Daggett, V., Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *The Journal of Physical Chemistry B* **1997,** *101* (25), 5051--5061.

252. Eisenberg, D.; Kauzmann, W., The structure and properties of waterOxford Univ. *Press, London* **1969**, 296.

ABSTRACT

CLASSICAL AND QUANTUM MECHANICAL SIMULATIONS OF CONDENSED SYSTEMS AND BIOMOLECULES

by

HEDIEH TORABIFARD

August 2017

Advisor: Dr. G. Andres Cisneros

Major: Chemistry

Degree: Doctor of Philosophy

This work describes the fundamental study of two enzymes of Fe(II)/α-KG super family enzymes (TET2 and AlkB) by applying MD and QM/MM approaches, as well as the development of multipolar-polarizable force field (AMOEBA/GEM-DM) for condensed systems (ionic liquids and water).

TET2 catalytic activity has been studied extensively to identify the potential source of its substrate preference in three iterative oxidation steps. Our MD results along with some experimental data show that the wild type TET2 active site is shaped to enable higher order oxidation. We showed that the scaffold stablished by Y1902 and T1372 is required for iterative oxidation. The mutation of these residues perturbs the alignment of the substrate in the active site, resulting in "5hmC-stalling" phenotype in some of the mutants. We provided more details on 5hmC to 5fC oxidation mechanism for wild type and one of the "5hmC-stallling" mutants (E mutant). We showed that 5hmC oxidizes to

5fC in the wild type via three steps. The first step is the hydrogen atom abstraction from hydroxyl group of 5hmC, while the second hydrogen is transferred from methylene group of 5hmC through the third transition state as a proton. Our results suggest that the oxidation in E mutant is kinetically unfavorable due to its high barrier energy. Many analyses have been performed to qualitatively describe our results and we believed our results can be used as a guide for other researchers.

In addition, two MD approaches (explicit ligand sampling and WHAM) are used to study the oxygen molecule diffusion into the active site of AlkB. Our results showed that there are two possible channels for oxygen diffusion, however, diffusion through one of them is thermodynamically favorable. We also applied multipolar-polarizable force field to describe the oxygen diffusion along the preferred tunnel. We showed that the polarizable force field can describe the behavior of the highly polarizable systems accurately.

We also developed a new multipolar-polarizable force field (AMOEBA/GEM-DM) to calculate the properties of imidazolium- and pyrrolidinium- based ionic liquids and water in a range of temperature. Our results agree well with the experimental data. The good agreement between our results and experimental data is because our new parameters provide an accurate description of non-bonded interactions. We fit all the non-bonded parameters against QM. We use the multipoles extracted from fitted electron densities (GEM) and we consider both inter- and intra-molecular polarization. We believe this method can accurately calculate the properties of condensed systems and can be helpful for designing new systems such as electrolytes.

AUTOBIOGRAPHICAL STATEMENT

HEDIEH TORABIFARD

Education

2012-2017 Doctor of Philosophy, Chemistry, Wayne State University, Detroit, MI, USA

2009-2011 Master Degree, Chemistry, Sharif University of Technology, Tehran, Iran.

2004-2008 Bachelor Degree, Chemistry, Shahid Beheshti University, Tehran, Iran.

Honors and Awards

2016: ACS Division of Computers in Chemistry, "Chemical Computing Group Excellence (CCG) Award for Graduate Students", ACS National Meeting, San Diego, CA.

Selected Publications

1. **Torabifard, H.,** Cisneros, G.A. "Computational Investigation of O2 Diffusion Through an Intra–molecular Tunnel in AlkB; Influence of Polarization on O2 Transport". Chem. Sci. Under review.

2. Liu, M.Y., **Torabifard, H.,** Crawford, D.J., DeNizio, J.E., Cao, X., Garcia, B.A., Cisneros, G.A., Kohli, R.M. "Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine". *Nat. Chem. Biol.*, **13**,181,2017.

3. Cisneros, G.A., Wikfeldt, K.Y., Ojamae, L., Lu, j., Xu, J., **Torabifard, H.,** Bartok, A.P., Csanyi, G., Molinero, V., Paesani, F. "Modeling molecular interactions in water: From pairwise to many-body potential energy functions" *Chem. Rev.*, **116**, 13:7501, 2016.

4. **Torabifard, H.,** Starovoytov, O.N., Ren, P., Cisneros, G.A. "Development of an AMOEBA water model using GEM distributed multipoles" *Theor. Chem. Acc.*, **134**, 101:1, 2015.

5. Starovoytov, O.N., **Torabifard, H.,** Cisneros, G.A. ``Towards accurate parameterization of imidazolium based ionic liquids potential" *J. Phys. Chem. B.*, **118**, 7156, 2014.