## **REACTIVE MOLECULAR DYNAMICS IN IONIC MEDIA**

A Dissertation Presented to The Academic Faculty

By

John P. Stoppelman

In Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the School of Chemistry and Biochemistry

Georgia Institute of Technology

December 2023

© John P. Stoppelman 2023

### **REACTIVE MOLECULAR DYNAMICS IN IONIC MEDIA**

Thesis committee:

Dr. Jesse G. McDaniel, Advisor School of Chemistry and Biochemistry *Georgia Institute of Technology* 

Dr. C. David. Sherrill School of Chemistry and Biochemistry *Georgia Institute of Technology* 

Dr. Thomas M. Orlando School of Chemistry and Biochemistry *Georgia Institute of Technology*  Dr. Joshua S. Kretchmer School of Chemistry and Biochemistry *Georgia Institute of Technology* 

Dr. Andrew J. Medford School of Chemical and Biomolecular Engineering *Georgia Institute of Technology* 

Date approved: July 26, 2023

To my parents, grandparents, siblings and fiancée Yi-Jung for their continuous love and support.

#### ACKNOWLEDGMENTS

First and foremost I need to thank my advisor, Dr. Jesse McDaniel, for his phenomenal mentorship over the last 5 years. I was one of the first graduate students within the group, and I think Jesse has done a great job maintaining a rigorous, productive and relaxed environment to perform science since I've started. He has been a great source of knowledge on a broad range of topics in theoretical chemistry, and his guidance and many suggestions have been invaluable for identifying new research plans and directions; additionally, he has helped me hone key skills as a scientist and really helped me develop my future plans and career goals. I'm sure he will continue to guide many students over his career.

I have had the privilege of working alongside some fantastic group members as well. It has been a pleasure to interact with Dr. Yi-Jung Tu (who I will thank more extensively in a few paragraphs), Dr. Shahriar Khan, Suehyun Park, John Hymel, Chloe Renfro, John Pederson, Shehan Parmar, Sam Delmerico and Archana Verma. It's been fun to spend time with all of you throughout the years and I've enjoyed hearing about all of your research progress. My knowledge of theoretical chemistry has definitely benefited from working with you all.

My thanks goes to Dr. Marcus Cicerone, who we collaborated with studying density scaling within propylene carbonate, and Dr. Angus Wilkinson, who we collaborated with on studying negative thermal expansion materials. Both projects were awesome opportunities to explore different areas of theoretical chemistry than what I worked on in my main project, and these collaborations make up two chapters of this thesis. I would also like to thank my committee members Dr. C. David Sherrill, Dr. Joshua Kretchmer, Dr. Thomas Orlando and Dr. Andrew Medford. I have appreciated your guidance in the development of this thesis.

I would also like to thank Dr. Sherrill, Dr. Kretchmer, Dr. Kamerlin and all the other grad students and postdocs I've come across here for helping facilitate a collaborative and friendly atmosphere within the Georgia Tech theoretical chemistry community. As each theory group at Georgia Tech works on different research areas, hearing about the work going on in the other labs has definitely enhanced my appreciation of theoretical chemistry. It has also been fun spending time with everyone outside of the lab at Cypress and Friday lunch, and I'll miss the community here after I've left. Beyond our theoretical division, I've felt that the Georgia Tech chemistry department has been a nice place to work and study, and I extend my thanks to all who have made that possible.

I would also like to thank Dr. Adam Moser and Dr. Lee-Ping Wang for initiating my interest in computational chemistry during my undergraduate years. Dr. Moser at Loras College introduced me to scientific research, which I had never really thought much about before entering undergrad. He helped prepare me as an applicant for grad school, and his guidance helped shape my academic path up to this point. I was able to work with Lee-Ping at University of California, Davis as part of the REU program. I had a great experience there, and the skills I gained from it have aided my research quite a few times in grad school. Lee-Ping has also helped me out along the way as well and I'm glad I got to work in his group.

Yi-Jung has been instrumental for my grad school experience (and life in general) for two reasons. First, and most importantly, she has been the best long distance girlfriend (and now fiancée) imaginable over the past 3.5 years. We both knew it would be difficult to maintain our relationship when she returned to Taiwan in 2020, and the pandemic only made this more challenging. Still, we always made time to call each other almost every day. She has been very encouraging and has been a source of support while I worked on finishing grad school. Second, she was an excellent postdoc in the group. Yi-Jung was one of the few people who started in the group before I did. She taught me a lot of lessons about the work ethic required to succeed in graduate school and research, and I'm not sure if I would have finished as many projects without her initial advice. She also helped give me a lot of comments for my defense presentation!

Finally I would like to thank my parents, siblings and grandparents for their love and support.

## TABLE OF CONTENTS

Acknow	vledgments	iv
List of 7	Tables	x
List of I	Figures	xi
Summa	<b>ry</b>	vi
Chapter	r 1: Theoretical Approaches for Modeling Chemical Reactions in Solution	1
1.1	Density Functional Theory	3
1.2	Ab initio molecular dynamics	7
1.3	Quantum Mechanics/Molecular Mechanics (QM/MM)	9
1.4	Empirical Valence Bond (EVB) Approach	10
1.5	Machine Learning Approaches	12
1.6	Thesis Organization	15
Chapte	r 2: Physics-Based, Neural Network Force Fields for Reactive Molecular Dy- namics: Investigation of Carbene Formation from [EMIM <sup>+</sup> ][OAc <sup>-</sup> ]	16
2.1	Abstract	16
2.2	Introduction	17
2.3	Methods	23

	2.3.1	General Functional Form	24
	2.3.2	Training Data Generation	26
	2.3.3	Force Field Components	28
2.4	Result	s and Discussion	34
	2.4.1	Intramolecular Neural Network: $(E_{Solute}^{Intra,NN})$	35
	2.4.2	Nonbonded Force Field $(E_{Solute,Solvent}^{Nonbonded,FF})$ and Intermolecular Neural Network $(E_{Solute}^{Inter,NN})$	36
	2.4.3	PB/NN Force Field Total Energy	40
	2.4.4	EMIM <sup>+</sup> /Acetate Reaction Free Energies with/without "Solvation" Environment	42
2.5	Conclu	ision	48
Chapter	r 3: <i>N-</i> I Elu	Ieterocyclic Carbene Formation in the Ionic Liquid [EMIM <sup>+</sup> ][OAc <sup>-</sup> ]:         cidating Solvation Effects with Reactive Molecular Dynamics	50
3.1	Abstra	ct	50
3.2	Introdu	ction	51
3.3	Metho		
		ds	54
	3.3.1	ds	54 55
	3.3.1 3.3.2	ds	54 55 58
	<ul><li>3.3.1</li><li>3.3.2</li><li>3.3.3</li></ul>	ds	54 55 58 60
3.4	<ul><li>3.3.1</li><li>3.3.2</li><li>3.3.3</li><li>Results</li></ul>	ds    Multistate PB/NN Hamiltonian      PB/NN Parameterization and Neural Network Architecture:      Reactive Molecular Dynamics Simulations      and Discussion	54 55 58 60 64
3.4	<ul> <li>3.3.1</li> <li>3.3.2</li> <li>3.3.3</li> <li>Results</li> <li>3.4.1</li> </ul>	ds	<ul> <li>54</li> <li>55</li> <li>58</li> <li>60</li> <li>64</li> <li>65</li> </ul>
3.4	<ul> <li>3.3.1</li> <li>3.3.2</li> <li>3.3.3</li> <li>Results</li> <li>3.4.1</li> <li>3.4.2</li> </ul>	ds	<ul> <li>54</li> <li>55</li> <li>58</li> <li>60</li> <li>64</li> <li>65</li> <li>74</li> </ul>

Chapter	4: Proton Transport in [BMIM <sup>+</sup> ][BF <sup>-</sup> <sub>4</sub> ] /Water Mixtures Near the Percolation Threshold	3
4.1	Abstract	3
4.2	Introduction	1
4.3	Methods	3
4.4	Results and Discussion	3
4.5	Conclusion	3
Chapte	<b>5: Excitations Follow (or Lead?) Density Scaling in Propylene Carbonate</b> 11	1
5.1	Abstract	l
5.2	Introduction	l
5.3	Methods $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $11^4$	1
	5.3.1 Molecular Dynamics Simulations	1
	5.3.2 Isomorph Search	5
	5.3.3 Identifying Excitations	3
5.4	Results	)
	5.4.1 Center of Mass Motion	)
	5.4.2 Excitations, Inherent State Motion, and Interbasin Crossing	3
5.5	Discussion	7
	5.5.1 IS and IB motion Involve Intramolecular Modes	7
	5.5.2 Scaling-Relevant Intramolecular and Collective Modes	3
5.6	Conclusions $\ldots \ldots 134$	1
Chapter	6: Equation of State Predictions for ScF <sub>3</sub> and CaZrF <sub>6</sub> with Neural Network- Driven Molecular Dynamics	5

6.1	Abstract
6.2	Introduction
6.3	Methods
	6.3.1 Training Set Generation and Model Training
	6.3.2 Neural Network-Driven, Molecular Dynamics Simulations
6.4	Results and Discussion
6.5	Conclusion
Chapter 7: Conclusions and Outlook	
7.1	Future Directions
7.2	Outlook
Referen	<b>ces</b>
Vita .	

### LIST OF TABLES

- 5.1  $\tilde{D}_T$  and  $1/\tilde{\tau}_{rot}$  for isomorph 1. Uncertainties in both  $\tilde{D}_T$  and  $1/\tilde{\tau}_{rot}$  are approximately 8% due to uncertainties in predicted density at a given T/P state point. . . . 121

## LIST OF FIGURES

2.1	Free energy as a function of C–H and O–H bond distance for the deprotonation of EMIM <sup>+</sup> by acetate in the gas-phase. Arrows are drawn to the reactants and products to indicate their respective location on the surface.	22
2.2	PBE-D3(BJ)/aug-cc-pVTZ vs. neural network energy predictions for each monomer involved in the reaction. We have added $E_{Solute}^{Morse}$ to the neural network prediction for EMIM <sup>+</sup> and acetic acid. Each plot shows 5000 randomly-selected structures from the test set; the MAE listed corresponds to the full test set for each monomer.	37
2.3	SAPTO vs. $E_{Solute}^{Inter,NN} + E_{Solvent}^{Nonbonded,FF}$ predictions for both $H_{11}$ and $H_{22}$ . Diabat 1, Transition state and Diabat 2 regions are defined with respect to their value of $c_1^2$ as discussed in the main text.	39
2.4	PBE-D3(BJ)/aug-cc-pVTZ energies vs the force field $H_{PB/NN}$ energies with C– H bond distances less than 1.8 Å and O–H bond distances less than 1.8 Å. The plot shows 5000 randomly-selected structures from the test set; the MAE listed corresponds to the full test set.	40
2.5	a) Diabatic expansion of PB/NN force field in terms of $c_1^2$ coefficient, compared to b) reaction free energy surface projected onto C–H and O–H bond distances Configurations are taken from dataset 5 in Table 2.1, and the lines drawn in the plot qualitatively separate the diabat 1, diabat 2, and transition state regions	41
2.6	Free energy surfaces for the EMIM <sup>+</sup> /acetate deprotonation reaction with/without the "model" solvation environment. PMFs computed from AIMD simulations are shown in a) and c) for the two ion and four ion systems, respectively. PMFs computed from the PB/NN force field are shown in b) and d) for the two ion and four ion systems, respectively.	44
3.1	Depiction of the chemical bond topology for the three diabatic states of the reacting complex. The dashed ovals depict the molecule associated with the reacting proton in each diabatic state.	56

3.2	Free energy surfaces for the proton transfer reaction between acetic acid and acetate ion, as computed from a) AIMD and b) PB/NN simulations. The free energy is plotted as a function of the distances between the acidic proton and closest oxygen atoms of the two acetate ions	59
3.3	a) Proton transfer free energy for EMIM <sup>+</sup> /OAc <sup>-</sup> dimer in the gas phase. b) Proton transfer free energy for EMIM <sup>+</sup> /(OAc <sup>-</sup> ) <sub>2</sub> trimer in the gas phase. The labels $R_1$ and $P_1$ correspond to the reactants and products of Equation 3.1. The $P_2$ label corresponds to the shared proton between two OAc <sup>-</sup> molecules in Equation 3.2	68
3.4	a) Proton transfer free energy within the bulk [EMIM <sup>+</sup> ][OAc <sup>-</sup> ] ionic liquid; b) Proton transfer free energy for EMIM <sup>+</sup> /(OAc <sup>-</sup> ) <sub>2</sub> trimer in the gas phase	71
3.5	Simulation snapshots depicting reacting complex intermediates. The top panel displays snapshots from the [EMIM <sup>+</sup> ][OAc <sup>-</sup> ]ionic liquid simulations, and the bottom panel displays snapshots for corresponding intermediates from the gas-phase, $EMIM^+/(OAc^-)_2$ trimer simulation. The labels match the corresponding locations on the reaction free energy profiles of Figure 3.4.	72
3.6	RDFs between the center of mass of the EMIM <sup>+</sup> ring and acetate oxygen atoms (" $R_1$ ") and between the center of mass of the NHC ring and acetate oxygen atoms (" $P_2$ "). The corresponding coordination number at a given distance is shown by the dashed lines.	73
3.7	a) EMIM <sup>+</sup> H3–O $\rho g(r)$ and H4–O $\rho g(r)$ . These RDFs were computed using all EMIM <sup>+</sup> molecules in the simulation. b) Reacting complex NHC H3–O $\rho g(r)$ and H4–O $\rho g(r)$ . The corresponding coordination number at a given distance is shown by the dashed lines.	75
3.8	Number density profiles at the $[EMIM^+][OAc^-]$ ionic liquid/vapor interface for a) the center of mass of the $EMIM^+$ methyl, ring and ethyl group and b) the center of mass of the $OAc^-$ carboxylate and methyl group. The dashed lines approximately denote the Gibbs dividing surface.	75
3.9	Free energy profiles for the proton transfer reaction a) at the $[EMIM^+][OAc^-]$ ionic liquid/vapor interface; and b) in bulk $[EMIM^+][OAc^-]$ ionic liquid	77
3.10	a) RDFs between reacting complex EMIM <sup>+</sup> center of mass/acetate oxygen atoms computed for both the liquid and the liquid/vapor interface. b) RDFs for EMIM <sup>+</sup> ring nonreactive protons/acetate oxygen atoms computed for both the liquid and the liquid/vapor interface.	78

3.11	Spatial distribution functions of acetate oxygen atoms surrounding the EMIM <sup>+</sup> ring for a) a liquid phase umbrella sampling simulation and b) a liquid/vapor interface umbrella sampling simulation. Depicted in purple are density regions with isovalue set to $0.532 \text{ nm}^{-3}$ . These configurations correspond to $R_1$ regions of the profile
4.1	Ionic conductivity $\sigma$ of H <sub>3</sub> O <sup>+</sup> /BMIM <sup>+</sup> /BF <sub>4</sub> <sup>-</sup> and BMIM <sup>+</sup> /BF <sub>4</sub> <sup>-</sup> aqueous solutions at varying water content; for H <sub>3</sub> O <sup>+</sup> /BMIM <sup>+</sup> /BF <sub>4</sub> <sup>-</sup> solutions, the proton concentration is set at 1:5.28 H <sub>3</sub> O <sup>+</sup> /H <sub>2</sub> O mole ratio. <sup>273</sup> Experimental conductivities taken from ref <sup>273</sup> are plotted as triangles, simulation predictions are plotted as circles, and dashed trend lines are to guide the eye
4.2	a) Hydronium, water and $BF_4^-$ diffusion coefficient and b) hydronium diffusion co- efficient and forward hop rate within aqueous $[BMIM^+][BF_4^-]$ mixtures at varying water volume fraction, $\phi_V^{H_2O}$ . The dashed lines reflect the linear trend and are to guide the eye. Error bars represent the range between the maximum and minimum calculated diffusion coefficients and forward hop rates from three independent sim- ulations
4.3	Hydrogen bond distribution of water molecules in aqueous $[BMIM^+][BF_4^-]$ mixtures as a function of water volume fraction, $\phi_V^{H_2O}$ . For comparison, the distribution for bulk water is shown as computed from similar simulations. A simulation snapshot of water clusters at $\phi_V^{H_2O} = 0.13$ is also shown
4.4	(a) Percolation probability $R(\phi_V^{H_2O}, L)$ and (b) weighted water cluster size $n_w$ (Equation 4.5) as a function of water volume fraction $\phi_V^{H_2O}$ in aqueous [BMIM <sup>+</sup> ][BF <sub>4</sub> <sup>-</sup> ] mixtures. Metrics are computed for different system sizes, and the percolation threshold is indicated by the vertical dashed line
4.5	Water cluster diameter "d" (Equation 4.4) distributions for a) $\phi_V^{H_2O} = 0.13$ and b) $\phi_V^{H_2O} = 0.25 H_3O^+/[BMIM^+][BF_4^-]$ /water mixtures. For each system, distributions are shown for all water clusters and those that contain a hydronium ion (H <sub>3</sub> O <sup>+</sup> cluster). The insets depict a magnified view of the tail ends of the distribution, where a secondary peak is observed for concentrations above the percolation threshold 104
4.6	Pairwise correlation between hydrogen atoms of hydronium ion and oxygen atoms of water molecules as a function of water volume fraction $\phi_V^{H_2O}$ in aqueous [BMIM <sup>+</sup> ][BF <sub>4</sub> <sup>-</sup> ]. Note that rather than plotting the radial distribution function, $g(r)$ , we plot $\rho g(r)$ normalized by the density of water to better compare between the different water concentration mixtures

4.7	Probability distribution of the MS-EVB coefficients for the principle $(c_1^2)$ and second largest $(c_2^2)$ states. The $c_1^2$ distribution is on the right towards larger coefficient values, and the $c_2^2$ distribution is the on the left towards smaller coefficient values. For comparison, we plot the corresponding distribution for hydronium in bulk water $(\phi_V^{H_2O} = 1.0)$ computed from an analogous simulation
4.8	Pairwise correlation g(r) for hydronium/BF <sub>4</sub> <sup>-</sup> ions in aqueous [BMIM <sup>+</sup> ][BF <sub>4</sub> <sup>-</sup> ] at $\phi_V^{H_2O} = 0.25$ water content. The distribution is decomposed into Eigen/Zundel contributions as discussed in the text, and is computed between hydrogen atoms of the hydronium ion and fluorine atoms of BF <sub>4</sub> <sup>-</sup> anions
5.1	Demonstration of $\rho^{\gamma_{TS}}/T = constant$ for the two isomorph curves, with $\gamma_{TS} = 3.7$ as determined from Pawlus et al. <sup>407</sup> plotted as dashed lines. The solid line denotes the liquid/glass boundary computed from state points taken from Bonetti and Dubois [416]
5.2	Center-of-mass structural and dynamic metrics at pseudo-isomorphic phase points compared to isobar phase points. Top: RDFs computed between propylene carbon- ate center of mass positions, and plotted as a function of reduced distance for a) all points along the 300 K pseudoisomorph; and b) 1 atm isobar points at several tem- peratures. Bottom: Comparison of $F(q,t)$ at selected wavevectors, computed based on center of mass positions for c) 300 K, 1 atm pseudoisomorph phase points; and d) 1 atm isobar points at several temperatures
5.3	Top) Scattering function $F_s(q, t)$ computed from MD simulation at 300K, 1atm based on H-atom positions. Bottom) Scattering function $F_s(q, t)$ measured by incoherent QENS from PC. <sup>403</sup> 124
5.4	$F(q,t)$ and fits. (Top), $F(q,t)$ at 300 K, 1 atm based on H-atom positions, plotted against $q^2$ at evenly spaced times between 0.5 ps and 6 ps. Black points are calculated from simulation, and dashed red lines are fits to (Equation 5.3). (Middle and Bottom) Time-dependence of model fits shown for each pseudoisomorph of 300 K, 1 atm
5.5	(Top) $F(q,t)$ computed from MD simulation at 300 K, 1 atm based on molecular centers-of-mass. Black points are calculated from simulation, and dashed red lines are fits to (Equation 5.3). (Middle and Bottom) Time-dependence of model fits shown for each pseudoisomorph of 300 K, 1 atm

5.6	Top Left) Polarizability and Raman response in PC. Black line: $\beta_{fast}$ polarizabil- ity spectra from OKE measurements. <sup>418</sup> Red dash line: OKE spectrum due to IS motion (known as librations in OKE literature). Blue dashed line: OKE spectrum due to IB crossing (known as collective motion in OKE literature). Red solid line: Low frequency component of a PC spontaneous Raman spectrum. (Right Top and Bottom) Power spectra of velocity autocorrelation functions for all atoms (Right Top) and centers of mass (Right Bottom) at 300 K and the pressures indicated in the middle panel. (Inset to Right Bottom panel) Grüneisen parameters calculated for each mode found in the power spectra. Left Bottom) Model fit to the 300 K 1 atm centers-of-mass power spectrum.	131
6.1	Thermal expansion of ScF <sub>3</sub> denoted by a) reduced lattice constant $a/a_0$ and b) CTE. Thermal expansion of CaZrF <sub>6</sub> denoted by c) reduced lattice constant $a/a_0$ and d) CTE. The experimental results for ScF <sub>3</sub> are from Greve <i>et al.</i> [443] and the experimental results for CaZrF <sub>6</sub> are from Hancock <i>et al.</i> [457] SCP and ISC predictions for ScF <sub>3</sub> are taken from Oba <i>et al.</i> [467]. The statistical uncertainty in our predictions is on par with the size of the symbols themselves, and thus error bars not shown.	145
6.2	ScF <sub>3</sub> atomic displacement parameters for a) Sc atoms $(U_{iso})$ and b) F atoms $(U_{11}$ and $U_{33})$ The experimental data is from Greve <i>et al.</i> [443]. The statistical uncertainty in our predictions is on par with the size of the symbols themselves, and thus error bars not shown.	149
6.3	$CaZrF_6$ atomic displacement parameter for a) Ca, b) Zr and c) F. The experimental data is from Hancock <i>et al.</i> [457]. The statistical uncertainty in our predictions is on par with the size of the symbols themselves, and thus error bars not shown	150
6.4	a) Experimental pressure vs. volume data of $ScF_3$ from Wei <i>et al.</i> [446] b) Pressure vs. volume from the cubic $ScF_3$ simulations.	151
6.5	a) $B_0$ parameter and b) $B'$ parameter from the Birch-Murnaghan equation of state from both simulation and experiment for ScF <sub>3</sub> . The experimental data is from Wei <i>et al.</i> [446]	151
6.6	CaZrF <sub>6</sub> Volume vs. Pressure data at 290K. The experimental data is from Hester <i>et al.</i> [458]	153
6.7	CaZrF <sub>6</sub> Bulk modulus at various temperatures from both simulation and experiment. The experimental data is from Hancock <i>et al.</i> [457]	154

#### **SUMMARY**

Chemical reactions are among the most fundamental phenomena within the field of chemistry. In many contexts, reactions are conducted or occur in condensed phase environments. Environmental effects can cause a host of complicated changes to a given chemical process, such as altering thermodynamic equilibrium, reaction rates or the associated mechanism. Solvents can thus be used to tune a given reaction. In particular, ionic media can cause substantial changes to a reaction due to the long-range Coulombic interactions between the reacting complex and solvent molecules, which, energetically, can be quite large in magnitude. Further study of reactions within ionic solvents would allow for modulating these interactions for selected applications. Theoretical approaches, such as quantum chemistry, represent one tract of methods that can be applied for this purpose. However, while quantum chemical techniques can effectively investigate many gas phase reactions, condensed phase reactions are much more challenging to investigate. The many degrees of freedom associated with the bulk solvent makes first principles modeling infeasible due to unfavorable scaling with respect to system size.

Force fields derived from *ab initio* methods specifically designed for simulating reactions can significantly enhance insight into solvent modulation of chemical reactions. A sufficiently accurate force field can be used to perform molecular dynamics at quantum chemistry-level accuracy within an external environment at a fraction of the cost. However, such reactive force fields have been challenging to parameterize and use, as typical physics-based expressions used in force fields are better suited for asymptotic interactions than describing short-ranged effects associated with chemical bond-breaking/formation. Recent machine learning approaches have proved effective at learning a wide range of physical interactions, however, and can potentially be combined with standard force fields in order to build an extensive framework for modeling chemical reactions. This thesis details our development of a reactive force field framework that combines these two methodologies.

Chapter 2 contains the framework of our reactive force field approach, named PB/NN. It is

in the form of an empirical valence bond (EVB) Hamiltonian and uses *ab initio* force fields for describing long-range effects and machine learning models for all short-range interactions. We demonstrate how to build each term in the model for the deprotonation of the ionic liquid cation  $\rm EMIM^+$  by acetate in order to form a *N*-heterocyclic carbene (NHC) and acetic acid. We show that the force field can produce reaction free energy profiles that are in semiquantitative agreement with AIMD for the gas phase dimer and a four ion model solvation environment.

Chapter 3 extends the model to simulations of the deprotonation reaction in the bulk ionic liquid, where proton sharing between acetic acid and acetate can occur. We investigate the effects of various solvation environments on the free energy required for formation of the NHC, including the gas phase, the bulk liquid and the air-liquid interface. We find that the bulk liquid lowers the free energy barrier for forming the products in comparison to the gas phase, partially due to local solvation of the reacting complex. We also find that there is stabilization of the NHC at the interface in comparison to the liquid, due to both solvophobic interactions with the NHC and destabilization of the ionic  $[EMIM^+][OAc^-]$  pair.

We use the MS-EVB reactive force field to simulate proton transport in  $[BMIM^+][BF_4^-]$ /water mixtures in Chapter 4. We analyze the effect varying water concentration has on proton transport rates and conductivity. We also examine the structure of the hydrogen bonded water network and determine at which concentration a percolating network forms in these systems. We find that proton diffusion increases linearly with water content along our studied concentration range and we find the  $BF_4^-$  anions stabilize the nonreactive Eigen form of  $H_3O^+$ .

In Chapter 5, we utilize an *ab initio* force field in order to examine various liquid properties of propylene carbonate. We examine the phenomenon known as density scaling in particular and connect it with intramolecular vibrations in the liquid. We also compare scattering functions computed from the force field with those from quasi-elastic neutron scattering experiments. While not a reaction, this connects with the previous chapters as it demonstrates the ability of *ab initio* force fields, which are a key part of our reactive force field methodology as explained in Chapters 2 and 3, to make predictions in agreement with experiment; it is useful to understand the level of

accuracy attained with such approaches with respect to experimental liquid properties.

In Chapter 6, we train neural network potentials to DFT data for the negative thermal expansion (NTE) materials  $ScF_3$  and  $CaZrF_6$ . We assess the accuracy between DFT and experiment for equations of state in these systems. We also quantify the extent to which nuclear quantum effects matter for predicting various NTE-related properties. We find these are significant at low temperature.

We conclude in Chapter 7 with future plans and directions for the work in this thesis.

#### **CHAPTER 1**

# THEORETICAL APPROACHES FOR MODELING CHEMICAL REACTIONS IN SOLUTION

Solvent effects on reactions are ubiquitous within chemistry. External environments play a key role from fields as diverse as biochemistry, where protein residues and solvent molecules influence numerous catalytic processes, to organic chemistry, where solvents modulate reaction rates and equilibria. It is well known that solvation can affect chemical reactivity through a number of methods, such as stabilization of reactants/products/transition states (TS), creation of new reaction pathways through interactions with the reacting molecules and alteration of product selectivity.<sup>1</sup> Of particular relevance to this thesis is the effect of ionic media on reactions. As Coulombic ion-ion interactions can rival the strength of a covalent chemical bond, ionic media can have a particularly strong influence. This is seen often in biochemistry, where charged groups aid in many catalytic steps necessary for sustaining life.<sup>2</sup> Nature has essentially optimized the use of ionic forces in these pathways through evolution. With further study, similar electrostatic interactions can be utilized by chemists to intelligently design new powerful reactions.

The effects a particular solvent will have on a given reaction may be difficult to predict. Empirical models based on concepts such as solvent polarity or other bulk solvent properties can often be useful for establishing qualitative relationships between solvation and reactivity. These approaches can sometimes lead to incorrect conjectures, however, and quantitative prediction is generally beyond the scope of such methods.<sup>3</sup> Thus more extensive modeling explicitly considering the molecular-scale interactions between the reacting complex (which we will refer to as the solute) and the solvent is required for further in-depth understanding of the relationship between solvent effects and reactions.

The explosive development of molecular simulation, namely molecular dynamics (MD) and Monte Carlo methods, over the past several decades has brought about its application in essen-

tially every branch of chemistry. Modern computers allow for the routine use of quantum mechanics (QM) and molecular mechanics (MM) for the study and discovery of chemical phenomena in fields as diverse as biochemistry to materials science, including the study of chemical reactions in environments. As chemical bond breaking/forming is an inherently quantum mechanical process, QM-based methods such as ab initio molecular dynamics (AIMD) and quantum mechanics/molecular mechanics (QM/MM) make up the most robust methods for the simulation of reactions in the condensed phase. The unfavorable scaling of both methods (with respect to system size for AIMD and with respect to the QM region for QM/MM) limits the amount of statistical sampling that can be performed with such simulations, however. This hinders examination of reactions that proceed over long time-scales and often prevents simulation of a sufficiently large solvent environment, which can introduce finite size effects.<sup>4</sup> For these reasons it is more desirable to use MM methods, which have a greatly reduced computational cost compared to the aforementioned QM approaches. Constructing a MM-based potential energy surface (essentially a force field,<sup>5</sup> which is a collection of parameters and potential functions, as we shall discuss later) that is sufficiently representative of the QM surface for a specific reaction has traditionally been a difficult and time-intensive process; this is mainly due to the lack of analytic classical analogs for all the effects associated with bond-breaking/formation. However, recent work in machine learning (ML) algorithms and methods has significantly aided in the development of such force fields and has already opened up exciting new possibilities for simulation within the condensed phase, as is partially described in this dissertation.

We present a method for the construction of reactive force fields based on physics-based and machine-learned potentials that can be applied to simulate reactions in diverse condensed phase environments; this is the focus of this work. A full introduction in this area would include a summary of all existing QM and MM approaches used for computational modeling of chemical reactions. We neglect such a thorough review here in the favor of brevity. Instead, we restrict ourselves to a selection of some of the most powerful existing simulation methods for the study of a reacting complex within explicit solvent and discuss their strengths and weaknesses. We start

with a brief summary of density functional theory (DFT), which both the QM and MM methods used here rely on, as we discuss. We follow with a description of AIMD and QM/MM, and then proceed with the empirical valence bond (EVB) model, a MM-based method which we use as a framework for our approach. We additionally present a short summary of some of the promising ML models developed in recent years for simulating reactions within solvent environments.

#### **1.1 Density Functional Theory**

In principle, the Schrödinger equation contains all the information needed for the study of a reaction within a solvent:

$$H\Psi(\mathbf{x}_1, \mathbf{x}_2...\mathbf{x}_N) = E\Psi(\mathbf{x}_1, \mathbf{x}_2...\mathbf{x}_N)$$
(1.1)

Equation 1.1 is the nonrelativistic time-independent Schrödinger equation for a *N*-electron system. In this equation,  $\Psi(\mathbf{x_N})$  is the wavefunction, with coordinates  $\mathbf{x_N}$  corresponding to 3*N* spatial coordinates  $(r^{3N})$  and *N* spin coordinates  $(s^N)$ , *E* is the energy of the system and  $\hat{H}$  is the electronic Hamiltonian operator and is given in atomic units by:

$$\hat{H} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ee} + \hat{V}_{nn} = \sum_{i=1}^N (-\frac{1}{2}\nabla^2) + \sum_{i=1}^N \sum_{\alpha=1}^M (-\frac{Z_\alpha}{r_{i\alpha}}) + \sum_{i(1.2)$$

where we have assumed stationary nuclei in comparison to the electrons under the Born-Oppenheimer approximation.<sup>6</sup>  $\hat{T}_e$  is the electron kinetic energy operator,  $\hat{V}_{en}$  is the electron-nucleus attraction operator,  $\hat{V}_{ee}$  is the electron-electron repulsion operator and  $\hat{V}_{nn}$  is the nucleus-nucleus repulsion operator for a system with N electrons and M nuclei. An exact solution would allow for the calculation of any observable property of the system through the wavefunction. However, Equation 1.1 is unsolvable for all but the simplest cases.

Fortunately, a host of useful approximations for Equation 1.1 have been developed. These approximate methods generally fall under two categories. In the first category, the wavefunction is the central quantity; this includes methods such as Hartree-Fock, many-body perturbation theory, configuration interaction and others.<sup>7</sup> The electron density,  $\rho(r)$ , is the central quantity in a second

category of approximations named density functional theory (DFT). Due to its favorable balance between accuracy and computational cost in comparison to wavefunction-based methods, DFT has had a meteoric rise in computational chemistry as attested by its share of the Nobel Prize in 1998 and the continuously growing number of publications using DFT in some form.<sup>8</sup> We shall give a brief review over the fundamental points of the methodology here, as we use it in various ways throughout this thesis.

First, we rewrite the electron-nuclear attraction term in Equation 1.2:

$$\hat{V}_{en} = \sum_{i=1}^{N} \sum_{\alpha=1}^{M} (-\frac{Z_{\alpha}}{r_{i\alpha}}) = \sum_{i=1}^{N} v(r_i)$$
(1.3)

where  $v(r_i)$  is defined as

$$v(r_i) = -\sum_{\alpha=1}^{M} \frac{Z_{\alpha}}{r_{i\alpha}}$$
(1.4)

The electron-nuclear attraction can be considered as an external potential on the electrons due to the nuclei. Rewriting Equation 1.2 with only the terms corresponding to the electrons, we have

$$\hat{H} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ee} = \sum_{i=1}^N (-\frac{1}{2}\nabla^2) + \sum_{i=1}^N v(r_i) + \sum_{i(1.5)$$

Looking at Equation 1.5, the first and third terms will be the same for any *N*-electron system; it is only the middle term that will be system specific as determined by the identity of the nuclei. Thus, knowledge of v determines the Hamiltonian for a given number of electrons. The first Hohenberg-Kohn theorem, produced in 1964, states that this external potential is uniquely determined by the electron density.<sup>9</sup> As the number of electrons can also be determined through integrating the density:

$$\int dr \rho(r) = N \tag{1.6}$$

both N and v can be calculated using  $\rho$ , which in principle allows calculation of the wavefunction from the Hamiltonian and thus any ground-state property. This is an immensely powerful statement, as it signifies that the three-dimensional electron density ultimately provides equivalent information to the many-dimensional wave function.

The second Hohenberg-Kohn theorem provides a variational principle for determining the electron density. As the density must determine all properties of the system per theorem 1, let us write the operators in Equation 1.5 as functionals of  $\rho$ :

$$E[\rho] = T_e[\rho] + V_{en}[\rho] + V_{ee}[\rho] = F[\rho] + V_{en}[\rho]$$
(1.7)

The  $F[\rho]$  term is said to be a universal functional and encapsulates the kinetic energy and electron-electron repulsion terms. The second Hohenberg-Kohn theorem states that any trial density (subject to  $\tilde{\rho}(r) \ge 0$ ,  $\int dr \tilde{\rho}(r) = N$  and the condition that  $\rho$  must correspond to some external potential v)<sup>10</sup> provides an upper bound to the ground state energy:

$$E_0[\rho_0] \le E[\tilde{\rho}] \tag{1.8}$$

Minimizing the energy with respect to the density thus provides a route to the ground-state energy. Note, however, we do not know what the form of the universal functional  $F[\rho]$  is, which prevents us from performing this procedure in practice. We do know that  $F[\rho]$  contains the kinetic energy and electron-electron repulsion interactions, so we can write:

$$F[\rho] = T_e[\rho] + J[\rho] + E_{QM}[\rho]$$
(1.9)

where  $T_e[\rho]$  is the electron kinetic energy,  $J[\rho]$  is classical Coulomb repulsion of a charge cloud with itself and  $E_{QM}[\rho]$  is nonclassical electron repulsion, which we discuss further in the next paragraph. The  $J[\rho]$  term can be written exactly, but not the kinetic energy or nonclassical repulsion terms.

Kohn-Sham DFT, published in 1965, provides some approximations for the unknown terms within the universal functional; this set the foundation for DFT as a practical method.<sup>11</sup> In Kohn-Sham DFT, a set of noninteracting orbitals  $\{\psi_i(r)\}$  are introduced. Note that the density corre-

sponding to these orbitals can be determined through:

$$\rho(r) = \sum_{i}^{N} |\psi_i(r)|^2$$
(1.10)

For such a system, the kinetic energy is

$$T_s[\rho] = \sum_i \langle \psi_i | -\frac{1}{2} \nabla^2 | \psi_i \rangle \tag{1.11}$$

While  $T_s[\rho]$  is not the exact kinetic energy functional  $T_e[\rho]$ , Kohn and Sham set up a problem where Equation 1.11 is the exact kinetic energy. Their treatment starts with rewriting Equation 1.9 as

$$F[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho]$$
(1.12)

with  $E_{xc}[\rho]$  defined as

$$E_{xc}[\rho] = T_e[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho]$$
(1.13)

Equation 1.13 is the exchange-correlation energy, and absorbs the residual kinetic energy from the difference of  $T_e[\rho]$  and  $T_s[\rho]$  (which is presumably relatively small), as well as the nonclassical electron-electron interactions. The total energy expression in Equation 1.7 is now written as:

$$E[\rho] = T_s[\rho] + V_{en}[\rho] + J[\rho] + E_{xc}[\rho]$$
(1.14)

Kohn-Sham DFT proceeds by defining an effective external potential:

$$v_{eff}(r) = v(r) + \int dr' \frac{\rho(r')}{|r - r'|} + v_{xc}(r)$$
(1.15)

which is a sum of the electron-nuclear attraction potential, the classical Coulomb potential and the exchange-correlation potential. Using this effective potential, a set of N one electron equations can

be written:

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}(r)\right]\psi_i = \epsilon_i\psi_i \tag{1.16}$$

These equations need to be solved iteratively through a self-consistent field (SCF) treatment, similar to the Hartree-Fock equations.<sup>7,12</sup> Equation 1.16 would be an exact treatment if Equation 1.13 was known. Of course, it is not, but many density functionals provide a good approximation, and can be used to carry out the procedure Equation 1.16. The reader is directed to the excellent article by Kaplan *et al.* [13] for a review of various exchange-correlation functionals and the constraints the exact exchange-correlation functional must obey.

DFT can be used to generate classical dynamic trajectories through AIMD and QM/MM, which we provide brief reviews of in the following two sections.

#### 1.2 Ab initio molecular dynamics

AIMD simulations propagate trajectories by evaluating the forces on nuclei as calculated by an electronic structure method, such as DFT.<sup>14</sup> As these forces are evaluated on the fly, they naturally allow for bond-breaking/forming events to occur. Modern software packages can perform AIMD simulations with  $O(10^2)$  up to  $O(10^3)$  atoms on picosecond timescales, which allows for the modeling of condensed phase environments.<sup>15–19</sup>

There are two types of schemes for running AIMD simulations.<sup>20</sup> The first is typically referred to as Born-Oppenheimer molecular dynamics. The total energy in these simulations is given by the DFT energy plus the nuclei kinetic energy:

$$E = E_{DFT}(r) + \sum_{i}^{3N_{atom}} \frac{1}{2}m_i v_i^2$$
(1.17)

where we have noted that the DFT energy is dependent on the system positions. The equation of motion is written from Newton's second law:

$$m_i \ddot{r}_i = f_i = -\frac{\partial E_{DFT}(r)}{\partial r_i}$$
(1.18)

Unless mentioned otherwise, this is the form of the standard equation of motion used for all methods mentioned in this Introduction. This formulation requires a full SCF calculation at each time step, which is typically chosen to be 1.0 fs.

The other type of AIMD is referred to as Car-Parrinello molecular dynamics.<sup>21</sup> This approach treats both the nuclei and electrons as dynamic variables by assigning a fictitious time-dependence to the orbitals. This allows the minimized orbitals to be propagated in response to the movement of the nuclei, which avoids performance of a SCF calculation at each time step. The Car-Parrinello Lagrangian is:

$$\mathcal{L} = \sum_{i}^{3N_{atom}} \frac{1}{2} m_i v_i^2 + \frac{1}{2} \mu \sum_{i} \int dr |\dot{\psi}_i(r,t)|^2 - E_{DFT}(r, \{\psi(r,t)\}) + \sum_{i,j} \Lambda_{i,j} [\int dr \psi_i^*(r,t) \psi_j(r,t) - \delta_{i,j}]$$
(1.19)

where the first term is the nuclei kinetic energy, the second term is the electron kinetic energy (with  $\mu$  the fictitious electron mass), the third term is the DFT potential energy (where we have noted that the energy will depend on both the positions and the time-dependent orbitals) and the last term is a Lagrange multiplier designed to keep the orbitals orthonormal. The equations of motion are now:

$$m_i \ddot{r}_i = f_i = -\frac{\partial E_{DFT}(r, \{\psi(r, t)\})}{\partial r_i}$$
(1.20)

for the nuclei and

$$\mu \ddot{\psi}_i(r,t) = -\frac{\partial}{\partial \psi_i^*(r,t)} E_{DFT}(r, \{\psi(r,t)\}) + \sum_k \Lambda_{i,k} \psi_k(r,t)$$
(1.21)

for the electrons. The CP approach requires a smaller time step than Born-Oppenheimer MD due to the electron degrees of freedom, with typical values set to 0.1 fs.

AIMD is powerful in that dynamics are allowed to evolve from first principles: no *a priori* division of the system between reacting complex and solution needs to be made, as we shall see is required in some of the methods we discuss later. This allows it to be used for either investigating

a particular reaction path or for the *ab initio* identification of new novel reactions.<sup>22</sup> However, even with system sizes including  $O(10^3)$  atoms, which is approximately the current limit that DFT-based AIMD can handle, finite size effects can still be introduced; additionally, these simulations are only able to achieve picosecond timescales. Many interesting reactions occur at timescales beyond these, which then requires biased sampling, and the associated computational expense may hinder obtaining converged statistics for some systems.

#### 1.3 Quantum Mechanics/Molecular Mechanics (QM/MM)

QM/MM is another method for studying reactions in the condensed phase. The work on this methodology was part of the 2013 Nobel Prize in Chemistry. The first QM/MM simulation was presented in 1976 by Warshel and Levitt.<sup>23</sup> The idea of QM/MM is to partition a system into a QM region, like a reacting complex, and an environmental region, which is modeled classically. This allows for restricting the expensive QM calculations to a relatively small number of degrees of freedom for the interesting part of the system while still accounting for the extended environment. The QM region can again be modeled by DFT, while the surrounding region is modeled by force fields, which are a collection of simple analytic potential functions and parameters that describe intramolecular bond vibrations, angle bending, classical electrostatic interactions, etc. There are many different formulations of QM/MM that exist.<sup>23–32</sup> We shall only discuss QM/MM with electrostatic embedding here.

The QM/MM Hamiltonian is:

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM}$$
(1.22)

The first two terms represent the energy corresponding to the QM region and the energy corresponding to the MM region. The last term is the key part and describes the coupling between QM and MM atoms.  $\hat{H}_{QM/MM}$  in electrostatic embedding contains van der Waals interactions between QM and MM atoms, which is modeled through the force field, and electrostatic interactions be-

tween a field of point charges from the MM atoms and the QM density. This modifies the DFT calculation. The new DFT energy incorporating this external set of point charges takes the form:

$$E[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] + V_{en}[\rho] + \int dr \rho(r) v_{MM}(r)$$
(1.23)

There is now an additional term so that the QM charge density interacts with the MM point charges. The integrals can be evaluated analogously to the  $V_{en}$  term. The difficulty lies in determining how to treat the set of MM charges surrounding the reacting complex. The most common approach is to use a cutoff, in which the point charges within a selected radial distance of the QM region are included and everything beyond the cutoff is truncated. However, very large cutoffs are generally required for converging the electrostatics, and the cutoff approach leads to some artifacts, particularly for ionic systems.<sup>31,33–35</sup> Our group has developed a method for exact electrostatic embedding using the particle-mesh Ewald (PME) approach commonly found in force field MD packages, which fixes the aforementioned truncation errors.<sup>31,36</sup>

One of the advantages of QM/MM is that it allows for simulation of a reaction within a large environment, as the QM calculation is generally the rate limiting step. Thus QM/MM can simulate reactions within larger proteins or within bulk solvents.<sup>2,37,38</sup> One limitation is that the QM part again generally limits QM/MM to picosecond timescales. Also, it is difficult to model diffusive processes using QM/MM simulations, as the reaction site is usually restricted to a fixed region. Extending QM/MM to studies where the reactive complex can change during a simulation, such as proton solvation and transport, requires adaptive schemes that either add artificial forces or are non-Hamiltonian.<sup>39–41</sup>

#### 1.4 Empirical Valence Bond (EVB) Approach

The empirical valence bond (EVB) model is not an *ab initio* approach like the previously discussed methods, but can be considered as a type of reactive force field. The EVB model was introduced by Warshel in 1980 and has been used to investigate reactions such as proton transfer, electron

transfer and others within solution and protein environments.<sup>42–51</sup> The basic idea of the method is to treat the reactant and product states in a single step reaction as diagonal elements of a matrix ("diabatic states"), which can be modeled as two standard force fields. The off-diagonal elements  $(H_{ij} = H_{ji})$  couple the two states together in order to model transition states.

$$H_{PB/NN} = \begin{bmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{bmatrix}$$
(1.24)

The energy is obtained from the lower eigenvalue of the matrix. Equation 1.24 has been used for obtaining free energies through a free energy perturbation/umbrella sampling procedure.<sup>47</sup> Alternatively, one can use the EVB model to run direct dynamics simulations. Forces from Equation 1.24 can be obtained from the Hellmann-Feynman theorem:

$$F_i = -\sum_{m,n} c_m c_n \frac{\partial H_{mn}}{\partial r_i} \tag{1.25}$$

with  $c_m$  and  $c_n$  obtained from the ground-state eigenvector of the Hamiltonian.

The ansatz of the EVB model is that the off-diagonal elements of the Hamiltonian are independent of the environment; this assumption has been empirically justified.<sup>38,47</sup> The result is that the Hamiltonian can be parameterized to reproduce a gas phase *ab initio* potential energy surface and still be transferable to different solution environments. The user is free to choose the form of the off-diagonal element, as there is no prescribed function. Different EVB-based models have employed a wide range of functions for this term, ranging from exponentials to Gaussians to splines and others.<sup>42,51,52</sup> In Chapters 2 and 3, we show how to implement a neural network for the off-diagonal element, which we show provides a fit to gas-phase potential energy surfaces within chemical accuracy.

In Chapter 4, we use the multistate empirical valence bond model (MS-EVB), extensively developed by Voth and coworkers,<sup>53,54</sup> in order to simulate proton solvation and transport within ionic liquid/water mixtures. An excess proton in water will convert between Eigen and Zundel

species as part of the Grotthuss mechanism, which gives the proton an exceptionally fast diffusion coefficient in comparison to other similarly sized cations.<sup>55</sup> Additionally, the proton is delocalized through multiple solvation shells due to the hydrogen bond water network. Appropriately modeling this effect requires large system sizes. The MS-EVB model is explicitly designed to simulate these processes, making it one of the most efficient methods available for modeling proton solvation and transport. It is a nontrivial extension to the EVB model, with a much larger matrix (Equation 1.24) in order to account for all possible diabats within three solvation shells of the excess proton. It uses a novel selection algorithm in order to identify every "hydronium" state that should be included within the Hamiltonian.<sup>53</sup>

There are a few key advantages of the EVB model. For one, the fact that it can be trained to gas phase *ab initio* data and then applied to different solvent environments makes it highly practical for investigating solvation effects on a given reaction at greatly reduced computational cost in comparison to AIMD or QM/MM. Also, the EVB approach allows for the use of standard nonreactive force fields for reactant and product states; building such force fields is fairly routine with modern software, and the "diabat" model is intimately connected to how we think of reactions in terms of reactants and products.<sup>56</sup> An important disadvantage is that high quality training data must be assembled in order to parameterize this model. As we are attempting to perform direct dynamics in this thesis, many individual QM calculations are required to properly fit the Hamiltonian. This can take a significant amount of time and makes application less straightforward than the previously mentioned *ab initio* methods. Another disadvantage is that the given reaction studied must be known *a priori*. AIMD and QM/MM allow for the identification of new reactions, as bondbreaking/forming events can occur on the fly. The EVB model is fit to a specific reaction, which restricts its use as a tool for chemical discovery.

#### **1.5 Machine Learning Approaches**

Machine learning (ML) as a technique has rapidly been adopted in theoretical chemistry within the last few years. The development of computer hardware (namely GPUs) and software packages like

PyTorch and Tensorflow have allowed for the fast construction of deep learning models that can be trained to QM data.<sup>57,58</sup> MD simulations performed using these potentials can obtain properties at nearly quantum chemistry accuracy at orders of magnitude reduced computational cost. The most popular deep learning approaches are neural networks and kernel methods. We focus on neural network-based models here, as they make up the deep learning approach used in this manuscript.

The work from Behler and Parrinello [59] in 2007, which demonstrated how to generate a feature representation suitable for training a neural network to DFT data for silicon, initiated increased interest in machine learning for chemistry. Since then, there has been an extraordinary amount of progress in the field, from building ML force fields to learning new density functionals.<sup>60–68</sup> Much of the work focused on training ML force fields has been focused on gas phase or crystalline systems than for liquids; modern QM software packages allow for routinely running QM calculations on a small number of gas phase molecules, and plane wave formulations of DFT can efficiently model crystal supercells.<sup>69–71</sup> There are a few ML packages that allow for training and simulation within liquid environments, however, and we discuss two of them below.

The DeePMD package has been used to model a diverse range of large-scale systems,<sup>68</sup> including liquid water,<sup>68,72,73</sup> molten salts,<sup>74</sup> and ionic liquids,<sup>75</sup> among others. It has also been used in order to model some chemical reactions within condensed phases, such as carboxylic acid deprotonation and reactions with  $CO_2$  in molten alkali carbonate-hydroxide salts.<sup>76,77</sup> DeePMD has an interface to AIMD packages, like CP2K and Quantum Espresso, that allow it to read in DFT energy, forces and stress tensor calculations. It employs a descriptor network to convert the environment around each atom into a set of features that respect translational, rotational and permutational invariance, which is then passed to a second fitting network that makes predictions for properties like the energy. This neural network, like many others in the literature, represents the total energy of a system as a sum of atom energies:

$$E = \sum_{i} E_i \tag{1.26}$$

This theoretically allows for the model to be extensive and be used for the simulation of larger systems than those the model has been trained on (see Chapter 6 for example). Note that the ability

of the partitioning in Equation 1.26 for modeling smooth potential energy surfaces for interactions between separate molecules has been called into question and should be investigated further.<sup>78</sup>

The second method we discuss here is called FieldSchNet.<sup>79</sup> SchNet is a convolutional graph neural network. Upon its introduction, it demonstrated state of the art predictions for energy and force calculations.<sup>60</sup> FieldSchNet was created in order to include the response to external fields within the ML model. FieldSchNet adds a predicted atomic dipole as one of the features to SchNet. A molecule's response to a field is modeled using message passing neural networks, incorporating both interactions for the atomic dipole with the applied electric field and between pairs of the constructed dipoles. As one of their examples, QM/MM simulations of the Claisen rearrangement were used to train the FieldSchNet model. The NN model was able to predict energies and forces within chemical accuracy of those from QM/MM, which makes this method highly promising for future studies.

ML is powerful in that it provides highly accurate simulations for a fraction of the computational cost of quantum chemistry, as mentioned. However, there are a few issues plaguing the field today. Creation of sufficient training data is again a significant problem as with any parameterized method. Often many data points are needed for a complete training set, as most NN architectures today are not aware of basic physical laws. This requires training data generation and curation over a broad region of phase space, and ensuring all necessary regions are covered within the training set is difficult. Active learning strategies are being developed to aid in this process.<sup>80</sup> In addition, the "cutoff" nature of many ML models – only neighbors within a certain radius of each atom are included in the model – can lead to notable problems such as incorrect asymptotic behavior and unsmooth forces. Another set of issues is related to their relative computational expense. ML models are often much more expensive than force fields, in some cases by an order of magnitude. Most optimized MD software packages do not support ML potentials yet, which further hinders sampling. Progress on all of these issues will likely be made in the coming years as the field continues to grow.

#### **1.6 Thesis Organization**

The remaining chapters of this thesis include the following topics:

Chapters 2 and 3 describe the development of our reactive force field approach, named PB/NN. We study the deprotonation of EMIM<sup>+</sup> by acetate to form a *N*-heterocyclic carbene in the gas phase and in condensed phase bulk liquid environments. Chapter 4 contains an application of a currently developed reactive force field, named MS-EVB, to simulate proton transport within  $[BMIM^+][BF_4^-]$  /water mixtures. We demonstrate that combining the MS-EVB approach with *ab initio* force field provides accurate predictions with respect to experiment for properties such as the conductivity. Next, in Chapter 5 we use an *ab initio* force field to study the liquid dynamics of propylene carbonate, a commonly used electrolyte. We further demonstrate here the ability of these force fields to make largely accurate experimental properties over a wide range of thermodynamic conditions. Finally, in Chapter 6 we study the negative thermal expansion materials ScF<sub>3</sub> and CaZrF<sub>6</sub> using neural networks trained to DFT. We obtain state-of-the art theoretical predictions for the experimental properties we examine, which serves as a demonstration of machine learning approaches for the condensed phase.

#### **CHAPTER 2**

# PHYSICS-BASED, NEURAL NETWORK FORCE FIELDS FOR REACTIVE MOLECULAR DYNAMICS: INVESTIGATION OF CARBENE FORMATION FROM [EMIM<sup>+</sup>][OAC<sup>-</sup>]

The following is reproduced from Ref. [81] and [82]

#### 2.1 Abstract

Reactive molecular dynamics simulations enable detailed understanding of solvent effects on chemical reaction mechanisms and reaction rates. While classical molecular dynamics using reactive force fields allows significantly longer simulation time scales and larger system sizes compared with ab initio molecular dynamics, constructing reactive force fields is a difficult and complex task. In this work, we describe a general approach following the Empirical Valence Bond (EVB) framework for constructing *ab initio* reactive force fields for condensed phase simulations by combining physics-based methods with neural networks (PB/NN). The physics-based terms ensure correct asymptotic behavior of electrostatic, polarization, and dispersion interactions, and are compatible with existing solvent force fields. Neural networks are utilized for versatile description of short-range orbital interactions within the transition state region, and accurate rendering of vibrational motion of the reacting complex. We demonstrate our methodology for a simple deprotonation reaction of the 1-ethyl-3-methylimidazolium (EMIM<sup>+</sup>) cation with acetate to form 1-ethyl-3-methylimidazol-2-ylidene and acetic acid. Our PB/NN force field exhibits  $\sim 1 \text{ kJ mol}^{-1}$ MAE accuracy within the transition state region for the gas-phase complex. To characterize solvent modulation of the reaction profile, we compute potentials of mean force (PMFs) for the gas-phase reaction as well as the reaction within a four ion cluster, and benchmark against *ab initio* molecular dynamics simulations. We find that the surrounding ionic environment significantly destabilizes formation of the carbene product, and we show that this effect is accurately captured by the reactive force field. By construction, the PB/NN potential may be directly employed for simulations of other solvents/chemical environments without additional parameterization.

#### 2.2 Introduction

Solvent effects on chemical reactions are ubiquitous in chemistry and biology, and are an important consideration when optimizing catalytic processes. The solvent may modulate the kinetics, mechanism, and/or the selectivity of a reaction, and unraveling such effects can be difficult, particularly in the presence of other catalytic species. Computational methods are useful in this regard, and approaches that involve a quantum mechanical (QM) treatment of the reacting species with implicit or continuum description of the solvent have become increasingly refined.<sup>83,84</sup> However, going beyond such a continuum description to a full atomistic representation of the solvent generally requires molecular dynamics (MD) simulations. MD simulations have provided significant insight into a variety of biological and chemical processes, and can, in principle, provide similar insights into the mechanisms associated with solution phase reactions.<sup>85-90</sup> However, MD simulations typically rely on classical force fields to describe the potential energy of the system, and these force fields generally do not describe bond-breaking/bond-forming processes, preventing their use for studying reactions. Quantum mechanical (QM) simulation methods, such as *ab initio* molecular dynamics (AIMD) or quantum mechanical/molecular mechanics (QM/MM), explicitly treat the electronic degrees of freedom involved in a reaction, but are computationally costly and are limited to small time scales and/or system sizes.<sup>19,37</sup> Furthermore, statistical sampling of the reaction coordinate and solvation environment is limited with these QM approaches, which may preclude their application to certain problems of interest.

Reactive force fields can potentially be used to simulate reactions in solution at a greatly reduced computational cost compared to either AIMD or QM/MM methods. A variety of such force fields exist including ReaxFF, COMB, REBO and others,<sup>5,91–94</sup> often utilizing the concept of bond-order to describe bond-breaking/forming. These force fields are designed with the goal of developing general parameters that allow for study and discovery of reactions in a variety of environments. Parameterization of such reactive force fields is intrinsically complex, however, due to the fact that transition states cannot be expressed as low-order expansions of either the reactant or product molecular properties. This is in contrast to standard (non-reactive) force fields, that make extensive use of harmonic approximations and expansions of molecular properties (charges, polarizabilities, etc.) to model intermolecular interactions. The consequence is that reactive force fields often utilize complex functional forms requiring numerous parameters, making parameter fitting/optimization difficult. Additionally, system-specific functions may be required, inhibiting the development of a general framework for modeling chemical reactions.<sup>95</sup> Achieving the desired level of accuracy/stability for these types of force fields is an ongoing challenge.<sup>96,97</sup>

The empirical valence bond (EVB) method, originally developed by Warshel,<sup>42,47</sup> represents another force field approach capable of simulating reactions within condensed phase environments. In contrast to previously discussed approaches, EVB Hamiltonians are typically parameterized on a reaction-specific basis, requiring new parameterization for each different reaction of interest. The intent is thus primarily to investigate solvent effects on a particular target reaction, and not to explore or discover new or unknown chemical reactions. Within the EVB approach, a Hamiltonian is constructed with diagonal elements that represent the reactant/product/intermediate states, and off-diagonal coupling elements mediate transitions between these states. The diagonal terms typically employ standard functional forms and may consist entirely of existing force fields, while the off-diagonal elements are system-specific and generally fit to reproduce the gas-phase potential energy surface. A common ansatz of the EVB approach is that solvent effects are included purely within the diagonal elements of the Hamiltonian; within this ansatz, an EVB Hamiltonian may thus be used to study a particular reaction within different chemical environments.<sup>38,47,98</sup> Note, however, that other approaches incorporate solvent degrees of freedom in the off-diagonal elements as well.<sup>99</sup> EVB and similar approaches have been used to investigate a variety of applications, such as enzyme catalysis,<sup>44,100</sup> proton transport in water,<sup>53,54,101,102</sup> and various other solution-phase reactions.<sup>47,103–106</sup> A particular difficulty is that there is no consensus for the functional form of the off-diagonal elements; previous work has employed constants, exponentials, distributed Gaussians
and spline fits.<sup>42,47,51–53</sup> Parameterization of EVB models to *ab initio* potential energy surfaces would require sufficiently versatile functional forms for the off-diagonal elements in order to generally achieve chemical accuracy within the transition state region.

Recent advances in machine learning and neural network parameterization offer new avenues for progress in reactive force field development. Particularly, these methods provide a general and versatile framework for accurately rendering "difficult-to-capture" regions of high dimensional, potential energy surfaces. Artificial neural networks (NNs) are universal function approximators, with demonstrated ability to accurately model chemical bond-breaking and forming processes.<sup>64</sup> Various neural network architectures have successfully reproduced reaction potential energy surfaces, such as PhysNet, DeepPot and others.<sup>62,107,108</sup> However, there are some key limitations associated with using neural network potentials to simulate reactions in solution. Most commonly, neural network potentials use a sum of atomic energies to represent the total energy of the system, generated from a set of descriptors that encode the atomic environment.<sup>109</sup> Descriptors are computed with a distance-based cutoff,<sup>109</sup> which may prevent encoding of the correct long-range physics, particularly for electrostatic interactions. Indeed short-range truncations in both physicsbased potentials,<sup>110-112</sup> as well as in machine learning contexts,<sup>113</sup> have been shown to cause unphysical artifacts in MD simulations. A related issue is the limited extrapolation capability of NNs, with unreliable predictions outside of the scope of the training data.<sup>64</sup>. A promising solution is to restrict the required range of the neural network by supplementing with physics-based terms, such as explicit Coulombic interactions.<sup>62</sup> Parameters for the physics-based terms can either be developed from direct quantum chemical calculations, or may themselves be predicted with machine learning techniques.<sup>114,115</sup> Further considerations for NN potentials include the large amount of training data required for parameterization, as well as simulation speed; while orders of magnitude faster than QM calculations, NNs are still generally an order of magnitude slower than the empirical reactive force fields mentioned previously.<sup>5</sup>

In this work, we demonstrate a general method for developing *ab initio*, reactive force fields by combining both physics-based and neural network terms within an EVB framework. The reaction

that we consider is N-heterocyclic carbene (NHC) formation from 1-ethyl-3-methylimidazolium (EMIM<sup>+</sup>) cation and acetate (OAc<sup>-</sup>) anion; these ions constitute the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid (IL). In this reaction, EMIM<sup>+</sup>undergoes deprotonation by acetate to form 1-ethyl-3-methylimidazol-2-ylidene (NHC) and acetic acid (AcOH):

$$EMIM^+ + OAc^- \Longrightarrow NHC + AcOH$$
 (2.1)

NHCs are highly reactive catalysts and rank among the most important tools in organocatalysis.<sup>116,117</sup> Imidazolium-based ionic liquids are commonly used solvents in organic chemistry due to their low vapor pressure and thermal/chemical stability;<sup>118</sup> because of the acidic proton of the imidazolium ring, there has been interest in using these ILs as both reagents and solvents.<sup>119</sup> A strong base or electrochemical reduction is often used to carry out the deprotonation of the imidazolium ring.<sup>120,121</sup> However, recent experimental work has suggested spontaneous NHC formation in ILs such as [EMIM<sup>+</sup>][OAc<sup>-</sup>] due to deprotonation of the imidazolium ring by weak base counterions.<sup>119,122–124</sup> While the formation of neutral species from two ions would seem unfavorable in ionic media, several NHC-catalyzed reactions have been shown to occur in imidazolium acetate ILs such as benzoin condensation, which may suggest carbene presence;<sup>125–127</sup> however, it has been pointed out there are alternative reaction pathways for formation of these products that do not involve direct carbene formation.<sup>123</sup> NHC content in these ILs has been characterized with cyclic voltammetry, yet interpretation of these results is somewhat inconclusive as the electrode can act as a carbene trap.<sup>123,128–130</sup> Due to the short lifetimes of any produced NHCs, it is difficult to devise experiments that directly measure carbene formation in ILs.

Understanding NHC formation within imidazolium-based ionic liquids is important for a variety of applications due to the innate reactivity of carbene species. As ILs are continually used as electrolytes in energy-storage devices, it is important to know the extent to which the liquid is chemically stable.<sup>131</sup> Presumably, more native NHC content would lead to reduced chemical stability due to facile degradation pathways catalyzed by carbene intermediates. This could be either beneficial or disadvantageous, depending on the application. For example, imidazolium ILs are used in hypergolic fuels,<sup>132</sup> and NHCs are thought to be involved in important steps of the spontaneous ignition process.<sup>133</sup> Better knowledge of the solvent/environmental conditions that lead to enhanced NHC content would be important for optimizing such processes. Beyond these energy and electrochemical applications, better characterization of NHC formation may allow ILs to find further use in novel organocatalysis applications.

Due to the difficulty of direct experimental measurement, computational approaches provide an important tool for investigating relative NHC stability in ILs and other environments. For example, equilibrium coefficients (e.g. for the reaction in Equation 2.1) can be derived from reaction free energies computed from enhanced sampling molecular dynamics approaches; this of course assumes an approach that allows modeling the chemical reaction in the condensed phase (vide supra). As an illustration, in Figure 2.1 we show the free energy surface for Equation 2.1 in the gas-phase, projected along the bond breaking (C-H distance) and bond forming (O-H distance) coordinates (see Results for calculation details). This free energy surface highlights several important aspects about the carbene formation reaction. The reaction is observed to be barrierless, as is typically true for similar proton transfer reactions (in absence of solvent effects). The reactants (EMIM<sup>+</sup> and acetate) are more stable than the products (NHC and acetic acid), but only by a relatively modest  $\sim 10 \text{ kJ mol}^{-1}$  which implies significant fraction of all species in the gas phase. The global minimum occurs at  $r_{CH} \approx 1.1$  Å and  $r_{OH} \approx 1.8$  Å, within a broad minima basin that spans  $\sim 0.75$  Å along the O–H coordinate. Because of the barrierless transition state, there is no local minima basin associated with the products, but rather they are inferred by approximate distance criteria of  $r_{CH}$  > 1.6 Å and  $r_{OH}$  ~ 1.1 Å. In the Supporting Information, we compare this free energy surface to the potential energy surface, and entropic effects are found to make only a relatively small quantitative contribution to the gas-phase reaction profile. Note that while Figure 2.1 indicates significant formation of carbene products in the gas-phase, large solvation energies of the reactant ions will substantially shift the equilibrium towards the reactants in solution. Developing a reactive force field that allows investigation of such solvent/environmental effects on the reaction



Figure 2.1: Free energy as a function of C–H and O–H bond distance for the deprotonation of EMIM<sup>+</sup>by acetate in the gas-phase. Arrows are drawn to the reactants and products to indicate their respective location on the surface.

free energy profile is the primary goal of this work.

In this paper, we develop a reactive force field for the carbene formation reaction (Equation 2.1) that combines both *physics-based* and *neural network* terms within an EVB approach. While neural networks and physics-based potentials have been combined before,<sup>134–136</sup> there are specific considerations for doing this within reactive force field development. Our approach, which we term "PB/NN potentials", is entirely general and is applicable to other chemical reactions in condensed phase environments. The ansatz of our PB/NN construction is the following: Physics-based terms are utilized for all long-range intermolecular interactions, which ensures correct asymptotic

behavior and facilitates transferability to various condensed-phase, solution environments. For the reacting complex, neural networks are utilized to describe all molecular vibrations in order to avoid inaccuracies associated with standard harmonic functional forms. In the transition state region, physics-based terms break down at close contact distances (e.g. Coulombic interactions), which is fixed by introducing short-range neural networks as residual corrections. Finally, the reactant and product "states" are coupled within an EVB-like Hamiltonian using a neural network for off-diagonal coupling which improves versatility and accuracy compared to analytic functional forms. The final PB/NN potential exhibits an accuracy of ~ 1 kJ mol<sup>-1</sup> MAE for the reacting complex within the transition state region, and can be directly used in condensed phase, molecular dynamics simulations. Overall, more than 1 million electronic structure calculations were used to parameterize the final PB/NN potential. It is expected that this amount of training/parameterization data could be greatly reduced in future work, as we made no effort to optimize training data sampling/generation.

The manuscript is structured as follows: In the Methods section, we explain the motivation and construction of the PB/NN force field and our procedure to generate the necessary training data for parameterization. In the Results section, we validate each component of the force field, and present accuracy benchmarks by comparing to free energy profiles generated by AIMD simulations. While the overarching motivation is to explore solvent/environment effects on the carbene formation reaction, in this work we restrict analysis to systems that can be benchmarked with AIMD. Thus the "solvent" environment considered is a simple ionic cluster consisting of two additional solvating ions in addition to the reactant species. Investigation of this reaction in more complex condensed-phase environments will be the subject of future work.

### 2.3 Methods

We first describe the general form of our PB/NN force field and the procedure for training and validating its individual components. We then discuss the electronic structure calculations used to generate the training data sets for the different force field components. Finally, we describe in

detail the functional forms of the individual components of the force field and specific process for fitting each component.

#### 2.3.1 General Functional Form

Our PB/NN force field has the generic functional form of a 2x2 Hamiltonian written in the basis of diabatic "reactant" and "product" states. This is similar to the empirical valence bond (EVB) method, multiscale-reactive molecular dynamics (MS-RMD) and other reactive force field approaches.<sup>42,47,50–52,137,138</sup> For a single step reaction, the total Hamiltonian and ground state energy are:

$$H_{PB/NN} = \begin{bmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{bmatrix}$$
(2.2)

$$E_g = \frac{1}{2} [(H_{11} + H_{22}) - \sqrt{(H_{11} - H_{22})^2 + 4H_{12}^2}]$$
(2.3)

The atomistic forces on each atom *i* are obtained using the Hellmann-Feynmann theorem in terms of the ground-state eigenvector  $(c_1, c_2)$ 

$$F_i = -\sum_{m,n} c_m c_n \frac{\partial H_{mn}}{\partial r_i} \tag{2.4}$$

In Equation 2.2, the diagonal elements,  $H_{ii}$ , correspond to the "reactant" and "product" diabatic states, and the off-diagonal element ( $H_{ij} = H_{ji}$ ) represents the coupling between these states. For the carbene formation reaction studied in this work, we refer to EMIM<sup>+</sup>/acetate configuration as diabat 1, represented with  $H_{11}$ , and NHC/acetic acid configurations as diabat 2, represented with  $H_{22}$ . Throughout the method discussion, we use the terminology "solute" to refer to the reacting complex, and "solvent" to refer to an arbitrary chemical environment encompassing all non-reacting species. The diagonal elements  $H_{11}$  and  $H_{22}$  consist of the following terms, with subscripts identifying whether the energy contribution is specific to the reacting "solute" or the "solvent", and superscript indicating the type of interaction:

$$H_{ii} = E_{Solute}^{Morse} + E_{Solute}^{Intra,NN} + E_{Solute}^{Inter,NN} + E_{Solvent}^{Bonded,FF} + E_{Solute,Solvent}^{Nonbonded,FF} + E_{Solute}^{\alpha}$$
(2.5)

The terms  $E_{Solute}^{Morse}$  and  $E_{Solute}^{Intra,NN}$  constitute the intra-molecular energy of the reacting complex in the specific diabatic topology;  $E_{Solute}^{Morse}$  is a Morse potential describing the chemical bond(s) that break/form during the reaction, and  $E_{Solute}^{Intra,NN}$  represents all (non-reactive) intra-molecular vibrational energy contributions described by neural networks. The term  $E_{Solvent}^{Bonded,FF}$  is the contribution from standard harmonic bond, angle, and dihedral force field terms for all solvent molecules. The  $E_{Solute,Solvent}^{Nonbonded,FF}$  refers to physics-based, force field terms accounting for all solvent-solvent, solutesolvent and solute-solute non-covalent, intermolecular interactions. These interactions must be corrected at short-range for the reacting species (e.g. transition state region) where the physicsbased terms break down, and this correction is given by  $E_{Solute}^{Inter,NN}$  modeled with a neural network. We note the subtle distinction between "Intra", "Inter" and "Bonded", "Nonbonded" labeling of the terms:  $E_{Solute}^{Intra,NN}$  refers to the full intra-molecular energy of the solute molecules (albeit with/without the bond dissociation  $E_{Solute}^{Morse}$  term), whereas  $E_{Solvent}^{Bonded,FF}$  is only part of the intramolecular energy of solvent molecules, as there is additionally the usual contribution from nonbonded interactions  $(E_{Solute,Solvent}^{Nonbonded,FF})$  for atoms separated by three or more bonds. Finally,  $E_{Solute}^{\alpha}$  is the gas-phase electronic energy of the isolated monomers at infinite separation. The rationale and specific expressions used for each of these terms will be explained in further detail. It is important to note that solvent effects on the reaction energetics are mediated entirely by the  $E_{Solute,Solvent}^{Nonbonded,FF}$ term, which in principle is given by an existing classical force field. This follows the ansatz of the EVB method, and implies that parameterization of the reactive force field (e.g. off-diagonal elements) can be done entirely based on the gas-phase potential energy surface of the reacting complex.38,47

For the  $H_{ij}$  off-diagonal coupling element, we utilize a neural network for enhanced accuracy and generality. We note that an alternative, analytic functional form would likely limit the extension of the model for more complex reactions, even if sufficient for the simple proton transfer reaction considered here.<sup>53</sup>. Utilizing a neural network for  $H_{ij}$  enables versatile description of complex energy dependence in a high dimensional coordinate space, and thus should be applicable for a wide variety of reaction coordinates. The exact structure of the  $H_{ij}$  neural network will be explained in detail in a later section.

### 2.3.2 Training Data Generation

Training data for the force field parameterization falls into three categories: 1) monomer DFT calculations, 2) dimer DFT calculations, and 3) dimer perturbation theory (SAPT0) calculations. Configurations were generated from (gas-phase) AIMD simulations of the EMIM<sup>+</sup>/acetate (NHC/acetic acid) reaction complex, which were then input to the three types of calculations. To parameterize a reactive force field, it is essential to generate sufficient training data in the transition state region. Because the carbene formation reaction is barrierless (Figure 2.1), the transition state (and proton transfer reactions) is adequately sampled by direct AIMD simulations of 1-10 ps trajectories. For a more complex reaction with significant reaction barrier, enhanced sampling approaches would be necessary to generate sufficient training data spanning the transition state region.

For the AIMD simulations, the QUICKSTEP method in the CP2k software package was used,<sup>19</sup> with PBE-D3(BJ) density functional and a mixed Gaussian/plane-wave basis set.<sup>139–141</sup> The basis set consisted of the aug-DZVP Gaussian atomic basis set, a cutoff of 280 Ry to expand the auxiliary electron density and Goedecker-Teter-Hutter pseudopotentials for the core electrons.<sup>142</sup> The Pulay mixing scheme is used for SCF convergence, with a convergence criterion of  $10^{-7}$  au.<sup>143</sup> We run NVT simulations using a Langevin thermostat at 300, 400, 500 and 600K of each of the individual monomers and the dimer, with the higher temperatures allowing sampling of higher energy configurations. A 0.005 fs<sup>-1</sup> friction coefficient was used for these simulations, with a 1.0 fs timestep for integrating the nuclear equations of motion.

Table 2.1: Summary of training data sets, including type and number of calculations in each set. Each data set is used to parameterize a particular neural network within the force field; the neural network architecture is listed as well as the number of atomic degrees of freedom that the neural network depends on (note that the total degrees of freedom is equal to 3N-6, with N equal to the number of atoms).

dataset no.	reference data	# of calculations	Neural Network	Architecture	Atomic DOF
1	EMIM <sup>+</sup> DFT energies/forces	403625	$CNN_{(EMIM^+)}$	SchNet	19
2	Acetate DFT energies/forces	406173	$CNN_{(OAc^{-})}$	SchNet	7
3	NHC DFT energies/forces	404705	$CNN_{(NHC)}$	SchNet	18
4	Acetic acid DFT energies/forces	421764	$CNN_{(AcOH)}$	SchNet	8
5	Dimer DFT energies/forces	226075	$H_{12}$	AP-Net	26
6	EMIM <sup>+</sup> /acetate SAPT0 energies	427004	$H_{11}: E_{Solute}^{Inter,NN}$	AP-Net	26
7	NHC/acetic acid SAPT0 energies	265343	$H_{22}: E_{Solute}^{Inter,NN}$	AP-Net	26

The Psi4 electronic structure package was then utilized to generate training data from these AIMD configurations.<sup>69</sup> For the DFT calculations, we use the PBE-D3(BJ)/aug-cc-pVTZ level of theory. Calculations employing the SAPT0/aug-cc-pVTZ flavor of symmetry adapted perturbation theory were conducted for geometries in the dimer training set; here, the dimers were partitioned into either diabat 1 (EMIM<sup>+</sup>/acetate) or diabat 2 (NHC/acetic acid) configurations for defining the SAPT0 monomer wavefunctions. Note that the commonly employed  $\delta E_{HF}$  term in SAPT0 is excluded, for reasons discussed later. Table 2.1 summarizes the different training datasets that were constructed, including the number of geometries/configurations in each dataset, the type of calculation, and whether energies/forces were calculated. In addition, this table lists the type/architecture of the neural network that each dataset was used to train, along with the number of atomic degrees of freedom built in to the corresponding neural network; such details are explained in the following sections. Throughout the remainder of the work, each dataset will be referenced by its corresponding number given in the first column of Table 2.1. We note that over 1 million monomer/dimer geometries/configurations were used in the training data generation. Because we made no attempt to optimize or reduce the amount of training data, it is likely that significant reductions could be achieved with more efficiently designed data generation approaches.

## 2.3.3 Force Field Components

# Bonded Force Field Terms ( $E_{Solute}^{Morse}$ and $E_{Solvent}^{Bonded,FF}$ )

The bond formation/dissociation that occurs during the reaction is modeled with Morse potential(s)

$$E_{Solute}^{Morse} = D(1 - e^{-a(r-r_o)})^2$$
(2.6)

In our case, we use two Morse potentials for the bonds involved in the deprotonation reaction; one is for the C–H bond of EMIM<sup>+</sup>in diabat 1 ( $H_{11}$ ), and the other is for the O–H bond of acetic acid in diabat 2 ( $H_{22}$ ). As will be discussed in Section 2.4.1, there are distinct advantages to separating out these bond dissociation/formation coordinates from the rest of the intramolecular energy (within  $E_{Solute}^{Intra,NN}$ ). The Morse potentials provide a physical way to account for this bond breaking/forming energy, with correct asymptotic behavior, while a neural network would not correctly extrapolate to the bond dissociation limit unless explicitly trained in that region. In short, using Morse potentials for dissociation coordinates restricts the remaining intra-molecular energy term  $E_{Solute}^{Intra,NN}$  to a much smaller region of phase space (non-dissociative vibrations) which is important for stability of the neural network (and force field). Fitting the Morse potentials involves straightforward bond dissociation scans, and the resulting parameters are found in Table S1 of the Supporting Information.

The bonded terms for the solvent molecules comprising  $E_{Solvent}^{Bonded,FF}$  are standard harmonic bond, angles, and dihedral terms and thus require no elaboration. In general, these may be taken from standard, existing force fields.<sup>144–146</sup> In this particular case, the solvent that we study is EMIM<sup>+</sup> and PF<sub>6</sub><sup>-</sup> ions (Section 2.4.4), and bonded parameters are taken from previous work.<sup>147,148</sup>

# Intramolecular Neural Network ( $E_{Solute}^{Intra,NN}$ )

For the reacting "Solute" complex, we use a neural network for each monomer to capture all intramolecular energetics besides the explicit bond dissociation (the latter accounted for by  $E_{Solute}^{Morse}$ ). The motivation for this is that standard harmonic bond, angle, and dihedral potentials introduce errors on the order of ~ 10 kJ mol<sup>-1</sup> or larger for these energetic contributions near the transition state region, which would likely preclude chemical accuracy (Section 2.4.1). In Figure S2 of the Supporting Information, we show an example of the error introduced when a standard bonded potential is alternatively used for the reacting complex. In contrast, the neural networks provide very accurate renderings of  $E_{Solute}^{Intra,NN}$ , with errors on the order of ~ 0.1 kJ mol<sup>-1</sup> as demonstrated in Section 2.4.1.

The SchNet architecture is used to train convolution neural networks (CNNs) for each monomer in the reacting complex.<sup>60</sup> The CNN architecture within SchNet has previously been shown to provide accurate potential energy surfaces of isolated molecules.<sup>60</sup> Specifically,  $E_{Solute}^{Intra,NN}$  for  $H_{11}$ and  $H_{22}$  are given by

$$H_{11}: E_{Solute}^{Intra,NN} = CNN_{(EMIM^+)}f(r_{CH}) + CNN_{(OAc^-)}$$

$$(2.7)$$

$$H_{22}: E_{Solute}^{Intra,NN} = CNN_{(NHC)} + CNN_{(AcOH)}f(r_{OH})$$
(2.8)

with

$$f(r) = \frac{1}{(e^{\beta(r-\mu r_o)} + 1)}$$
(2.9)

The output of the CNN is thus multiplied by a Fermi-Dirac like damping function for the monomers with dissociative bonds. The argument of the damping function is the length of the dissociating bond, e.g.  $r_{CH}$  for EMIM<sup>+</sup> and  $r_{OH}$  for acetic acid. The motivation is that these neural networks will predict uncontrolled values in the dissociated regime which is outside the scope of the training data. Introduction of the damping function does not adversely effect the accuracy of the force field; in the damped regime, the bond dissociation energy  $E_{Solute}^{Morse}$  is very large, and thus the contribution of this diabat to the total energy will be small. The parameters of the damping function are found in Table S2 of the Supporting information. Specifically,  $r_o$ , is the equilibrium bond length, and  $\mu$  is a "scaling factor" for which we assign the value  $\mu = 2.25$ , so that the damping occurs at just over twice the equilibrium bond length.

Datasets 1-4 in Table 2.1 are used to train the monomer CNNs in Equations Equation 2.7 and

Equation 2.8. Before training, all energies are converted to relative values by subtracting off the energy of the minimized monomer geometry. For the monomers with an  $E_{Solute}^{Morse}$  contribution, we fit to the energy difference between the QM energy and Morse energy. Because the damping function doesn't "turn on" until approximately twice the equilibrium bond length, the  $CNN_{(EMIM^+)}$  and  $CNN_{(AcOH)}$  which involve bond dissociation coordinates must be fit to an expanded dataset. To fit these specific monomers, Datasets 1 and 4 are supplemented with "dissociation" configurations from Dataset 5 in which the appropriate bond is stretched to distances longer than twice the equilibrium bond length. The importance of adding such additional "dissociation" data will be further discussed in the Results section. The hyperparameters used to train the CNNs are similar to default SchNet values,<sup>60</sup> and are explicitly discussed in the Supporting Information. We use the combined energy and force mean-squared error loss function from Schütt *et al.* [60] in order to train the neural networks (shown in Equation S1 in the Supporting Information).

# Nonbonded Force Field ( $E_{Solute,Solvent}^{Nonbonded,FF}$ ) and Intermolecular Neural Network ( $E_{Solute}^{Inter,NN}$ )

In principle, any standard force field could be utilized for the diabatic  $E_{Solute,Solvent}^{Nonbonded,FF}$  terms, which contain contributions from all nonbonded interactions. We choose to use the SAPT-FF force field<sup>149,150</sup> for  $E_{Solute,Solvent}^{Nonbonded,FF}$ , which has been previously developed based on symmetry adapted perturbation theory calculations. The SAPT-FF force field utilizes Drude oscillators to model electronic polarization, meaning that polarization is explicitly included in the  $E_{Solute,Solvent}^{Nonbonded,FF}$  terms; the complete functional form can be found in the work of McDaniel and Schmidt [150]. SAPT-FF has been demonstrated to be very accurate for organic ions constituting ionic liquids,<sup>151,152</sup> such as the species studied in this work.

The SAPT-FF parameters for each monomer are listed in Tables S3-S6 of the Supporting Information. A virtual site was added to acetate in order to better describe the electrostatic interactions, as described in the Supporting Information. The EMIM<sup>+</sup> and NHC monomers share the same parameters, except for those associated with the reactive  $sp^2$  carbon; we adjusted these parameters in order to better fit NHC/acetic acid SAPT0 energies, which we show in Figure S10 of the Supporting Information. All intra-molecular, nonbonded interactions within the reacting "Solute" complex are purposely excluded, as the full intra-molecular energy is captured in the  $E_{Solute}^{Intra,NN}$  term. For the solvent, 1-4 and greater intra-molecular nonbonded interations are included in the standard way. It is well-known that Drude oscillator models exhibit "polarization catastrophe" at short distances/large electric fields.<sup>153</sup> Without modification, this would inevitably be problematic for a reactive force field in the transition state region at short atom-atom contact distances. To fix this issue, we introduce an anharmonic restraining potential on each Drude oscillator in order to avoid polarization catastrophe, as described in Huang *et al.* [154]. Further details of the Drude oscillator approach are given in the Supporting Information.

Standard force fields account for long-range interactions (i.e. electrostatics, dispersion), and short range repulsion, and are designed to be accurate at van der Waals (VDWs) contact distances. At distances much shorter than the VDWs contact distance, these force fields will inevitably break down due to their limiting functional forms.<sup>155,156</sup>. To ensure accuracy of the diabatic  $H_{11}$  and  $H_{22}$  terms within the transition state region, we thus supplement  $E_{Solute,Solvent}^{Nonbonded,FF}$  with a "correction" term,  $E_{Solute}^{Inter,NN}$ . The  $E_{Solute}^{Inter,NN}$  term is only for the reacting "Solute" species, and is modeled by a neural network; it is designed to compensate for breakdown of the analytic functional forms within  $E_{Solute,Solvent}^{Nonbonded,FF}$  that occur at distances significantly shorter than the VDWs contact distance (i.e. within transition state region).

Care must be taken to parameterize  $E_{Solute}^{Inter,NN}$  consistently with the diabatic representation ansatz of the force field (Equation 2.2). As  $E_{Solute}^{Inter,NN}$  enters into the diagonal elements ( $H_{11}$ ,  $H_{22}$ ), it should not include orbital interactions related to breaking/forming chemical bonds, the latter of which are accounted for by the off-diagonal coupling,  $H_{12}$ . The standard approach for parameterizing diabatic Hamiltonian terms is using constrained DFT.<sup>38,47,157</sup> Here, we propose to use symmetry adapted perturbation theory instead, in the form of SAPT0. Our choice of SAPT0 is for two reasons: First, it is consistent with the SAPT-FF force field used for  $E_{Solute,Solvent}^{Nonbonded,FF}$ , and second, it provides a mechanism for interpreting solvent effects on the reaction profile utilizing the explicit energy decomposition.<sup>158</sup> We note that SAPT0 generally incorporates a variational "delta-Hartree Fock" ( $\delta E_{HF}$ ) correction to the interaction energy.<sup>159</sup> Here, we exclude the  $\delta E_{HF}$  term to avoid variational energy contributions in the diabatic parameterization.<sup>156</sup> The criteria for a "good" diabatic state definition in the context of a reactive force field is the capability of accurately capturing solvation effects on the reaction profile; this is benchmarked for our approach in Section 2.4.4.

The  $E_{Solute}^{Inter,NN}$  term is then parameterized to the residual difference between SAPT0 energies and the energy contribution of  $E_{Solute,Solvent}^{Nonbonded,FF}$ . Specifically datasets 6 and 7 listed in Table 2.1 are employed for this parameterization. A subtle issue is that SAPT0 was used to generate datasets 6 and 7, while DFT-SAPT was primarilly used in prior parameterization of SAPT-FF for the  $E_{Solute,Solvent}^{Nonbonded,FF}$  term;<sup>147,150</sup> this leads to a minor discrepancy between levels of theory. For the calculations in this work, we have utilized SAPT0 purely for reasons of computational efficiency, and the small errors introduced relative to a higher-level method will be discussed within the context of analyzing our results (Section 2.4.4).

We utilize a different neural network architecture for  $E_{Solute}^{Inter,NN}$  than was used for the  $E_{Solute}^{Inter,NN}$  terms in Section 2.3.3.2. The AP-Net neural network architecture<sup>78</sup> is used for  $E_{Solute}^{Inter,NN}$ , which partitions the total energy into contributions from *atom pairs* (rather than *individual atoms*) and thus is compatible with typically used pairwise interaction energy expansions. Importantly, we find that AP-Net provides a much smoother potential energy surface for intermolecular interactions compared to other neural network architectures (see Supporting Information), which is an essential requirement for use in MD simulations. We modify the AP-Net output slightly from the original work,<sup>78</sup> to explicitly restrict the energy contribution to configurations where the molecules are at close range ( $\leq 4$ Å):

$$E_{Solute}^{Inter,NN} = \sum_{i \in A, j \in B} (E_{ij} f_c(r_{ij})) f(r_{bond})$$
(2.10)

$$f_c(r_{ij}) = \begin{cases} \frac{1}{2} (\cos(\frac{\pi r_{ij}}{r_c}) + 1) & r_{ij} \le r_c \\ 0 & r_{ij} > r_c \end{cases}$$
(2.11)

Here,  $E_{ij}$  is the pairwise energy contribution predicted by the neural network (equivalent to  $\Delta E_{ab}$  labeled in Glick *et al.* [78]) and the cutoff function  $f_c(r_{ij})$  is applied to pairs of atoms in the dimer and is the same form as that commonly used for gathering the atomic neighbors in many neural network frameworks.<sup>78,109</sup> For example, if an atom pair has a distance larger than  $r_c$ , then the particular pairwise energy is zero. This restricts  $E_{Solute}^{Inter,NN}$  to only correct the interaction energy between dimers at relatively close distances. The  $f(r_{bond})$  is the same Fermi-Dirac damping function as in Equation 2.9, and is a function of the dissociating bond length of the particular diabat. This additional damping function zeroes the  $E_{Solute}^{Inter,NN}$  energy contribution when the bond has stretched too far (and the system has switched to the opposite diabat). The parameters for the Fermi-Dirac functions used for the  $E_{Solute}^{Inter,NN}$  neural networks are given in Table S7 of the Supporting Information.

The majority of the hyperparameters used in the construction of the  $E_{Solute}^{Inter,NN}$  neural network are the default values given in Glick *et al.* [78]. The small differences in our present implementation are as follows. While the ReLU activation function was used in the original AP-Net paper, we use the shifted softplus activation function as implemented in SchNet.<sup>60</sup> We set  $\rho$  in Equation S1 to 1, as we are only fitting energies in this case. We also use a batch size of 100 for training as we used for training the intramolecular neural networks, and we used the same learning rate decay procedure as mentioned earlier, except we start with an initial learning rate of 5.0 x  $10^{-4}$  due to observing improved results. Our train-test-validation split was the same as was used for the intramolecular neural networks.

## *H*<sub>12</sub> Neural Network

We lastly discuss the  $H_{12}$  coupling term, which is also modeled with a neural network. As noted by Chang and Miller [138], it is only necessary to define  $H_{12}$  for the intermediate transition state region between reactants and products; the system is well-described by the diagonal elements of the Hamiltonian for asymptotic reactant and product configurations, and  $H_{12}$  may be set to zero. We use a similar pairwise representation and neural network architecture as AP-Net to model  $H_{12}$ ; however, we no longer restrict the pairs of atoms to be in different monomers, as there is no way to distinguish between monomers while the reaction is occurring.

$$H_{12} = \left( \left| \sum_{i,j \neq i} E_{i,j} \right| \right) f_c(r_{CH}) f_c(r_{OH})$$
(2.12)

Here, labels 'i' and 'j' indicate any atom in the reacting complex. The absolute value of the sum of pairwise energies from the neural network is taken in order to ensure that  $H_{12}$  is strictly positive. The neural network output is also multiplied by Fermi-Dirac functions with both bond distances,  $r_{CH}$  and  $r_{OH}$ . This ensures that the off-diagonal element will be zero and that the total energy will follow the lowest energy diabatic state in the asymptotic limit. The parameters for the Fermi-Dirac function are the same as those used for  $E_{Solute}^{Inter,NN}$  Our procedure for identifying configurations from the DFT dimer dataset 5 to use for training  $H_{12}$  is described in the Supporting Information Figure S8, and we also explain how to obtain energies and forces from these configurations suitable for training  $H_{12}$  in the Supporting Information. The structures in dataset 5 have geometries that fall within the range of the Fermi-Dirac damping functions i.e. have C–H distances less than the cutoff distance for the C–H damping function and O–H distances less than the cutoff for the O– H damping function. We use the same loss function as was used for training  $E_{Solute}^{Intra,NN}$ , and the hyperparameters for the neural network are the same as were used for training  $E_{Solute}^{Intra,NN}$ .

#### 2.4 Results and Discussion

We organize the discussion of our results as follows. In Sections subsection 2.4.1 and subsection 2.4.2, we first discuss accuracy benchmarks of the different force field components. Within this discussion, the rationale for our particular choice of force field components is elucidated by comparison to alternative choices/models. In Section 2.4.4, we evaluate the accuracy of the total

PB/NN force field within the transition state region of the reacting complex. Lastly in Section 2.4.4, we evaluate the ability of the force field to accurately capture solvent effects on the reaction profile. This is benchmarked by computing reaction free energies with and without the presence of a "model" solvation environment, and comparing to explicit AIMD free energy predictions.

# 2.4.1 Intramolecular Neural Network: $(E_{Solute}^{Intra,NN})$

In Figure 2.2, we show the energy predictions of the intramolecular neural networks compared to DFT relative energies for each monomer involved in the reaction. The test sets have approximately 20,000 structures for each monomer sampled from datasets 1-4 in Table 2.1, with the test data not included in the training data. The energy from the Morse potential is added to  $E_{Solute}^{Intra,NN}$  for EMIM<sup>+</sup> and acetic acid to obtain the total energy. The energy MAEs for each figure are listed in the figure; they are also listed in Table S8 along with the force MAEs. The largest energy MAE is 0.34 kJ mol<sup>-1</sup> for EMIM<sup>+</sup>, as well as the largest force MAE of 0.85 kJ mol<sup>-1</sup> Å<sup>-1</sup>. These errors are comparable to those computed by SchNet neural networks for similarly sized molecules.<sup>60</sup> For acetate and NHC, the two monomers that do not have a reacting bond, there are no significant outliers. In contrast, there are a few outliers for EMIM<sup>+</sup> and acetic acid that occur for configurations with the C–H and O–H bond dissociated, respectively. For these configurations, the Fermi-Dirac functions in Equation 2.9 damp the neural network prediction and lead to the associated errors. Note however that these outliers are at very high energy (above 400 kJ mol<sup>-1</sup>), and thus will not affect the accuracy of the full PB/NN potential (the system will have switched to the opposite diabatic state).

As mentioned, there is a significant improvement in accuracy for  $E_{Solute}^{Intra,NN}$  when using neural networks compared to typical force field expressions. This is demonstrated in Figure S2 of the Supporting Information. Standard force field expressions (harmonic bonds, angles, dihedrals) would introduce errors of at least several to tens of kJ mol<sup>-1</sup> even after optimizing the parameters. This introduced error would have one of two effects; it would either hinder the total accuracy of the force field, or these errors would be compensated when fitting the  $H_{12}$  coupling term to the

adiabatic surface. If the latter case, then  $H_{12}$  would have diminished physical meaning for the diabatic representation. At this point, one can only speculate how this would effect the ability to capture solvent effects in the condensed phase and/or the stability of the neural network.

The need for the Morse potential for EMIM<sup>+</sup> and acetic acid monomers is shown in Figure S4 and S5. We compare predicted energies for bond dissociation scans of EMIM<sup>+</sup> and acetic acid for two different neural networks trained to datasets 1 and 4 in Table 2.1; one has been trained to residual energies including the Morse potential, and the other is trained "stand-alone". While the stand-alone neural network accurately predicts the QM energy for short bond distances, it produces severe, uncontrolled errors for longer bond lengths toward the dissociation limit. Clearly, such error would be unacceptable in a working reactive force field. While presumably this issue could be fixed by introducing more training data in the dissociation region, there are practical difficulties such as SCF instability, etc., and supplementing with a Morse potential is likely the best solution. The final intra-molecular potential including the Morse potential and damped CNN accurately reproduce the bond dissociation scans, as shown in Figure S5.

We note that it is important to include stretched geometries sampled from the dimer AIMD simulations to the training datasets 1-4 in Table 2.1 for all monomers. As mentioned, the intramolecular neural networks for EMIM<sup>+</sup> and acetic acid are damped at approximately twice the equilibrium bond length, and so training data spanning this regime is included in datasets 1 and 4 (Table 2.1) to ensure stability of the neural networks.

# 2.4.2 Nonbonded Force Field $(E_{Solute,Solvent}^{Nonbonded,FF})$ and Intermolecular Neural Network $(E_{Solute}^{Inter,NN})$

As discussed in the Methods section, the  $E_{Solute}^{Inter,NN}$  term is designed to be a short-range correction to the SAPT-FF intermolecular interactions ( $E_{Solute,Solvent}^{Nonbonded,FF}$ ) for the reacting complex. In Figures S9 and S10, we show the accuracy of SAPT-FF for describing intermolecular interactions for *nonreactive configurations* of the dimer (see Supporting Information for details). It is clear that the accuracy for these configurations is fairly good, on par with previous work for this specific force field.<sup>147,150</sup> However, the accuracy of SAPT-FF breaks down for close range "reactive" configura-



Figure 2.2: PBE-D3(BJ)/aug-cc-pVTZ vs. neural network energy predictions for each monomer involved in the reaction. We have added  $E_{Solute}^{Morse}$  to the neural network prediction for EMIM<sup>+</sup> and acetic acid. Each plot shows 5000 randomly-selected structures from the test set; the MAE listed corresponds to the full test set for each monomer.

tions in dataset 5 (Table 2.1) generated from the AIMD simulations. This breakdown is expected due to limitations in functional form, as these configurations are significantly closer than the VDWs contact distance. The motivation for adding the  $E_{Solute}^{Inter,NN}$  term is to improve the diabatic energy (as compared to SAPT0) for such close distances within the transition state regime.

In Figure 2.3, we benchmark the total intermolecular interaction energy,  $E_{Solute}^{Inter,NN} + E_{Solute}^{Nonbonded,FF}$ , compared to test data taken from datasets 6 and 7 in Table 2.1. For greater insight, we have divided each test set into 3 different subsets to illustrate the accuracy at distinct regions of the potential energy surface. The three subsets are generated by computing the diabatic expansion coefficient " $c_1$ " for each configuration utilizing the final Hamiltonian. The first subset corresponds to the "Diabat 1" region, categorized by  $c_1^2 > 0.8$ ; the second subset is the "Transition State" region, corresponding to  $0.2 < c_1^2 < 0.8$ ; and the third subset is the "Diabat 2" region, corresponding to  $c_1^2 < 0.2$ . In Figure 2.3, we show the MAE of  $E_{Solute}^{Inter,NN} + E_{Solute}^{Nonbonded,FF}$  for the three regions, for both  $H_{11}$  and  $H_{22}$  force field elements. In Figure S11 of the Supporting Information, we show corresponding plots for the stand-alone force field energy  $(E_{Solute}^{Nonbonded,FF})$ ; thus comparison of Figure 2.3 and Figure S11 indicates the contribution of the neural network term,  $E_{Solute}^{Inter,NN}$ . The  $E_{Solute}^{Inter,NN}$  term greatly improves the agreement with the SAPT0 energies compared to the test sets. The effect is most significant in the transition state region and dissociation limit, where the force field is clearly expected to breakdown. Figure S11 indicates that the stand-alone force field terms  $(E_{Solute}^{Nonbonded,FF})$  of both  $H_{11}$  and  $H_{22}$  lose predictive ability in the transition state region. The neural network  $(E_{Solute}^{Inter,NN})$  substantially improves accuracy in all regions of the potential energy surface, for both  $H_{11}$  and  $H_{22}$  force field elements. The most pertinent MAE is for the particular subset in which the diabatic region and Hamiltonian element correspond, i.e.  $H_{11} \Leftrightarrow$  "Diabat 1", and  $H_{22}$  $\Leftrightarrow$  "Diabat 2"; here, the MAEs are 1.32 kJ mol<sup>-1</sup> and 0.50 kJ mol<sup>-1</sup> respectively. Furthermore, corresponding MAEs within the transition state region are 1.34 kJ mol<sup>-1</sup> and 0.73 kJ mol<sup>-1</sup>, which is also well within chemical accuracy. The only significant breakdown occurs in regions where the Hamiltonian element and diabat do not correspond, e.g.  $H_{22} \Leftrightarrow$  "Diabat 1", which has MAE = 8.28 kJ mol<sup>-1</sup>. However, for this region the contribution to the total energy from  $H_{22}$  diminishes



Figure 2.3: SAPTO vs.  $E_{Solute}^{Inter,NN} + E_{Solvent}^{Nonbonded,FF}$  predictions for both  $H_{11}$  and  $H_{22}$ . Diabat 1, Transition state and Diabat 2 regions are defined with respect to their value of  $c_1^2$  as discussed in the main text.

to the extent that the error is largely inconsequential.

The potential energy surface generated by  $E_{Solute}^{Inter,NN}$  must be smooth in order to run MD simulations. As noted by Glick *et al.* [78], neural network potentials that utilize a sum of atom energies to represent the total energy often do not generate smooth potential surfaces for intermolecular interactions, leaving them unsuitable for running MD simulations. While our final working  $E_{Solute}^{Inter,NN}$  terms employ AP-Net type neural network architectures (Section 2.3.3.3), the SchNet architecture was also tested in preliminary stages of the force field development. In Figure S12 and S13 of the Supporting Information, we compare scans of  $E_{Solute}^{Inter,NN}$  for [EMIM<sup>+</sup>][OAc<sup>-</sup>] as predicted with either the SchNet or AP-Net architecture. For the SchNet architecure, the scan is not smooth and exhibits somewhat uncontrolled behavior at longer distances. In contrast, the



Figure 2.4: PBE-D3(BJ)/aug-cc-pVTZ energies vs the force field  $H_{PB/NN}$  energies with C–H bond distances less than 1.8 Å and O–H bond distances less than 1.8 Å. The plot shows 5000 randomly-selected structures from the test set; the MAE listed corresponds to the full test set.

 $E_{Solute}^{Inter,NN}$  term constructed from the AP-Net architecture exhibits smooth behavior over the full scan region. The final,  $E_{Solute}^{Inter,NN} + E_{Solute}^{Nonbonded,FF}$  term thus utilizes the AP-Net architecture, with the  $E_{Solute}^{Inter,NN}$  contribution damped at longer distances, such that the total intermolecular interaction energy asymptotically matches SAPT0 *ab initio* values.

### 2.4.3 PB/NN Force Field Total Energy

In the previous Sections 2.4.1 and 2.4.2, the intra- and inter-molecular components of the diagonal  $H_{11}$  and  $H_{22}$  force field terms were benchmarked and discussed. The final stage in the PB/NN force field development is parameterization of the  $H_{12}$  off-diagonal coupling to the adiabatic potential energy surface (dataset 5, Table 2.1). As discussed in Section 2.3.3.4, the  $H_{12}$  term consists of an AP-Net type neural network architecture. In Figure 2.4, we compare the total PB/NN energy (Equation 2.3) to test set energies from dataset 5 in Table 2.1; this comparison effectively demonstrates the accuracy of the  $H_{12}$  neural network. We note that in this comparison, the energies of the isolated monomers in their equilibrium geometries have been subtracted out (i.e. relative energies). The MAE for these configurations is 1.10 kJ mol<sup>-1</sup>, which is within chemical accuracy. Furthermore, the force field matches well with DFT energies for high energy configurations, ap-



Figure 2.5: a) Diabatic expansion of PB/NN force field in terms of  $c_1^2$  coefficient, compared to b) reaction free energy surface projected onto C–H and O–H bond distances Configurations are taken from dataset 5 in Table 2.1, and the lines drawn in the plot qualitatively separate the diabat 1, diabat 2, and transition state regions.

proximately 250 kJ mol<sup>-1</sup> above the minimum. We list the energy MAE with the force MAE in Table S9 in the Supporting Information. The force MAE is 2.55 kJ mol<sup>-1</sup> Å<sup>-1</sup>, demonstrating high fidelity in force predictions as well. In Figure S14 (Supporting Information) we also show DFT vs. PB/NN energy comparison from an MD trajectory, which gives a similar MAE of 1.14 kJ mol<sup>-1</sup>.

The primary motivation for the diabatic representation utilized within our PB/NN force field is to enable direct incorporation of solvent effects in condensed-phase simulations. In this context, it is insightful to analyze the diabatic decomposition of the final force field after  $H_{12}$  parameterization (Figure 2.4). In Figure 2.5, we reproduce the AIMD free energy surface of the gas-phase reaction alongside a plot of the expansion coefficient  $c_1^2$  of the reactant (EMIM<sup>+</sup>/acetate) state. The configurations in Figure 2.5 are from dataset 5 in Table 2.1, and the reaction coordinate is projected onto the C–H and O–H bond distances. It is evident that the diabatic representation of the force field matches the respective "reactant", "transition state", and "product" regions of the reaction free energy surface. The wide minimum in the free energy profile (dark blue region, Figure 2.5b) corresponds to the EMIM<sup>+</sup>/acetate "reactant" state with O–H distance approximately greater than 1.5 Å. For this region, the force field predicts values of  $c_1^2 > 0.8$ , indicating that the system is mostly composed of the EMIM<sup>+</sup>/acetate "reactant" state. Indeed, at the center of this minimum region, the force field predicts values close to  $c_1^2 = 1$  corresponding to pure EMIM<sup>+</sup>/acetate reactant. Figure 2.5a) clearly indicates the transition between diabatic states in the force field representation as the reaction proceeds through the transition state region. At intermediate values for both C–H and O–H bond lengths, the force field predicts a mixed state consisting of significant fraction of both "reactant" and "product" diabats. For example, the center of the "transition state" region with  $c_1^2 = 0.5$  (and  $c_2^2 = 0.5$ ) is roughly located at  $r_{OH} = 1.25$  Å and  $r_{CH} = 1.35$  Å. As the reaction continues to longer C–H distances (and shorter O–H distances), the force field gives small  $c_1^2$  values, indicating that the system has largely transitioned to the product diabatic state. For  $r_{CH} > 1.8$ Å, the system is essentially entirely composed of the product diabat, with  $c_1^2 = 0$  (and  $c_2^2=1$ ).

The analysis in Figure 2.5 confirms that the force field gives a physically meaningful description of the reaction profile in terms of "reactant" and "product" diabatic states. The diabatic representation was parameterized to SAPTO calculations (subsubsection 2.3.3), and the realistic nature of the force field's diabatic expansion lends merit to this approach. However, the essential test of the diabatic representation is the ability for the force field to accurately capture solvent effects on the reaction profile. This criteria is benchmarked next, for which we consider a "model" solvation environment that is amenable to explicit AIMD benchmark calculations.

## 2.4.4 EMIM<sup>+</sup>/Acetate Reaction Free Energies with/without "Solvation" Environment

We now benchmark the ability of the PB/NN force field to capture solvent effects on the EMIM<sup>+</sup>/ acetate reaction profile. For this purpose, we consider a "model" solvation environment consisting of a gas-phase ion cluster. This choice enables tractable AIMD simulations, whereas explicit AIMD benchmarks would be intractable for alternative liquid phase environments. Our "model" solvation environment is a four ion cluster, created by adding two additional ions, EMIM<sup>+</sup> cation and  $PF_6^-$  anion, to the EMIM<sup>+</sup>/acetate reacting dimer. The  $PF_6^-$  anion is not basic and cannot deprotonate EMIM<sup>+</sup>, which is the motivation for adding this species rather than another acetate anion. While this "model" solvation environment is clearly not the same as a liquid or condensed phase system, the surrounding ions will create large electric fields felt by the reacting complex. We thus believe that this provides a good test of the force field's ability to capture modulation of the reaction profile by external electric fields, while simultaneously enabling tractable AIMD benchmarks.

We compute free energy reaction profiles with both the PB/NN force field and also explicit AIMD simulations. Potentials of mean force (PMFs) are computed as a function of two coordinates, namely the C-H and O-H (reacting proton) distances, using umbrella sampling. The umbrella sampling utilized harmonic potentials with a 0.01 au force constant. The C-H and O-H distances at which the umbrella potentials are centered are described in the Supporting Information. Each simulation window was started with an initial configuration from an equilibrated simulation of the previous window. The Weighted Histogram Analysis Method (WHAM) was used in order to compute the PMF from the umbrella sampling simulations.<sup>160</sup> Free energy profiles were computed for both the isolated reacting complex (two ion system), as well as with the "model" solvation environment (four ion system). For the two ion system, each window consisted of 40-50 ps simulations, with 5 ps used as equilibration. For the four ion system, the simulations for each window were run for  $\sim 200$  ps, with 10 ps used as equilibration. The AIMD simulations were run at 300K in CP2k, using the same simulation settings as mentioned in the Methods section. The force field simulations were run using an ASE calculator;<sup>161,162</sup> the calculator combined the energies and forces of the force field expressions, computed using the OpenMM software package,<sup>163</sup> and the neural network energies and forces, which are computed using PyTorch.<sup>57</sup> The simulations were also run at 300K with a Langevin integrator. A 0.005  $fs^{-1}$  friction coefficient was used, and a 0.5 fs time step was used to integrate the nuclear equations of motion. The 0.5 fs time step is needed for good energy conservation with the anharmonic restraining potential used for the Drude oscillators (discussed in the Supporting Information).<sup>164</sup> These simulations were performed *in vacuo*, as were the AIMD simulations. A benchmark of the energy conservation of the PB/NN force field for a NVE simulation of the two-ion system is shown in the Supporting Information Figure S15.



Figure 2.6: Free energy surfaces for the EMIM<sup>+</sup>/acetate deprotonation reaction with/without the "model" solvation environment. PMFs computed from AIMD simulations are shown in a) and c) for the two ion and four ion systems, respectively. PMFs computed from the PB/NN force field are shown in b) and d) for the two ion and four ion systems, respectively.

Figure 2.6a and Figure 2.6b show the free energy reaction profiles of the two ion system as computed from AIMD and the PB/NN force field respectively; for the two-ion PMF from the force field, we added an arbitrary shift of 3 kJ mol<sup>-1</sup> in order to better compare its features to the AIMD result. Overall, there is good agreement between the free energy profiles, and the PB/NN force field predicts the reaction free energy with close to quantitative accuracy as compared to AIMD. Both PB/NN and AIMD predict a broad minimum basis for the EMIM<sup>+</sup>/acetate reactants, and a barrierless transition state for the deprotonation reaction. There is no local minimum corresponding to the NHC/acetic acid products, and both PB/NN and AIMD predict that at 1.75 Å C–H distance, the NHC/acetic acid products are ~ 8-10 kJ/mol higher in free energy than the reactants. There do exist small discrepancies between the PB/NN and AIMD free energy profiles, in terms of both the width of the reactant minimum along the C–H distance, and also the center of the reactant minimum; these discrepancies will be explained after first discussing the solvation effect.

Solvent effects on the reaction profile are analyzed by comparing predicted free energy surfaces of the "model solvent" four ion system, to corresponding profiles of the isolated reacting complex (two ion system). Free energy surfaces for the four ion, model solvent system are shown in Figures Figure 2.6c and Figure 2.6d as computed from AIMD and with the PB/NN force field respectively. Comparing Figures Figure 2.6a)-d), several observations are made concerning the solvent effects on the reaction profile. First, the electrostatic interactions from the solvating ions broaden the minimum region associated with the reactant species (EMIM<sup>+</sup>/acetate). This is seen by comparing e.g. Figure 2.6a) and Figure 2.6c), where the EMIM<sup>+</sup>/acetate reactant minima (blue region) is significantly broader when the reacting complex is solvated by the additional ions (four ion system, Figure 2.6c). Because the electrostatic interactions with "solvent" are much stronger for the EMIM<sup>+</sup>/acetate ions in the "reactant" state compared to neutral NHC/acetic acid molecules in the product state, there is a substantial shift in the relative reactant/product free energies for the four ion compared to two ion systems. For example, the NHC/acetic acid products are ~ 40-50 kJ mol<sup>-1</sup> higher in free energy than EMIM<sup>+</sup>/acetate reactions in presence of the two "solvating" ions, whereas the difference was only ~ 10 kJ mol<sup>-1</sup> for the isolated complex (note the difference in

energy scales between Figures Figure 2.6a,b and Figures Figure 2.6c,d). This qualitative conclusion that carbene formation from EMIM<sup>+</sup>/acetate ions is destabilized in the presence of solvent, due to the large solvation energy of the reactants, is intuitive and is expected to be general to any condensed phase, chemical environment.

The PB/NN force field reproduces the solvation effect on the reaction profile in semi-quantiative agreement with the AIMD benchmarks. Note that exact quantitative agreement is not expected, for reasons discussed subsequently. The PB/NN force field predicts similar changes in the reaction profile with/without solvating ions (Figure 2.6b vs Figure 2.6d) relative to the AIMD benchmarks (Figure 2.6a vs Figure 2.6c). Although this benchmark is for a "model" solvation environment consisting of two additional solvating ions, the external electric field on the reacting complex is large and causes pronounced shifts of the reaction profile and resulting reactant/product relative free energies. We thus expect that the PB/NN force field should provide comparable accuracy for predicting solvation effects on the reaction profile in arbitrary condensed phase environments that exhibit strong electrostatic interactions. Such quantitative predictions would be important and necessary for evaluating the presence of trace, catalytic carbene species in imidazolium-based ionic liquids. Recall that similar to the EVB framework, the ansatz of the PB/NN force field construction is that solvent effects enter only through the diagonal Hamiltonian elements corresponding to the pure diabatic states (Section 2.3.3). Therefore, the accuracy of the PB/NN force field in capturing the solvent modulation of the reaction profile validates the diabatic decomposition of the force field, which was parameterized to symmetry adapted perturbation theory.

We finally discuss the subtle quantitative differences between free energy profiles predicted by the PB/NN force field (Figure 2.6b,d) and AIMD (Figure 2.6a,c). As always, there is the possibility that discrepancies are due to parameterization deficiencies; note, however, the very high accuracy of the energetic benchmarks of force field components in Figures Figure 2.2, Figure 2.3, and Figure 2.4. A more likely source of discrepancy is the fundamental difference in electronic structure theory: *the level of theory between PB/NN and AIMD is consistent only in the transition state region, but not in the asymptotic region of dissociated reactants/products*. The PB/NN force

field was trained to match (adiabatic) PBE-D3(BJ) energies/forces in the transition state region, but asymptotically matches SAPT description of intermolecular interactions in reactant/product regions. Even more subtle is that the  $E_{Solute}^{Inter,NN}$  term is parameterized at SAPT0 level of theory, while force field interactions contributing to  $E_{Solute,Solvent}^{Nonbonded,FF}$  were previously parameterized to DFT-SAPT (Section 2.3.3.3). Thus the asymptotic reactant and product states in the free energy profiles (Figures Figure 2.6a-d) fundamentally differ in electronic structure description between the PB/NN and AIMD Hamiltonians. We believe this is the primary reason for why the solvent destabilization of the carbene products is predicted to be more extreme by PB/NN compared to AIMD. The question of which prediction (PB/NN or AIMD) is more accurate requires consideration of the relative accuracy of PBE-D3(BJ) vs SAPT level of theory, which is beyond the scope of this work.

In future work, it may be profitable to explore other electronic structure choices for parameterizing the diabatic (Section 2.3.3.3) and coupling elements (Section 2.3.3.4) of the PB/NN potential. At the intersection of the pure diabat (reactant/product) and transition state region, the force field switches from SAPT0 to PBE-D3(BJ) parameterization. Due to differences in these electronic structure theories, there is no rigorous guarantee that the diabatic state will be strictly higher energy (or equal) compared to the underlying PBE-D3(BJ) adiabatic description (this would be rigorously true for equivalent levels of theory). Inspecting the free energy profile of the (two ion) reacting complex predicted by PB/NN (Figure 2.6b), slight error is observed in the center of the minimum (blue region) corresponding to the EMIM<sup>+</sup>/acetate reactant at 1.2 Å C–H distance/1.8 Å O–H distance. This error is most likely due to the imperfect coupling of SAPT0/PBE-D3(BJ) descriptions as the force field transitions from the pure diabatic state. Alternative utilization of constrained DFT for diabatic parameterization would resolve this problem, but would likely sacrifice accuracy for asymptotic interactions for which SAPT is known to be very accurate.<sup>159</sup> It may be that a different flavor of SAPT, particularly DFT-SAPT, would lead to better coupling of the diabat/transition state region, and this will be explored in future work.

### 2.5 Conclusion

We have described a new approach for developing physics-based/neural network (PB/NN) reactive force fields with the central goal of predicting solvent effects on reaction free energies with chemical accuracy. This approach follows an ansatz similar to the widely utilized EVB framework, but proposes novel combination of neural network and physics-based terms that follow a general *ab initio* parameterization approach. We have applied this methodology to study the deprotonation reaction between EMIM<sup>+</sup> and acetate ions to form NHC and acetic acid, and analyze electric field modulation of the reaction free energy profile from a "model solvent" environment. Free energy profiles predicted by PB/NN are in good agreement with AIMD benchmarks, demonstrating that our PB/NN framework can accurately predict reaction free energy profiles in different chemical environments. In agreement with physical intuition, we show that the surrounding ionic environment significantly destabilizes carbene formation relative to the gas-phase, due to the large solvation energies of the ionic reactant species. Our newly developed PB/NN force field should allow prediction of reaction free energies and equilibrium constants for carbene formation within various ionic liquid and solvent environments, to evaluate when and if trace amounts of these catalytic species are present.

While this work has focused on one specific reaction, our PB/NN approach is entirely general and can be applied to a wide variety of reactions. Additionally, as solvent interactions with the reacting complex are incorporated through physics-based force field terms, one can readily study the effects of a large number of different solvents for which force fields either already exist or can be constructed utilizing well-established methodology. This is the primary advantage of our PB/NN approach as compared to stand-alone neural network potentials, as solvent effects are not easily incorporated within neural network potentials in a general way. We have proposed the utilization of symmetry-adapted perturbation theory for parameterizing diabatic states of the PB/NN force field, as opposed to more traditionally utilized constrained DFT approaches. The advantage is that SAPT exhibits rigorous connection to intermolecular interaction terms of a force field and is highly accurate for describing non-covalent interactions. The disadvantage is the lack of rigorous connection between SAPT and the variational approach (DFT) used to parameterize the adiabatic reaction surface; however in this work, such introduced error is found to be small. It is possible that different flavors of SAPT, particularly DFT-SAPT may enhance accuracy and this will be explored in future work.

There are two aspects of our training data generation that will require greater attention when developing PB/NN force fields for more complex reactions. The first is adequately sampling the transition state region of the reaction surface; for the deprotonation reaction studied in this work, this was straightforward (direct AIMD) due to the lack of reaction barrier. However, reactions exhibiting significant barriers would require biased sampling approaches (with AIMD) to generate training data spanning the entire reaction coordinate, particularly in the transition state region. The second consideration is the total amount of training data required for full force field parameterization. Over one million quantum chemistry calculations were utilized for training, with the vast majority used for neural network parameterization. A more efficient approach would focus the training data by identifying the regions of configuration space where the neural networks exhibit the most uncertainty, using e.g. clustering or active learning approaches. [165–167] It is important to note however that this is a general issue for any neural network potential; in fact the issue is actually simplified for our PB/NN framework, as all of the employed neural networks are shortrange, and confined to minimal regions of phase space. We believe that restricting neural networks to small regions of phase space, and utilizing physics-based terms for asymptotic intermolecular interactions, is a robust approach for incorporating machine learning techniques within force field development.

## **CHAPTER 3**

# *N*-HETEROCYCLIC CARBENE FORMATION IN THE IONIC LIQUID [EMIM<sup>+</sup>][OAC<sup>-</sup>]: ELUCIDATING SOLVATION EFFECTS WITH REACTIVE MOLECULAR DYNAMICS

The following is reproduced from Ref. [168]

## 3.1 Abstract

Recent experimental and theoretical work has debated whether N-heterocyclic carbenes (NHCs) are natively present in imidazolium-based ionic liquids (ILs) such as 1-ethyl-3-methylimidazolium acetate ([EMIM<sup>+</sup>][OAc<sup>-</sup>]) at room temperature. Since NHCs are powerful catalysts, determining their presence within imidazolium-based ILs is important, but experimental characterization is difficult due to the transient nature of the carbene species. Because the carbene formation reaction involves acid-base neutralization of two ions, ion solvation will largely dominate the reaction free energy, and thus must be considered in any quantum chemical investigation of the reaction. To computationally study the NHC formation reaction, we develop physics-based, neural network reactive force fields to enable free energy calculations for the reaction in bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>]. Our force field explicitly captures the formation of NHC and acetic acid by deprotonation of a EMIM<sup>+</sup> molecule by acetate, and in addition describes the dimerization of acetic acid and acetate. Using umbrella sampling, we compute reaction free energy profiles within the bulk IL and at the liquid/vapor interface to understand the influence of the environment on ion solvation and reaction free energies. Compared to reaction of the EMIM<sup>+</sup>/OAc<sup>-</sup> dimer in the gas phase, the bulk environment destabilizes formation of the NHC as expected, due to the large ion solvation energies. Our simulations reveal a preference for the product acetic acid to share its proton with an acetate in solution and at the interface. We predict NHC content in bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] to be on the order of parts-per-million (ppm) levels, with order of magnitude enhancement of NHC

concentration at the liquid/vapor interface. The interfacial enhancement of NHC content is due to both poorer solvation of the ionic reactants as well as solvophobic stabilization of the neutral NHC molecule at the liquid/vapor interface.

### 3.2 Introduction

*N*-heterocyclic carbenes (NHCs) are important catalysts in organic chemistry.<sup>117,169,170</sup> The synthesis of the first isolable NHC by Arduengo and coworkers<sup>171–173</sup> has led to an enormous number of applications in diverse fields, such as heterogenous catalysis,<sup>174–176</sup> formation of metal clusters and nanoparticles,<sup>177–183</sup> development of pharmaceutical compounds,<sup>184–186</sup> organocatalysis<sup>116,187–190</sup> and more.<sup>117,170,191</sup> While a variety of methods exist for NHC synthesis, deprotonation of imidazole or imidazolium cations represents one of the most common procedures.<sup>192,193</sup> In addition to being precursors to NHCs, imidazolium cations are very commonly the cationic component of room-temperature ionic liquids (ILs).<sup>194–199</sup> Because ILs are utilized both as solvents in organic synthesis<sup>118,200,201</sup> and electrolytes in electrochemical applications,<sup>196</sup> the formation/presence of NHCs in such ILs would have important consequences for catalysis, reactivity, and stability.<sup>202</sup> The presence of NHCs in ILs could be both beneficial (catalysis) or detrimental (electrolyte degradation), but regardless is essential to characterize for mechanistic determination.

As imidazolium cations have high pKa values (~22),<sup>198</sup> formation of NHCs from these precursors typically requires addition of medium to strong bases<sup>203,204</sup>, electrochemical reduction,<sup>205</sup> and/or elevated temperatures<sup>133,206,207</sup>. Furthermore formation of NHCs from imidazolium cations and base must overcome the large solvation energies of the ionic reactants, the magnitude of which may depend on chemical environment. However, recent work suggests spontaneous NHC formation in imidazolium acetate ILs, despite the acetate ion being only a weak/mild base. For example, several NHC-catalyzed reactions have been found to occur in the ILs [EMIM<sup>+</sup>][OAc<sup>-</sup>] and [BMIM<sup>+</sup>][OAc<sup>-</sup>], such as benzoin condensation,<sup>125</sup> formation of imidazol-2-thiones,<sup>208</sup> polymerization of epoxy resins<sup>209</sup> and others, which provides indirect evidence.<sup>126,127,210</sup> While mass spectrometry measurements show existence of NHCs in vaporized samples of [EMIM<sup>+</sup>][OAc<sup>-</sup>] and related ILs,<sup>211,212</sup> the vapor phase may not be representative of the liquid due to the expected large contribution of ion solvation to the reaction free energy. No direct experimental observation of NHC formation in the liquid phase has been made, as their inherently short lifetime makes such characterization challenging. Chiarotto *et al.* [129] performed cyclic voltammetry experiments and observed an oxidation peak corresponding to a NHC in [BMIM<sup>+</sup>][OAc<sup>-</sup>] at elevated temperatures (120 °C), but it was noted that the electrode itself may act as a carbene trap and thus influence the reaction.<sup>123</sup> The lack of direct evidence has led to several works questioning the extent to which free carbenes exist in these systems.<sup>119,123,124,213</sup> Alternative reaction pathways that do not require NHC content have been proposed for many of the aforementioned reactions, casting further doubt on whether they are spontaneously formed in these ILs at ambient conditions.<sup>213</sup>

Various computational/theoretical methods have been employed in order to rationalize these experimental observations. Reaction free energies calculated using implicit solvent models for both [MMIM<sup>+</sup>][OAc<sup>-</sup>] and [EMIM<sup>+</sup>][OAc<sup>-</sup>] indicate that forming the NHC is unfavorable for dielectric constant values corresponding to these ILs.<sup>214,215</sup> Additionally, Gehrke *et al.* [124] calculated a free energy profile of deprotonation of EMIM<sup>+</sup> by acetate using *ab initio* molecular dynamics (AIMD) simulations of a 26 ion pair [EMIM<sup>+</sup>][OAc<sup>-</sup>] system. The calculated free energy (26 kcal mol<sup>-1</sup>) indicates that the formation of NHCs is minimal at room temperature. Additional AIMD simulations of [EMIM<sup>+</sup>][OAc<sup>-</sup>] from Brehm *et al.* [216] seemed to indicate that the deprotonation of EMIM<sup>+</sup> by acetate does not occur readily; however, further simulations of a free NHC in [EMIM<sup>+</sup>][OAc<sup>-</sup>] showed an interesting C····H–C bond between the carbene carbon and the EMIM<sup>+</sup> methyl group that may stabilize NHCs in solution.<sup>122</sup> It has been suggested that the presence of impurities or gas molecules such as CO<sub>2</sub> may influence carbone formation in [EMIM<sup>+</sup>][OAc<sup>-</sup>] and similar ILs, due to the possible formation of adducts.<sup>217</sup>

In this work, we further investigate carbene formation in the  $[EMIM^+][OAc^-]$  IL through reactive molecular dynamics simulations.  $[EMIM^+][OAc^-]$  is a common ionic liquid, that has been widely studied for use in biomass processing,<sup>218–221</sup> among other applications. The motivation to develop a reactive force field to study NHC formation in  $[EMIM^+][OAc^-]$  is based on the expected importance of ion solvation energetics on the reaction free energy, and the corresponding difficulty of adequate statistical sampling of the viscous ionic liquid environment. For example, our developed approach enables much enhanced statistical sampling compared to alternative computational approaches such as *ab initio* molecular dynamics (AIMD) or hybrid quantum mechanics/molecular mechanics (QM/MM) methods.<sup>14,222</sup> Our reactive force field combines physics-based expressions with neural networks and is termed "PB/NN",<sup>81</sup> and constructed with an ansatz similar to the Empirical Valence Bond (EVB) approach.<sup>47</sup> In previous work,<sup>81</sup> we demonstrated the ability of the PB/NN force field to capture solvation effects on the deprotonation of EMIM<sup>+</sup> by acetate, but this prior work was limited to one reactive ion pair and only ion clusters (not condensed phase environments) were studied. In this work, we extend the PB/NN approach for computing NHC formation reaction free energies in the bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid.

Gehrke *et al.* [124] have previously investigated the deprotonation of EMIM<sup>+</sup> by acetate using AIMD simulations. This and studies of related systems show that there are a variety of possible proton transfer processes, including shuttling of the proton between acetic acid and acetate molecules.<sup>223,224</sup> In [EMIM<sup>+</sup>][OAc<sup>-</sup>], the acetic acid formed from the deprotonation of EMIM<sup>+</sup> may share the proton with other acetate ions, with the acetic acid/acetate "dimer" being lower in free energy than a bare acetic acid molecule. To this end, our PB/NN force field is explicitly designed to model both of the following reactions:

$$EMIM^+ + OAc^- \Longrightarrow NHC + AcOH$$
 (3.1)

$$AcOH + OAc^{-} \rightleftharpoons [AcO^{-} \cdots H^{+} \cdots OAc^{-}] \rightleftharpoons OAc^{-} + AcOH$$
 (3.2)

Equation 3.2 takes into account the possible sharing of the proton between acetate molecules. We note that our PB/NN force field does not explicitly model possible deprotonation of EMIM<sup>+</sup> by the newly formed NHC, since the purpose of the study is to investigate the initial carbene formation.

In addition to the bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid, we also investigate the NHC formation reaction at the ionic liquid/vapor interface. The motivation is that, since ion solvation is an important contribution to the reaction free energy, the equilibria might be significantly altered (shifted towards NHC product) due to decreased solvation at the liquid/vapor interface relative to the bulk liquid. Indeed, prior work has found that a variety of reactions involving ions may have increased propensity to occur at the liquid/vapor interface,<sup>225–227</sup> and the liquid/vapor interface will have pronounced importance for organic synthesis conducted in ionic liquid microdroplets.<sup>228,229</sup> Whether there is enhanced formation of NHCs in [EMIM<sup>+</sup>][OAc<sup>-</sup>] at the liquid/vapor interface is of fundamental interest in light of recent reports of accelerated reactions in microdroplets and thin films,<sup>230–236</sup> and thus may have practical importance for applications involving this IL.

Our free energy simulations show two minima in the proton transfer reaction profile, corresponding to the ionic reactants and the NHC product state with the proton shared as an acetic acid/acetate (AcOH/OAc<sup>-</sup>) dimer. Largely due to the solvation energies of the reactant ions, the computed reaction free energy is endergonic by  $\sim 70$  kJ mol<sup>-1</sup> in the bulk ionic liquid. We compare the proton transfer reaction in the bulk to the analogous reaction in the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> ion trimer, which is the smallest cluster enabling formation of both the NHC species and AcOH/OAc<sup>-</sup> dimer with shared proton. This comparison allows elucidation of both transition state geometric effects and long-range solvation effects on the reaction free energy profile. We additionally compute the proton transfer reaction at the [EMIM<sup>+</sup>][OAc<sup>-</sup>] liquid/vapor interface, and predict order of magnitude enhancement of interfacial NHC resulting from solvation modulation of the interfacial environment. After describing in detail the computational PB/NN reactive force field approach in Section 3.3, our simulation results are presented in Section 3.4.

### 3.3 Methods

We describe our 3x3 multistate PB/NN Hamiltonian utilized to simulate both reactions in Equation 3.1 and Equation 3.2 in the bulk ionic liquid, and discuss the energy partitioning of matrix elements. We then briefly discuss functional forms and parameterization approach, building off of our previous work.<sup>81</sup>. Finally, we detail the algorithms and software implementation utilized for our reactive molecular dynamics simulations and free energy sampling.
#### 3.3.1 Multistate PB/NN Hamiltonian

We utilize a multistate Hamiltonian in order to simulate both reactions in Equation 3.1 and Equation 3.2 in the condensed phase; note that similar multistate Hamiltonians have been constructed in other EVB-like force fields for reactive molecular dynamics simulations.<sup>49,52–54,102,104,106,237–240</sup> Our multistate PB/NN Hamiltonian used here is of 3x3 form, with matrix elements composed of both physics-based terms and neural networks as will be discussed:

$$H_{PB/NN} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix}$$
(3.3)

Following the EVB ansatz,<sup>42,47,50,52</sup> the diagonal terms correspond to diabatic states that have direct chemical interpretation as reactant and product species. In this case, the three different diabatic states correspond to different chemical bond topologies involving the single acidic/reactive proton, as schematically depicted in Figure 3.1. The off-diagonal elements couple the diabatic states to mediate the chemical reaction(s), in this case proton transfer, between the reactant(s) and product(s). Note that in the ionic liquid [EMIM<sup>+</sup>][OAc<sup>-</sup>], while every EMIM<sup>+</sup> cation has an acidic proton that in principle could react, only a single tagged EMIM<sup>+</sup> cation is considered "reactive" in the simulation, along with its two closest acetate ions, and this choice defines the reacting complex and diabatic states (Figure 3.1).

Our parameterization procedure for the multi-state Hamiltonian utilizing both physics-based and neural network terms has been described in our prior work.<sup>81,82</sup> The reader is referred to the previous publication for more in-depth discussion,<sup>81</sup> while we summarize the key aspects of the methodology here. The diagonal elements of the Hamiltonian  $H_{ii}$  are represented by the following energy partitioning

$$H_{ii} = E_{Solute}^{Morse} + E_{Solute}^{Intra,NN} + E_{Solute}^{Inter,NN} + E_{Solvent}^{Bonded,FF} + E_{Solute,Solvent}^{Nonbonded,FF} + E_{Solute}^{electronic}$$
(3.4)



Figure 3.1: Depiction of the chemical bond topology for the three diabatic states of the reacting complex. The dashed ovals depict the molecule associated with the reacting proton in each diabatic state.

In Equation 3.4, each energy term is labeled according to whether it applies to the reacting complex (depicted in Figure 3.1 and referred to as the "Solute") or to non-reacting solvent molecules (labeled as "Solvent"). Equation 3.4 employs both physics-based terms and neural networks. The physics-based terms are:  $E_{Solute}^{Morse}$  is a Morse potential modeling bonds involving the reactive proton (which can break and form);  $E_{Solvent}^{Bonded,FF}$  are standard harmonic bond, angle, and dihedral terms<sup>144,241</sup> describing intramolecular flexibility of solvent molecules;  $E_{Solute,Solvent}^{Nonbonded,FF}$  encompasses physics-based, nonbonded interactions between solute/solute, solute/solvent, and solvent/solvent molecules. While in principle  $E_{Solute,Solvent}^{Nonbonded,FF}$  could consist of generic Coulomb plus Lennard-Jones terms using parameters from standard force fields, in our implementation we utilize the SAPT-FF force field<sup>149,242</sup> parameterized on the basis of symmetry-adapted perturbation theory (SAPT), so as to be consistent with our *ab initio* diabatization scheme (*vide infra*). The constant term  $E_{Solute}^{elcetronic}$ is the electronic energy of the non-interacting Solute molecules in the reacting complex, for the particular chemical bond topology of the diabatic state (and minimum energy molecular geometries).

The two remaining terms in Equation 3.4,  $E_{Solute}^{Intra,NN}$  and  $E_{Solute}^{Inter,NN}$ , are described by neural networks.  $E_{Solute}^{Intra,NN}$  is composed of energetics from monomer-based neural networks, one for each molecule in the reacting complex, accounting for the intramolecular vibrational energetics.

These neural networks replace bonded force field terms for these molecules (e.g. bond/angle/dihedral potentials), the motivation being that the latter analytic expressions are generally not flexible enough to describe intramolecular energetics with chemical accuracy, particularly near a transition state and far from the equilibrium monomer geometries. Note that  $E_{Solute}^{Intra,NN}$  describes all *non-dissociative, intramolecular vibrations*, while bonds involving the reactive proton are described by the Morse potential,  $E_{Solute}^{Morse}$  (the motivation for this partitioning is to remedy asymptotic instabilities of the neural networks, as discussed in our previous work<sup>81</sup>). The final term  $E_{Solute}^{Inter,NN}$  is composed of energetics from short-range, dimer-based neural networks, that account for breakdown of the physics-based interaction terms  $E_{Solute,Solvent}^{Nonbonded,FF}$  at very close monomer/monomer distances. Note that the diabatic decomposition of the Hamiltonian (Equation 3.3) inherently leads to inaccurate intermolecular interactions at unphysically close distances for one or more of the diabats, for which a standard force field will exhibit uncontrolled behavior/divergence.  $E_{Solute}^{Inter,NN}$  is meant to correct this divergent behavior, without modifying long-range physics-based interactions (e.g. electrostatics, induction, dispersion), and is discussed in detail in our previous work.<sup>81</sup>

The off-diagonal terms of the Hamiltonian (Equation 3.3),  $H_{ij}$ , are symmetric ( $H_{ij} = H_{ji}$ ) and describe the coupling between diabatic states along the reaction coordinate. While prior EVB approaches typically have employed analytic functional forms for these off-diagonal elements,<sup>47,137</sup> we choose to utilize neural networks for the  $H_{ij}$  to provide a more versatile description of the diabatic coupling, and enable accurate rendering of the adiabatic potential energy surface (PES) for general dimensionality. The neural networks utilized for  $H_{ij}$  depend on coordinates of the dimer (acid/base) for which the bond topology changes between diabats *i* and *j*. Note that there is no dependence of  $H_{ij}$  on the solvent; this is in line with the ansatz of previous EVB approaches,<sup>46,47</sup> and means that solvent interactions with the reacting complex are only explicitly incorporated into the diagonal terms. The success of this ansatz for capturing solvent modulation of reaction free energies has been previously demonstrated by others,<sup>42,47</sup> as well as in our recent work.<sup>81</sup>

## 3.3.2 PB/NN Parameterization and Neural Network Architecture:

The terms  $E_{Solute}^{Morse}$ ,  $E_{Solute}^{Intra,NN}$ ,  $E_{Solvent}^{Bonded,FF}$ ,  $E_{Solute,Solvent}^{Nonbonded,FF}$ , and  $E_{Solute}^{electronic}$  have been parameterized in our previous work,<sup>81</sup> and will only be briefly discussed. The monomer neural networks constituting  $E_{Solute}^{Intra,NN}$  utilize the SchNet architecture,<sup>61</sup> and are trained to AIMD generated data at the PBE-D3(BJ)/aug-cc-pVTZ level of theory, as consistent with our prior work.<sup>81</sup> We note that prior studies have benchmarked DFT functionals for their ability to accurately model the electronic structure and reactivity of various carbene species.<sup>243–245</sup> In this work, we have refined the original parameterization of  $E_{Solute}^{Intra,NN}$  terms to better capture the *anti* configuration of acetic acid in the condensed phase (see Supporting Information).  $E_{Solvent}^{Bonded,FF}$  handles all bonded terms within the solvent. The EMIM<sup>+</sup> parameters are taken from prior work,<sup>147</sup> and the acetate parameters were fit using the ForceBalance package.<sup>56</sup> These force field parameters can be found within the Supporting Information.  $E_{Solute,Solvent}^{Nonbonded,FF}$  handles nonbonded intramolecular interactions in solvent molecules, all intermolecular solvent-solvent and solvent-solute interactions, and all asymptotic intermolecular interactions between solute molecules. We use the SAPT-FF force field for this term,<sup>242</sup> as parameterized previously.<sup>81</sup> The final term in the diagonal Hamiltonian elements (Equation 3.4) that needs to be parameterized is  $E_{Solute}^{Inter,NN}$ ; as mentioned, this is effectively a correction to the intermolecular interaction at small separation distances. We utilize the AP-Net architecture for these neural networks, as AP-Net has been explicitly designed for modeling intermolecular interactions.<sup>78</sup>  $E_{Solute}^{Inter,NN}$  is then trained to the difference between SAPTO *ab initio* data<sup>159</sup> and the  $E_{Solute,Solvent}^{Nonbonded,FF}$  energy (SAPT-FF) for dimer configurations. For this fitting, the so-called "delta Hartree Fock" term is not included, so that the energy is non-variational and purely based on intermolecular perturbation theory. Thus the "diabatization" defining the diagonal elements (Equation 3.4) is based on perturbation theory, and the pros/cons of this approach are discussed in our previous work,<sup>81</sup> with elaboration provided in the Supporting Information. In this work, we develop new parameterization for the acetic acid/acetate dimer interaction in  $E_{Solute}^{Inter,NN}$  with comprehensive detail given in the Supporting Information.  $E_{Solute}^{electronic}$  is the gas phase energy of the isolated monomers within each diabat as computed previously at PBE-D3(BJ)/aug-cc-pVTZ



Figure 3.2: Free energy surfaces for the proton transfer reaction between acetic acid and acetate ion, as computed from a) AIMD and b) PB/NN simulations. The free energy is plotted as a function of the distances between the acidic proton and closest oxygen atoms of the two acetate ions.

level of theory.<sup>81</sup>

For the off-diagonal elements of the Hamiltonian (Equation 3.3), we utilize neural networks consisting of a modified AP-Net architecture.<sup>78,81</sup> For this work, the specific off-diagonal element  $H_{ij}$  mediating the acetic acid/acetate proton transfer was trained (i.e. coupling the diabats involving acetic acid bond topologies), while the other off-diagonal elements were based upon previous work.<sup>81</sup> For training, PBE-D3(BJ)/aug-cc-pVTZ energies and forces for the (gas-phase) acetic acid/acetate dimer were generated along the (adiabatic) proton transfer reaction coordinate. The off-diagonal  $H_{ij}$  element is fit through minimization of the appropriate 2x2 subblock of the Hamiltonian (Equation 3.3) to the *ab initio* data for the adiabatic (gas-phase) reaction surface. Details of the *ab initio* data generation, neural network architecture and parameters, and fitting approach are given in the Supporting Information. Figure S2 shows the final force field fit to the DFT PES, with an MAE of 0.81 kJ mol<sup>-1</sup> for the total test set. As a further test for the fidelity of modeling the acetic acid/acetate proton transfer reaction, we computed reaction free energy surfaces using umbrella sampling, with both AIMD and the PB/NN reactive force field. Computational details are given in the Supporting Information. The resulting free energy surfaces are shown in Figure 3.2, with the trained PB/NN predictions in excellent agreement with AIMD.

An important aspect of the method is that, for each neural network involved in modeling the atoms in a dissociating bond (this includes  $E_{Solute}^{Intra,NN}$  neural networks for EMIM<sup>+</sup> and acetic acid, the  $E_{Solute}^{Inter,NN}$  and  $H_{ij}$  neural networks), the output is multiplied by a Fermi-Dirac damping function (Equations S1 and S5) that takes as input the value of the respective dissociating bond distance. For the off-diagonal terms,  $H_{ij}$ , the neural network(s) are multiplied by two Fermi-Dirac functions, as there are two possible dissociating bonds for the reactant and product. As neural networks are incapable of extrapolation outside of the training set, the damping functions restrict each neural network output to a well-defined region of phase space, preventing uncontrolled predictions. The specific details of the damping functions have been slightly modified from our previous work,<sup>81</sup> as discussed in the Supporting Information.

We finally note the pros/cons of our choices for functional form of the Hamiltonian matrix elements in Equation 3.3. Our PB/NN model incorporates both a finer degree of energy partitioning and more complex terms (e.g. neural networks) compared to more standard EVB models,<sup>47</sup> which makes parameterization more complex/tedious and requires a large number of *ab initio* calculations (*vide infra*). The benefit of the added complexity/neural networks is the versatile, high-accuracy rendering of both diabatic and adiabatic PES, in general dimensionality, relative to the underlying *ab initio* PES. This enables direct application of the Hamiltonian (Equation 3.3) for molecular dynamics simulation in the condensed phase. For example in both this, and previous work,<sup>81</sup> we show that our PB/NN approach produces an adiabatic PES with chemical accuracy relative to the underlying density functional theory (DFT), *ab initio* description, and enables prediction of reaction free energies in different chemical environments in quantitative agreement with AIMD simulations.<sup>81</sup> We include timings for the various terms in the PB/NN Hamiltonian in the Supporting Information.

### 3.3.3 Reactive Molecular Dynamics Simulations

Utilizing our multistate PB/NN Hamiltonian (Equation 3.3), reactive molecular dynamics simulations are conducted to compute reaction free energies in bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] and at its liquid/vapor interface. We utilize a hybrid software implementation, with both OpenMM<sup>163</sup> and PyTorch<sup>57</sup> libraries used to calculate energies/gradients of the different Hamiltonian matrix elements (Equation 3.3) as defined by their energy partitioning (Equation 3.4). The adiabatic atomistic forces are obtained from the ground state eigenvector of Equation 3.3 and derivatives of each Hamiltonian matrix element, based on the Hellmann-Feynman theorem.<sup>81</sup> The ASE calculator is utilized to run the MD simulations,<sup>81,161</sup> with Plumed<sup>246</sup> utilized for umbrella sampling simulations. Note that all workflow is wrapped in Python, employing the Python APIs of these software libraries. Our implementation can be found on our github site and in the Supporting Information.<sup>162</sup>

Simulations were run for three different systems/environments. The first system is the isolated reaction complex, consisting of one EMIM<sup>+</sup> cation and two acetate anions in the gas-phase without periodic boundary conditions. The second system contains 40 EMIM<sup>+</sup>/acetate ion pairs and is designed to represent the bulk liquid. Periodic boundary conditions were used, with cubic box dimensions of 22.08 Å for each side. One EMIM<sup>+</sup> cation is randomly selected to be "reactive", with the two closest acetate anions defining the reacting complex (Figure 3.1). The third system that we simulate is the [EMIM<sup>+</sup>][OAc<sup>-</sup>], vapor/liquid interface. A vacuum gap is added to the previous bulk system, creating a new simulation box with dimensions of 22.08 Å x 22.08 Å x 66.24 Å. For the latter vapor/liquid interface system, we utilize a standard 3D Ewald/PME treatment of electrostatics, without slab correction.<sup>247</sup> This is because the slab system is symmetric, with no net average dipole moment (and vacuum gap of ~ 44 Å) for which any such slab correction is expected to be insignificant.<sup>247,248</sup>

The reacting complex is made up of one EMIM<sup>+</sup> and two acetate anions. We use umbrella sampling with MD simulations in order to calculate free energy profiles encompassing the two different proton transfer processes shown in Equation 3.1 and Equation 3.2. The collective variables utilized for umbrella sampling are

$$CV_1 = r_{CH} - r_{OH_{\min}} \tag{3.5a}$$

$$CV_2 = r_{OH_{min}} - r_{OH_{min,2}}$$
 (3.5b)

The CVs are differences of distances: The first CV (Equation 3.5a) is the difference between the EMIM<sup>+</sup> carbon-reactive proton distance  $(r_{CH})$  and the shortest acetate oxygen-reactive proton distance  $(r_{OH_{\min}})$ . The second CV (Equation 3.5b) is the difference between the shortest acetate oxygen-reactive proton distance  $(r_{OH_{min}})$  and the second shortest acetate oxygen-reactive proton distance  $(r_{OH_{\min,2}})$ . All acetate oxygens are considered in the  $r_{OH}$  values computed in Equation 3.5, with the condition that the O atom in  $r_{OH_{min,2}}$  must be on a different molecule than the O atom in  $r_{OH_{min}}$ . The reaction (different protonation states of all three molecules) is then sampled by applying umbrella potentials in the 2D space spanned by these CVs. Each umbrella potential  $(\frac{1}{2}kx^2)$  uses a 200 kJ/mol/Å<sup>2</sup> force constant. For all systems/environments, umbrellas were applied spanning values for  $CV_1$  ranging from -1.0 to 3.5 Å, increasing in increments of 0.5 Å. The value of  $CV_2$ ranged from -2.5 to 2.5 Å, increasing in increments of 0.25 - 0.3 Å. This results in approximately  $\sim 200$  windows per umbrella sampling simulation, and the WHAM procedure is subsequently used to construct 2D PMFs from each set of umbrella sampling simulations.<sup>160</sup> Each window is simulated for 40 ps in the NVT ensemble at 300K, controlled by a Langevin thermostat with friction coefficient of 1  $ps^{-1}$ . The simulation time step is set to 0.5 fs for all simulations. For the liquid and liquid/vapor interface simulations, particle mesh Ewald (PME) is used for long-range electrostatics,<sup>36</sup> and van der Waals interactions are truncated/cutoff at 1.1 nm, as necessitated by the system box sizes. Drude oscillators are utilized to model polarization within the  $E_{Solute,Solvent}^{Nonbonded,FF}$ terms, with Drude oscillator positions (adiabatically) optimized using the "DrudeSCFIntegrator" in OpenMM.<sup>163</sup> An anharmonic restraining potential is utilized to prevent Drude oscillator divergence at short contact distance, as described by Huang et al.<sup>154</sup> For the liquid/vapor interface simulations, we apply a restraint of approximately 20 kJ mol<sup>-1</sup> to the center of mass of the reacting complex, centered at the Gibbs dividing surface (Figure S9).

We finally discuss practical issues that inherently arise from the creation of a "reacting complex" or "active zone" in the liquid simulations. In our case, the "reacting complex" is defined as the specified EMIM<sup>+</sup> ion and its two closest acetate anions. During the simulation, it is of course possible for this complex to change identity if a solvent acetate ion displaces one of the two closer acetate ions in the EMIM<sup>+</sup> coordination, so that the reacting complex must be redefined as a function of simulation time. This is analogous to similar issues dealt with in QM/MM simulations of solution-phase chemical reactions, for which "adaptive" QM/MM schemes have been developed.<sup>39–41,249</sup> In adaptive QM/MM schemes, the goal is to switch between quantum mechanical and molecular mechanics descriptions for molecules moving in/out of the active zone, while minimizing introduction of spurious forces. Because definition of an active zone is usually distance based, discontinuities will arise at the boundary due to application of different Hamiltonian terms in the inner and outer regions. If a smoothing or switching function is introduced as an attempted remedy, then a spurious force term will arise from the switching function.

We describe how we deal with the issue of redefining the reacting complex as a function of simulation time for our PB/NN model. If the identity of acetate ions changes in the reacting complex (e.g. from a solvent acetate ion displacing a closer coordination acetate ion), the PB/NN Hamiltonian is "reinitialized" to the new reacting complex identity. As discussed below, each "reinitialized" term will have different effects on the energy conservation for the PB/NN Hamiltonian. It is important to note that during such "reinitialization", the PB/NN Hamiltonian (Equation 3.3) has essentially collapsed to a 2x2 matrix, in terms of its description of the adiabatic ground state. This is because the acetate ions that switch in/out of the reacting complex are sufficiently far away from the acidic proton that the corresponding protonated diabatic state is very high in energy (due to the Morse potential) relative to the other diabatic states. This minimizes adiabatic force discontinuities, as discontinuities in a (high energy) diabat may not propagate to the adiabatic forces. For example, switching the Morse potential (with a harmonic bond) and off-diagonal coupling  $H_{ij}$  applied to different acetate ions during "reinitialization" should have little-to-no effect on the adiabatic forces due to the mentioned rationale.

Force discontinuities will, however, arise from the intramolecular energetic terms of the acetate ions. This is because the acetate ions in the reacting complex are modeled with a neural network  $(E_{Solute}^{Intra,NN})$ , while solvent acetate ions are modeled with a standard force field  $(E_{Solvent}^{Bonded,FF})$  for these energetics. There will thus be a jump/discontinuity in energy/forces when acetate ions are switched between these models, an issue that cannot be straightforwardly corrected with a smoothing/switching function (which would introduce spurious forces). We thus apply the neural network description for intramolecular energetics ( $E_{Solute}^{Intra,NN}$ ) to a broader set of acetate ions (specifically the 6 closest acetate ions to the acidic proton) than just the two acetate ions in the reacting complex. When one of the acetate ions previously considered a solvent molecule moves into the reacting complex, this term thus does not need to be reinitialized. We find that applying the  $E_{Solute}^{Intra,NN}$ term to the 6 acetate molecules closest to the reacting EMIM<sup>+</sup> is sufficient for ensuring no acetate molecule with the  $E_{Solvent}^{Bonded,FF}$  intramolecular energy description diffuses into the reacting complex during the timescale of our simulations.

More details for the "reinitialization" of the reacting complex, and subsequent effects on energy conservation, are discussed in the Supporting Information (note there is an effect from the  $E_{Solute}^{Inter,NN}$  terms, as discussed in the SI). Energy conservation is the rigorous test for a conservative forcefield implementation (i.e. no cutoffs, jumps/discontinuities, drift), and we thus we run PB/NN simulations in the NVE ensemble to benchmark the above treatment/approximations. Energy conservation benchmarks are shown in Figures S3-S5 in the Supporting Information. We do observe energy jumps of order ~ 10 kJ mol<sup>-1</sup>, which we note are most likely due to switching the  $E_{Solute}^{Inter,NN}$  terms. In this regard, there are two important points to note: First, the magnitude of these jumps are much smaller than typical energy fluctuations caused by an NVT thermostat. Second, the total energy drift is comparable to that of a baseline (non-reactive) OpenMM simulation, as limited by energy drift from the Drude oscillators (imperfect convergence). For these reasons, it is expected that our simulation predictions for reaction free energies and other thermodynamic properties are largely unaffected by this issue.

#### 3.4 Results and Discussion

In Section 3.4.1, we first compare the free energies for the proton transfer of  $\text{EMIM}^+/\text{OAc}^-$  dimer and  $\text{EMIM}^+/(\text{OAc}^-)_2$  trimer in the gas phase, and then analyze the proton transfer reaction in the bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid. In Section 3.4.2, we analyze the proton transfer reaction occurring at the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid/vapor interface, and compare/contrast interfacial solvation effects relative to solvation in the bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid.

# 3.4.1 Proton Transfer Free Energies for EMIM<sup>+</sup>/OAc<sup>-</sup>, EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub>, and Bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] Ionic Liquid

Because proton transfer between EMIM<sup>+</sup> and OAc<sup>-</sup> is substantially influenced by the solvation energies of the ions, it is insightful to analyze how the proton transfer free energy changes from gas-phase ion clusters (dimers, trimers) to the bulk ionic liquid. We have previously computed the proton transfer free energy for the EMIM<sup>+</sup>/OAc<sup>-</sup> dimer in the gas phase.<sup>81</sup> Here we compute the corresponding reaction free energy for the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer (using the extended 3x3 Hamiltonian, Equation 3.3), and compare the dimer/trimer reaction free energies in Figure 3.3. The umbrella sampling procedure utilized for free energy calculations of the dimer is the same as in our prior work and is detailed in the Supporting Information.

Figure 3.3a shows the proton transfer free energy for EMIM<sup>+</sup>/OAc<sup>-</sup> dimer, as a function of CVs " $r_{CH}$ " and " $r_{OH_{min}}$ ". Figure 3.3b shows the proton transfer free energy for the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer, as a function of the CVs defined in Equation 3.5. The " $R_1$ " and " $P_1$ " labels in both Figures Figure 3.3a and Figure 3.3b correspond to the points of the profile in which the reactant and product in Equation 3.1 are formed, while the " $P_2$ " label in Figure 3.3b corresponds to the AcOH/OAc<sup>-</sup> dimer with shared proton configuration (Equation 3.2). It is important to note that the free energy scales of Figure 3.3a and Figure 3.3b are very different; the proton transfer for the EMIM<sup>+</sup>/OAc<sup>-</sup> dimer is relatively isoenergetic, so that Figure 3.3a is plotted with a narrow free energy range (~ 25 kJ mol<sup>-1</sup>). In contrast, the additional ion in the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer complex modulates the proton transfer such that the equilibrium is substantially shifted towards ionic reactants, and correspondingly the free energy profile in Figure 3.3b encompasses a much larger range (~ 160 kJ mol<sup>-1</sup>).

For the gas-phase EMIM<sup>+</sup>/OAc<sup>-</sup> dimer, proton transfer to form the carbene species occurs rapidly, on a thermal energy scale ( $\leq 10 \text{ kJ mol}^{-1}$ ). Figure 3.3a shows a broad low energy basin

for  $R_1$  (EMIM<sup>+</sup>/OAc<sup>-</sup>) at short  $r_{CH}$  distances (~ 1.1 Å) and longer  $r_{OH,min}$  distances ( $\geq 1.25$  Å). The formation of  $P_1$  (NHC/AcOH) at  $r_{CH} = 1.75$  Å and  $r_{OH,min} = 1.0$  Å is essentially a barrierless reaction for the EMIM<sup>+</sup>/OAc<sup>-</sup> dimer, with a small free energy increase of ~ 10 kJ mol<sup>-1</sup>. At this point, there is a stabilizing O–H···C hydrogen bond between the formed NHC and acetic acid, and there is an accompanied increase in free energy as the  $r_{CH}$  distance increases and this hydrogen bond is broken. Rapid proton transfer converts the EMIM<sup>+</sup>/OAc<sup>-</sup> dimer between the two different protonation state topologies on ~ ps timescales, as observed previously in AIMD simulations.<sup>81</sup>

In contrast, proton transfer to form the carbene species incurs a significantly higher free energy cost for the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> gas-phase trimer, as shown in Figure 3.3b. As will be discussed, this is due to the "solvation" energy of the ionic reactants due to the additional acetate anion in the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> complex. The reaction in Equation 3.1 can proceed with either acetate in the reacting complex, so we show two labels that correspond to  $P_1$ . A value of 0 on the y-axis of Figure 3.3b indicates the proton is equidistant to the acceptor oxygen on both acetate molecules, while positive or negative values indicate that one acetate molecule is closer to the proton than the other. Similar to the dimer system, the trimer global free energy minimum corresponds to the ionic reactants (i.e. Figure 1 panel 1). As the minimum free energy state is located at x-axis values of  $\sim -1$  Å and a broad range of values along the y-axis, there appears to be little preference for the specific orientation of the acetates with respect to the EMIM<sup>+</sup> ring; rather, the preference is for each species in the reacting complex to remain ionic, with substantial electrostatic stabilization. The minimum free energy pathway from  $R_1$  to  $P_1$  proceeds along y-axis values  $\geq 2$  and  $\leq -2$  in Figure 3.3b, indicating Reaction Equation 3.1 is further inhibited if the nonreacting acetate is too close to the EMIM<sup>+</sup> and OAc<sup>-</sup> engaged in the deprotonation reaction. The favorable electrostatic interactions with the "spectator" acetate anion leads to a free energy cost of  $\sim 90 \text{ kJ mol}^{-1}$  to form the neutral products  $(P_1)$  from the ions, which is substantially larger than for the gas phase dimer.

Figure 3.3b shows that the proton transfer within the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer proceeds through a transition state, before reaching the local free energy minimum at  $P_2$  corresponding to the NHC

species and AcOH/OAc<sup>-</sup> dimer with shared proton. The free energy of this NHC "product" state  $(P_2)$  is ~ 70 - 80 kJ mol<sup>-1</sup> higher than the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> ionic reactant state. The transition state occurs in a region between  $P_1$  and  $P_2$  with a configuration depicted by simulation snapshots in Figure S12. At the transition state, the nonreacting "spectator" acetate is oriented above the plane of the imidazolium ring and hydrogen bonded to one of the ring nonreactive protons, as the formed NHC/AcOH engage in hydrogen bonding (Figure S12c and Figure S12d). The transition state between  $P_1$  and  $P_2$  has a relative free energy of 130 kJ mol<sup>-1</sup>, which is caused by unfavorable geometries initiating the formation of the AcOH/OAc<sup>-</sup> dimer close to the NHC ring. The local free energy minimum at  $P_2$  corresponding to the NHC species and AcOH/OAc<sup>-</sup> dimer product only occurs when the distance between the AcOH/OAc<sup>-</sup> dimer is sufficiently far away from the NHC ring.

There are two major takeaways from comparing the gas-phase, proton transfer reactions for the EMIM<sup>+</sup>/OAc<sup>-</sup> dimer and EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer. First, there is a clear influence of "solvation" energy in stabilizing the ionic reactants (EMIM<sup>+</sup> and OAc<sup>-</sup> species) relative to carbene product (NHC and AcOH species). With the additional "spectator"  $OAc^{-}$  anion in the EMIM<sup>+</sup>/( $OAc^{-}$ )<sub>2</sub> trimer, the NHC products " $P_1$ " are shifted roughly ~ 80 kJ mol<sup>-1</sup> higher in free energy, relative to the similar reaction coordinate for the EMIM<sup>+</sup>/OAc<sup>-</sup> dimer. This is because the spectator  $OAc^{-}$  anion "solvates" the ionic EMIM<sup>+</sup>/ $OAc^{-}$  reactants, substantially lowering the free energy of the ionic state  $(R_1)$ . There is no such solvation for the gas-phase EMIM<sup>+</sup>/OAc<sup>-</sup> dimer, which is the reason why the ionic/neutral (reactant/product) states are nearly isoenergetic for (only) the dimer complex. As we will show, the proton transfer reaction profile in the [EMIM<sup>+</sup>][OAc<sup>-</sup>] bulk ionic liquid is qualitatively similar to that of the EMIM<sup>+</sup>/ $(OAc^{-})_2$  trimer; this is interesting, as it indicates that the single "spectator" OAc- anion captures a "large chunk" of the actual solvation energy in the liquid. The second major takeaway from Figure 3.3b is the importance of the shared proton, AcOH/OAc<sup>-</sup> dimer configuration. When the NHC species is formed, the stable proton state is the AcOH/OAc<sup>-</sup> dimer complex, and not a bare AcOH molecule; the importance of this bonded topology was the motivation for the 3x3 Hamiltonian developed for the simulations. Thus



Figure 3.3: a) Proton transfer free energy for EMIM<sup>+</sup>/OAc<sup>-</sup> dimer in the gas phase. b) Proton transfer free energy for EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer in the gas phase. The labels  $R_1$  and  $P_1$  correspond to the reactants and products of Equation 3.1. The  $P_2$  label corresponds to the shared proton between two OAc<sup>-</sup> molecules in Equation 3.2.

the solvation energy of the AcOH/OAc<sup>-</sup> dimer is an important contributor to the overall reaction free energy for NHC formation, as will be analyzed in more detail for the bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid.

We next discuss our simulations of the proton transfer reaction in the bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid. As discussed above, the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer is the better reference system to provide context for proton transfer in the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid, and thus we directly compare reaction free energies for these systems. In Figure 3.4 we show the reaction free energy profile computed for the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid and compared to the previous profile computed for the gas-phase, EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer. The immediate observation is the qualitative similarity between the proton transfer reaction profiles computed in the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid (Figure 3.4a) and gas-phase, EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer (Figure 3.4b). The reaction profile in the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid shows similar free energetic trends going from the ionic reactants at  $R_1$  to the transition state region between  $P_1$  and  $P_2$ , to the product state at  $P_2$  encompassing the NHC species and AcOH/OAc<sup>-</sup> dimer complex. There are, however, quantitative differences between these free energy profiles resulting from the extended solvation environment of the liquid state. Both the "transition state" between  $P_1$  and  $P_2$ , as well as  $P_2$  product are significantly lower in free energy in the ionic liquid phase. This indicates greater stabilization of these intermediates relative to the ionic reactants ( $R_1$ ) in the extended ionic liquid compared to the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer cluster.

In Figure 3.5, we show snapshots of the various intermediates along the reaction profile from both the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid and gas-phase, EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer simulations, with more snapshots shown in Figures S11 and S12. As previously mentioned, the ionic reactants exhibit a broad free energy (global) minimum basin spanning  $R_1$ , with free energy largely insensitive to complex orientation. This is demonstrated in Figures S11a-b and S12a-b, which depict visually different reacting complex configurations that are approximately isoenergetic. The free energy along the reaction path proceeding from  $R_1$  to  $P_2$  is, however, significantly different for the ionic liquid compared to the EMIM<sup>+</sup>/ $(OAc^{-})_2$  trimer cluster. This is largely due to differences in solvation and/or geometrical configuration of the key intermediates along the reaction coordinate. Figure 3.5 and Figures S11-S12 show simulation snapshots depicting configurations of the reaction intermediates, within both the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid and gas-phase, EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer. For the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer, as mentioned previously, the "spectator" acetate anion hydrogen bonds to one of the protons on the EMIM<sup>+</sup> ring and then orients itself above the ring plane during the course of the proton transfer reaction. In the liquid phase, additional solvent acetate anions instead participate in this hydrogen bonding with EMIM<sup>+</sup>, allowing the corresponding "spectator" acetate anion to position closer to the EMIM<sup>+</sup> ethyl group in the reacting complex configuration (Figure 3.5 " $P_1$ " panel and Figure S11c). This is a key difference, requiring breaking of a hydrogen bond to reach  $P_1$  for the gas-phase EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer, but not when the reaction proceeds in the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid. The result is that the intermediate  $P_1$  preceding the transition state is  $\sim 35 \text{ kJ mol}^{-1}$  lower in free energy in the liquid phase compared to the trimer cluster. This free energy difference propagates along the reaction profile towards the transition state, so that the transition state is  $\sim 20 - 40 \text{ kJ mol}^{-1}$  lower in relative free energy in the ionic liquid compared to the gas-phase,  $EMIM^+/(OAc^-)_2$  trimer. Simulation snapshots depicting the

transition state configuration in both the liquid and trimer cluster are shown in Figures S11d and S12d.

To more quantitatively demonstrate the differing reacting complex configurations, in Figure S14 we plot a histogram of the distance between the NHC ring center of mass and the AcOH/OAc<sup>-</sup> dimer center of mass for several points along the reaction coordinate spanning  $P_1$  to  $P_2$ ; also shown are histograms of the angle between the NHC carbon engaged in the deprotonation reaction, the NHC ring center of mass and the AcOH/OAc<sup>-</sup> dimer center of mass, which provides information about the orientation of the dimer with respect to the imidazole plane. The distance histograms show that the dimer is farther away from the NHC center of mass in the liquid phase than the gas phase trimer for all points along the reaction coordinate. The distribution of the angle histograms shows that the AcOH/OAc<sup>-</sup> dimer is in plane with the imidazolium ring in the liquid phase (angles close to  $0^{\circ}$ ) and above this plane in the gas phase trimer (angles close to  $90^{\circ}$ ). These results further indicate that the extended solvation environment distinctly modulates the reacting complex configurations. The reactive trimer is packed closer together in the gas phase, leading to higher free energies while coordinating the individual proton transfers.

Our prediction for the free energy of the initial proton transfer (Reaction 3.1) in the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid is in semi-quantitative agreement with prior AIMD studies. Gehrke *et al.* [124] calculated a free energy of ~ 100 kJ mol<sup>-1</sup> for the deprotonation reaction (Equation 3.1) in the liquid phase; this free energy corresponds to the transition state region of our profile slightly past  $P_1$ . We compute a free energy of 90 ± 10 kJ mol<sup>-1</sup> for this region of the reaction profile in the liquid phase, which is close to the prior AIMD result. We note that any asymmetry in our computed free energy profile spanning  $CV_2 = 0$  (along the vertical axis) is due to statistical uncertainty, and is the origin of the stated uncertainty in our prediction. Such statistical uncertainty arises from the viscous nature of the ionic liquid (and correspondingly slow dynamics), making converged statistical sampling difficult.<sup>250</sup> Indeed we have observed that the finite simulation time leads to different distributions of solvent molecules around the imidazolium ring for symmetric values of  $CV_2$ , indicating somewhat incomplete sampling. We show the varying solvent orientations through spatial distribution



Figure 3.4: a) Proton transfer free energy within the bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid; b) Proton transfer free energy for EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer in the gas phase.

functions (SDFs), computed with TRAVIS and plotted with ChimeraX in Figure S16.251,252

We next characterize the solvation structure around the reacting complex in the liquid phase though analysis of radial distribution functions (RDFs). In Figure 3.6, we plot  $\rho q(r)$  and running coordination numbers for the EMIM<sup>+</sup> ring/acetate oxygen at location " $R_1$ " in Figure 3.4a and the NHC ring/acetate oxygen at location " $P_2$ ". These points of the profile correspond to the initial reactant and final product from our simulations. To improve statistics, these RDFs are computed from a total of 360 ps of simulation for the specific configuration along the reaction coordinate. From inspection of the RDFs, it is clear that the anion coordination surrounding the reacting complex is reduced upon formation of the neutral NHC species. In the " $R_1$ " RDF, the first peak at 4 Å corresponds to oxygen atoms present in three different acetate anions, with  $N_{coord} = 4$ . The first acetate is part of the reacting complex and has two oxygen atoms that are close to equidistant to the EMIM<sup>+</sup> center of mass (see panel  $R_1$  Figure 3.5 for an example of the configuration). The second acetate also belongs to the reacting complex, but has one oxygen oriented toward the reacting proton and the other oxygen oriented toward the surrounding solvent, so that only one oxygen from this anion contributes to the peak. The third acetate is hydrogen bonded to one of the nonreactive protons, and also contributes one oxygen atom to the coordination number. There is also a peak at  $\sim 9~\text{\AA}$  which is indicative of coordination from acetate anions in a secondary solvation shell.



Figure 3.5: Simulation snapshots depicting reacting complex intermediates. The top panel displays snapshots from the [EMIM<sup>+</sup>][OAc<sup>-</sup>]ionic liquid simulations, and the bottom panel displays snapshots for corresponding intermediates from the gas-phase,  $EMIM^+/(OAc^-)_2$  trimer simulation. The labels match the corresponding locations on the reaction free energy profiles of Figure 3.4.



Figure 3.6: RDFs between the center of mass of the EMIM<sup>+</sup> ring and acetate oxygen atoms (" $R_1$ ") and between the center of mass of the NHC ring and acetate oxygen atoms (" $P_2$ "). The corresponding coordination number at a given distance is shown by the dashed lines.

For the NHC " $P_2$ " RDF, in contrast to EMIM<sup>+</sup>, the acetates are positioned farther away from the ring. There is a peak at ~ 5.5 Å comprised of the reacting complex oxygen atoms and various other oxygens surrounding the NHC, but it is smaller in magnitude than the peak seen in the  $R_1$  RDF. There is also no secondary peak at 9 Å as seen for the Reactant, which suggests the absence of longer range structure in the acetate coordination.

The solvent acetates hydrogen bond to both nonreactive imidazolium protons, but the two hydrogen bonds are unequal in strength. We plot the oxygen-proton RDFs for both nonreactive EMIM<sup>+</sup> ring protons in Figure 3.7a, which we label as H3 and H4 (H3 is closer to the methyl group and H4 is closer to the ethyl group, see Figure S20). Figure 3.7a shows the computed  $\rho g(r)$ with all EMIM<sup>+</sup> molecules (solvent + reacting complex) in the simulation while Figure S17 plots similar RDFs with only the reacting complex EMIM<sup>+</sup>. The first peak in both RDFs in Figure 3.7a is due to the oxygen hydrogen bonded directly to the respective proton. There is a stronger hydrogen bond/higher RDF peak at the H3 proton. This is likely due to sterics, with less steric constraint for acetate to hydrogen bond to the proton close to the smaller methyl group than the ethyl group. Interestingly, Figure S17 indicates that the reacting complex EMIM<sup>+</sup> H4 proton may form a stronger hydrogen bond compared to the H3 proton; however, a definitive conclusion is difficult given the statistical uncertainty.

Upon formation of the NHC, the hydrogen bond strength to these nonreacting imidazolium protons is significantly reduced. Figure 3.7b shows the NHC nonreactive ring protons-oxygen  $\rho g(r)$ ; we show the same plot in Figure S19 with the corresponding  $\rho g(r)$  for the reacting complex EMIM<sup>+</sup>. The RDF was computed from a simulation in the " $P_2$ " region of the reaction profile. There is little coordination at the H3 proton, and the H4 proton shows almost no coordination at 2-2.5 Å. This conclusion is consistent with Figure 3.6, which indicated fewer coordinating acetates around the NHC due to the reduced electrostatics of the neutral molecule. The lower acetate coordination for the NHC leads to less hydrogen bonding with the nonreactive, ring hydrogen atoms.

These RDFs help explain the varying free energies observed in Figure 3.4a. The solvent stabilizes the ionic  $R_1$  state, as the  $P_2$  state has significantly less coordination around the ring (Figure 3.6); this inhibits formation of this species through Reaction 3.1 and 3.2. The RDFs in Figure 3.7 show that the solvent acetate molecules hydrogen bond to the H3 and H4 EMIM<sup>+</sup> ring protons, which we discussed earlier in connection with the [EMIM<sup>+</sup>][OAc<sup>-</sup>] bulk liquid and EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer free energy profiles in Figure 3.4. While the OAc<sup>-</sup> not involved with the initial EMIM<sup>+</sup> deprotonation is allowed to hydrogen bond to these protons in the gas phase EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer, the solvent prevents this in the liquid phase, leading to the varying free energies and geometries observed in Figure 3.4 and Figure 3.5.

## 3.4.2 Proton Transfer Free Energy at [EMIM<sup>+</sup>][OAc<sup>-</sup>]Ionic Liquid/Vapor Interface

We next analyze the proton transfer reaction at the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid/vapor interface. Because of the important solvation energy contribution to the reaction free energy (as previously discussed), it is interesting to investigate differences between the reaction at the liquid/vapor interface compared to the bulk liquid. Before discussing the interfacial reaction, we first discuss the structure of the ionic liquid/vapor interface. In Figure 3.8, we plot cation and anion density profiles (decomposed by select functional groups) that span the liquid/vapor interface. As dis-



Figure 3.7: a) EMIM<sup>+</sup> H3–O  $\rho g(r)$  and H4–O  $\rho g(r)$ . These RDFs were computed using all EMIM<sup>+</sup> molecules in the simulation. b) Reacting complex NHC H3–O  $\rho g(r)$  and H4–O  $\rho g(r)$ . The corresponding coordination number at a given distance is shown by the dashed lines.



Figure 3.8: Number density profiles at the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid/vapor interface for a) the center of mass of the EMIM<sup>+</sup> methyl, ring and ethyl group and b) the center of mass of the OAc<sup>-</sup> carboxylate and methyl group. The dashed lines approximately denote the Gibbs dividing surface.

cussed in Section 3.3.3, we are limited to small system sizes (40 ion pairs), and thus there are significant fluctuations in the interfacial structure; the liquid/vapor density profiles are generated from a 100 ns (non-reactive) simulation for better statistical convergence. We divide the cell into half along the z dimension of the simulation box and shift the system center of mass along the zdimension to the origin for each frame, and average the density profile over both interfaces; this analysis procedure is similar to previous work.<sup>253</sup> Figure 3.8a shows the number density profiles for the EMIM<sup>+</sup> methyl group (-CH<sub>3</sub>), ethyl group (-C<sub>2</sub>H<sub>5</sub>), and the imidazolium ring. Figure 3.8b shows the number density profiles for the OAc<sup>-</sup> carboxylate (-COO) and methyl (-CH<sub>3</sub>) groups. A dashed line is used to denote the Gibbs dividing surface, which is computed from the total number density profile (Figure S9). Inspection of Figure 3.8 indicates that nonpolar groups of both the cation and anion have a preference to reside near the vacuum side of the interface. The EMIM<sup>+</sup> ethyl group is positioned closest to the interface, with the methyl group of EMIM<sup>+</sup> farther into the liquid side of the liquid/vapor interface. Similarly, the methyl group of acetate is positioned closer to the vacuum while the carboxylate group remains buried in the liquid. Both of these results match with prior simulations of imidazolium ionic liquids, which found the terminal ethyl carbon of various EMIM<sup>+</sup> cation ILs was oriented towards the vacuum at the liquid/vapor interface;<sup>253,254</sup> additionally, an AIMD study of [EMIM<sup>+</sup>][OAc<sup>-</sup>] droplets by Brehm and Sebastiani [255] found that the methyl group of acetate orients itself toward the vacuum.

We now discuss the free energy of the proton transfer reactions at the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid/vapor interface. As discussed in Section 3.3.3, this calculation was performed by constraining the reaction complex near the Gibbs dividing surface of the liquid/vapor interface during the free energy calculation, as similar to previously employed procedures.<sup>226</sup> In Figure 3.9a, we show the reaction free energy computed at the liquid/vapor interface, as compared to the corresponding reaction free energy previously computed for the bulk ionic liquid (Figure 3.9b). *A priori*, one might anticipate that the energetics of the reaction at the liquid/vapor interface would be intermediate between that computed for the bulk liquid and the gas-phase ion clusters, based on solvation considerations. However, we (somewhat surprisingly) find that the reaction free energy profile



Figure 3.9: Free energy profiles for the proton transfer reaction a) at the [EMIM<sup>+</sup>][OAc<sup>-</sup>]ionic liquid/vapor interface; and b) in bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>]ionic liquid.

is qualitatively similar as computed for the liquid/vapor interface and the bulk ionic liquid, with only some quantitative differences. The global minimum, which corresponds to the ionic reactants in Figure 3.1a, is almost identical in the two free energy profiles, and so is the free energy of intermediate " $P_1$ " (Figure 3.9a). There are, however, quantitative differences between the free energy profiles in the reaction coordinate region spanning " $P_1$ " to " $P_2$ ", which encompasses the transition state and product basins. In general, the proton transfer is better facilitated at the liquid/vapor interface, with a reduced transition state barrier and a lower free energy cost for forming NHC and AcOH/OAc<sup>-</sup> dimer products; the relative product free energies are  $\sim 60 \text{ kJ mol}^{-1}$  at the liquid/vapor interface compared to  $\sim 70 \text{ kJ mol}^{-1}$  for the bulk ionic liquid. One reason for the similarity in the reaction free energy profiles, is that the reacting complex configurations are much more similar to one another in the liquid and at the liquid/vapor interface, compared to the gas-phase EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer complex. This is observed in simulation snapshots shown in Figures S11 and S13. In Figure S15, we show histograms of the center of mass NHC / center of mass AcOH/OAc<sup>-</sup> dimer distance, and it is observed that the histograms largely overlap as computed in the two different environments. As was discussed in Section 3.4.1, differences in the geometry of the reacting complex lead to higher reaction free energies for the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer complex; these configurational differences are not observed when comparing the reaction at



Figure 3.10: a) RDFs between reacting complex EMIM<sup>+</sup>center of mass/acetate oxygen atoms computed for both the liquid and the liquid/vapor interface. b) RDFs for EMIM<sup>+</sup> ring nonreactive protons/acetate oxygen atoms computed for both the liquid and the liquid/vapor interface.

the liquid/vapor and bulk liquid environments. Interestingly, Figure 3.9 indicates better statistics for the free energy calculation at the liquid/vapor interface, as the reaction profile better matches the required symmetry about  $CV_2=0$  (vertical axis). Evidently, there are different timescales involved with the solvation coordinate(s) within the bulk liquid and liquid/vapor interface, leading to differences in statistical sampling; a more detailed analysis of such solvation timescales is beyond the focus of this work.

We analyze the coordination of the reacting complex at the liquid/vapor interface. At short distances, the reacting complex EMIM<sup>+</sup> at the liquid/vapor interface only has marginally reduced acetate coordination in comparison to the bulk liquid. In Figure 3.10, we compare the reacting complex EMIM<sup>+</sup> ring/acetate oxygen RDF at the liquid/vapor interface to that in the bulk liquid. The coordination number is slightly reduced through the first peak in the RDF for the liquid/vapor interface as compared to the bulk ionic liquid. At longer distance, the RDFs and coordination numbers become dissimilar, reflecting the different liquid/vapor and bulk liquid environments. In Figure 3.10b, we show RDFs between nonreactive EMIM<sup>+</sup> protons and acetate oxygen atoms, for both the liquid/vapor and bulk liquid environments. In proton the liquid/vapor and bulk liquid environments. In the liquid/vapor and bulk liquid environments. In the liquid/vapor and bulk liquid environments. In the liquid phase, there are noticeable peaks in the RDF for both of these protons. At the interface, there is a significant peak for the H3 proton



Figure 3.11: Spatial distribution functions of acetate oxygen atoms surrounding the EMIM<sup>+</sup> ring for a) a liquid phase umbrella sampling simulation and b) a liquid/vapor interface umbrella sampling simulation. Depicted in purple are density regions with isovalue set to 0.532 nm<sup>-3</sup>. These configurations correspond to  $R_1$  regions of the profile.

(positioned on the side of the imidazolium ring with the methyl group), while there is only a small peak for the H4 proton (positioned on the side of the ring with the ethyl group). As was shown in Figure 3.8a, the ethyl group of EMIM<sup>+</sup> is positioned closer to the vacuum than the methyl group, which leads to fewer hydrogen bonds with acetate molecules at the H4 proton. To further demonstrate this distribution of solvent acetates, we show spatial distribution functions (SDFs) in Figure 3.11 of oxygen atoms surrounding the reactant EMIM<sup>+</sup> for both environments. These SDFs indicate similar distribution of acetate oxygen atoms around the ring, except at the H4 position which is largely uncoordinated at the liquid/vapor interface.

As discussed, both the transition state and product free energies for the proton transfer reaction are more favorable/lower by ~ 10 kJ mol<sup>-1</sup> at the liquid/vapor interface compared to the bulk ionic liquid. This is caused by two factors, the first being the reduced solvation around the EMIM<sup>+</sup> cation at the liquid/vapor interface, which leads to less stabilization and higher free energy of the ionic reactants ( $R_1$  state). The second factor is a direct stabilization of the products at the liquid/vapor interface. Because the product NHC molecule is a neutral species, there is a solvophobic force within the bulk ionic liquid that expels this NHC species to the liquid/vapor interface. This results in an enhanced propensity for NHC at the liquid/vapor interface, corresponding to a more favorable/lower product (" $P_2$ ") free energy in the reaction profile of Figure 3.9a. To quantify the solvophobic driving force for NHC to reside at the liquid/vapor interface, we compute a potential of mean force (PMF) for the NHC molecule as a function of distance from the liquid/vapor interface. This PMF is shown in Figure S21, and demonstrates that the NHC is stabilized by 15-20 kJ mol<sup>-1</sup> at the Gibbs dividing surface of the interface relative to its solvation in the bulk ionic liquid. Thus this solvophobic force is substantial, and is an important underlying cause for the quantitative free energy difference of the " $P_2$ " product state in Figures Figure 3.9a and Figure 3.9b. To summarize, the proton transfer reaction is facilitated at the ionic liquid/vapor interface relative to the bulk liquid due to both poorer solvation of the ionic reactants as well as stabilization of the NHC product molecule at the interfacial environment.

## 3.5 Conclusion

We have developed a PB/NN reactive force field to simulate proton transfer reactions of *N*-heterocyclic carbene formation in the [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid. Reaction free energy profiles were computed for the bulk ionic liquid, its liquid/vapor interface, and additionally EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> ion clusters. Our results indicate that the reaction free energy depends on both the geometry/structure of the local reacting complex, as well as the surrounding solvation environment. The EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> ion trimer is the smallest cluster for which the proton transfer reaction proceeds qualitatively (*al-beit not quantitatively*) similar to as in the bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid. This is because the product of the proton transfer reaction is an NHC species and AcOH/OAc<sup>-</sup> dimer with a shared proton, and furthermore the third ion electrostatically stabilizes the ionic reactants, reminiscent of the bulk liquid solvation energy; this is distinct from the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> ion trimer compared to the bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] ionic liquid. This is due to the constrained reacting complex geometry as well as nonexistent long-range ion solvation for the EMIM<sup>+</sup>/(OAc<sup>-</sup>)<sub>2</sub> trimer, both of which lead to more favorable NHC formation within the bulk ionic liquid.

In previous literature, there has been debate as the extent of NHC formation/concentration

within [EMIM<sup>+</sup>][OAc<sup>-</sup>] and similar ionic liquids. Our computed reaction free energy profiles provide an estimate of the concentration of NHC within the bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>] and at its liquid/vapor interface. Given a reaction free energy of  $\sim 70 \text{ kJ mol}^{-1}$  (Figure 3.4) we estimate that NHC species exist in the bulk [EMIM<sup>+</sup>][OAc<sup>-</sup>]at parts-per-million (ppm) levels; note that such concentrations are catalytically relevant in certain contexts.<sup>118,170</sup> As mentioned in the Introduction, our PB/NN reactive force field neglects possible proton-sharing/dimerization of the NHC with EMIM<sup>+</sup> cations. If such interactions are significant, an even higher concentration of NHC would be expected than the ppm levels predicted here. Our results thus support previous conjectures of NHC content in these ionic liquids, which speculated that NHCs are formed in low concentrations but lead to rapid reactions.<sup>123</sup> We have additionally predicted that NHC concentration is significantly enhanced at the ionic liquid/vapor interface. This is due to both poorer solvation of the ionic reactants as well as solvophobic stabilization of the neutral NHC molecule at the interface. The concentration enhancement at the interface is predicted to be an order of magnitude based on the reaction free energy (Figure 3.9), with an even larger enhancement predicted based solely on the NHC potential of mean force (Figure S21). The estimated NHC content has important implications for [EMIM<sup>+</sup>][OAc<sup>-</sup>]as a solvent and/or electrolyte, due to the catalytic activity of carbene species. Additionally, the NHC content may affect the chemical/thermal stability of the IL, as the carbene could potentially catalyze decomposition reactions.

We finally comment on the utilization of our PB/NN reactive force field for this study, compared to alternative computational approaches. Ionic liquids are highly viscous with liquid structure consisting of long range charge oscillations,<sup>256</sup> such that long simulation times with fairly large simulation boxes are required to minimize statistical uncertainty and finite size effects. While our predictions do exhibit statistical uncertainty (Figure 3.4) and likely finite size artifacts as well (Section 3.3.3), our PB/NN reactive force field allows longer simulations for larger systems compared to either AIMD or QM/MM approaches. The key drawback is the significant effort required to parameterize the PB/NN Hamiltonian (Section 3.3.1). However, similar to related EVB methods, the advantage of PB/NN is that, once parameterized, the reactive potential is transferable to arbitrary solvation environments. For example, interesting future directions could include investigation of NHC formation in [EMIM<sup>+</sup>][OAc<sup>-</sup>] mixtures and/or at solid/liquid interfaces. Continued work on improving the accuracy and ease of construction of the PB/NN potentials will be a key route of further research.

# CHAPTER 4

# PROTON TRANSPORT IN [BMIM<sup>+</sup>][BF<sup>-</sup><sub>4</sub>] /WATER MIXTURES NEAR THE PERCOLATION THRESHOLD

This chapter reproduces the work in Ref. [257].

#### 4.1 Abstract

The incorporation of ionic liquids into existing proton exchange membrane (PEM) materials has been shown to enhance thermal stability and improve conductivity at reduced water content. Because proton transport is dictated by both vehicular diffusion and the Grotthuss mechanism, it is expected that the nanoscale structure of the resulting ionic liquid/water networks will sensitively influence transport properties. In this work, we study proton transport in  $[BMIM^+][BF_4^-]$ /water mixtures of systematically varying water volume fraction, focusing on concentrations near the percolation threshold in which water networks are connected over macroscopic lengthscales. We utilize reactive molecular dynamics within the multistate empirical valence bond (MS-EVB) framework to explicitly model Grotthuss hopping processes. Excellent agreement with experimental conductivity data is obtained within the Nernst-Einstein approximation, indicating that proton transport proceeds in a largely uncorrelated manner even at pH values < 0. We additionally study the changing topology of the hydrogen-bonded water network in these mixtures using percolation and graph theory analysis. We find that the proton diffusion coefficient and forward hop rate increase linearly with water content at concentrations ranging from dilute through the percolation threshold; surprisingly, we find no deviation in this trend at the percolation transition. The high concentration of  $BF_4^-$  anions inherently alters the fraction of Eigen and Zundel proton states, producing a net detrimental effect on proton transport rates relative to bulk water. This mechanistic insight is useful for selecting ideal ionic liquid candidates and determining the optimal ionic liquid concentration to incorporate into PEM materials.

# 4.2 Introduction

Proton exchange membrane fuel cells (PEMFCs) are promising devices for clean and efficient energy production.<sup>258–261</sup> In PEMFCs, hydrogen gas is oxidized at the anode, and protons are transported across the proton exchange membrane to the cathode where oxygen gas is reduced to water. Due to the intrinsic water content and thermal stability of commonly used membrane materials, e.g. Nafion,<sup>260,262–264</sup> operating temperatures of PEMFCs are limited to  $\leq 80$  ° C, which introduces several challenges to fuel cell operation.<sup>263,265</sup> For example, trace CO in the hydrogen input gas can poison the commonly used Pt catalysts at these temperatures; additionally, as 40-50% of the energy from PEMFCs is produced as heat, expensive cooling systems are required.<sup>266,267</sup> Higher operating temperatures ( $\geq 120$  ° C) can reduce or eliminate these limitations.<sup>263,265</sup> With this temperature range in mind, one proposed modification to the PEMFC framework is to incorporate room-temperature ionic liquids (ILs)<sup>197,199,268–270</sup> into existing membrane materials; this would significantly lower the membrane water content, with additional benefits from the intrinsic thermal stability and conductivity of the IL.<sup>264,271–273</sup>

A variety of ILs incorporated in Nafion have been tested for use in PEMFCs, and it has been shown that their total conductivities rise with increasing temperatures above 80 ° C, while dry Nafion's total conductivity decreases.<sup>264,271–273</sup> Thermal stability is improved due to the very low vapor pressures of the neat ILs, and also because ILs favorably absorb water and are thus difficult to dry out.<sup>274,275</sup> Due to their large cohesive energies, ionic liquids exhibit high viscosities, and thus increased temperatures serve to enhance IL conductivity. Viscosity is further decreased when ionic liquids are mixed with water, and the conductivity maximum of IL/water mixtures generally occurs for dilute solutions of ~ 10-20 % ion content.<sup>276–278</sup> However, optimizing IL/water concentration in PEM materials requires special consideration: while low water content mixtures are better for thermal stability considerations, sufficient water content is required for facile proton transport. This is because water serves as the vehicle for Grotthuss transport, which gives the excess proton remarkable diffusion rates in comparison to similarly sized ions.<sup>55</sup> Thus a membrane material that

retains a low amount of water for Grotthuss transport at elevated temperatures may result in a highly useful PEMFC.

The Grotthuss transport mechanism is highly sensitive to underlying water nanostructure, and it is thus important to investigate how this mechanism is altered in IL/water mixtures at varying water concentration. Unique water networks may promote faster (or slower) proton transport rates than in bulk water; an example is water wires within carbon nanotubes which exhibit an order of magnitude faster proton transport compared to bulk water.<sup>279–284</sup> In the Grotthuss mechanism, protons experience structural diffusion by hopping along hydrogen-bonded water networks, accompanied by transitions between the Eigen  $(H_9O_4^+)$  and Zundel  $(H_5O_2^+)$  states of hydronium;<sup>285</sup> in addition to this structural diffusion, the total proton transport mechanism is dictated by standard vehicular diffusion through the principle hydronium species as well.<sup>286,287</sup> The total proton conduction mechanism may be altered within IL/water mixtures due to differences in water structure as well as the strong Coulomb interactions imposed by the ionic liquid. Hydrogen-bonded water networks in these mixtures exhibit unique and dynamically changing structures, and system-specific differences arise from intrinsic hydrophobicity differences among ionic liquids.<sup>288–294</sup> Different ILs possess a range of hydrophilic and hydrophobic interactions, e.g. alkyl chains on imidazolium cations and -CF<sub>3</sub> groups on certain anions, leading to classes of ILs that range from completely miscible to immiscible/phase-separated when mixed with water.<sup>288,294</sup> This is ultimately caused by electrostatic screening from the ILs, and offers the potential for the varying hydrophilic and hydrophobic interactions to be leveraged for the creation of ideal water networks for proton transport.<sup>288</sup> Formation of water wires, as observed to form in carbon nanotubes at small pore widths, are of particular interest for fast proton diffusion.<sup>279–284</sup> In carbon nanotubes, proton transport is enhanced by an order of magnitude because hydronium ions primarily exist as distorted, reactive Zundel complexes,<sup>281,295</sup> while the Eigen form is destabilized/nonexistent in these highly confined environments. As ILs will also provide a confining effect on water structure, we explore how similar mechanisms affect proton transport in IL/water mixtures.

Theoretical and computational studies can significantly aid interpretation of environment ef-

fects on proton transport. Ab initio and reactive molecular dynamics (MD) simulations have provided a multitude of insights into the transport mechanisms of excess protons in diverse systems ranging from bulk water to interfaces to biological and synthetic channels.<sup>53,284,295–310</sup> The multistate empirical valence bond (MS-EVB) model developed by Voth and coworkers has been particularly successful at enhancing theoretical understanding of proton transport.<sup>53,54,101,102,304,311</sup> The MS-EVB methodology allows for explicit and efficient simulation of the bond breaking and forming process that accompanies proton structural diffusion that accompanies the Grotthuss mechanism, allowing for an assessment of solvent and environmental effects on proton transport mechanisms.<sup>295,312</sup> MS-EVB simulations have been applied to evaluate proton solvation and transport motifs in ionic and/or confined systems that may possess similar characteristics to IL/water mixtures, such as Nafion,<sup>313,314</sup> carbon nanotubes,<sup>281,284</sup> biological/synthetic channels<sup>300,302,315–317</sup> and salt solutions.<sup>312,318</sup> For example, while an order of magnitude faster diffusion is observed for proton transport in model hydrophobic pores and carbon nanotubes of sufficiently low channel radius,<sup>281,284</sup> MS-EVB simulations have shown that more complex cavities have varied effects on the Grotthuss mechanism; while narrower channels still favor the Zundel cation, substituents located inside pores such as those found in cell membranes can serve to aid or hinder proton transport depending on the local solvated environment of the excess proton.<sup>300,302,315–317</sup> Additionally, MS-EVB simulations of protons in chloride salt solutions<sup>312,318</sup> indicate decreased proton transport rates with increasing ion concentration, partly due to a disruption in the hydrogen-bonded water structure needed for Grotthuss transport and also due to increased electrostatic interactions.

Computational work using MS-EVB and other simulation methodologies has investigated proton transport mechanisms in Nafion<sup>319</sup>, the prototypical membrane used for PEMFCs. Nafion exhibits both hydrophilic and hydrophobic regions due to the sulfonic acid groups and perfluorinated carbon backbone. It has been found that the sulfonate groups of Nafion, once deprotonated, can stabilize the Zundel configuration through a solvent-separated ion pair.<sup>313,314,320–323</sup> Additionally, theoretical calculations have determined that the electrostatic potential of the SO<sub>3</sub><sup>-</sup> groups creates a high activation barrier that hinders proton diffusion from near the SO<sub>3</sub><sup>-</sup> anions to neighboring water molecules;<sup>324</sup> proton transport thus occurs in these systems by "passing" excess protons between the electrostatic potential wells of multiple sulfonate groups through the hydrophilic regions of the membrane.<sup>325–327</sup> These are all examples of complex environmental mediation of proton transport. We expect that incorporation of ILs into Nafion membranes may alter these mechanisms by modulating existing confinement effects on the hydrogen-bonded water network, introducing asymmetric charge distributions and adding additional hydrophilic/hydrophobic driving forces. In lieu of explicit Nafion/IL simulations, in this work we focus on characterizing proton transport mechanisms in the IL/water mixtures themselves.

The percolation threshold of IL/water mixtures is particularly important for PEM applications.<sup>322,328</sup> The percolation threshold is the minimum water concentration required for connected hydrogen bond networks to fully span the system volume. Above the percolation threshold, water networks are interconnected over macroscopic lengthscales, with obvious implications for mass and ion transport. Percolation thresholds can be estimated from computer simulations employing periodic boundary conditions, and are rigorously quantified from large system-size limits.<sup>329</sup> Due to the mentioned anomalous proton transport for water wire topologies,<sup>279–284</sup> it is important to investigate whether proton diffusion rates are altered near the percolation threshold of mixtures when macroscopic water networks are formed. Graph theory analysis<sup>330–332</sup> can provide significant insight into the nature of hydrogen-bonded water networks within ion/water mixtures below and above the percolation threshold.<sup>333–335</sup> To our knowledge, the influence of mixture percolation on proton transport rates has not yet been rigorously characterized, and this is one of the purposes of our study.

In this work, we use MS-EVB simulations to study proton transport in aqueous mixtures of the 1-butyl-3-methylimidazolium tetrafluoroborate ( $[BMIM^+][BF_4^-]$ ) IL at various concentrations.  $[BMIM^+][BF_4^-]$  is hydrophilic and fully miscible with water at all concentrations, and we focus on IL/water mixtures near the percolation threshold as these are most relevant for PEM applications. To characterize proton transport, we study IL/water mixtures in which one BMIM<sup>+</sup> cation is replaced with a hydronium cation; extrapolation to lower pH conditions is done using Nernst-

Einstein relations and validated *a posteriori* by comparison to experimental conductivity studies. We find that statistically converged diffusion coefficients for the excess proton are obtainable with  $\sim 20$  ns of simulation. Graph theory and percolation analysis are employed to characterize how the hydrogen-bonded water network changes as a function of the water concentration, and we elucidate correlation between structure and transport properties. We find that water tends to form isolated clusters in [BMIM<sup>+</sup>][BF<sup>-</sup><sub>4</sub>] at concentrations below the percolation threshold, while one large, spanning cluster accompanied by smaller water clusters is observed above this value. Proton transport rates are slower in these IL/water mixtures as compared to bulk water, which is due to less connected water networks and anion stabilization of the Eigen hydronium state relative to the Zundel state. The validity of the Nernst-Einstein approximation indicates that proton transport occurs in a largely uncorrelated manner, and surprisingly we find no significant alteration of proton diffusion near the percolation threshold of the IL/water mixture. These findings on anion effects on proton solvation and transport should aid the rational design of solvents for use with proton exchange membranes.

#### 4.3 Methods

We perform molecular dynamics (MD) simulations of  $[BMIM^+][BF_4^-]$  /water mixtures, with and without hydronium ions, at water volume fraction of  $\phi_V^{H_2O} = 0.08$ , 0.13, 0.19, 0.22, 0.25 and 0.30; these concentrations span just above and below the percolation threshold for  $[BMIM^+][BF_4^-]$  /water (*vide infra*). Simulation systems consist of 220 ion pairs and 200, 350, 550, 643, 750 and 1000 water molecules (larger systems were also studied for percolation analysis, see Supporting Information). MS-EVB simulations were employed to explicitly capture the reactive proton hopping process;<sup>53,54,102,311</sup> due to computational expense, the IL/water mixtures were first equilibrated using standard MD in the NPT ensemble in absence of hydronium ions. Equilibration was conducted for 5 ns at 300K and 1 bar utilizing the OpenMM 7.3.1 software package,<sup>336</sup> employing a Langevin thermostat with a 0.1 ps<sup>-1</sup> friction coefficient and Monte Carlo barostat. Subsequently, 20 ns simulations were run in the NVT ensemble for all analysis not involving the hydronium ions.

For starting configurations for the MS-EVB simulations, one BMIM<sup>+</sup> cation was removed from each system and replaced with a hydronium ion. MS-EVB simulations were conducted with an in-house code,<sup>302</sup> following the algorithm described by Voth and coworkers.<sup>53,54,102,311</sup> The particlemesh Ewald (PME) method was used for electrostatics,<sup>36</sup> van der Waals interactions were truncated at 1.4 nm cutoff distance, and a Langevin thermostat was used with a 0.1 ps<sup>-1</sup> friction coefficient. Comparison to benchmark NVE ensemble simulations verified that this thermostat choice does not significantly alter predictions of proton transport dynamics. For each system, three separate NVT simulations were run at 300K, each for 20 ns. Four additional 5 ns simulations at  $\phi_V^{H_2O} = 0.04$  were also run to compare static properties for the low water content system.

A hybrid force field was utilized to enable the reactive, MS-EVB simulations. All proton/water interactions were modeled with the MS-EVB 3.254 and SPC/Fw337 force field combination, developed by Voth and coworkers. The SAPT-FF-UA model developed by Son et al.<sup>338</sup> was used to model  $[BMIM^+][BF_4^-]$ ; this is a non-polarizable, united atom model with explicit hydrogen atoms only on the imidazolium ring (denoted ringH\_nopol in ref<sup>338</sup>). A non-polarizable model was required for practical reasons within the MS-EVB software, but we note that explicitly polarizable simulations are expected to be more accurate for ILs.<sup>152,339</sup> Because the IL and H<sub>3</sub>O<sup>+</sup>/water force fields were developed independently and employ different functional forms, new cross-term interactions were explicitly parameterized to *ab initio* calculations in this work; these parameters are given in Table S1. For parameterization, density functional theory-based symmetry-adapted perturbation theory (DFT-SAPT)<sup>340</sup> calculations were conducted using the Psi4 software package,<sup>341</sup> and force fields were fit following previously described techniques.<sup>150</sup> To generate parameters for IL/water cross-terms, DFT-SAPT calculations were performed for 2000 water dimers in randomly generated configurations, and individual terms were fit to the SAPT energy decomposition (the total fit is shown in Figure S2a).<sup>339</sup> Combination rules were then used to generate IL/water interactions using these new water parameters in combination with the existing IL potential (Supporting Information). For IL/H<sub>3</sub>O<sup>+</sup>cross terms, we focus parameterization on the H<sub>3</sub>O<sup>+</sup>/BF<sub>4</sub><sup>-</sup> (cation/anion) interactions; 1000 close-contact  $H_3O^+/BF_4^-$  dimers were sampled from a 10 ns NVT simulation with an approximate hydronium force field, and DFT-SAPT interaction energies were computed for these dimers. The force field fit to the SAPT energy decomposition is shown in Figure S1; previously described protocol was followed to generate non-polarizable potentials.<sup>338</sup> The newly generated non-polarizable force field was benchmarked against all-atom, polarizable simulations for water and the IL as discussed in the Supporting Information.

The ionic conductivity for the  $H_3O^+/BMIM^+/BF_4^-$ /water mixtures is computed as

$$\sigma_{tot} = c_{BMIM^+} \cdot \Lambda_{BMIM^+} + c_{H_3O^+} \cdot \Lambda_{H_3O^+} + c_{BF_4^-} \cdot \Lambda_{BF_4^-}$$
(4.1)

where  $c_X$  is the concentration and  $\Lambda_X$  is the molar conductivity of ion type 'X' in the mixture. Because diffusion of hydronium ions in aqueous mixtures proceeds by a very different mechanism (Grotthuss) than other ions, the molar conductivities were computed using different approximations. Explicit Green-Kubo relations were used to compute the concentration dependent  $\Lambda_{BMIM^+}$  and  $\Lambda_{BF_4^-}$  contributions, and Nernst-Einstein expressions were used in combination with MS-EVB calculated diffusion coefficients to compute the concentration dependent  $\Lambda_{H_3O^+}$  term. The molar conductivity of BMIM<sup>+</sup>/BF<sub>4</sub><sup>-</sup>/water mixtures is computed without hydronium ions as

$$\Lambda_{BMIM^+/BF_4^-} = \lim_{t \to \infty} \frac{1}{6tNk_BT} \sum_{i,j} \langle (q_i[\mathbf{R}_i(t) - \mathbf{R}_i(0)]) \cdot (q_j[\mathbf{R}_j(t) - \mathbf{R}_j(0)]) \rangle$$
(4.2)

where "N" is the number of ion pairs in the mixture,  $q_i$  and  $\mathbf{R}_i(t)$  are the charge and position of ion "i", the sums run over all ions, and  $\langle ... \rangle$  denotes an ensemble average. Due to changes in viscosity and ion correlation, the molar conductivity for the IL depends sensitively on ion concentration,<sup>278</sup> and thus  $\Lambda_{BMIM^+/BF_4^-}$  is computed for different ion concentration mixtures from independent simulations. We make the approximation that  $\Lambda_{BMIM^+/BF_4^-}$  does not depend on the exact type of ions in the mixture (e.g. fraction of H<sub>3</sub>O<sup>+</sup>to BMIM<sup>+</sup>cations ), and compute this conductivity from pure BMIM<sup>+</sup>/BF<sub>4</sub><sup>-</sup>/water mixtures. Because of significant ion correlation at low water content,<sup>278</sup> there exists no rigorous decomposition into separate cation and anion contributions, and so we make the simplest approximation,  $\Lambda_{BMIM^+} = \Lambda_{BF_4^-} = \Lambda_{BMIM^+/BF_4^-}/2$ .
The molar conductivity of hydronium ions is computed using the Nernst-Einstein expression

$$\Lambda_{H_3O^+} = \frac{1}{k_B T} D_{H_3O^+} \tag{4.3}$$

Due to the Grotthuss mechanism, the diffusion coefficient of hydronium ions  $(D_{H_3O^+})$  depends sensitively on water content and thus is computed from separate MS-EVB simulations for each particular ion concentration, with one hydronium ion in each simulation. The diffusion coefficient is calculated from the mean square displacement of the principal H<sub>3</sub>O<sup>+</sup>diabat in the MS-EVB formalism.<sup>54</sup> Our implicit approximation that diffusion and conductivity of H<sub>3</sub>O<sup>+</sup>ions only depends on total ion/water content and not hydronium concentration/pH is validated *a posteriori* by comparison against experimental conductivity measurements (Figure 4.1).

Graph theory analysis is utilized to characterize hydrogen-bonding/water structure of  $[BMIM^+][BF_4^-]$ /water mixtures, employing the NetworkX software package.<sup>342</sup> Graphs of water networks were constructed based on standard hydrogen-bond criteria, namely two water molecules are connected by a hydrogen bond if the O-O distance is less than 0.36 nm, the O-H distance is less than 0.245 nm and the H-O-O angle is less than  $30^{\circ}$ .<sup>343,344</sup> Each water molecule is then defined as a node of the graph, and hydrogen bonds represent edges between nodes. The number of hydrogen bonds per water molecule is calculated from the degree distribution,<sup>345</sup> which provides the number of edges for each node. The spatial extent of hydrogen bond networks is characterized by the diameter (*d*) of each subgraph (i.e. hydrogen-bonded water cluster) given by<sup>346</sup>

$$d = \max_{i,j \in G} L_{min}(i,j) \tag{4.4}$$

where  $L_{min}$  is the minimum distance (number of nodes) between nodes *i*, *j* of the graph (G), and the maximum is taken over all pairs of nodes in the graph. The IL/water mixtures are generally composed of numerous, unconnected subgraphs, with each subgraph exhibiting a characteristic diameter *d*. The system is then described by the histogram/probability distribution, P(d), characterizing the probability of a subgraph to have a particular diameter. We note that if a connected water network percolates the entire system, then P(d) will inherently be system size dependent, and this subtlety will be discussed within the context of the analysis. In addition to the diameter distribution, the size of the water clusters is also characterized by the average weighted cluster size of the water subgraphs,  $n_w$ , given as<sup>347</sup>

$$n_w = \sum_n \frac{n^2 m_n}{N^2} \tag{4.5}$$

Here, *n* is the number of water molecules in a cluster,  $m_n$  is the number of clusters containing *n* water molecules and *N* is the total number of water molecules (note our normalization is slightly different by a factor of *N* compared to Ref<sup>347</sup>).

Percolation analysis is conducted to determine the probability for formation of a systemspanning hydrogen-bonded network as a function of IL/water concentration. The probability that a system contains a percolating water network is denoted  $R(\phi_V^{H_2O}, L)$ , and depends on the water volume fraction  $\phi_V^{H_2O}$  of the system, and the simulation box length L.<sup>329,348,349</sup> We compute the percolation probability using a recursive search algorithm as described by Edvinsson *et al.* [350] In the limit of infinite system size,  $R(\phi_V^{H_2O}, L)$  approaches a step function with the zero/one transition occurring at the percolation threshold.<sup>329,348,349</sup> As described in the Supporting Information, we conduct percolation analysis on different sized systems to better determine the percolation threshold; all percolation analysis was performed on IL/water MD simulations with no hydronium ions for better statistics.

Unless otherwise noted, all analysis involving the hydronium ion is conducted assuming that the location of the hydronium ion is defined by the principle diabatic state within the MS-EVB state vector. This is done for simplicity, rather than employing previous definitions based on the center-of-excess charge.<sup>53,54</sup>

### 4.4 Results and Discussion

Incorporation of ILs into PEMs is expected to alter both the thermal stability and ion transport properties of the membrane material. Thermal stability is straightforward to characterize experimentally,<sup>275</sup> and is not the focus of this work; we focus on how incorporation of ILs alters the ionic conductivity and proton transport of the membrane solution. Unlike other electrolyte applications,<sup>351–353</sup> the net proton transport rather than the total ionic conductivity is the important performance metric. Depending on the membrane solution composition, the proton transport may represent only a portion of the total ionic conductivity; indeed this fractional contribution will change as ILs are incorporated into the membrane. For example, a Nafion membrane may be interpreted as an idealized system in which protons are the only free ions in aqueous solution, while sulfonate anion groups are tethered to rigid polymer backbones.<sup>258,354</sup> In this case, the ionic conductivity results entirely from proton transport. However, if ionic liquids are then incorporated into the membrane, the ionic conductivity will contain contributions from non-hydronium ions, and thus may not directly reflect proton transport ability. Additional characterization is then necessary to determine the fractional contribution of proton transport to the total ionic conductivity.

Experimentally, the fractional contribution of proton transport to the total conductivity of a mixed ionic/aqueous solution may be inferred from separate conductivity measurements at different pH.<sup>273</sup> From a computer simulation, the proton conductivity contribution may be directly computed assuming uncorrelated behavior (e.g. Equation 4.3), but it is not *a priori* clear whether this is a good approximation for IL/water mixtures. To test such ansatz, we compute the total ionic conductivity for  $H_3O^+/BMIM^+/BF_4^-$  aqueous mixtures at varying ion concentration and pH, employing Equations Equation 4.1-Equation 4.3. We compare our predicted conductivities to experimental conductivity measurements of aqueous [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] and HBF<sub>4</sub>/[BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] aqueous mixtures by<sup>273</sup> Figure 4.1 presents a comparison of computed conductivities to experimental measurements. For all  $H_3O^+/BMIM^+/BF_4^-$  aqueous mixtures, the proton concentration was set at 1:5.28  $H_3O^+/H_2O$  mole ratio to enable direct comparison with experimental measurements of Yu



Figure 4.1: Ionic conductivity  $\sigma$  of H<sub>3</sub>O<sup>+</sup>/BMIM<sup>+</sup>/BF<sub>4</sub><sup>-</sup> and BMIM<sup>+</sup>/BF<sub>4</sub><sup>-</sup> aqueous solutions at varying water content; for H<sub>3</sub>O<sup>+</sup>/BMIM<sup>+</sup>/BF<sub>4</sub><sup>-</sup> solutions, the proton concentration is set at 1:5.28 H<sub>3</sub>O<sup>+</sup>/H<sub>2</sub>O mole ratio.<sup>273</sup> Experimental conductivities taken from ref<sup>273</sup> are plotted as triangles, simulation predictions are plotted as circles, and dashed trend lines are to guide the eye.

*et al.* [273] It is evident that the conductivity of the aqueous  $[BMIM^+][BF_4^-]$  solutions increases with water content, as the viscosity of the mixture decreases upon addition of water to the pure ionic liquid.<sup>355</sup> Within experimental uncertainty,<sup>273,276,356</sup> our predicted conductivities agree well with experiment, and also agree well with previous simulation predictions for the conductivity of aqueous  $[BMIM^+][BF_4^-]$  mixtures.<sup>278</sup> As a benchmark of the non-polarizable IL/water model used in our simulations, we compare predicted conductivities to those from simulations using a polarizable force field, as shown in Figure S5; we find that the agreement is quite good.

Figure 4.1 shows that the trend in ionic conductivity changes significantly upon addition of protons to the  $[BMIM^+][BF_4^-]$  /water mixtures. The conductivity of the aqueous  $H_3O^+$  /  $BMIM^+$  /  $BF_4^-$  mixtures nearly exactly follows a quadratic increase with proton/water content; note that because all solutions exhibit a 1:5.28  $H_3O^+/H_2O$  mole ratio, the proton content increases proportionally with water concentration. The stark difference between the conductivity trends with/without

excess protons in the aqueous  $[BMIM^+][BF_4^-]$  mixtures results from the unique and facile proton transport enabled by the proton hopping mechanism in aqueous environments. In Table S1, we provide the decomposition of the total ionic conductivity into different ion contributions corresponding to Equation 4.1. Above proton concentrations of 2 M, the total conductivity of the  $H_3O^+/[BMIM^+][BF_4^-]$  /water solution is at least double that of the  $[BMIM^+][BF_4^-]$  /water solution, indicating the primary contribution from the protons; note the molar conductivity of the protons is larger than that of the other ions for all concentrations, Table S1. The larger contribution of protons to the conductivity is due to both increasing proton *and* water concentrations; while the dependence on proton concentration is simply captured in Equation 4.1, the dependence on water content is less obvious and will be discussed in detail.

The excellent agreement between our predicted conductivities and the experimental conductivities for the aqueous  $H_3O^+/BMIM^+/BF_4^-$  mixtures (Figure 4.1) suggests a clear physical picture for the mechanism of proton transport in these solutions. In evaluating Equation 4.1, the molar conductivity of hydronium ions was computed using the Nernst-Einstein approximation (Equation 4.3), and thus neglects ion correlation between hydronium ions and the  $BMIM^+/BF_4^-$  ions. Furthermore, the hydronium diffusion coefficient employed in Equation 4.3 was computed from MS-EVB simulations with a single hydronium ion, for aqueous [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] mixtures at the total ion content of the solutions in Figure 4.1. Both of these approximations are validated a posteriori by the good agreement with experiment shown in Figure 4.1. Thus even at pH values < 0, proton transport in aqueous  $H_3O^+/BMIM^+/BF_4^-$  solutions proceeds largely by independent proton hopping events, with the microscopic proton transport mechanism seemingly unaffected by the precise solution pH within the range studied here. Proton transport in aqueous environments is primarily affected by water volume fraction of the solution ( $\phi_V^{H_2O}$ ), and the dependence on water content is the reason for the quadratic conductivity trend shown in Figure 4.1, as will be discussed below. Note that the aqueous [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] solutions of water volume fraction  $\phi_V^{H_2O} \le 0.30$  studied in this work exhibit significant ion correlation,<sup>278</sup> and thus the apparently minor influence of ion correlation on hydronium ion transport is unique to the proton structural diffusion mechanism.

Figure 4.1 indicates that the ionic conductivity of aqueous  $H_3O^+/BMIM^+/BF_4^-$  solutions is primarily due to proton transport above proton concentrations of 2 M. Thus the quadratic trend in ionic conductivity observed in Equation 4.1 reflects changing proton transport rates with water content. In Figure 4.2a we plot the computed diffusion coefficient of hydronium ion  $(D_{H_3O^+})$  in aqueous  $[BMIM^+][BF_4^-]$  solutions at varying water volume fraction (the mean square displacements used to compute these diffusion coefficients are shown in Figure S7). We find that  $D_{H_3O^+}$ increases nearly linearly with water content, similarly to the trend observed in hydrochloric acid solutions.<sup>312</sup> The quadratic trend of the ionic conductivity in Equation 4.1 is explained by the linear increase in  $D_{H_3O^+}$  with water content (Figure 4.2a), when substituted into the Nernst-Einstein expression. The quadratic dependence results from the factor of concentration in Equation 4.1 along with the linear dependence of the diffusion coefficient on concentration in Equation 4.3.

The increase in the hydronium diffusion coefficient with water concentration is partly because sufficient water content is required for formation of water networks that mediate the structural diffusion mechanism of excess protons. Indeed, our computed value of  $D_{H_3O^+} = 0.5 \cdot 10^{-5} \text{ cm}^2/\text{s}$ for the  $\phi_{\rm V}^{\rm H_2O}$ = 0.30 water content system is  $\sim$  8 times smaller than the diffusion coefficient of hydronium ions in bulk water (as computed with a similar force field<sup>54</sup>), indicating the general detrimental effect of lower water concentration on proton transport. For comparison, Figure 4.2a also shows the water  $(D_{H_2O})$  and  $BF_4^-(D_{BF_4^-})$  diffusion coefficients at each concentration; Figure S8 shows that the calculated values of  $D_{H_2O}$  are in reasonable agreement with experimental values measured in [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] /water mixtures.<sup>357</sup> At  $\phi_V^{H_2O}$ = 0.30,  $D_{H_2O}$  = 0.6 ·10<sup>-5</sup> cm<sup>2</sup>/s; this is ~ 4 times smaller than the diffusion coefficient for bulk water, indicating that  $D_{H_2O}$  is less inhibited in comparison to  $D_{H_3O^+}$ . Interestingly, in contrast to bulk water, in which  $D_{H_3O^+}$  in the MS-EVB model is ~ 1.6 times greater than  $D_{H_2O}$ ,  $D_{H_3O^+}$  is smaller than the corresponding  $D_{H_2O}$  at each concentration.<sup>54</sup>  $D_{H_2O}$  increases linearly with water concentration as does  $D_{H_3O^+}$ ; however,  $D_{H_2O}$ experiences a slower rate of increase with water concentration in comparison to  $D_{H_3O^+}$ . This suggests that, as expected, strong Coulombic interactions associated with the  $[BMIM^+][BF_4^-]$  ions have a greater effect on the diffusion of hydronium than they do on water. As seen in Figure 4.2a,  $D_{BF_4^-}$  is significantly smaller than  $D_{H_3O^+}$  at all concentrations. The differences between H<sub>3</sub>O<sup>+</sup> and BF<sub>4</sub><sup>-</sup> diffusion, which are relatively similarly sized, indicate the existence of the unique proton hopping mechanism in these environments.

The transport of hydronium ions in water proceeds by two mechanisms: standard "vehicular" diffusion, common to all molecules in solution, and structural diffusion/proton hopping, which makes hydronium diffusion much faster in bulk water compared to similarly sized ions.<sup>55</sup> The relative importance of the different mechanisms is inferred by analysis of the so-called "forward hop function", h(t), which is a measure of how frequently a proton shuttles between distinct water molecules.<sup>53</sup> Essentially, h(t) counts the number of times a proton migrates to a new water molecule, correcting for back-and-forth, "oscillatory" shuttling that does not contribute to diffusive transport.<sup>53</sup> In Figures S9-S10, we show the forward hop function h(t) for a hydronium ion in the varying concentration, aqueous [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] mixtures. For sufficiently long timescales, h(t) grows linearly with time, and the slope of the function gives an average forward hop rate. The forward hop rate as a function of water content is shown in Figure 4.2b along with the hydronium diffusion coefficient. Within statistical uncertainty, Figure 4.2b indicates that the forward hop rate of hydronium ions in the concentrated [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] /water solutions increases linearly with water content. This linear trend is similar to the dependence of the hydronium diffusion coefficient,  $D_{H_3O^+}$ , on water content. While  $D_{H_3O^+}$  is eight times smaller at  $\phi_V^{H_2O} = 0.30$  in comparison to its value in bulk water, the forward hop rate for the  $\phi_V^{H_2O}$  = 0.30 mixture is ~30 ns<sup>-1</sup>, which is five times smaller than the corresponding rate in bulk water. This implies that a substantial portion of the concentration dependence of  $D_{H_3O^+}$  is due to modulation of the proton shuttling mechanism/rate. However, the fact that these ratios are different also indicates that long-range water structure and/or vehicular diffusion rates may contribute to the concentration dependence as well. The diffusion coefficients follow a similar trend as is observed in Nafion, with  $D_{H_3O^+}$ falling below  $D_{H_2O}$  at low water concentrations.<sup>303</sup> In Nafion, this occurs because the vehicular mechanism is the dominant mode of proton transport at low water concentrations, with structural diffusion notably suppressed; we speculate that a similar effect occurs in these systems. Figure



Figure 4.2: a) Hydronium, water and  $BF_4^-$  diffusion coefficient and b) hydronium diffusion coefficient and forward hop rate within aqueous  $[BMIM^+][BF_4^-]$  mixtures at varying water volume fraction,  $\phi_V^{H_2O}$ . The dashed lines reflect the linear trend and are to guide the eye. Error bars represent the range between the maximum and minimum calculated diffusion coefficients and forward hop rates from three independent simulations.

S11 contains trajectories of the principal hydronium and the initial oxygen that the excess proton is bound to. As can be seen, the initial oxygen diffuses to a much greater area than the principal hydronium at  $\phi_V^{H_2O} = 0.08-0.13$ , suggesting that structural diffusion is less important to overall proton diffusion at these lower water concentrations. The principal hydronium clearly diffuses to a much larger area relative to the initial oxygen above these concentrations, which implicates a change in the relative importance of vehicular vs. structural diffusion. Further evidence of a change in the dominant mode of proton transport in these systems comes from computing the hop distance of the excess proton using the jump-diffusion model.<sup>358</sup> Using  $D_{H_3O^+}$  as the diffusion coefficient and 1/forward hop rate as the hop time at each concentration, we obtain a proton hopping distance of 2.4 Å at  $\phi_V^{H_2O} = 0.08$  and a proton hopping distance of 3.1 Å at  $\phi_V^{H_2O} = 0.30$ . The clear concentration suggests a change in the dominant mechanism of proton transport with water concentration. We interpret this to be an increased importance of structural diffusion at greater water concentrations as seen in Nafion, but further investigations beyond the scope of this work need to be performed in order to definitively identify the exact mechanistic character of proton transport for these systems.

The concentration dependence of the forward hop rate is due to changes in the water networks

that shuttle the proton(s). The hydrogen-bonded water network significantly influences the frequency of Grotthuss hopping, and we analyze the structure of the underlying water network to better explain the concentration dependence observed in Figure 4.2. We first analyze the distribution of hydrogen bonds per water molecule as a function of water content (see Methods for analysis details). This hydrogen bond distribution is shown in Figure 4.3 for select concentrations of the aqueous  $[BMIM^+][BF_4^-]$  mixtures, and the full concentration range is shown in Figure S12. At  $\phi_{\rm V}^{\rm H_2O}$  = 0.04 water volume fraction, the majority of water molecules exhibit one hydrogen bond, and a significant fraction are solvated entirely by ions and exhibit zero hydrogen bonds; this observation is consistent with experimental NMR studies.<sup>359</sup> As water concentration is increased from  $\phi_{\rm V}^{\rm H_2O}$  = 0.04 to  $\phi_{\rm V}^{\rm H_2O}$  = 0.13 volume fraction, the majority of water molecules form two hydrogen bonds, and there are relatively few individually solvated water molecules (zero hydrogen bonds). This indicates that water molecules are beginning to cluster together and form connected networks at concentrations near  $\phi_{V}^{H_2O} \sim 0.13$  water volume fraction. The percolation threshold of the aqueous  $[BMIM^+][BF_4^-]$  mixtures occurs near  $\phi_V^{H_2O} \sim 0.18$  (vide infra); for water concentrations above this threshold, system-spanning connected water networks are formed. At concentrations higher than the percolation threshold, the average hydrogen bond number continues to increase, as indicated by the increased probability for three and four hydrogen bonds from  $0.13 \le \phi_V^{H_2O} \le 0.22$ . The formation and growth of more "bulk-like" water networks with increasing  $\phi_V^{H_2O}$  is the reason for the concentration dependence of the proton forward hopping rate observed in Figure 4.2b. We note that water wires which may facilitate efficient proton transport<sup>281,284</sup> are not prominently formed in these solutions, as this would entail a sharp peak at 2 hydrogen bonds per water molecule.

At sufficient water content, percolating water networks fully span the system volume over macroscopic lengthscales. The percolation threshold is an important metric for proton diffusion, as more connected water networks will enable longer-ranged Grotthuss transport. Above and below the percolation threshold the water structure will be fundamentally different, with localized water clusters appearing below the percolation threshold and a system spanning cluster/network appearing above. The probability that the system exhibits a percolating water network is denoted



Figure 4.3: Hydrogen bond distribution of water molecules in aqueous [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] mixtures as a function of water volume fraction,  $\phi_V^{H_2O}$ . For comparison, the distribution for bulk water is shown as computed from similar simulations. A simulation snapshot of water clusters at  $\phi_V^{H_2O}$ = 0.13 is also shown.

 $R(\phi_V^{H_2O}, L)$ , and depends on water volume fraction  $\phi_V^{H_2O}$  and simulation box size "L"; in the limit of an infinite system,  $R(\phi_V^{H_2O}, L)$  will be a step function, with the abrupt transition occurring at the percolation threshold.<sup>329,348</sup> To better locate the percolation threshold in the macroscopic system limit, we compute  $R(\phi_V^{H_2O}, L)$  for different system sizes with L ~ 4.5 and 9 nm (see Supporting Information).

Figure 4.4a shows the probability  $R(\phi_V^{H_2O}, L)$  for a percolating network to form, as a function of the water volume fraction in aqueous  $[BMIM^+][BF_4^-]$  mixtures for the different system sizes. The curves are sigmoidal and approach a step function as  $L \rightarrow \infty$ , and the percolation threshold occurs at  $\phi_V^{H_2O} \sim 0.18$  where the curves intersect (indicated by the dashed vertical line in Figure 4.4a). Note that a previous study estimated the percolation threshold in  $[BMIM^+][BF_4^-]$  /water to be  $\phi_V^{H_2O} = 0.25$  by qualitative inspection.<sup>360</sup> The percolation threshold of  $\phi_V^{H_2O} \sim 0.18$  corresponds to  $\sim 10$  M water concentration, which interestingly coincides with the onset of significantly greater conductivity of aqueous  $H_3O^+/BMIM^+/BF_4^-$  solutions compared to aqueous  $[BMIM^+][BF_4^-]$  as shown in Figure 4.1. Thus above  $\phi_V^{H_2O} \sim 0.18$ , connected water networks extend over macroscopic lengthscales and proton transport contributes a significant fraction of the total ionic conductivity



Figure 4.4: (a) Percolation probability  $R(\phi_V^{H_2O}, L)$  and (b) weighted water cluster size  $n_w$  (Equation 4.5) as a function of water volume fraction  $\phi_V^{H_2O}$  in aqueous [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] mixtures. Metrics are computed for different system sizes, and the percolation threshold is indicated by the vertical dashed line.

(for 1:5.28 H<sub>3</sub>O<sup>+</sup>/H<sub>2</sub>O mole ratio mixtures). Taken together, these properties suggest that aqueous H<sub>3</sub>O<sup>+</sup>/BMIM<sup>+</sup>/BF<sub>4</sub><sup>-</sup> solutions near  $\phi_{V}^{H_2O} \sim 0.18$  water volume fraction could be ideally suited for utilization in PEM applications.

Near the percolation threshold, large water clusters are formed with significant fluctuations in cluster size.<sup>347</sup> To better analyze these clusters, we compute the weighted cluster size,  $n_w$  (Equation 4.5), which is shown in Figure 4.4b. The weighted cluster size is system size dependent below the percolation threshold; near the percolation threshold, predictions from different system sizes are similar. The plot of  $n_w$  vs  $\phi_V^{H_2O}$ (Figure 4.4b) shows a very similar concentration dependence as the percolation probability,  $R(\phi_V^{H_2O}, L)$  (Figure 4.4a), implying strong correlation between the weighted cluster size metric and percolation, as noted by Geiger *et al.* [347] Below the percolation threshold, there is a tendency for many small, similarly sized water clusters to form. As the water concentration approaches the percolation threshold, there is a steep increase in the weighted cluster size signifying growth of larger clusters which contain a significant fraction of the total water present in the system; however, numerous small water clusters still exist in the mixture near the percolation threshold.

It might be expected that a marked change in the hydronium diffusion rate would occur near the percolation threshold due to the intrinsic changes the underlying water networks; however, this is apparently not the case. As shown in Figure 4.2b, there is a linear relationship between the hydronium diffusion coefficient and proton forward hopping rate with increasing water content, with no discernible change in trend near the percolation threshold of  $\phi_V^{H_2O}$  = 0.18. This suggests that the percolation transition does not affect proton forward hopping rates, beyond the general linear concentration-dependent trend. The reason for this is most likely because of separation in time scale between proton transport and water network rearrangement. For example, hydronium diffusion coefficients (Figure 4.2) and mean-square displacements (Figures S6-S7) indicate that hundreds of picoseconds are required for protons to migrate nanometer distances within the aqueous  $[BMIM^+][BF_4^-]$  mixtures. On hundreds of picosecond to nanosecond timescales, nanometer water networks will break, form, and rearrange. This "loss of memory" of nanometer network structure over the timescale for proton diffusion is the reason for the insignificant influence of the percolation threshold on hydronium diffusion rates. We note that in membrane systems that exhibit a high degree of nanoconfinement, and/or when employing a more viscous ionic liquid, the water dynamics/rearrangement would be slower and there could potentially be a more significant effect of percolation on proton dynamics. However, in the aqueous [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] mixtures, proton diffusion is most affected by the local water structure around the hydronium ion.

Hydronium ions serve as nucleation sites for larger water clusters and networks within aqueous  $[BMIM^+][BF_4^-]$ , and thus partially facilitate their own proton shuttling. This is indicated through analysis of the size/diameter of water clusters (Equation 4.4), by separately computing distributions for clusters that do or do not contain an excess proton. This analysis is shown in Figure 4.5 for concentrations both below ( $\phi_V^{H_2O}$ = 0.13) and above ( $\phi_V^{H_2O}$ = 0.25) the percolation threshold (similar graphs for all concentrations are shown in Figure S13). Below the percolation threshold at  $\phi_V^{H_2O}$ = 0.13, the majority of clusters are small and many isolated water molecules are present in the mixture; this interpretation is consistent with the previously discussed hydrogen bond distribution (Figure 4.3). Note that the weighted cluster size analysis in Figure 4.4b is comparatively

skewed towards larger clusters as the weighting depends on the number of molecules per cluster, whereas Figure 4.5 represents a per-cluster histogram. Interestingly, as shown in Figure 4.5, water clusters that contain a hydronium ion are generally much larger, consisting of a greater number of water molecules. This indicates that hydronium ions reside in larger-than-average water clusters, and serve as nucleation sites for water clusters and networks. This is important, because it indicates that hydronium ions may effectively nucleate their own water networks to facilitate Grotthuss transport, and that the percolation threshold of aqueous ionic liquid solutions will most likely be pH dependent. The motif that hydronium ions nucleate their own water clusters/transport networks is similar to that observed in studies of other systems: for example, Peng et al. [361] have shown that an excess proton can create its own water wires in confined environments such as carbon nanotubes and other hydrophobic cavities.<sup>281,284,362</sup> We note that previous work on proton transport in surfactant-based lyotropic liquid crystals emphasized the importance of interactions of the hydronium oxygen atom with hydrophobic moities on the surfactants.<sup>301,302</sup> We have computed correlation functions between the oxygen atom of the hydronium ion and aliphatic chains of the BMIM<sup>+</sup>cation, which are shown in Figure S14; while there is tendency for coordination with the hydrophobic BMIM<sup>+</sup> tail, concentration effects are subtle and appear relatively minor.

The cluster diameter distributions fundamentally change for concentrations above the percolation threshold. Above the percolation threshold, there is a bimodal distribution observed in Figure 4.5, with a clear secondary peak formed at significantly larger diameter (for our system sizes, this occurs at  $d \approx 50$ ). Figure S13 indicate that a bimodal distribution exists for all concentrations above the percolation threshold ( $\phi_V^{H_2O}$ = 0.19, 0.22, 0.25, and 0.30), while only single modal distributions are observed for concentrations below the percolation threshold ( $\phi_V^{H_2O}$ = 0.08 and 0.13). Therefore the secondary peak in the distribution represents the formation of the percolating water cluster. Note that because only one minimum image of each water molecule is used in the graph theory analysis (Equation 4.4), the size of the percolating water cluster is calculated as finite rather than infinite. A more marked change is apparent in the clusters that contain hydronium ions (H<sub>3</sub>O<sup>+</sup>Cluster, Figure 4.5). The hydronium ion(s) primarily reside in the large water clus-



Figure 4.5: Water cluster diameter "d" (Equation 4.4) distributions for a)  $\phi_V^{H_2O} = 0.13$  and b)  $\phi_V^{H_2O} = 0.25 H_3O^+/[BMIM^+][BF_4^-]$ /water mixtures. For each system, distributions are shown for all water clusters and those that contain a hydronium ion (H<sub>3</sub>O<sup>+</sup> cluster). The insets depict a magnified view of the tail ends of the distribution, where a secondary peak is observed for concentrations above the percolation threshold.

ter that percolates/spans the system, as indicated by the large probability of the second peak in the distribution. This indicates that proton transport primarily occurs within the large percolating water network. The conclusion that the hydronium ion nucleates/resides in larger water clusters is general to all concentrations as shown in Figure S13. Because the percolating water network consists of a large fraction of the total water molecules for the  $\phi_V^{H_2O}$  = 0.25 system (Figure 4.4b), it is highly improbable for the hydronium ion to not reside within the percolating water network at this concentration.

There is a strong energetic driving force for a hydronium ion to hydrogen bond to three water molecules within aqueous  $[BMIM^+][BF_4^-]$  mixtures, even at low water content, which is the reason that hydronium ions tend to nucleate larger water clusters. Note that the hydronium ion is effectively amphiphilic,<sup>225,297,318,361,363,364</sup> as its protons are strong hydrogen-bond donors, but its oxygen atom has only a small probability of accepting a hydrogen-bond due to the positive charge of the ion; this is why hydronium generally makes three and not four hydrogen bonds.<sup>54</sup> The energetic driving force for formation of three hydrogen bonds between hydronium and water molecules is indicated by the hydronium/water pairwise correlation function. This pairwise cor-

relation function is shown in Figure 4.6 as a function of water content, and correlation functions for all concentrations are given in Figure S15 (note that we plot  $\rho q(r)$  rather than q(r) itself for consistent normalization across concentrations). In Figure 4.6, the first two sharp peaks at  $\sim 1.4$ and  $\sim 3.0$  Å correspond to water molecules in the first solvation shell, and the smaller magnitude and broader shoulder peaks at  $\sim$  3.5 and  $\sim$  5.0 Å correspond to water molecules in the second solvation shell around the hydronium. Two correlation peaks per water molecule are observed; the closer peak reflects the proton involved in the O-H hydrogen bond, and the farther peak reflects the two farther protons of the hydronium ion. Interestingly, the magnitudes of the first solvation shell peaks do not change significantly with water volume fraction; for all concentrations, a coordination number of three is obtained when integrating the first peak of the pair distributions in Figure 4.6, corresponding to the three water molecules receiving hydrogen bonds from the hydronium ion. Thus the hydronium ion is solvated by (at least) three water molecules even at the lowest water content [BMIM<sup>+</sup>][BF<sup>-</sup><sub>4</sub>] solution, evidently due to the strongly favorable energetics of these interactions. With increasing water content, water molecules fill in the second solvation shell around the hydronium ion, as indicated by the increasing magnitudes of the peaks at  $\sim 3.5$ and  $\sim 5.0$  Å in Figure 4.6. This is consistent with the increasing diameter of hydronium/water clusters with water content, as seen by comparing Figure 4.5 and Figure S13. By comparing these metrics with the water-water hydrogen bond distribution in Figure 4.3, it is clear that at low water content the hydronium ion forms more hydrogen bonds than does the average water molecule. This is the reason why the hydronium ion acts as a nucleation site for water networks/clusters, manifesting in the shifted diameter distribution of hydronium/water clusters relative to pure water clusters (Figure 4.5, Figure S13).

In addition to the size and topology of the water networks, strong Coulombic interactions from the ionic liquid modulate the local structure and transport of hydronium ions in the aqueous  $[BMIM^+][BF_4^-]$  mixture. As mentioned in the Introduction, hydronium ions exist in two primary motifs: the Eigen form, a H<sub>9</sub>O<sub>4</sub><sup>+</sup> configuration with a localized hydronium ion solvated by three water molecules, and the Zundel form, a H<sub>5</sub>O<sub>2</sub><sup>+</sup> configuration in which the excess proton is shared



Figure 4.6: Pairwise correlation between hydrogen atoms of hydronium ion and oxygen atoms of water molecules as a function of water volume fraction  $\phi_{\rm V}^{\rm H_2O}$  in aqueous [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>]. Note that rather than plotting the radial distribution function, g(r), we plot  $\rho g(r)$  normalized by the density of water to better compare between the different water concentration mixtures.

equally between two water molecules.<sup>53</sup>. The distribution and interconversion between the Eigen and Zundel motifs play an important role in the mechanism of Grotthuss transport.<sup>53,295</sup> The ions in the concentrated [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] solutions alter the relative Eigen/Zundel distribution relative to the distribution in bulk water; this is shown in Figure 4.7. In Figure 4.7 we plot the probability of the MS-EVB state vectors<sup>53</sup> for different water concentration mixtures, focusing on the principle and second largest state component (Figure S16 provides similar analysis for all concentrations). The Eigen motif corresponds to peaks in the distribution at  $c_1^2 \approx 0.64$  and  $c_2^2 \approx 0.13$ , while the Zundel motif is identified by near equality of the first two expansion coefficients, i.e.  $c_1^2 \approx c_2^2 \approx$  $0.45.^{53,295}$  While the distributions in aqueous [BMIM<sup>+</sup>][BF<sup>-</sup><sub>4</sub>] exhibit the same general shape as for hydronium in bulk water, the relative probability of the Eigen and Zundel motifs is different due to the ionic content. At the low water concentration of  $\phi_V^{H_2O} = 0.04$ , the Eigen structure is noticeably more probable relative to the distribution in bulk water, while the probability for the Zundel structure is lower. The relative Eigen/Zundel probability begins to converge to the bulk water distribution with increasing water concentration, but generally there is higher probability for Eigen motifs within all the aqueous [BMIM<sup>+</sup>][BF<sup>-</sup><sub>4</sub>] solutions. The greater probability for the Eigen structure helps explains the slower self-diffusion of hydronium in aqueous [BMIM<sup>+</sup>][BF<sup>-</sup><sub>4</sub>]



Figure 4.7: Probability distribution of the MS-EVB coefficients for the principle  $(c_1^2)$  and second largest  $(c_2^2)$  states. The  $c_1^2$  distribution is on the right towards larger coefficient values, and the  $c_2^2$  distribution is the on the left towards smaller coefficient values. For comparison, we plot the corresponding distribution for hydronium in bulk water ( $\phi_V^{H_2O} = 1.0$ ) computed from an analogous simulation.

relative to bulk water, as the Grotthuss mechanism invokes the Zundel configuration as a "transition state" for proton transfer.<sup>295</sup>

The Eigen form of the hydronium ion is stabilized by strong Coulombic attraction with BF<sub>4</sub><sup>-</sup> anions in the aqueous [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] mixtures. Note that the Eigen form of hydronium has localized charge density as the proton primarily resides on one water molecule, whereas the Zundel form delocalizes the excess charge over two water molecules;<sup>53</sup> therefore it is expected that the Eigen motif exhibits stronger interaction with anions. To confirm this, we compute a "decomposed" radial distribution function, g(r), between the hydronium cation and the BF<sub>4</sub><sup>-</sup> anions. This distribution is decomposed into contributions from the Eigen and Zundel states to illuminate how the correlation with the anion changes as a function of proton localization. Explicitly, we compute the coordinate  $\delta = |\mathbf{r}_{0^*H^*} - \mathbf{r}_{0_{1x}H^*}|$ , where O<sup>\*</sup> is the principal hydronium oxygen, H<sup>\*</sup> is the principal hydronium hydrogen and O<sub>1x</sub> is the closest water molecule present in the first solvation shell of the hydronium.<sup>295,308</sup> Eigen structures are defined as  $\delta > 0.4$  and Zundel structures are defined as  $\delta < 0.1$ .<sup>295,308</sup> The correlation function is then decomposed by separately collecting statistics when the hydronium ion exists in either of these motifs. The decomposed pairwise correlation between the hydronium ion and  $BF_4^-$  anion is shown in Figure 4.8 for the aqueous  $[BMIM^+][BF_4^-]$  mixture at  $\phi_V^{H_2O} = 0.25$  water content (Figure S16 shows the distributions for other concentrations). Note that these distributions are inherently noisy as there is only one hydronium ion in the system and also many structures are not used for statistical averaging (those with  $0.1 < \delta < 0.4$ ); however, comparison across different concentrations in Figure S17 indicates that qualitative conclusions are robust in lieu of statistical noise.

The hydronium/BF<sub>4</sub> pairwise distribution shown in Figure 4.8 exhibits a small peak near ~ 5.0 Å and a closer shoulder at 3.0 Å. Notably, both the primary peak and shoulder have higher intensity when the hydronium ion is in the Eigen form than when hydronium is in the Zundel form. This signifies that the BF<sub>4</sub> anions coordinate more strongly to the Eigen motif, which is due to its more localized charge distribution.<sup>318</sup> It is important to note that the BF<sub>4</sub> coordination does not disrupt first-solvation shell water interactions, but rather substitutes for water molecules in the second solvation shell. This is indicated by the constant coordination (Figure 4.6), as well as the fact that the peaks in hydronium/BF<sub>4</sub> g(r) exist at roughly the same positions as the two peaks/shoulders corresponding to the second solvation shell water molecules (Figure 4.6). The enhanced BF<sub>4</sub> coordination, which provides electrostatic stabilization of the hydronium ion in the Eigen motif, is the reason for the enhanced Eigen populations in the aqueous [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] mixtures (Figure 4.7). This hydronium/BF<sub>4</sub><sup>-</sup> interaction likely contributes to the decreased hydronium diffusion in comparison to water as observed in Figure 4.2.

#### 4.5 Conclusion

Utilizing reactive molecular dynamics simulations, we have thoroughly characterized proton transport in aqueous  $[BMIM^+][BF_4^-]$  mixtures at concentrations near the percolation threshold. We have provided an ansatz for predicting/interpreting the pH-dependent, ionic conductivity of H<sub>3</sub>O<sup>+</sup>/  $BMIM^+/BF_4^-/$  water mixtures that demonstrates excellent agreement with previous experimental conductivity studies. Despite the inaccuracy of Nernst-Einstein conductivity relations for neat



Figure 4.8: Pairwise correlation g(r) for hydronium/BF<sub>4</sub><sup>-</sup> ions in aqueous [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] at  $\phi_V^{H_2O}$  = 0.25 water content. The distribution is decomposed into Eigen/Zundel contributions as discussed in the text, and is computed between hydrogen atoms of the hydronium ion and fluorine atoms of BF<sub>4</sub><sup>-</sup> anions.

ionic liquids due to neglect of ion correlation,<sup>278</sup> such expressions work well in combination with MS-EVB computed diffusion coefficients to predict proton conductivity of the mixtures. This is presumably due to the unique structural diffusion mechanism of the protons, and implies that proton transport occurs primarily in an uncorrelated fashion, depending mainly on local water networks. For  $\phi_V^{H_2O} = 0.30$  water volume fraction [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] mixtures, we find that diffusion of hydronium ions is ~ 8 times slower than rates in bulk water. This is partly due to the stabilization of the Eigen form of the hydronium ion from electrostatic interactions with BF<sub>4</sub><sup>-</sup> anions, which reduces proton transport rates. From the concentrated ionic liquid, we find that the proton diffusion coefficient increases linearly with water content due to a corresponding linear increase in proton forward hopping rates.

Consistent with previous studies,<sup>361</sup> we have shown that the hydronium ion nucleates larger water clusters and networks within concentrated ionic solutions, facilitating its own structural diffusion. A full first-solvation shell of water molecules is formed around hydronium ions for water content as low as  $\phi_V^{H_2O} = 0.04$  volume fraction, due to the strong energetic driving force for hydronium to hydrogen bond to three water molecules; for reference, water molecules form only one hydrogen bond on average at this concentration. This means that water networks and clusters

within concentrated  $[BMIM^+][BF_4^-]$  /water mixtures will be significantly pH dependent. Indeed, we have shown that water clusters and networks incorporating a hydronium ion tend to be much larger and contain more water molecules than pure water clusters within the mixture. Thus hydronium ions are tightly coupled with water solvation structures under low water conditions, and therefore the perturbation of water network structures by hydronium ions is an important consideration in addition to the intrinsic water networks of aqueous  $[BMIM^+][BF_4^-]$  solutions themselves.

We have analyzed the percolation behavior of aqueous  $[BMIM^+][BF_4^-]$  mixtures, which is particularly relevant for membrane applications in which water networks should extend over macroscopic lengthscales. The percolation threshold of aqueous [BMIM<sup>+</sup>][BF<sub>4</sub><sup>-</sup>] mixtures is predicted to occur at water volume fraction of  $\phi_V^{H_2O} \sim 0.18$ , but as mentioned, this value will most-likely be pH dependent. Combined with the fact that proton transport makes up a significant contribution of the total ionic conductivity for  $\phi_V^{H_2O} \sim 0.18$ , aqueous  $H_3O^+/BMIM^+/BF_4^-$  mixtures at 1:5.28  $H_3O^+/H_2O$  mole ratio, we suggest that this approximate IL/water concentration may be optimal for utilization in PEM materials. Somewhat surprisingly, we found that the percolation transition of the mixture had no significant effect on proton transport rates, which we attribute to differences in time scale between proton diffusion and water network rearrangement. However, it is yet to be explored whether this is true in nanoconfined systems (membranes) for which water dynamics would be slower, and percolation may play a bigger role in influencing macroscopic transport rates. It is possible that other ionic liquids may be more optimal for tuning the percolation threshold and or water networks, for example  $[BMIM^+][BF_4^-]$  may be potentially too miscible with water, forming smaller isolated clusters, and a more hydrophobic (but also miscible) ionic liquid may be preferred.<sup>289</sup> Additionally, different choice of anion may favor the Zundel configuration instead of the Eigen configuration, with benefits for proton transport rates. The many previous examples of enhanced proton transport in low-water content systems<sup>281,302,365,366</sup> suggest much room for further optimization of ionic liquid-based proton conducting solutions.

## CHAPTER 5

# EXCITATIONS FOLLOW (OR LEAD?) DENSITY SCALING IN PROPYLENE CARBONATE

This chapter reproduces the work in Ref. [367].

## 5.1 Abstract

Structural excitations that enable interbasin (IB) barrier crossings on a potential energy landscape (PEL) are thought to play a facilitating role in the relaxation of liquids. Here we show that the population of these excitations exhibits the same density scaling observed for  $\alpha$  relaxation in propylene carbonate (PC), even though they are heavily influenced by intramolecular modes. We also find that IB crossing modes exhibit a Grüneisen parameter ( $\gamma_G$ ) that is approximately equivalent to the density scaling parameter  $\gamma_{TS}$ . This observations suggests that the well-documented relationship between  $\gamma_G$  and  $\gamma_{TS}$  may be a direct result of the pressure dependence of the frequency of unstable (relaxation) modes associated with IB motion.

## 5.2 Introduction

For many decades there has been evidence that discrete molecular rearrangements, so-called "hops", contribute significantly to relaxation in viscous liquids.<sup>368,369</sup> At reduced temperatures it is thought that molecules become trapped in a cage formed by their neighbors so that relaxation requires collective rearrangements of neighborhood particles<sup>370</sup> that are thought to be facilitated by structural excitations.<sup>371,372</sup>

The case for the importance of excitations in influencing liquid dynamics is supported through their ability to account for super-Arrhenius relaxation,<sup>373</sup> the breakdown of the Stokes-Einstein relation,<sup>374</sup> the appearance of dynamic heterogeneity,<sup>375,376</sup> and the behavior of specific heat at the

calorimetric glass transition.<sup>377</sup> If excitations are a causal factor in liquid dynamics, one would expect also a connection to the ubiquitous scale-invariance observed in liquid dynamics,<sup>378–392</sup> but this connection has not been explored.

As pointed out by Roland et al.,<sup>388</sup> expectation of scale invariance arises when the potential that controls system behavior is well-approximated by a pairwise additive inverse power law (IPL).<sup>389,392</sup> Dyre and coworkers<sup>379,390,392,393</sup> have expanded these ideas, demonstrating pressure and temperature invariance of dynamic, structural, and thermodynamic properties of model and real liquids when length and energy are properly scaled.<sup>379</sup> These properties are invariant along "isomorphs" of constant  $\rho^{\gamma}/T$ , where  $\gamma = n/3$  and n is the exponent of the IPL pair potential. In systems where only the IPL potential is important, the isomorph scaling exponent  $\gamma$  is related to instantaneous fluctuations in energy and the first pressure virial.<sup>379,390–392</sup>

When the dominating interactions in a liquid do not conform to IPL potentials the entire gamut of isomorph behavior can be lost. This is observed in systems with multiple interaction length-scales, dominant electrostatic forces, or those with distinct directionality such as hydrogen bond-ing.<sup>392</sup> For example, MD simulations of the SPC water model have shown that there is essentially no virial-energy correlation, and that isomorphic state points do not exist.<sup>379,393</sup> Similar results are found in experimental studies of heterogeneous hydrogen bonding (e.g. polyalcohol) systems.<sup>394,395</sup>

When only minor motional modes of a system (such as intramolecular or long-range intermolecular degrees of freedom) do not conform to IPL potentials, some aspects of isomorph behavior are lost, but many other aspects may be retained. These systems exhibit what is sometimes referred to as "pseudoisomorphs,"<sup>396</sup> and properties associated with long timescales such as  $\alpha$  relaxation may exhibit density scaling in that they are invariant at constant values of  $\rho^{\gamma_{TS}}/T$  where  $\gamma_{TS}$  is an empirical density scaling exponent. Density scaling has been experimentally verified for a wide variety of systems, including van der Waals, and other non-associating liquids, polymers, ionic liquids, and some associating and hydrogen bonding liquids.<sup>380–388</sup> Several theoretical and experimental works<sup>392,397,398</sup> have demonstrated a connection between  $\gamma_{TS}$  and the volume dependence of vibrational frequency for spatially extended collective modes of the liquid through the Grüneisen parameter.

In these pseudoisomorph systems, pressure-energy correlations are broken since energy can partition to the non-IPL modes without strong coupling to changes in volume or pressure. The pressure of these fluids does not depend strongly on forces arising from angle vibrations and torsional rotations of at least some intramolecular modes.<sup>399</sup> Olsen et al.<sup>400</sup> found in model diatomic and oligomeric systems with harmonic intramolecular bonds that vibrational eigenmodes representing center-of-mass motion scale properly and could be used to find the scaling exponent, but that the highest frequency eigenmodes did not scale. The number of non-scaling eigenmodes was equal to the number of intramolecular bonds, so all of the intramolecular modes for the diatomic model, and at least 1/3 of the highest frequency intramolecular modes for the oligomeric system did not scale.

Thus, while low-frequency vibrational modes and slow dynamic processes follow density scaling, there appears to be some cutoff in timescale or lengthscale below which dynamic processes in the liquid no longer scale. Tölle et al. found that relaxation processes in ortho-terphenyl on timescales of 1 ns and longer follow density scaling.<sup>378</sup> Hansen et al.<sup>382</sup> and Riberio et al.<sup>401</sup> demonstrated invariance down to 100 ps and 10 ps respectively. Puosi et al.<sup>383</sup> found density scaling at timescales of the "fast  $\beta$ " ( $\beta_{fast}$ ) relaxation, roughly 1 ps. In that work, Puosi et al. quantified the mean-squared displacement  $\langle u^2 \rangle$  of particles in course-grained polymer systems, finding excellent correspondence between  $\gamma_{TS}$  obtained for  $\langle u^2 \rangle$  and  $\alpha$  relaxation, but only when a small term quadratic in ( $T\rho^{-\gamma_{TS}}$ ) was added to fit the latter.

We have recently shown that the Angstrom-lengthscale and picosecond timescale motion that has traditionally been classified as  $\beta_{fast}$  relaxation and measured through  $\langle u^2 \rangle$  contains signatures of two separate processes.<sup>402,403</sup> The faster of the two corresponds to elastic deformations of equilibrium local atomic structure, corresponding to so-called "inherent state" (IS) dynamics on a potential energy landscape (PEL).<sup>404</sup> The slower of the two processes represents<sup>403,405</sup> rearrangements of local atomic structure corresponding to interbasin (IB) barrier crossings on a PEL that lead to a new IS, and the instantaneous population of IB crossing events is equivalent to the excitation population.<sup>406</sup> We have further found evidence that intramolecular modes may figure prominently in excitations,<sup>405</sup> leading us to question whether excitations and all the components of  $\beta_{fast}$  should be expected to have the same  $\gamma_{TS}$  as the  $\alpha$  process, or whether involvement of intramolecular modes in these short-timescale, local motions break this scaling.

To investigate to what extent excitations obey density scaling, we have conducted molecular dynamics simulations of propylene carbonate (PC) over a range of pressures and temperatures. PC was previously shown to follow density scaling,<sup>398,407</sup> and provides an interesting case study for this investigation due to its strong electrostatic interactions ( $\epsilon \sim 65$ ), molecular anisotropy, and the considerable flexibility of its ring, all of which affect its liquid state properties.<sup>408–410</sup>

Our simulations utilize an *ab initio*, polarizable force field which predicts properties that are in excellent agreement with experimental measurements near STP. By searching its phase diagram, we identify two P-T curves in PC that exhibit largely invariant structural and dynamical properties. We then quantify excitation populations at these phase points, utilizing a previously benchmarked approach.<sup>402,403</sup> We conclude that excitations fundamentally involve the atomic-scale intramolecular structure of PC, and are not well characterized by center of mass motion alone. Furthermore, we demonstrate that excitation populations in this liquid obey density scaling, as they are constant across the P-T curves exhibiting invariant structure and dynamics. We show also that motion associated with excitations are the fastest dynamic processes that can be considered to follow density scaling.

## 5.3 Methods

#### 5.3.1 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were conducted for propylene carbonate (PC) utilizing the OpenMM simulation package.<sup>163</sup> We utilize the atomistic, polarizable SAPT-FF force field, which is an *ab initio* force field developed on the basis of symmetry adapted perturbation theory (SAPT) calculations.<sup>150</sup> We supplement these existing force field parameters with atomic charges and dihe-

dral potentials that are specifically parameterized for PC utilizing additional DFT calculations, as described in the Supporting Information. Polarization is incorporated via a Drude oscillator model utilizing a dual-Langevin thermostat as implemented in OpenMM, with a 1 ps<sup>-1</sup> friction coefficient used for both thermostats.<sup>411</sup> The particle-mesh Ewald method (PME)<sup>36</sup> is used for electrostatics and van der Waals (VDWs) interactions are computed up to a 1.4 nm cutoff. Each simulation consisted of 400 propylene carbonate molecules initially constructed using Packmol.<sup>412</sup> Equilibration consisted of 1 ns NPT simulations conducted with a Monte Carlo Barostat, followed by subsequent 5 ns NVT simulations at the equilibrated density. All simulations utilized a 1 fs timestep. Because OpenMM utilizes a Monte Carlo Barostat, the virial is not directly accessible. We have thus utilized Gromacs to conduct an additional MD simulation in the NVT ensemble at 300 K to compute energy-virial correlation, shown in Supplementary Fig. S1.<sup>413</sup> The simulation details for the Gromacs simulation are largely identical to the OpenMM simulations, except that in the former Drude oscillator positions are treated strictly adiabatically (minimized at every timestep). See the Supporting Information for more details about the simulation settings.

## 5.3.2 Isomorph Search

Finding isomorph families for Lennard Jones (LJ) fluids and other simple systems for which pairwise IPL potentials dominate is accomplished by analyzing the virial-energy correlation function.<sup>379</sup> For IPL potentials, the virial and potential energy are perfectly correlated, so that the normalized correlation function is unity.<sup>393</sup> This is not possible for systems with significant intramolecular degrees of freedom. For organic liquids with molecular constituents of increasing size and complexity, it is expected that isomorphic behavior will not be found based on analysis of virial-energy correlation.<sup>382,383,396</sup> This is because *intramolecular* rather than *intermolecular* interactions dominate the potential energy of the system, with the former growing as 3N-6 as the size "N" of the molecules gets larger. The intramolecular energy are not expected to correlate with virial fluctuations. The latter is inferred from the fact that the pressure of fluids does not depend on forces arising from angle vibrations and torsional rotations.<sup>399</sup> For molecules of large size and complexity, near zero correlation of the virial-energy is expected due to the dominance of the intramolecular interactions. Experimentally, Hansen et al. have shown that density scaling breaks down for dynamic signatures indicative of intramolecular motion.<sup>382</sup>

We have computed the virial-energy correlation from an MD simulation of PC at 300 K, 1 atm; the results are shown in Suplemental Fig. S1. We plot the data as normalized fluctuations of each quantity, in analogy to previous work.<sup>383,393,396</sup> As expected from the dominance of the intramolecular energetics of PC and similar to results for other flexible molecules,<sup>383,396</sup> we find essentially no correlation between the virial and energy fluctuations.

We note that if it *were* possible to explicitly separate the *intramolecular* and *intermolecular* energy contributions, then one could analyze virial correlations with just the latter. However, such a separation is generally not tractable within a molecular dynamics code. The reason being that electrostatic interactions, which contribute to intramolecular energy for 1-4 and longer atom separations, are computed with an Ewald/PME sum and not easily separated on a pairwise interaction basis. Furthermore, the electrostatic energy is formally many-body for explicitly polarizable force fields, as utilized in this work.

Because of these considerations, in this work we utilize a brute-force approach to search for isomorph curves. We searched for two isomorph curves in the P-T phase diagram, shown in Figure 5.1. We initially ran a 1 ns NPT simulation of 400 PC molecules at temperature/pressure values of 300 K/1 atm; simulation settings are the same as noted above. We then performed a 20 ns NVT simulation. We calculated the diffusion coefficient ( $D_T$ ) and the inverse rotational correlation time ( $1/\tau_{rot}$ ) for this state point.  $D_T$  was calculated using the Einstein relation:

$$D_T = \lim_{t \to \infty} \frac{1}{6} \frac{\partial \langle |\boldsymbol{r}(t) - \boldsymbol{r_0}|^2 \rangle}{\partial t}$$
(5.1)

and  $\tau_{rot}$  was calculated by integrating the rotational correlation function:

$$\tau_{rot} = \int_0^{t_{max}} dt \langle \mathbf{z}_{CO}(0) \cdot \mathbf{z}_{CO}(t) \rangle$$
(5.2)

with  $t_{max}$  equal to 1 ns and  $\mathbf{z}_{CO}$  equal to the carbonyl bond vector. As isomorph properties are only apparent in reduced units, the calculated values of  $D_T$  and  $\tau_{rot}$  were converted to reduced unit values  $\tilde{D}_T$  and  $\tilde{\tau}_{rot}$ .<sup>379</sup> We performed a series of 1 ns NPT/20 ns NVT simulations at temperatures of 325, 350, 375 and 400 K and a range of pressures.  $\tilde{D}_T$  and  $1/\tilde{\tau}_{rot}$  were computed at each state point. A point was classified as an isomorph point if the calculated values agreed with  $\tilde{D}_T$  and  $1/\tilde{\tau}_{rot}$  at 300 K/1 atm (within statistical uncertainty).

We followed a very similar procedure for the second isomorph. Here, we initially ran a simulation at a temperature/pressure value of 250 K/1 atm. As the dynamics are noticeably slower at this state point compared to 300 K/1 atm, we ran 50 ns NVT simulations after the initial 1 ns NPT simulation; both  $\tilde{D}_T$  and  $1/\tilde{\tau}_{rot}$  were then computed at this P-T point from the NVT simulation. We then followed a similar procedure as above, running simulations at varying pressures at the temperature points 275, 300, 325 and 350 K. The temperature/pressure points making up both isomorphs along with their calculated values of  $\tilde{D}_T$  and  $1/\tilde{\tau}_{rot}$  are listed in Table 5.1.

Isomorph and density scaling behavior present in reduced units of time, length and energy as described by Gnan et al.<sup>379</sup> The reduced units of length and time used here are  $\rho^{-1/3}$  and  $\sqrt{m/k_BT}/\rho^{1/3}$  respectively. Here we take the mass 'm' as the molecular mass of PC, but note that most scaling arguments are derived for atomic fluids. We compute scaled diffusion coefficients as  $\tilde{D} = \left(\rho^{1/3}\sqrt{m/k_BT}\right) D$ . We compute scaled time constants  $\tau$  using the reduced time unit. Dynamic scattering functions are given in real length/time units when comparing to experiment, and reduced length/time units otherwise, and this convention will be explicitly stated in context of discussion.

We note that experimental characterization of density scaling has primarily utilized *real* rather than *reduced* units. Formally, the latter should always be used, but in practice use of real units is largely inconsequential when relaxation times are plotted over many orders of magnitude. For example, Pawlus et al.<sup>407</sup> demonstrated density scaling for propylene carbonate by characterizing relaxation times over ten orders of magnitude, at different phase points. While relaxation times were plotted in *real* units, the conversion to *reduced* units differs by order unity for the investigated

phase space, and is thus not observable over plots that span many decades in time. We show a plot comparing reduced vs. real units for  $D_T$  vs.  $\tau$  in Fig. S2 for the PC pseudoisomorph points.

## 5.3.3 Identifying Excitations

In computer simulations, excitation populations are generally quantified by searching for particles that undergo sufficiently large spatial displacements that persist in time, usually on the order of a picosecond.<sup>372,414</sup> In such an analysis, one must average out fast vibrations that correspond to motion within the same inherent structure, and one must justify the spatial and temporal cutoff values applied to discriminate hops.<sup>372,414</sup> Instead, we quantify excitation populations through analysis of the self intermediate scattering function  $F(q, t)^{402,403}$  by modeling it as a double-Gaussian in q:

$$F(q,t) = [1 - \Phi(t)]e^{-(q\pi\sigma_{IS})^2} + \Phi(t)e^{-(q\pi\sigma_{IB})^2}$$
(5.3)

This approach has the advantage of being applicable to both experimental and computational characterization, and requires no *a priori* assumptions about lengthscales or timescales of the excitations, since the lengthscale and timescale signatures of excitations and their associated hops come directly from fitting to the model. Here,  $\Phi(t)$  signifies the fraction of particles (H atoms or molecular centers of mass) that have undergone an excursion of lengthscale  $\sigma_{IB}$  up to time t.  $\Phi(t \approx 1 ps)$  is the instantaneous population of particles at an IB barrier, and thus the instantaneous excitation population. Also at  $t \approx 1ps$ ,  $\sigma_{IS}$  and  $\sigma_{IB}$  are the characteristic lengthscales for inherent state elastic deformation and for hopping respectively. In (Equation 5.3) we have neglected a diffusive term, so, at later times,  $\sigma_{IS}$  and  $\sigma_{IB}$  begin to grow slowly as they acquire diffusive character. This form provides excellent fits for experimental and simulation data, as is shown in Figure 5.3, Figure 5.4, and Figure 5.5. The time dependence of fit parameters  $\sigma_{IS}$ ,  $\sigma_{IB}$  and  $\Phi$  is shown.

#### 5.4 Results

We first briefly discuss the accuracy of the PC force field utilized for the MD simulations, as this dictates the accuracy of all simulation results presented in the manuscript. Compared to experiment, errors in the predicted density and enthalpy of vaporization are 3% and 9% respectively, at 300 K, 1 atm; such accuracy is comparable to previous property predictions of organic liquids with SAPT-FF force field.<sup>415</sup> Additionally, computed diffusion coefficients and rotational correlation times are within 10% of the experimental values over the temperature range of 250-350 K at 1.0 atm (Fig. S8). Because there are no empirical parameters in the model, similar accuracy is expected throughout the studied temperature-pressure phase diagram.

### 5.4.1 Center of Mass Motion

Utilizing a brute-force screening approach (Section 5.3.2), we were able to locate two curves on the temperature-pressure phase diagram that display largely invariant structural and dynamical properties (computed in reduced units). The curves that we located are shown in Figure 5.1, in which we have mapped the pressure to the equilibrium density. Also shown in Figure 5.1, for perspective, is the glass transition curve.<sup>416</sup> One pseudo-isomorphic curve starts at 300 K, 1 atm and extends to higher temperature/pressure, and the other starts at 250 K, 1 atm; by intention, both curves were chosen to be far away from the glass transition.

Our isomorphic curves display identical density scaling to that previously determined experimentally for propylene carbonate. Pawlus et al.<sup>407</sup> found that relaxation times and dc conductivity in propylene carbonate solutions measured across the phase diagram, collapsed to a common curve when plotted against  $V^{-3.7}/T$ .<sup>407</sup> This scaling exponent of  $\gamma_{TS} = 3.7$  agrees with our computed isomorph curves, as shown in Figure 5.1. All points on a given isomorph curve were fit to the form  $\rho^{\gamma_{TS}}/T = constant$ . A best fit of our data gave a value of  $\gamma_{TS} = 3.75 \pm 0.09$ , consistent with the findings of Pawlus.

In Table 5.1, we give calculated diffusion coefficients  $(\tilde{D}_T)$  and inverse rotational relaxation



Figure 5.1: Demonstration of  $\rho^{\gamma_{TS}}/T = constant$  for the two isomorph curves, with  $\gamma_{TS} = 3.7$  as determined from Pawlus et al.<sup>407</sup> plotted as dashed lines. The solid line denotes the liquid/glass boundary computed from state points taken from Bonetti and Dubois [416].

Isomorph Point	$\tilde{D}_T (10^{-3}))$	$1/\tilde{\tau}_{rot}(10^{-1})$
300 K/1 atm	4.1	0.95
325 K/1000 atm	4.1	0.98
350 K/3200 atm	3.9	1.02
400 K/4600 atm	3.9	1.03
250 K/1 atm	1.2	0.31
275 K/1300 atm	1.2	0.32
300 K/2400 atm	1.1	0.32
325 K/4000 atm	1.1	0.32
350 K/5600 atm	1.1	0.33

Table 5.1:  $\tilde{D}_T$  and  $1/\tilde{\tau}_{rot}$  for isomorph 1. Uncertainties in both  $\tilde{D}_T$  and  $1/\tilde{\tau}_{rot}$  are approximately 8% due to uncertainties in predicted density at a given T/P state point.

times  $(1/\tilde{\tau}_{rot})$ , both in reduced units, for the two pseudo-isomorphic curves. These computed dynamical properties are invariant along each pseudoisomorph to within statistical uncertainty. In addition to invariance to these dynamic measures, we expect structural properties of the fluid to be invariant along pesudoisomorphs. To analyze the liquid structure, we compute radial distribution functions (RDFs) between propylene carbonate molecules in the liquid as a function of center-ofmass positions. The RDFs are shown in Figure 5.2a) (top panel) and in Supplementary Fig. S3 as computed for all phase points on each pseudoisomorph curve. The RDFs for all points on each pseudoisomorph collapse to a single curve, when the lengthscale is plotted in reduced units. In the inset, we highlight the similarity across the phase points with a high resolution scale of the first RDF peak. As a comparison, we show RDFs computed from 1 atm isobar phase points spanning temperature 300K-350K in Figure 5.2b). It is clear from the insets in Figure 5.2a)-b) that there is structural deviation (albeit small) between the isobar points, in contrast with the identical structure among pseudoisomorph phase points.

In the bottom panel of Figure 5.2 we show the *self* part of the intermediate scattering function F(q, t), defined as

$$F_s(q,t) = \frac{1}{N} \langle \sum_{i=1}^{N} exp\left(i\mathbf{q} \cdot \left(\mathbf{r}_i(t) - \mathbf{r}_i(0)\right)\right) \rangle$$
(5.4)

where **q** is the momentum transfer wavevector, and  $\langle ... \rangle$  denotes an ensemble average. In Fig-



Figure 5.2: Center-of-mass structural and dynamic metrics at pseudo-isomorphic phase points compared to isobar phase points. Top: RDFs computed between propylene carbonate center of mass positions, and plotted as a function of reduced distance for a) all points along the 300 K pseudoisomorph; and b) 1 atm isobar points at several temperatures. Bottom: Comparison of F(q,t) at selected wavevectors, computed based on center of mass positions for c) 300 K, 1 atm pseudoisomorph phase points; and d) 1 atm isobar points at several temperatures.

ure 5.2c) we show  $F_s(q, t)$  computed for molecular center-of-mass (COM) positions, and in Fig. S3 we show  $F_s(q, t)$  calculated for hydrogen atom (H-atom) positions (i.e. the sum in Eq Equation 5.4 runs over COM or H-atom indices respectively). Here and for the remainder of the manuscript, all reference to the scattering function concerns the *self* part (and not the *distinct* contribution), and the quantity in Equation 5.4 will be referred to simply as F(q, t) ("s" subscript dropped). The qvalues shown describe dynamics on lengthscales ranging from 70% to 2.5 times the average intermolecular spacing ( $q_{max,COM} = 5.76$ ). At each of these lengthscales, COM dynamics among all the phase points investigated are identical to within uncertainty. These combined dynamic and structural analyses indicate that our identified phase point curves (Figure 5.1) indeed behave as pseudoisomorphs. For comparison, in Figure 5.2d) we show  $F_s(q, t)$  computed from 1 atm isobar phase points spanning temperature 300K-350K (computed for COM positions) As expected, there is significant temperature dependence of dynamic relaxation for the isobar points; in contrast, the pseudoisomorph phase points exhibit essentially identical COM dynamics as indicated by  $F_s(q, t)$ in Figure 5.2c).

Based on our simulation analysis, we find that propylene carbonate exhibits density scaling such that structural and dynamic properties are functions of  $\rho^{3.7}/T$ , in agreement with previous experimental findings.<sup>407</sup>

## 5.4.2 Excitations, Inherent State Motion, and Interbasin Crossing

Interbasin barrier crossing events in atomic model systems are local cooperative particle reorganizations<sup>370,417</sup> and represent fundamental relaxation events. We have recently shown that motion associated with IB crossing and IS basin exploration can be quantified from the scattering function.<sup>403</sup> Specifically, we showed that experimental  $S(q, \omega)$  (the time Fourier transform of F(q, t)) from quasi-elastic neutron scattering (QENS) on propylene carbonate is optimally fitted with three Lorenzians, having q-dependent amplitudes consistent with a diffusive process and two localized processes. The latter two processes had time-dependent lengthscale characteristics corresponding to localized processes, consistent with inherent state (IS) vibrational motion ( $\sigma_v$ ) and interbasin



Figure 5.3: Top) Scattering function  $F_s(q,t)$  computed from MD simulation at 300K, 1atm based on H-atom positions. Bottom) Scattering function  $F_s(q,t)$  measured by incoherent QENS from PC.<sup>403</sup>

(IB) hopping motion ( $\sigma_h$ ) on a PEL. In order to characterize IB and IS motions in this study we focus on F(q, t) calculated from Equation 5.4 For comparison with QENS data from PC that were previously acquired by one of us,<sup>403</sup> we sum over H atoms because incoherent scattering from hydrogen atoms dominates the QENS signal in simple organics.

In Figure 5.3, we compare the scattering function computed from our simulations to that previously measured by QENS.<sup>403</sup> There is generally good agreement between the simulated and experimental scattering functions, attesting to the accuracy of the employed force field. Since we do not include coherent scattering in Equation 5.4, we expect deviations of a few percent at low q values which are not visible in Figure 5.3. However, there are prominent discrepancies between simulated and experimental F(q, t) at  $q > q_{max} = 1.1$  Å<sup>-2</sup> and at short times and high q values.

These discrepancies are not surprising, considering the approximations of the intramolecular interaction potential (harmonic bonds, angles, etc.) used in the simulations (Supporting Information). For example, the largest plotted wavevector  $q^2 = 5$  Å<sup>-2</sup> corresponds to less than 3Å real space distance, which is smaller than the molecular diameter of PC. Thus F(q, t) at this wavevector is expected to be quite sensitive to the torsional potentials and even methyl group rotation, the latter of which is not expected to be accurately modeled by the force field. In contrast, the scattering at  $q < q_{max}$  (distances longer than intermolecular separation) is dictated largely by intermolecular forces, and the intermolecular potential utilized in the MD simulations is expected to be quite accurate, as confirmed by the good agreement with experiment for F(q, t) in this regime.

We obtain parameters for excitations, IB and IS motion by fitting F(q, t) with (Equation 5.3). A straight line in a ln(F(q, t)) vs  $q^2$  plot signifies a single relaxation process that is Gaussian in q. The bi-linear form of ln(F(q, t) vs  $q^2$  illustrates the two-step relaxation process. From a data-fitting perspective,  $\Phi(t)$  serves as a time-dependent non-Gaussian parameter. From a particle dynamics perspective,  $\Phi(t)$  represents the fraction of particles that have "hopped" (undergone a non-reversing IB crossing) up to time t. Bearing in mind that hops constitute significantly larger excursions than IS cage distortions, it is clear that until a molecule executes a hop, the q-dependence of its scattering signature will be characteristic of only small lengthscale  $\sigma_{IS}$ . Once a hop occurs, that signature will change to larger lengthscale for all subsequent times.

Figure 5.4 shows F(q, t) calculated from hydrogen atoms, fits to (Equation 5.3) and fit parameters. As shown in the top panel, the model fits quite well, fully parameterizing F(q, t) for this q and t range. The middle and bottom panels in Figure 5.4 show excellent agreement in the time-dependence of model fits for each of the 300 K, 1 atm pseudoisomorphs. In the middle panel,  $\Phi(t)$  increases monotonically as expected,<sup>403</sup> with (1- $\Phi$ ) exhibiting an initial rapid drop during the period of quasi-ballistic motion when atoms are exploring the cage formed by their neighbors. As we have explained elsewhere,<sup>406</sup> just after this initial drop in (1- $\Phi$ ),  $\Phi(t \approx 1 ps)$  represents the instantaneous fraction of H atoms involved in large-lengthscale motion characteristic of structural



Figure 5.4: F(q,t) and fits. (Top), F(q,t) at 300 K, 1 atm based on H-atom positions, plotted against  $q^2$  at evenly spaced times between 0.5 ps and 6 ps. Black points are calculated from simulation, and dashed red lines are fits to (Equation 5.3). (Middle and Bottom) Time-dependence of model fits shown for each pseudoisomorph of 300 K, 1 atm.
excitations. After this initial period of cage exploration,  $(1-\Phi)$  decays almost exponentially as excitations visit new regions of the sample and facilitate hops of an increasingly large fraction of particles. In the bottom panel, the characteristic lengthscales  $\sigma_{IS}$  and  $\sigma_{IB}$  initially show signatures of localized motion during the period of cage exploration, reaching quasi-asymptotic values of  $\approx 0.1$  and  $\approx 0.2$  respectively, consistent with IS vibration and IB crossing.<sup>417</sup> The values of these fit parameters continue to grow, taking on a diffusive character at longer times in this model because, for simplicity's sake, we do not explicitly include diffusion through a separate parameter.

The high degree of overlap among the fit parameters at each of the pseudo-isomorphic phase points signifies excellent correspondence between F(q,t) for H atoms at each phase point, and demonstrates that excitation populations as well as dynamic timescales and lengthscales associated with IS vibrations and IB crossing obey density scaling. We have previously shown that the IS and IB modes are related in a simple way to  $\alpha$  relaxation and translational diffusion,<sup>402,403,405</sup> so the fact that they exhibit scaling is consistent with Hansen's conclusion (based on longer timescale data) that motion coupled to  $\alpha$  relaxation should exhibit density scaling.<sup>382</sup>

#### 5.5 Discussion

# 5.5.1 IS and IB motion Involve Intramolecular Modes

While the dynamics encoded by the diffusion coefficient, rotational relaxation time (Table 5.1) and F(q,t) for COM integrate over intramolecular motion, the atomic scattering function formally encompasses dynamical processes over *all* temporal and spatial scales relevant to IB and IS motion. The fast timescales and short lengthscales of IS and IB motion make them candidates for being associated with intramolecular modes. Comparing the top panels of Figure 5.4 and Figure 5.5 provides evidence that IS and IB modes in PC involve significant intramolecular motion. The top panel of Figure 5.5 shows  $F_{COM}(q,t)$  calculated from molecular centers of mass for the 300 K, 1 atm phase point. Comparing with  $F_H(q,t)$  from H atoms at the same phase point in Figure 5.4, it is quite clear that the COM motion is of lower amplitude overall since, for all q and t,  $F_H(q,t) \leq F_{COM}(q,t)$ . A similar relationship was found between COM and H-atom F(q,t) for glycerol.<sup>405</sup> Molecular rotation will cause some differential decay of  $F_H(q, t)$  compared to its COM counterpart, but a detailed comparison of the two scattering functions suggests another significant contribution. The two characteristic lengthscales for motion are similar for H-atom and COM fits, suggesting that the molecule moves in a *quasi*-rigid fashion, but  $\Phi_H \ge \Phi_{COM}$  at all times. This pronounced muting of the non-Gaussian nature of the COM F(q, t) with respect to the H-atom F(q, t) directly demonstrates that IS and IB motion executed by PC are largely intramolecular in nature (although they could in principle involve parts of more than one molecule). Hopping of individual H-atoms - parts of the molecule - precedes similar lengthscale motion of the entire molecule reflected in the COM.

#### 5.5.2 Scaling-Relevant Intramolecular and Collective Modes

We turn now to the question of which intramolecular modes are likely to contribute to IB and IS motions, and why, if they contribute significantly, do they not spoil the density scaling demonstrated in Figure 5.4. The origin of the effective IPL appears to be collective in nature since strong energy-virial correlations are observed only when averaging over the ensemble in rigid particle models.<sup>392</sup> Thus, we expect that if any intramolecular modes do contribute to density scaling, the contributing modes would most likely low-frequency ones that involve significant portions of the molecule.

In the top left panel of Figure 5.6 we overlay Raman spectra of PC from two sources. The red solid trace shows the lowest frequency intramolecular vibrational modes, obtained through spontaneous Raman scattering. These Raman modes at 191, 318, and 451  $cm^{-1}$  are all molecularly extended ring deformation modes. The black trace is the Fourier transform of the  $\beta_{fast}$  component from time-domain optical polarizability measured by Optical Kerr effect spectroscopy.<sup>418</sup> The  $\beta_{fast}$  OKE signal contains a collective and librational part, which we demonstrated to have the same temperature dependent frequencies and relative amplitudes as IB barrier crossing and IS motion measured by neutron scattering.<sup>418</sup> The blue dashed line represents the spectrum of IB motion and the red dashed line represents the IS spectrum. The black line is the sum of the IS and IB



Figure 5.5: (Top) F(q,t) computed from MD simulation at 300 K, 1 atm based on molecular centers-of-mass. Black points are calculated from simulation, and dashed red lines are fits to (Equation 5.3). (Middle and Bottom) Time-dependence of model fits shown for each pseudoisomorph of 300 K, 1 atm.

spectrum. In the PEL picture, both the IS and IB modes are coupled to relaxation in the sense that the IS modes assist in basin exploration but the molecular reorganizations that make up relaxation events proceed directly along IB mode coordinates. The intramolecular vibrational modes with frequencies below 500  $cm^{-1}$  overlap spectrally with, and thus *may* couple to the IS and IB modes.

In the right top and bottom panels of Figure 5.6 we overlay power spectra of velocity autocorrelation functions for ring atoms and molecular centers of mass respectively at the three simulated pressures indicated. The right top panel shows that the pressure dependence of the intermolecular mode characteristic frequencies drops with increasing frequency, indicating that higher frequency peaks contribute less to density scaling. Consistent with this, we found that eliminating the much higher frequency bond stretching modes by constraining all bond lengths led to no change in density scaling (Supplementary Fig. S4). Also consistent with these findings, Raman modes found below 800  $cm^{-1}$  in organic molecules with more than one backbone atom typically involve molecular deformations typically have larger Grüneisen parameters than do higher-frequency modes involving bond stretches and H-X bending.<sup>419–421</sup>

The broad  $\beta_{fast}$  features below  $\approx 150 \ cm^{-1}$  in the right top panel are due to rotational and translational motion, but arise only from translational motion in the right bottom panel. In this spectral region, the IS and IB responses are not well-separated, but their characteristic frequencies appear to have relatively strong pressure dependencies. We quantify the characteristic frequencies of the IB and IS motion by fitting the COM power spectra to functions commonly used to characterize OKE data<sup>418</sup> as described in the Supplementary Information. We find acceptable fits using functions corresponding to IB and IS motion and an additional Gaussian process at slightly higher frequency than the IS mode. In the left bottom panel of Figure 5.6 we show the model fit to the 300 K, 1 atm COM power spectrum. Analogous fits for 1900 atm and 4600 atm are shown in Supplementary Figure S5 and are of similar quality.

In the inset to the right bottom panel of Figure 5.6 we show Grüneisen parameters for each of the modes found in the power spectra, calculated as:



Figure 5.6: Top Left) Polarizability and Raman response in PC. Black line:  $\beta_{fast}$  polarizability spectra from OKE measurements.<sup>418</sup> Red dash line: OKE spectrum due to IS motion (known as librations in OKE literature). Blue dashed line: OKE spectrum due to IB crossing (known as collective motion in OKE literature). Red solid line: Low frequency component of a PC spontaneous Raman spectrum. (Right Top and Bottom) Power spectra of velocity autocorrelation functions for all atoms (Right Top) and centers of mass (Right Bottom) at 300 K and the pressures indicated in the middle panel. (Inset to Right Bottom panel) Grüneisen parameters calculated for each mode found in the power spectra. Left Bottom) Model fit to the 300 K 1 atm centers-of-mass power spectrum.

$$\gamma_{G,i} = -\partial ln(\omega_i) / \partial ln(V) \tag{5.5}$$

where *i* represents the  $i^{\underline{th}}$  vibrational mode. Among the modes characterized, we see a continuous progression of increasing  $\gamma_G$  with decreasing frequency. We emphasize that, although the IB and IS modes are distinct from intramolecular eigenmodes, the differences in  $F_H(q,t)$  and  $F_{COM}(q,t)$  suggest they involve intramolecular motion and thus likely couple to these modes.

Finally, we note that  $\gamma_{IB} \approx \gamma_{TS}$ . This suggests that it may be generally possible to find  $\gamma_{TS}$  in a simulated liquid by simply finding  $\gamma_{IB}$ . It also suggests that the relationship between  $\gamma_{TS}$  and the Grüneisen parameter may be more direct than previously thought. Associations between  $\gamma_{TS}$  and  $\gamma_G$  have been made theoretically through the latter's connection to configurational entropy,<sup>397</sup> or through the kinetic energy derivative of the pressure in the harmonic limit,<sup>392</sup> but the striking result in the inset to Figure 5.6 suggests that the Grüneisen parameter may influence  $\gamma_{TS}$  directly through  $\gamma_{IB}$ .

The Grüneisen parameter formally characterizes the volume dependence of normal mode frequencies in a crystal, but, of course, there is no relaxation in an ideal crystal. Similarly, Debye-Waller (DW) factors strictly describe only harmonic vibrations but are used to describe liquid relaxation through the mean square displacement  $\langle u^2 \rangle$ . An extensive literature, going back three decades<sup>402,422-425</sup> has explored the relationship between  $\alpha$  relaxation and  $\langle u^2 \rangle$  measured at  $\approx 1$  ps, the time of the  $\beta_{fast}$  process. In these approaches, it is assumed either that the amplitude of harmonic motion describes a localization volume for particles trapped in a cage of their neighbors or is proportional to the probability of a local relaxation event. Recently, we have made a more direct connection between motions that constitute  $\langle u^2 \rangle$  in amorphous systems and relaxation processes by demonstrating that  $\langle u^2 \rangle$  contains signatures of both IS basin exploration and elemental relaxation events - the IB barrier crossing. Thus, when properly parsed, dynamic signatures of the elemental relaxation process are obtained directly from the  $\beta_{fast}$  relaxation of  $\langle u^2 \rangle$  or F(q, t).<sup>402,403</sup>

In analogy with the idea that  $\langle u^2 \rangle$  is directly related to relaxation through its reporting on IB barrier crossings, we propose that a causal connection exists between  $\gamma_G$  and  $\gamma_{TS}$  through  $\gamma_{IB}$ as follows: The normal modes of the liquid that couple to the IB process are by definition found at the Frenkel frequency,<sup>426</sup>  $\nu_F = 1/\tau$ . At this frequency, the liquid supports longitudinal modes, but transverse modes are overdamped since the liquid can reorganize on the timescale of these local shear deformations, and this motion directly couples to relaxation. In an amorphous system, longitudinal and transverse modes will also be coupled since there is no well-defined extended lattice. Since motion along these mode coordinates is coupled to elemental relaxation (IB crossing) events, the resonant frequency of the mode can be thought of as a relaxation attempt frequency. Thus, for a fixed barrier height, small changes in the frequency of these modes would lead to directly proportional changes in the rate of elemental relaxation events. We have previously demonstrated that, for  $T > T_B$  ( $T_B$  being where the  $\alpha$  and Johari-Goldstein  $\beta$  relaxations bifurcate), the timescales for IB crossing events and  $\alpha$  relaxation have the same temperature dependence.<sup>403</sup> Thus, for this temperature range, the Grüneisen parameter of the Frenkel modes ( $\gamma_{IB}$ ), which describes the pressure dependence of their frequency, also describes the pressure dependence of the  $\alpha$  relaxation time  $(\gamma_{TS})$  through the pressure dependence of the elemental relaxation time. As we see in Figure 5.6,  $\gamma_{IB} \approx \gamma_G.$ 

Above we showed that high-frequency modes associated strictly with bond stretching have no impact on density scaling. Beyond arguing that the pressure dependence of lower frequency intramolecular modes and their spectral overlap with IB modes suggest that the former may contribute to IB barrier crossings and to density scaling, we can only speculate as to how important they may be. As we have shown above, hops associated with IB transitions appear to occur at the atomic rather than molecular level. Thus, intramolecular modes that conserve molecular center of mass may well contribute to hops of constituent atoms. However, to quantify this would require a series of simulations in which we systematically constrain these various intramolecular modes. While standard MD simulations are done using Cartesian equations of motion, an alternative approach for exploring the role of intramolecular modes on liquid dynamics would be to perform molecular dynamics in general/internal coordinates (or hybrid) to enable explicit interrogation of specific internal modes.<sup>427–429</sup> This direction will be explored in future work.

#### 5.6 Conclusions

While evidence of density scaling in  $\beta_{fast}$  at ~1 ps timescales has been demonstrated previously, the relationship involved a quadratic correction term.<sup>383</sup> Given that  $\beta_{fast}$  relaxation has contributions from distinct modes of motion, some of them potentially intramolecular, and given that intramolecular vibrational modes can destroy the virial-energy correlation that is often the formal basis for density scaling arguments, it was unclear whether all aspects of  $\beta_{fast}$  motion would couple, or whether the necessity of a quadratic correction term signaled a partial decoupling. In this work we have shown that the IB barrier crossing aspect of the  $\beta_{fast}$ , a fundamental structural excitation mode, seems to obey the same density scaling, whereas the higher frequency IS mode is associated with a nearly two-fold smaller scaling factor. Because structural excitations correspond to fundamental hopping (intrabasin barrier crossing) processes on the liquid potential energy landscale (PEL), our results suggest a novel microscopic perspective for density scaling. Interpretation of density scaling based on excitations may prove a promising general perspective applicable to molecular liquids for which virial-energy correlation or single inverse power law arguments may not hold due to the intra-molecular modes or due to multiple effective potentials.

For propylene carbonate, we have demonstrated that the characteristic lengthscales  $\sigma_{IS}$  and  $\sigma_{IB}$  have the same values in F(q, t) computed for COM and to H-atom motion, but that  $\Phi(t)$  calculated for COM always trails that of H atoms. We interpret this as demonstrating that excitations must thus be interpreted as essentially *atomic* rather than *molecular* displacement events. This important conclusion has implications for generalizing excitation theories of atomic and model fluids<sup>372,414</sup> to real molecular liquids. The fact that the Grüneisen parameter for collective IB modes is approximately equal to  $\gamma_{TS}$ , suggests that finding these modes in molecular liquids will be the key to connecting excitation approaches in these liquids to those of atomic model systems.

# **CHAPTER 6**

# EQUATION OF STATE PREDICTIONS FOR SCF<sub>3</sub> AND CAZRF<sub>6</sub> WITH NEURAL NETWORK-DRIVEN MOLECULAR DYNAMICS

This work is reproduced from [430].

# 6.1 Abstract

The crystalline materials ScF<sub>3</sub> and CaZrF<sub>6</sub> are known to exhibit negative thermal expansion (NTE) over a broad temperature range. A variety of theoretical methods including ab initio calculations have been utilized to better understand the NTE behavior of these materials. However, explicit application of *ab initio* molecular dynamics (AIMD) approaches is challenging due to the high computational cost. In this work, we develop neural network (NN) potentials for both ScF<sub>3</sub> and CaZrF<sub>6</sub> trained to extensive density functional theory (DFT) data, that enable direct molecular dynamics (MD) prediction of equation of state behavior over a broad temperature/pressure range. The NN-driven MD simulations enable predictions with full incorporation of phonon anharmonicity, which is a key contribution to NTE. From these simulations, we compute coefficients of thermal expansion, atomic displacement parameters, bulk moduli, and other equation of state behavior. The computational efficiency of the neural network potential (relative to AIMD) enables path integral MD simulations that incorporate nuclear quantum effects (NQE) which may be important for high-frequency phonon modes, and we characterize the contribution of NQE to equation of state behavior. Overall, our equation of state predictions are in semiquantitative agreement with experiment, with the quality of predictions depending on system and property of interest. We show that NQE have a significant effect on NTE at low temperature, but are of diminishing importance at increasing temperature. The experimentally observed pressure-induced softening effect in ScF<sub>3</sub> is elusive and not captured in our simulations of the cubic phase, in contrast to a previous computational study.

# 6.2 Introduction

Negative thermal expansion (NTE) materials are of significant interest due to both their unusual physics and their potential applications. Numerous applications require matching coefficients of thermal expansion between two or more materials, and the discovery of new NTE materials and elucidation of their behavior provides greater tunability in this design space. Examples of possible uses for NTE materials include fuel cells, mirrors in space telescopes, optics, thermoelectric materials and more.<sup>431–433</sup> The discovery of new NTE materials as well as better fundamental understanding of documented NTE behavior are both important goals. Regarding the latter, predicting/rationalizing NTE behavior from first-principles is often challenging and may serve as a stringent test of solid-state theories, particularly those focused on incorporating phonon anharmonicity.

A variety of mechanisms can cause NTE, and thus it is observed in several classes of compounds, such as ferroelectric materials,<sup>434</sup> MOFs,<sup>435</sup> Prussian Blue analogues<sup>436</sup> and other openframework materials.<sup>431,437</sup> The work of Mary *et al.* [438] on ZrW<sub>2</sub>O<sub>8</sub> documenting NTE over a wide temperature range, launched interest into framework-type materials.<sup>439</sup> Due to their simpler structure in comparison to ZrW<sub>2</sub>O<sub>8</sub>, ReO<sub>3</sub>-type materials are of interest for fundamental understanding of NTE behavior in open framework structures.<sup>440–443</sup> While ReO<sub>3</sub> itself only shows small NTE at low temperatures, Greve *et al.* [443] found that ScF<sub>3</sub>, which possesses a ReO<sub>3</sub>-type structure, displays strong NTE from 10 K up to ~ 1100 K with a coefficient of thermal expansion (CTE) of  $\alpha_L \sim -14$  ppm/K at 100 K. In most metal trifluorides, the rhombohedral ( $R\bar{3}c$ ) phase is more stable than the cubic ( $Pm\bar{3}m$ ) phase at low temperatures and elevated pressures, which leads to positive thermal expansion; in contrast, for ScF<sub>3</sub>, the cubic phase is more stable than the rhombohedral at low temperature and ambient pressure, which leads to NTE.<sup>443–445</sup> The discovery of strong NTE in ScF<sub>3</sub> has led to further research into its other properties, such as pressure-induced softening,<sup>446</sup> which causes a material to become more compressible with increasing pressure, and methods for controlling its thermal expansion. Yang *et al.* [447] observed a reduction of NTE upon formation of ScF<sub>3</sub> crystals with grain sizes of 80 nm. Similarly, Hu *et al.* [448] synthesized ~ 6 nm ScF<sub>3</sub> crystals for which only positive thermal expansion was observed. Further strategies for tuning thermal expansion with ScF<sub>3</sub> include the formation of solid solutions through isovalent cation substitution,<sup>449–453</sup> redox intercalation of cations into the ScF<sub>3</sub> A-sites<sup>454</sup> and the introduction of excess fluoride through aliovalent cation substitution.<sup>455,456</sup> CaZrF<sub>6</sub> is another open-framework material with ReO<sub>3</sub>-type structure that displays NTE, and exhibits a more negative CTE ( $\alpha_L \sim -18$  ppm/K at 100 K) than ScF<sub>3</sub>.<sup>457</sup> The NTE behavior of CaZrF<sub>6</sub> can be modulated in various ways; for example, incorporation of helium under high-pressure gas has been shown to create a defect perovskite (He<sub>2-x</sub> $\Box_x$ )(CaZrF<sub>6</sub>)<sup>458</sup> and a stoichiometric hybrid perovskite [He<sub>2</sub>][CaZrF<sub>6</sub>],<sup>459</sup> with different equations of state.

Theoretical methods and computational modeling have aided physical understanding of NTE behavior, suggesting mechanisms dictating the CTE of various materials.<sup>442,460–464</sup> The rigid unit vibrational mode (RUM) model is commonly used for rationalizing NTE in framework-type materials. The RUM model assumes that the framework  $MX_n$  polyhedra remain rigid with increasing temperature, with NTE caused by thermally-modulated tilting motion of these rigid polyhedra which contracts the crystal volume.<sup>442,461,465</sup> However, the validity of the assumptions behind models such as these have been called into question.<sup>463,466</sup> Ab initio approaches usually based on density functional theory (DFT) have also been used to model NTE materials. Several works have investigated NTE in ScF<sub>3</sub> with DFT-based approaches, utilizing either vibrational free energy predictions or ab initio molecular dynamics (AIMD) simulations. Oba et al. [467] evaluated quasiharmonic approximation  $(QHA)^{468}$  free energy expressions for ScF<sub>3</sub>, and found that QHA predicts qualitatively incorrect behavior. This is likely due to the fundamental importance of anharmonicity in NTE systems, requiring extensions beyond harmonic treatments.<sup>469</sup> Methods that go beyond the QHA have been used to study ScF<sub>3</sub>, such as the self-consistent phonon theory (SCP), which includes quartic anharmonicity,<sup>470,471</sup> and the improved self-consistent phonon theory (ISC), which additionally treats cubic anharmonicity.<sup>472</sup> These procedures require parameterization of high-order force constants from first-principles calculations.<sup>473,474</sup> In contrast, AIMD is straightforward and naturally

takes into account anharmonicity. The limitations of AIMD are its computational expense, as well as classical description of phonons/vibrations; path integral approaches can be used to incorporate nuclear quantum effects (NQE), but this imposes further computational expense. AIMD simulations have been conducted for ScF<sub>3</sub>, providing insight into its NTE behavior.<sup>469,475–477</sup> Utilizing AIMD, Lazar *et al.* [475] were able to reproduce the NTE effect over a temperature range of 200K - 800K using a 2x2x2 supercell of the cubic phase. Bocharov *et al.* [476] found that at least a 4x4x4 supercell for ScF<sub>3</sub> is needed to converge AIMD predictions with system size, due to the long wavelength phonons with negative Grüneisen parameters accounting primarily for NTE behavior.<sup>448,476</sup> As CaZrF<sub>6</sub> is a more complicated material, there are fewer *ab initio* studies in the literature.<sup>478,479</sup> Gupta *et al.* [478] used the QHA to study CaZrF<sub>6</sub> and related systems, and again demonstrated the importance of anharmonicity for modeling thermal expansion effects.

While AIMD is a powerful technique, its high computational cost limits its utilization, particularly for large crystal supercells. As discussed, larger supercells are often necessary to avoid finite size effects associated with long-wavelength/low frequency phonons that make important contributions to NTE.<sup>457</sup> Additionally, nuclear quantum effects (NQE) may be important, requiring computationally expensive path integral simulations. As an example, consider the phonon spectra of CaZrF<sub>6</sub> discussed by Hancock *et al.* [457]. The high frequency phonons typically have positive Grüneisen parameters ( $\gamma$ ), whereas low frequency modes exhibit negative  $\gamma$ , the latter contributing to NTE. While a classical treatment is expected to work well for low frequency (e.g.  $\leq 200 - 300$ cm<sup>-1</sup>) modes, quantization will be important for high frequency modes; a classical treatment will incorrectly assign energy equipartition to these high frequency modes and thus tend to underestimate NTE effects at lower temperatures.<sup>467</sup> Combining path integral approaches with AIMD may often be computationally intractable for systems of interest. Alternatives include fitting molecular mechanics force fields to quantum mechanical data.<sup>446</sup> For materials of increasing complexity, developing sufficiently accurate force fields is both a challenging and time consuming task.

Machine learning (ML) methods have been increasingly adopted in the chemistry and materials science community to simulate materials with *ab initio* accuracy at orders of magnitude lower computational cost as compared to AIMD.<sup>63,480,481</sup> As long as a training set of high-quality reference data is available, usually consisting of *ab initio* energies/forces for the system of interest, ML techniques such as neural networks, kernel methods or other tools can be used to construct ML potentials.<sup>482</sup> These ML potentials can then be used to run molecular dynamics (MD) or Monte Carlo simulations to predict physical properties of interest. In particular, ML models have been used to study a large variety of materials, such as silicon,<sup>59,483</sup> various metal systems,<sup>484-486</sup> zeolites,<sup>487</sup> MOFs<sup>488</sup> and more.<sup>480,489</sup> ML techniques have also been used to screen materials such as MOFs, zeolites and perovskites for selected applications.<sup>490-492</sup> In this work, we develop similar ML potentials for the NTE materials ScF<sub>3</sub> and CaZrF<sub>6</sub> to predict equations of state from direct MD simulations. These ML potentials allow for efficient MD simulations that explicitly incorporate NQE.

In this work, we train NN potentials to a training set of DFT energy, force, and stress tensor calculations for ScF<sub>3</sub> and CaZrF<sub>6</sub> to enable direct MD prediction of the equations of state. We compare predictions for CTE, atomic displacement parameters, the bulk modulus, and additional equation of state data against both experiment and prior theoretical predictions. Utilizing PIMD simulations, we additionally quantify the contribution of NQE to the equations of state of these materials over a wide temperature range. Overall, the predictions from our NN-driven, MD simulations are generally in semiquantitative agreement with experimental equation of state data. Our predictions underestimate the extent of negative thermal expansion, even when NQE are fully incorporated, and we speculate that this is likely due to deficiencies in the underlying density functional (training data). We find that NQE modulate the CTE by 30-60% at low temperature (100 K) for both materials, which is qualitatively consistent with findings from previous studies.<sup>467</sup> A particularly apparent discrepancy of our simulations is the missing pressure-induced softening effect for cubic ScF<sub>3</sub> which is observed experimentally.<sup>446</sup> Overall our study demonstrates the utility and expected accuracy of ML-driven MD simulations for equation of state predictions for NTE materials.

### 6.3 Methods

# 6.3.1 Training Set Generation and Model Training

We first describe our procedure for constructing a training dataset of DFT energies, forces and stress tensors for ScF<sub>3</sub> and CaZrF<sub>6</sub>. The training sets for both crystals were generated from *ab initio* geometry optimizations and AIMD simulations using the Quantum ESPRESSO package.<sup>18</sup> We used the Atomic Simulation Environment (ASE) package<sup>161</sup> to generate a 4x4x4 supercell for ScF<sub>3</sub> and a 2x2x2 supercell for CaZrF<sub>6</sub> from the cubic (primitive) unit cells obtained from the Materials Project.<sup>493</sup> The PBE density functional and PAW pseudopotentials were used for all calculations,<sup>139,494</sup> with  $\Gamma$ -point sampling of the band structure in all cases. Suitable values for the plane wave cutoffs (kinetic energy/density) were determined from convergence tests of the stress tensor (which is typically harder to converge than energy/forces). These convergence tests are shown in Figures S1 and S2, and the final cutoff values employed were 180 Ry/ 1152 Ry for the kinetic energy/density for ScF<sub>3</sub> and 200 Ry/1200 Ry for the kinetic energy/density for CaZrF<sub>6</sub>.

AIMD simulations were conducted for both materials over a range of temperatures and pressures to generate training data. For ScF<sub>3</sub>, we ran numerous ~1 ps NPT simulations of the 4x4x4 supercell over a temperature range of 300 to 1600K and a pressure range of 0 to 800 MPa.<sup>495</sup> This resulted in a training set of energies, forces, and stress tensors for approximately 8500 ScF<sub>3</sub> cubicphase structures of different coordinates and lattice parameters. We additionally generated training data for the rhombohedral phase of ScF<sub>3</sub> to explore how the inclusion of this training data altered the NN predictions (*vide infra*). We thus ran additional AIMD simulations for a 2x2x2 supercell of the rhombohedral conventional unit cell structure at 300 K and pressures ranging between 0-1000 MPa (the unit cell was obtained from the Materials Project<sup>493</sup>). This added training data for approximately 1200 ScF<sub>3</sub> rhombohedral structures of different coordinates and lattice parameters. For CaZrF<sub>6</sub>, training data was generated for the cubic phase only. AIMD simulations of 2x2x2 CaZrF<sub>6</sub> supercells were run over a temperature range of 300 - 1400 K and a pressure range of 0 -500 MPa. This resulted in a training set of energies, forces, and stress tensors for approximately 6000 CaZrF<sub>6</sub> cubic structures of different coordinates and lattice parameters.

The DeepMD architecture was utilized for the NN potentials,<sup>68,107</sup> and was trained to the energies, forces, and stress tensors comprising the training data for each material. Within the DeepMD architecture, an initial descriptor network converts the local environment of each atom into a set of embeddings that obey translational, rotational and permutational invariance, and a second network utilizes this embedding to predict the energy, forces, and virial of the system.<sup>68</sup> DeepMD's 'se\_-2\_a' descriptor was used for the embedding network, which incorporates both radial and angular information.<sup>496</sup> Separate neural networks were trained for both ScF<sub>3</sub> and CaZrF<sub>6</sub>. The descriptor deep neural network was made up of three hidden layers using 25, 50 and 100 neurons. Neighbors within 8 Å were included in the local environment for each atom. For ScF<sub>3</sub>, the fitting net consisted of three hidden layers with 240 neurons each; for CaZrF<sub>6</sub>, the fitting net was reduced to 40 neurons each in order to balance accuracy and computational cost. A multi-target loss function was used in order to fit all three properties (see Wang et al. [68] and the Supplementary Material). The initial learning rate for both neural networks was set to  $1 \times 10^{-3}$  and ended at  $3.5 \times 10^{-8}$ , with 5000 decay steps. A 80:20 split was used to construct a training and validation set from the total dataset; the test set was assembled from MD simulations using the final neural networks (the settings used for the simulations are described in the next section). Each model was trained for  $10^6$  steps for training and validation. The input files with all hyperparameters used to build the DeepMD models are included in the Supplementary Material.

Neural network potentials may become unstable during simulations, since spurious forces will be predicted if the system drifts far outside of the configuration space included in the training set.<sup>497</sup> We observed that the initially trained neural networks for both materials were unstable for high temperature (>1000 K) simulations after  $\sim 10$  ps. To fix this initial instability, we added additional training data consisting of Quantum ESPRESSO computed energies, forces, and stress tensors for structures taken from the NN simulation snapshots. The DeepMD NN potential was then re-trained to the expanded training set. This procedure was done iteratively until there were no observed instabilities while running MD with the NN potentials. In total, additional training data for approximately 1000 structures was added to the original training set for both materials within this iterative procedure.

The final NN potentials were then tested as follows. MD simulations using these NN potentials were performed over a temperature range of 300 - 1000K and a pressure range of 0 to 300 MPa. A 4x4x4 supercell was used for the ScF<sub>3</sub> simulations and a 2x2x2 supercell was used for the CaZrF<sub>6</sub> simulations. Quantum ESPRESSO was used to compute the DFT energy, forces, and stress tensor for approximately 300 snapshots from these simulations. A comparison of the predicted neural network energy vs. DFT energy on the test set for both materials is shown in Figures S3 of the Supplementary Material. For ScF<sub>3</sub>, the energy mean absolute error (MAE) is 0.042 eV (0.1 meV/atom) and the forces MAE is 0.02 eV/Å; to test the accuracy of the stress tensor, we computed the internal pressure from the stress tensor from both DFT and the neural network. The MAE is  $6.6 \times 10^{-5} \text{ eV/Å}^3$  (approximately 10 MPa). For CaZrF<sub>6</sub>, the energy MAE is slightly worse at 0.18 eV (0.7 meV/atom); however, as can be seen in Figure S3, there is little scatter in the predicted energies. The forces MAE is 0.033 eV/Å and the pressure MAE is  $7.2 \times 10^{-5} \text{ eV/Å}^3$  (approximately 11 MPa).

## 6.3.2 Neural Network-Driven, Molecular Dynamics Simulations

We perform MD simulations with the final NN potentials to predict equations of state for the ScF<sub>3</sub> and CaZrF<sub>6</sub> materials. Both classical MD and path integral PIMD simulations were run in order to examine the impact of NQE; we explicitly denote which predictions correspond to each simulation type when discussing the results. All simulations were conducted using ASE with the DeepMD calculator.<sup>68,161</sup> The DeepMD architecture allows for simulations of supercells of arbitrary size, since the total energy of the system is represented as a sum of atomic energies. The majority of our calculations were performed on 5x5x5 supercells of ScF<sub>3</sub> and 3x3x3 supercells of CaZrF<sub>6</sub> (both in the cubic phase), while larger supercells were simulated in order to estimate finite size effects. For computing the coefficient of thermal expansion ( $\alpha_L$ ), a series of simulations were run in the NPT ensemble using the isotropic Berendsen barostat.<sup>498</sup> A Berendsen thermostat was

used for the temperature coupling. The simulations were run for  $\sim 4 - 5$  ns at a pressure of 0 MPa and temperatures ranging from 100 - 1200 K. A time step of 1.0 femtosecond was used with a barostat coupling constant of 1.0 ps<sup>-1</sup> and a thermostat coupling constant of 1.0 ps<sup>-1</sup>. NPT simulations were conducted to compute pressure vs. volume curves for ScF<sub>3</sub>, starting from the 5x5x5 cubic supercell geometry. These simulations were run in the i-Pi package interfaced with ASE. Each simulation was run for 1 - 2 ns and performed over a pressure range of 0 to 500 MPa and temperature range of 55 to 240 K. A Langevin thermostat was used along with the MTK barostat.<sup>499</sup> A time constant of 1 ps<sup>-1</sup> was used for both the thermostat and barostat, and the time step was set to 1.0 fs.

Similar NPT simulations were run for CaZrF<sub>6</sub>. Three main sets of properties were calculated for CaZrF<sub>6</sub>. Thermal expansion was investigated between 100 and 1000 K at 0 MPa, similar to the ScF<sub>3</sub> simulations. Then, volume vs. pressure curves were obtained from a set of simulations run at a temperature of 290 K and pressures from 0 to 300 MPa. Finally, the bulk modulus of CaZrF<sub>6</sub> was computed over a temperature range of 300 to 500 K. A set of three simulations from 0 to 200 MPa was performed for each temperature. A linear fit to the average volume vs. pressure data from these simulations was used to compute the bulk modulus at each temperature. All CaZrF<sub>6</sub> simulations were run for ~ 1-2 ns. The Berendsen barostat with a 1.0 ps<sup>-1</sup> coupling constant was used for all simulations, along with a Langevin thermostat, also with a 1.0 ps<sup>-1</sup> coupling constant.

The i-Pi software package interfaced with ASE was utilized to run the path integral MD simulations.<sup>500,501</sup> Isotropic path integral NPT simulations were run for both systems. Simulations were run over a temperature range of 100 - 1200 K for cubic  $ScF_3$  and a temperature range of 100 - 1000 K for cubic  $CaZrF_6$ , and the pressure was fixed to 0 MPa. A smaller time step of 0.5 fs was used here. The barostat implementation details can be found in Bussi *et al.* [502] and Ceriotti *et al.* [500] A time constant of 5 ps<sup>-1</sup> was used for the barostat. A standard Langevin thermostat was used for temperature coupling, with a time constant of 5 ps<sup>-1</sup>. Depending on temperature, between 5 to 15 beads were utilized for the path integral simulations of both  $ScF_3$  and  $CaZrF_6$ (with greater number of beads used for lower temperatures). Convergence tests of the lattice parameter with

respect to number of beads are shown in Figure S5 and S11. Each PIMD simulation was run for 300-500 ps.

### 6.4 **Results and Discussion**

Our initial target is to investigate the extent to which the NN-driven MD simulations correctly predict qualitative trends in the CTE for both ScF<sub>3</sub> and CaZrF<sub>6</sub>. Because the NN potentials exhibit essentially DFT-level accuracy (Section 6.3.1), our results should be interpreted as the accuracy to which the underlying density functional (PBE) describes the material properties.. In Figure 1, we plot the predicted thermal expansion behavior for ScF<sub>3</sub> and CaZrF<sub>6</sub> as computed from both classical MD (labeled as "NN") and PIMD simulations (labeled as "PI NN") utilizing the NN potentials. We plot both the temperature-dependent reduced lattice constant ( $a/a_0$ ) and linear coefficient of thermal expansion: CTE =  $\frac{1}{a} \frac{da}{dT}$ ; these are plotted in Figure 6.1a and Figure 6.1b for ScF<sub>3</sub> and Figure 6.1c and Figure 6.1d for CaZrF<sub>6</sub>. In all cases, we plot corresponding experimental data and prior theoretical/computational predictions for comparison, where available. The curves are third order polynomials fit to the data points, with  $a_0$  taken as the extrapolated OK value for *a* from the fitted polynomial for each set of data.

Inspection of Figure 6.1 indicates that the NN simulations indeed predict NTE for both systems, in qualitative agreement with experiment. Furthermore, the simulations correctly predict the experimental trend that the CTE is more negative for CaZrF<sub>6</sub> than ScF<sub>3</sub>. Experimentally, the CTE values are  $\alpha_{100K,L} = -18$  ppm/K for CaZrF<sub>6</sub> and  $\alpha_{100K,L} = -14$  ppm/K for ScF<sub>3</sub> at 100 K. For comparison, the classical MD ("NN") simulations predict a CTE of  $\alpha_{100K,L} = -9.8$  ppm/K for CaZrF<sub>6</sub> and  $\alpha_{100K,L} = -4.3$  ppm/K for ScF<sub>3</sub> at 100 K. Incorporation of NQE within the PIMD simulations ("PI NN") brings the CTE values into closer agreement with experiment, with these simulations predicting  $\alpha_{100K,L} = -14.3$  ppm/K for CaZrF<sub>6</sub> and  $\alpha_{100K,L} = -8.7$  ppm/K for ScF<sub>3</sub> at 100 K. Thus NQE lower the CTE by ~ 30-60% at 100 K, while the influence of NQE diminishes at higher temperatures and largely disappears by temperatures of 700 K to 800 K. Overall, these results indicate that the MD simulations driven by DFT-trained NNs predict NTE behavior of ScF<sub>3</sub>



Figure 6.1: Thermal expansion of ScF<sub>3</sub> denoted by a) reduced lattice constant  $a/a_0$  and b) CTE. Thermal expansion of CaZrF<sub>6</sub> denoted by c) reduced lattice constant  $a/a_0$  and d) CTE. The experimental results for ScF<sub>3</sub> are from Greve *et al.* [443] and the experimental results for CaZrF<sub>6</sub> are from Hancock *et al.* [457] SCP and ISC predictions for ScF<sub>3</sub> are taken from Oba *et al.* [467]. The statistical uncertainty in our predictions is on par with the size of the symbols themselves, and thus error bars not shown.

and  $CaZrF_6$  in semi-quantitative to qualitative agreement with experiment.

Our CTE predictions for CaZrF<sub>6</sub> are comparatively better than for ScF<sub>3</sub> in that there is closer quantitative agreement with experiment. For CaZrF<sub>6</sub> the predicted CTE at 100 K is within a factor of 1.3 of the experimental value when incorporating NQE with the PIMD simulations (and a factor of 1.8 without NQE). There is no obvious reason for why the predictions for CaZrF<sub>6</sub> agree somewhat better with experiment compared to the predictions for ScF<sub>3</sub> (although as discussed later, the pressure-induced softening effect for ScF<sub>3</sub> is also not reproduced by our simulations). The influence of NQE on the CTE is consistent with the trend expected from analysis of the phonon spectrum. For both CaZrF<sub>6</sub> and ScF<sub>3</sub> materials, low frequency phonon modes exhibit negative Grüneisen parameters associated with NTE, while higher frequency phonons exhibit positive Grüneisen parameters.<sup>457,503</sup> A classical treatment will incorrectly assign energy equipartition to these high frequency modes and overpredict their contribution to (positive) thermal expansion at low temperatures.<sup>467</sup> Further inconsistencies of our simulation predictions with the experimental NTE behavior are predicted transitions from negative to positive thermal expansion at lower temperatures than those experimentally measured. As seen in Figure 6.1, our simulations for  $ScF_3$ predict positive thermal expansion above temperatures of 600 K/700 K for MD/PIMD respectively, while the experimental crossover temperature is near 1100 K. For CaZrF<sub>6</sub>, a similar effect is observed, with the transition occurring at 900 K in both the classical MD and PIMD predictions, while solely NTE is experimentally observed over the full characterized temperature range.

We discuss possible reasons for the quantitative discrepancies between our thermal expansion predictions and the experimental data. Important considerations for *ab initio* equation of state prediction for these materials include: 1) phonon anharmonicity; 2) NQE; 3) finite-size effects; and 4) accuracy of the underlying density functional. Regarding 1), predictions from MD simulations explicitly incorporate full anharmonicity of phonon modes, in contrast to theoretical free energy models (*vide infra*). For 2), we have explicitly evaluated the influence of NQE, and NQE do not account for the remaining discrepancy with experiment; Figures S5 and S11 show that NQE are largely converged with respect to the number of beads used in the PIMD simulations. Regarding 3)

finite-size effects, we have tested convergence of our predictions with respect to simulated supercell size. Figures S4 and S10 show that CTE predictions for ScF<sub>3</sub> and CaZrF<sub>6</sub> with our simulation systems are converged with respect to supercell size. The remaining consideration is the accuracy of the underlying density functional, which is PBE in this study. In this regard, our predictions for ScF<sub>3</sub> thermal expansion are similar to the AIMD results from Bocharov *et al.* [476] that utilize the PBEsol functional (Figure S6), although predictions with PBEsol show a higher temperature for transition from negative to positive thermal expansion, in better agreement with experiment. Dispersion corrected functionals would potentially be expected to yield better physical predictions. In fact, these were the initial functionals chosen for this study, but our initial NN models trained to dispersion corrected functionals all proved to exhibit dramatic instabilities when utilized in MD simulations (see Supplementary Material); similar issues have been noted before.<sup>504</sup> An interesting direction for future work would be to examine an improved treatment of electron correlation through methods such as the random-phase approximation (RPA),<sup>504–506</sup> which may be achieved by either completely rebuilding the neural network/training set or possibly through a transfer learning procedure.<sup>507</sup>

Despite the quantitative errors, our NN-driven MD simulations represent the "state-of-the-art" in *ab initio*, equation of state predictions for ScF<sub>3</sub> and CaZrF<sub>6</sub>. Figure 6.1b compares alternative theoretical predictions for the CTE of ScF<sub>3</sub>. Specifically, we compare to predictions from the SCP and ISC phonon theories from Oba *et al.* [467]. The SCP theory goes beyond the quasi-harmonic approximation (QHA) by incorporating quartic anharmonicity, and the ISC incorporates cubic anharmonicity. The predictions of SCP and ISC theories are based on quantum mechanical free energy functions, and thus explicitly take into account NQEs.<sup>467</sup> Furthermore classical limits can be derived within the SCP framework, and we label such results from Oba *et al.* [467] as "Classical SCP" in Figure 6.1b. As indicated in Figure 6.1b, the shape of the classical SCP CTE curve looks similar to our classical MD and PIMD results; however the predictions exhibit significant quantitative deviation from experiment as they display essentially no NTE over the reported temperature range. Comparing the classical and quantum SCP results indicates a very similar con-

tribution of NQEs as predicted by our classical MD/PIMD simulations. Of the theoretical models, the ISC predictions exhibit the best quantitative agreement with experiment, and show a similar temperature-dependent trend for CTE as compared to experiment. Overall, the ISC predictions and our "PI NN" predictions for CTE exhibit similar quantitative accuracy as compared to experiment. This indicates that including cubic and quartic anharmonicity within the ISC theory provides similar predictive accuracy as PIMD simulations for ScF<sub>3</sub> and similar materials. The methods should be viewed as complementary; PIMD provides a straightforward approach for property predictions utilizing standard DFT machinery (energy, forces and stress tensor), while the ISC phonon theory provides enhanced physical understanding of anharmonicity contributions to the equation of state behavior.<sup>467</sup>

We next compare atomic displacement parameters (ADPs), which measure the mean-square displacement of an atom from its crystal lattice position.<sup>508</sup> There are six unique components to the anisotropic atomic displacement tensor (three diagonal and three off-diagonal).<sup>508</sup> The meansquare displacement of each atom is straightforward to compute from simulations, e.g.  $U_{xx}$  =  $\langle u_x^2 \rangle$ , where u is an instantaneous atomic displacement from its mean position and x is a Cartesian coordinate. These can be compared to corresponding experimental values as measured by neutron or X-ray diffraction.<sup>508</sup> Within the ScF<sub>3</sub> crystal lattice, the Sc atoms undergo isotropic displacement,  $U_{iso} = U_{11} = U_{22} = U_{33}$ ;<sup>446</sup> for F atoms, the displacement parameters corresponding to transverse motion  $(U_{11} = U_{22})$  are different from the displacement parameter corresponding to longitudinal motion  $(U_{33})$ . In Figure 6.2, we compare the atomic displacement parameters for ScF<sub>3</sub> as predicted by our simulations to experimental values.<sup>443</sup> We first focus on the scandium atoms (Figure 6.2a). At low temperatures, the agreement between simulation and experiment is quite good. At higher temperatures, the simulations predict somewhat larger atomic displacement parameters for Sc than what is observed in experiment. For the fluorine atoms (Figure 6.2b) there is very good agreement between the predicted and experimental ADPs over the full temperature range. The transverse  $U_{11}$  parameter for fluorine is much larger, with a stronger temperature dependence compared to the longitudal  $U_{33}$  parameter. As discussed previously,<sup>446</sup> this is the expected



Figure 6.2: ScF<sub>3</sub> atomic displacement parameters for a) Sc atoms ( $U_{iso}$ ) and b) F atoms ( $U_{11}$  and  $U_{33}$ ) The experimental data is from Greve *et al.* [443]. The statistical uncertainty in our predictions is on par with the size of the symbols themselves, and thus error bars not shown.

behavior for the cubic crystal structure. We note that there is no significant difference between the "NN" and "PI NN" results for all predicted ADP values, indicating a negligible influence of NQE.

The ADPs for CaZrF<sub>6</sub>, shown in Figure 6.3, display similarly good agreement with experiment.<sup>457</sup> In this case,  $U_{iso} = U_{11} = U_{22} = U_{33}$  for both Ca and Zr. The fluorine ADPs are similar to those in ScF<sub>3</sub>, with two ADPs corresponding to transverse motion and one ADP corresponding to longitudinal motion. The fluorine ADPs within CaZrF<sub>6</sub> (Figure 6.3c) are somewhat larger than the corresponding ADPs in ScF<sub>3</sub>(note the different temperature scales of Figure 6.2 and Figure 6.3); this correlates with the more substantial NTE in CaZrF<sub>6</sub> as compared to ScF<sub>3</sub>. The agreement between the Ca and Zr ADPs in Figure 6.3a and b between simulation and experiment is good, although there is some slight disagreement at 100 K for Ca (Figure 6.3a). Analagous to ScF<sub>3</sub> there is no significant difference between the classical and path integral ADP predictions for CaZrF<sub>6</sub>, indicating negligible influence of NQE.

We next discuss pressure-volume equation of state predictions, as compared with experiment. ScF<sub>3</sub> is reported to undergo pressure-induced softening, which is an anomalous phenomenon that has been hypothesized to be driven by similar phonon mechanisms as NTE.<sup>446</sup> The Birch-Murnaghan equation of state provides information about the extent of pressure-induced softening



Figure 6.3: CaZrF<sub>6</sub> atomic displacement parameter for a) Ca, b) Zr and c) F. The experimental data is from Hancock *et al.* [457]. The statistical uncertainty in our predictions is on par with the size of the symbols themselves, and thus error bars not shown.

observed over a given region of the phase diagram. The third-order Birch-Murnagahn equation of state is given by:

$$P(V) = \frac{3B_0}{2} [(\frac{V_0}{V})^{7/3} - (\frac{V_0}{V})^{5/3}] \times \{1 + \frac{3}{4}(B'-4)[(\frac{V_0}{V})^{2/3} - 1]\}$$
(6.1)

There are three fitted quantities in Equation 6.1:  $B_0$  is the value of the bulk modulus at zero pressure,  $V_0$  is the value of the volume at zero pressure and B' is the first derivative of the bulk modulus with respect to pressure. A negative value of B' indicates pressure-induced softening. Figure 6.4a shows the pressure-volume curves from experiment,<sup>446</sup> and Figure 6.4b shows our simulation predictions for the cubic phase of ScF<sub>3</sub>. Each data set was subsequently fit to Equation 6.1, with the fits corresponding to solid lines in Figure 6.4 and the fitted parameters  $B_0$  and B'plotted in Figure 6.5. Note that here we only show predictions from classical MD, as we observed little change in the results with PIMD.

We first discuss the predictions from our simulations of the  $ScF_3$  cubic phase. As seen in Figure 6.4b, the Birch-Murnaghan equation of state fits our simulation data well. The fitted  $B_0$  values from the equation of state, shown in Figure 6.5, fall within the range of those fit to the experimental data. The simulation predicted volumes are approximately 3% smaller than the corresponding experimental volumes, which is largely due to errors in DFT predicted bond lengths. However, there are other non-trivial differences between our simulation curves and the experimental curves.



Figure 6.4: a) Experimental pressure vs. volume data of  $ScF_3$  from Wei *et al.* [446] b) Pressure vs. volume from the cubic  $ScF_3$  simulations.



Figure 6.5: a)  $B_0$  parameter and b) B' parameter from the Birch-Murnaghan equation of state from both simulation and experiment for ScF<sub>3</sub>. The experimental data is from Wei *et al.* [446]

The experimental P-V curves (Figure 6.4a) display a significant curvature at higher pressure, corresponding to the reported pressure-induced softening effect.<sup>446</sup> The pressure-induced softening observed experimentally is most apparent for the low temperature data. In contrast, there are no signatures of pressure-induced softening in our simulations of cubic phase ScF<sub>3</sub>. As shown in Figure 6.4b, the simulated P-V curves all exhibit a close to linear relationship between pressure and volume, even at lower temperatures. This is reflected in the value of B' plotted in Figure 6.5b, in which simulation values are all close to 0, reflecting no pressure-induced softening; in contrast, the experimental values are all significantly negative. The high pressure bound of the experimental P-V data in Figure 6.4a corresponds to a phase transition of  $ScF_3$  from the cubic to rhombohedral phase.<sup>443</sup>. It has thus been hypothesized that the pressure-induced softening effect is related to the proximity to this phase transition, and/or local fluctuations involving ScF<sub>6</sub> octahedral rotations that resemble motifs of the rhombohedral phase.<sup>446</sup> We thus speculated that the reason simulated P-V curves do not display pressure-induced softening was because the NN was trained solely to ScF<sub>3</sub> cubic phase data, and does not "extrapolate" to such structural motifs. We hypothesized that by adding rhombohedral phase structures to the NN training set, the NN would "learn" about such rhombohedral-like, local ScF<sub>6</sub> octahedral rotations, and possibly improving agreement with experiment. However, this hypothesis turned out to be false (at least in terms of improving agreement with experiment). Upon adding significant training data encompassing rhombohedral structures, and retraining the neural network, the simulation predictions were essentially unchanged (Figure S8).

At this point, we can only speculate on the discrepancy with experiment. The simulations of the  $ScF_3$  cubic phase may not be capturing local fluctuations involving  $ScF_6$  octahedral rotations that resemble motifs of the rhombohedral phase, and are an important mechanism for pressure induced softening. Of course, the (unknown) phase diagram of the DFT Hamiltonian is likely quantitatively different than the physical/experimental phase diagram of  $ScF_3$ . If indeed the pressure-induced softening is related to close proximity to the phase-transition, then any discrepancy between DFT and experimental phase behavior would affect predictions of this phenomena. For example, if the



Figure 6.6: Ca $ZrF_6$  Volume vs. Pressure data at 290K. The experimental data is from Hester *et al.* [458]

cubic to rhombohedral transition occurred at much higher pressures on the DFT phase diagram, then the simulated pressures in Figure 6.4b may be relatively far from the phase-transition, possibly explaining the lack of pressure-induced softening. This is entirely speculative, and would require characterization of the DFT-predicted phase behavior; however, such free energy calculations are beyond the scope of the present work. We note that previous MD simulations utilizing simple bond/angle potentials have predicted pressure-induced softening, in good qualitative agreement with experiment.<sup>446</sup> However, the CTE predicted by these simulations were in large quantitative error, in contrast to our present *ab initio* predictions.

We next discuss equation of state and bulk modulus predictions for CaZrF<sub>6</sub>. In Figure 6.6, we show experimental volume vs. pressure data for CaZrF<sub>6</sub> at 290 K from Hester *et al.* [458], as well as our corresponding predictions from the NN-driven MD simulations (only classical MD results are shown, as NQE are minor here). There is very good agreement in the P-V trend as predicted by simulation compared to experiment. The absolute unitcell volumes from the NN simulations differ from the experimental values by about 3%, again due to corresponding error in DFT predicted bond lengths. Additionally, it is seen that there is essentially no pressure-induced softening for CaZrF<sub>6</sub> bulk modulus to experimental data over a wide temperature range.<sup>457</sup> At STP, the predicted



Figure 6.7: CaZrF<sub>6</sub> Bulk modulus at various temperatures from both simulation and experiment. The experimental data is from Hancock *et al.* [457]

bulk modulus is 39 GPa, which is in excellent agreement with the experimental value of  $\sim$  37 GPa. The experimental bulk modulus indicates some thermal softening at higher temperatures, and our NN predictions follow a similar trend. Overall, the simulation predictions are in good agreement with experiment over the full temperature range.

#### 6.5 Conclusion

We have demonstrated the accuracy attainable with DFT-trained NNs for predicting equation of state properties of the materials ScF<sub>3</sub> and CaZrF<sub>6</sub>. Our benchmarks have covered a representative set of properties for these systems, including the coefficient of thermal expansion, atomic displacement parameters, equations of state and the bulk modulus. Our NN-driven, MD simulation predictions largely follow the experimentally-observed trends. The advantage of the NN potentials is that their development/parameterization is straightforward utilizing modern software libraries,<sup>68,107</sup> and sufficient training data is generated from relatively short AIMD simulations. The enhanced computational efficiency of the NNs enable longer simulations of larger supercells compared to AIMD, and additionally allow application of path integral approaches for incorporating NQE.

Examining the accuracy of our predictions for the CTE and pressure-volume properties such as the bulk modulus, it appears that the DFT/NN prediction is somewhat more accurate for the latter

than the former. We have quantified the significant contribution of NQE to NTE at low temperatures; in general, classical simulations underpredict NQE in  $ScF_3$  and  $CaZrF_6$  at low temperature due to unphysical equipartition in high-frequency phonon modes with positive Grüneisen parameters. The NN potentials correctly predict that  $CaZrF_6$  undergoes NTE to a greater extent than  $ScF_3$ , and overall the predictions exhibit semi-quantitative to qualitative agreement with experiment. The most apparent discrepancy with experiment is the lack of pressure-induced softening in our simulaton predictions for cubic  $ScF_3$ . In future work, it would be interesting to investigate whether a higher-level of *ab initio* theory, such as RPA, could improve our predictions. This may be most easily achieved through "transfer learning" of the present NN potentials to higher accuracy training data.<sup>507</sup>

Concerning future outlook, our workflow based on NN potentials provides the opportunity to screen promising new materials for pronounced NTE. The present workflow can be automated and extended to a variety of similar compounds, and can help identify promising candidates for further experimental investigation. A database of the results and structures from these NN simulations would help identify and analyze the important features necessary for strong NTE, which could also guide experiment. Additionally, it is possible to extend our simulations in order to incorporate defects or impurities into the framework, which has been done experimentally in order to tune NTE behavior.<sup>449-451,459,462</sup> Through these efforts, we anticipate that NN-driven, MD simulations will continue to play an increasingly important role in "first-principles" materials property predictions.

# CHAPTER 7 CONCLUSIONS AND OUTLOOK

This thesis has recorded our efforts to develop and implement force fields for investigation of reactions within ionic media. Our method combines conventional force field approaches with neural network methodologies, providing an extensible approach for studying reactions within the condensed phase. We demonstrated how to build the force field on the example reaction between EMIM<sup>+</sup> and acetate to form a N-heterocyclic carbene and acetic acid, and then showed how to incorporate additional reactions within the Hamiltonian. Computed free energy profiles with our model exhibit *ab initio* accuracy with respect to gas phase AIMD profiles at greatly reduced computational cost. We also included examples of the utility of the individual components of our reactive force field for modeling condensed phase systems. We incorporated ab initio-based force fields to handle hydronium and water interactions with the ionic liquid within the existing MS-EVB reactive force field. The accurate asymptotic interactions between solvent molecules and the reacting proton were key for producing conductivity values close to those from experiment. The accuracy of these force fields for modeling solvent dynamics was also demonstrated in Chapter 5 of this thesis, which showed propylene carbonate properties were modeled accurately over a wide temperature and pressure range. We concluded with an example of fitting a neural network to simulate the crystalline materials  $ScF_3$  and  $CaZrF_6$ , which require precise modeling of their various phonon modes in order to produce predictions in accord with experiment. Many of our predictions were state of the art for these systems, demonstrating the utility of the ML models.

Future work within computational chemistry will undoubtedly feature further applications of both neural networks and *ab initio* calculations. We share future directions for the projects included in this work and share some outlook on the field.

# 7.1 Future Directions

The future work involved with our reactive force field involves simplifying the parameterization procedure. Currently, the most difficult aspect to work with is constructing the diagonal elements of the Hamiltonian, which contains two sets of neural networks: one describing the intramolecular interactions for each reacting complex molecule and one describing intermolecular interactions between pairs of reacting complex molecules. This has certain advantages in that we have constructed the training set for these molecules so it is very clear which terms model which interactions. However, this setup requires additional training data, additional training time and significantly adds to the complexity of the model.

Future plans involve minimizing the number of steps required to build the Hamiltonian. The intermolecular neural networks are the most difficult to fit as they require a diabatization scheme; ab *initio* diabats are not rigorously defined except for simple cases, and our method for parameterizing these terms (symmetry adapted perturbation theory) includes 2 QM calculations per diabat, which adds to the size of the training data set required. One possibility that we are currently exploring is employing a "machine learned" diabatization scheme. We plan to do this by fitting the intermolecular neural networks and off-diagonal elements of 2x2 blocks of the Hamiltonian simultaneously. Fitting these terms to a dataset composed of reacting dimer configurations within the gas phase and within a solvent may allow for the intermolecular NN terms to adjust to the different environments, forming an approximate diabat within the training process. This would remove the need to take two separate fitting steps for each diagonal term, greatly reducing the training time. The gas phase training data can still be obtained with AIMD, and the solvent data can be obtained from QM/MM. The additional benefit of fitting the Hamiltonian to QM/MM is that, if the PB/NN and QM/MM energy differences are small, it will be possible to switch the PB/NN Hamiltonian with the QM/MM Hamiltonian if the PB/NN Hamiltonian uncertainty grows too large. The QM/MM Hamiltonian can then be used to gather more training data for fitting the PB/NN model. This switching process would not be possible for PB/NN condensed phase simulations if it was only trained to AIMD.

# 7.2 Outlook

The advent of machine learning has brought about a kind of revolution within quantum chemistry and molecular modeling over the last few years. Broader society is experiencing its own type of AI revolution currently thanks to the introduction of models like ChatGPT. It is forecast that billions of dollars will be spent on AI and machine learning within the next few years, with the goal of advancing current hardware and software. Theoretical chemistry will almost certainly benefit from further progress brought about by the broader ML and computing fields. Innovations made in the design of GPUs and CPUs will reduce several key limitations involved with ML today, aiding faster dataset generation and model training/evaluation. Algorithm development will also be key within the next several years, with the goal of reducing the computational cost of the neural network Advances in data efficiency and model speed may eventually allow for a stable ML potential to be parameterized by short *ab initio* simulations, which will significantly ease the training process. Along with adoption of neural networks into current simulation packages, machine learning methods may become as commonly used for molecular simulation as force fields are currently. In companion with faster ML models, novel ML-enhanced biased sampling approaches will allow for the discovery and potentially the generation of new reaction pathways and mechanisms, which will provide another robust complement to the molecular simulation community. All of these advances together, along with further application to systems within ionic media, will enhance our current knowledge of how to develop targeted reactions for key new technologies.

#### REFERENCES

- [1] J. J. Varghese and S. H. Mushrif, "Origins of complex solvent effects on chemical reactivity and computational tools to investigate them: A review," *React. Chem. Eng.*, vol. 4, no. 2, pp. 165–206, 2019.
- [2] A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu, and M. H. M. Olsson, "Electrostatic Basis for Enzyme Catalysis," *Chem. Rev.*, vol. 106, no. 8, pp. 3210–3235, 2006.
- [3] C. Reichardt, "Solvation Effects in Organic Chemistry: A Short Historical Overview," J. Org. Chem., vol. 87, no. 3, pp. 1616–1629, 2022.
- [4] J. Blumberger, L. Bernasconi, I. Tavernelli, R. Vuilleumier, and M. Sprik, "Electronic Structure and Solvation of Copper and Silver Ions: A Theoretical Picture of a Model Aqueous Redox Reaction," J. Am. Chem. Soc., vol. 126, no. 12, pp. 3928–3938, 2004.
- [5] J. A. Harrison, J. D. Schall, S. Maskey, P. T. Mikulski, M. T. Knippenberg, and B. H. Morrow, "Review of force fields and intermolecular potentials used in atomistic computational materials research," *Appl. Phys. Rev.*, vol. 5, no. 3, p. 31 104, 2018.
- [6] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln," Ann. Phys., vol. 389, no. 20, pp. 457–484, 1927.
- [7] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, First. Mineola: Dover Publications, Inc., 1996.
- [8] R. O. Jones, "Density functional theory: Its origins, rise to prominence, and future," *Rev. Mod. Phys.*, vol. 87, no. 3, pp. 897–923, 2015.
- [9] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Phys. Rev.*, vol. 136, no. 3B, B864–B871, 1964.
- [10] R. G. Parr and W. Yang, Density-Functional Theory of Atoms and Molecules (International Series of Monographs on Chemistry). Oxford University Press, USA, 1994, ISBN: 0195092767.
- [11] W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Phys. Rev.*, vol. 140, no. 4A, A1133–A1138, 1965.
- [12] S. Lehtola, F. Blockhuys, and C. Van Alsenoy, "An Overview of Self-Consistent Field Calculations Within Finite Basis Sets," *Molecules*, vol. 25, no. 5, p. 1218, 2020.

- [13] A. D. Kaplan, M. Levy, and J. P. Perdew, "The Predictive Power of Exact Constraints and Appropriate Norms in Density Functional Theory," *Annu. Rev. Phys. Chem.*, vol. 74, no. 1, pp. 193–218, 2023.
- [14] M. E. Tuckerman, "Ab initio molecular dynamics: Basic concepts, current trends and novel applications," *J. Phys. Condens. Matter*, vol. 14, no. 50, 2002.
- [15] S. Seritan *et al.*, "TeraChem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 11, no. 2, e1494, 2021.
- [16] A. Sisto, D. R. Glowacki, and T. J. Martinez, "Ab Initio Nonadiabatic Dynamics of Multichromophore Complexes: A Scalable Graphical-Processing-Unit-Accelerated Exciton Framework," Acc. Chem. Res., vol. 47, no. 9, pp. 2857–2866, 2014.
- [17] F. Shao, J. K. Wong, Q. H. Low, M. Iannuzzi, J. Li, and J. Lan, "In situ spectroelectrochemical probing of CO redox landscape on copper single-crystal surfaces," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 119, no. 29, e2118166119, 2022.
- [18] P. Giannozzi *et al.*, "Quantum ESPRESSO toward the exascale," J. Chem. Phys., vol. 152, no. 15, p. 154 105, 2020.
- [19] T. D. Kühne *et al.*, "CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations," *J. Chem. Phys.*, vol. 152, no. 19, p. 194 103, 2020.
- [20] R. Iftimie, P. Minary, and M. E. Tuckerman, "Ab initio molecular dynamics: Concepts, recent developments, and future trends," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 19, pp. 6654–6659, 2005.
- [21] R. Car and M. Parrinello, "Unified Approach for Molecular Dynamics and Density-Functional Theory," *Phys. Rev. Lett.*, vol. 55, no. 22, pp. 2471–2474, 1985.
- [22] L.-P. Wang, A. Titov, R. McGibbon, F. Liu, V. S. Pande, and T. J. Martínez, "Discovering chemistry with an ab initio nanoreactor," *Nat. Chem.*, vol. 6, no. 12, pp. 1044–1048, 2014.
- [23] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," J. Mol. Biol., vol. 103, no. 2, pp. 227–249, 1976.
- [24] M. Liu, Y. Wang, Y. Chen, M. J. Field, and J. Gao, "QM/MM through the 1990s: The First Twenty Years of Method Development and Applications," *Isr. J. Chem.*, vol. 54, no. 8-9, pp. 1250–1263, 2014.

- [25] E. G. Kratz, A. R. Walker, L. Lagardère, F. Lipparini, J.-P. Piquemal, and G. Andrés Cisneros, "LICHEM: A QM/MM program for simulations with multipolar and polarizable force fields," *J. Comput. Chem.*, vol. 37, no. 11, pp. 1019–1029, 2016.
- [26] U. C. Singh and P. A. Kollman, "A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH3Cl + Cl exchange reaction and gas phase protonation of polyethers," J. *Comput. Chem.*, vol. 7, no. 6, pp. 718–730, 1986.
- [27] M. J. Field, P. A. Bash, and M. Karplus, "A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations," *J. Comput. Chem.*, vol. 11, no. 6, pp. 700–733, 1990.
- [28] J. Gao and X. Xia, "A Priori Evaluation of Aqueous Polarization Effects through Monte Carlo QM-MM Simulations," *Science*, vol. 258, no. 5082, pp. 631–635, 1992.
- [29] M. A. Thompson and G. K. Schenter, "Excited States of the Bacteriochlorophyll b Dimer of Rhodopseudomonas viridis: A QM/MM Study of the Photosynthetic Reaction Center That Includes MM Polarization," J. Phys. Chem., vol. 99, no. 17, pp. 6374–6386, 1995.
- [30] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, "ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for DielsAlder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition," *J. Phys. Chem.*, vol. 100, no. 50, pp. 19357–19363, 1996.
- [31] J. P. Pederson and J. G. McDaniel, "DFT-based QM/MM with particle-mesh Ewald for direct, long-range electrostatic embedding," *J. Chem. Phys.*, vol. 156, no. 17, p. 174 105, 2022.
- [32] X. Pan, K. Nam, E. Epifanovsky, A. C. Simmonett, E. Rosta, and Y. Shao, "A simplified charge projection scheme for long-range electrostatics in ab initio QM/MM calculations," *J. Chem. Phys.*, vol. 154, no. 2, p. 024 115, 2021.
- [33] R. N. Tazhigulov and K. B. Bravaya, "Free Energies of Redox Half-Reactions from First-Principles Calculations," *J. Phys. Chem. Lett.*, vol. 7, no. 13, pp. 2490–2495, 2016.
- [34] C. L. Brooks III, B. M. Pettitt, and M. Karplus, "Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids," J. Chem. Phys., vol. 83, no. 11, pp. 5897–5908, 1985.
- [35] H. Schreiber and O. Steinhauser, "Taming cut-off induced artifacts in molecular dynamics studies of solvated polypeptides: The reaction field method," *J. Mol. Bio.*, vol. 228, no. 3, pp. 909–923, 1992.

- [36] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *J. Chem. Phys.*, vol. 103, no. 19, pp. 8577–8593, 1995.
- [37] G. Jindal and A. Warshel, "Exploring the Dependence of QM/MM Calculations of Enzyme Catalysis on the Size of the QM Region," J. Phys. Chem. B, vol. 120, no. 37, pp. 9913– 9921, 2016.
- [38] G. Hong, E. Rosta, and A. Warshel, "Using the constrained DFT approach in generating diabatic surfaces and off diagonal empirical valence bond terms for modeling reactions in condensed phases," J. Phys. Chem. B, vol. 110, no. 39, pp. 19570–19574, 2006.
- [39] A. Heyden, H. Lin, and D. G. Truhlar, "Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations," *J. Phys. Chem. B*, vol. 111, no. 9, pp. 2231–2241, 2007.
- [40] A. W. Duster, C.-H. Wang, C. M. Garza, D. E. Miller, and H. Lin, "Adaptive quantum/molecular mechanics: what have we learned, where are we, and where do we go from here?" *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 7, no. 5, e1310, 2017.
- [41] A. W. Duster, C. M. Garza, B. O. Aydintug, M. B. Negussie, and H. Lin, "Adaptive Partitioning QM/MM for Molecular Dynamics Simulations: 6. Proton Transport through a Biological Channel," *J. Chem. Theory Comput.*, vol. 15, no. 2, pp. 892–905, 2019.
- [42] A. Warshel and R. M. Weiss, "An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes," J. Am. Chem. Soc., vol. 102, no. 20, pp. 6218–6226, 1980.
- [43] L. Mones, W.-J. Tang, and J. Florián, "Empirical Valence Bond Simulations of the Chemical Mechanism of ATP to cAMP Conversion by Anthrax Edema Factor," *Biochemistry*, vol. 52, no. 15, pp. 2672–2682, 2013.
- [44] J. Åqvist and A. Warshel, "Simulation of Enzyme Reactions Using Valence Bond Force Fields and Other Hybrid Quantum/Classical Approaches," *Chem. Rev.*, vol. 93, no. 7, pp. 2523–2544, 1993.
- [45] A. Warshel, "Dynamics of reactions in polar solvents. Semiclassical trajectory studies of electron-transfer and proton-transfer reactions," J. Phys. Chem., vol. 86, no. 12, pp. 2218– 2224, 1982.
- [46] S. C. Kamerlin, J. Cao, E. Rosta, and A. Warshel, "On unjustifiably misrepresenting the evb approach while simultaneously adopting it," *J. Phys. Chem. B*, vol. 113, no. 31, pp. 10905–10915, 2009.
- [47] S. C. L. Kamerlin and A. Warshel, "The empirical valence bond model: theory and applications," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 1, no. 1, pp. 30–45, 2011.
- [48] A. Warshel and J.-K. Hwang, "Simulation of the dynamics of electron transfer reactions in polar solvents: Semiclassical trajectories and dispersed polaron approaches," J. Chem. Phys., vol. 84, no. 9, pp. 4938–4957, 1986.
- [49] D. R. Glowacki, A. J. Orr-Ewing, and J. N. Harvey, "Non-equilibrium reaction and relaxation dynamics in a strongly interacting explicit solvent: F + CD3CN treated with a parallel multi-state EVB model," *J. Chem. Phys.*, vol. 143, no. 4, p. 044 120, 2015.
- [50] B. Hartke and S. Grimme, "Reactive force fields made simple," *Phys. Chem. Chem. Phys.*, vol. 17, no. 26, pp. 16715–16718, 2015.
- [51] H. B. Schlegel and J. L. Sonnenberg, "Empirical valence-bond models for reactive potential energy surfaces using distributed Gaussians," *J. Chem. Theory Comput.*, vol. 2, no. 4, pp. 905–911, 2006.
- [52] C. Knight, G. E. Lindberg, and G. A. Voth, "Multiscale reactive molecular dynamics," *J. Chem. Phys.*, vol. 137, no. 22, 2012.
- [53] Y. Wu, H. Chen, F. Wang, F. Paesani, and G. A. Voth, "An Improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transport," J. Phys. Chem. B, vol. 112, no. 2, pp. 467–482, 2008.
- [54] R. Biswas, Y.-L. S. Tse, A. Tokmakoff, and G. A. Voth, "Role of Presolvation and Anharmonicity in Aqueous Phase Hydrated Proton Solvation and Transport," *J. Phys. Chem. B*, vol. 120, no. 8, pp. 1793–1804, 2016.
- [55] D. Marx, "Proton Transfer 200 Years after von Grotthuss: Insights from Ab Initio Simulations," *ChemPhysChem*, vol. 7, no. 9, pp. 1848–1870, 2006.
- [56] L. P. Wang, T. J. Martinez, and V. S. Pande, "Building force fields: An automatic, systematic, and reproducible approach," *J. Phys. Chem. Lett.*, vol. 5, no. 11, pp. 1885–1891, 2014.
- [57] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., 2019, pp. 8024–8035.
- [58] Martín Abadi *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015.
- [59] J. Behler and M. Parrinello, "Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces," *Phys. Rev. Lett.*, vol. 98, no. 14, p. 146401, 2007.

- [60] K. Schütt, P.-J. Kindermans, H. E. S. Felix, S. Chmiela, A. Tkatchenko, and K.-R. Müller, "SchNet: A continuous-filter convolutional neural network for modeling quantum interactions," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 991–1001.
- [61] K. T. Schütt, H. E. Sauceda, P. J. Kindermans, A. Tkatchenko, and K. R. Müller, "SchNet -A deep learning architecture for molecules and materials," *J. Chem. Phys.*, vol. 148, no. 24, 2018.
- [62] O. T. Unke and M. Meuwly, "PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges," J. Chem. Theory Comput., vol. 15, no. 6, pp. 3678– 3693, 2019.
- [63] J. Behler, "Perspective: Machine learning potentials for atomistic simulations," J. Chem. *Phys.*, vol. 145, no. 17, p. 170901, 2016.
- [64] J. Behler, "First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems," *Angew. Chem.*, vol. 56, no. 42, pp. 12828–12840, 2017.
- [65] O. T. Unke *et al.*, "Machine Learning Force Fields," *Chem. Rev.*, vol. 121, no. 16, pp. 10142–10186, 2021.
- [66] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, "Bypassing the Kohn-Sham equations with machine learning," *Nat. Commun.*, vol. 8, no. 1, p. 872, 2017.
- [67] M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller, and K. Burke, "Quantum chemical accuracy from density functional approximations via machine learning," *Nat. Commun.*, vol. 11, no. 1, p. 5223, 2020.
- [68] H. Wang, L. Zhang, J. Han, and W. E, "DeePMD-kit: A deep learning package for manybody potential energy representation and molecular dynamics," *Comput. Phys. Commun.*, vol. 228, pp. 178–184, 2018.
- [69] D. G. Smith *et al.*, "Psi4 1.4: Open-source software for high-throughput quantum chemistry," *J. Chem. Phys.*, vol. 152, no. 18, p. 184 108, 2020.
- [70] Y. Shao *et al.*, "Advances in molecular quantum chemistry contained in the Q-Chem 4 program package," *Mol. Phys.*, vol. 113, no. 2, pp. 184–215, 2015.
- [71] P. Giannozzi *et al.*, "Advanced capabilities for materials modelling with Quantum ESPRESSO," *J. Condens. Matter Phys.*, vol. 29, no. 46, p. 465 901, 2017.

- [72] X.-T. Fan, X.-J. Wen, Y.-B. Zhuang, and J. Cheng, "Molecular insight into the GaP(110)–water interface using machine learning accelerated molecular dynamics," J. Energy Chem., vol. 82, pp. 239–247, 2023.
- [73] M. S. Gomes-Filho, A. Torres, A. Reily Rocha, and L. S. Pedroza, "Size and Quality of Quantum Mechanical Data Set for Training Neural Network Force Fields for Liquid Water," J. Phys. Chem. B, vol. 127, no. 6, pp. 1422–1428, 2023.
- [74] Y. Xie, M. Bu, G. Zou, Y. Zhang, and G. Lu, "Molecular dynamics simulations of CaCl2–NaCl molten salt based on the machine learning potentials," *Sol. Energy Mater. Sol. Cells*, vol. 254, p. 112 275, 2023.
- [75] Y. Ling *et al.*, "Revisiting the structure, interaction, and dynamical property of ionic liquid from the deep learning force field," *J. Power Sources*, vol. 555, p. 232 350, 2023.
- [76] A. Mondal, D. Kussainova, S. Yue, and A. Z. Panagiotopoulos, "Modeling Chemical Reactions in Alkali Carbonate–Hydroxide Electrolytes with Deep Learning Potentials," *J. Chem. Theory Comput.*, 2022.
- [77] A. S. Raman and A. Selloni, "Modeling the Solvation and Acidity of Carboxylic Acids Using an Ab Initio Deep Neural Network Potential," J. Phys. Chem. A, vol. 126, no. 40, pp. 7283–7290, 2022.
- [78] Z. L. Glick, D. P. Metcalf, A. Koutsoukas, S. A. Spronk, D. L. Cheney, and C. D. Sherrill, "AP-Net: An atomic-pairwise neural network for smooth and transferable interaction potentials," J. Chem. Phys., vol. 153, no. 4, p. 044 112, 2020.
- [79] M. Gastegger, K. T. Schütt, and K.-R. Müller, "Machine learning of solvent effects on molecular spectra and reactions," *Chem. Sci.*, vol. 12, no. 34, pp. 11473–11483, 2021.
- [80] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and W. E, "Active learning of uniformly accurate interatomic potentials for materials simulation," *Phys. Rev. Mater.*, vol. 3, no. 2, p. 023 804, 2019.
- [81] J. P. Stoppelman and J. G. McDaniel, "Physics-based, neural network force fields for reactive molecular dynamics: Investigation of carbene formation from [EMIM+][OAc]," J. *Chem. Phys.*, vol. 155, no. 10, p. 104 112, 2021.
- [82] J. P. Stoppelman and J. G. McDaniel, "Erratum: "Physics-based, neural network force fields for reactive molecular dynamics: Investigation of carbene formation from [EMIM<sup>+</sup>] [OAc<sup>-</sup>]" [J. Chem. Phys. 155, 104112 (2021)]," J. Chem. Phys., vol. 156, no. 2, p. 029 901, 2022.
- [83] C. J. Cramer and D. G. Truhlar, "Implicit solvation models: Equilibria, structure, spectra, and dynamics," *Chem. Rev.*, vol. 99, no. 8, pp. 2161–2200, 1999.

- [84] J. Tomasi, B. Mennucci, and R. Cammi, "Quantum mechanical continuum solvation models," *Chem. Rev.*, vol. 105, no. 8, pp. 2999–3094, 2005.
- [85] C. D. Snow, H. Nguyen, V. S. Pande, and M. Gruebele, "Absolute comparison of simulated and experimental protein-folding dynamics," *Nature*, vol. 420, no. 6911, pp. 102– 106, 2002.
- [86] J.-E. Shea and C. L. Brooks III, "FROM FOLDING THEORIES TO FOLDING PRO-TEINS : A Review and Assessment of Simulation Studies of Protein Folding and Unfolding," Annu. Rev. Phys. Chem., vol. 52, no. 1, pp. 499–535, 2001.
- [87] W. Yang, Y. Q. Gao, Q. Cui, J. Ma, and M. Karplus, "The missing link between thermodynamics and structure in F1-ATPase," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, no. 3, pp. 874– 879, 2003.
- [88] J. W. Ponder and D. A. Case, "Force fields for protein simulations," Adv. Protein Chem., vol. 66, pp. 27–85, 2003.
- [89] F. Dommert, K. Wendler, R. Berger, L. Delle Site, and C. Holm, "Force fields for studying the structure and dynamics of ionic liquids: A critical review of recent developments," *ChemPhysChem*, vol. 13, no. 7, pp. 1625–1637, 2012.
- [90] S. Cui, J. Liu, M. E. Selvan, D. J. Keffer, B. J. Edwards, and W. V. Steele, "A molecular dynamics study of a nafion polyelectrolyte membrane and the aqueous phase structure for proton transport," *J. Phys. Chem. B*, vol. 111, no. 9, pp. 2208–2218, 2007.
- [91] Y. Han, D. Jiang, J. Zhang, W. Li, Z. Gan, and J. Gu, "Development, applications and challenges of ReaxFF reactive force field in molecular simulations," *Front. Chem. Sci. Eng.*, vol. 10, no. 1, pp. 16–38, 2016.
- [92] R. Slapikas, I. Dabo, and S. B. Sinnott, "Optimized utilization of COMB3 reactive potentials in LAMMPS," *J. Chem. Phys.*, vol. 152, no. 22, p. 224702, 2020.
- [93] D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott, "A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons," *J. Phys. Condens. Matter*, vol. 14, no. 4, pp. 783–802, 2002.
- [94] T. Liang *et al.*, "Reactive Potentials for Advanced Atomistic Simulations," *Annu. Rev. Mater. Res.*, vol. 43, no. 1, pp. 109–129, 2013.
- [95] A. C. Van Duin, A. Strachan, S. Stewman, Q. Zhang, X. Xu, and W. A. Goddard, "ReaxFF-SiO reactive force field for silicon and silicon oxide systems," *J. Phys. Chem. A*, vol. 107, no. 19, pp. 3803–3811, 2003.

- [96] L. W. Bertels, L. B. Newcomb, M. Alaghemandi, J. R. Green, and M. Head-Gordon, "Benchmarking the Performance of the ReaxFF Reactive Force Field on Hydrogen Combustion Systems," J. Phys. Chem. A, vol. 124, no. 27, pp. 5631–5645, 2020.
- [97] D. Furman and D. J. Wales, "Transforming the Accuracy and Numerical Stability of ReaxFF Reactive Force Fields," *J. Phys. Chem. Lett.*, pp. 7215–7223, 2019.
- [98] J.-K. Hwang, S. Creighton, G. King, D. Whitney, and A. Warshel, "Effects of solute-solvent coupling and solvent saturation on solvation dynamics of charge transfer reactions," *J. Chem. Phys.*, vol. 89, no. 2, p. 859, 1998.
- [99] M. Higashi and D. G. Truhlar, "Combined Electrostatically Embedded Multiconfiguration Molecular Mechanics and Molecular Mechanical Method: Application to Molecular Dynamics Simulation of a Chemical Reaction in Aqueous Solution with Hybrid Density Functional Theory," J. Chem. Theory Comput., vol. 4, no. 7, pp. 1032–1039, 2008.
- [100] S. C. Kamerlin and A. Warshel, "The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions," *Faraday Discuss.*, vol. 145, no. 0, pp. 71–106, 2010.
- [101] U. W. Schmitt and G. A. Voth, "Multistate empirical valence bond model for proton transport in water," *J. Phys. Chem. B*, vol. 102, no. 29, 1998.
- [102] T. J. F. Day, A. V. Soudackov, M. Čuma, U. W. Schmitt, and G. A. Voth, "A Second Generation Multistate Empirical Valence Bond Model for Proton Transport in Aqueous Systems," J. Chem. Phys., vol. 117, no. 12, pp. 5839–5849, 2002.
- [103] G. A. Voth, "Computer simulation of proton solvation and transport in aqueous and biomolecular systems," *Acc. Chem. Res.*, vol. 39, no. 2, pp. 143–150, 2006.
- [104] M. Čuma, U. W. Schmitt, and G. A. Voth, "A multi-state empirical valence bond model for weak acid dissociation in aqueous solution," *J. Phys. Chem. A*, vol. 105, no. 12, pp. 2814– 2823, 2001.
- [105] C. M. Maupin, K. F. Wong, A. V. Soudackov, S. Kim, and G. A. Voth, "A multistate empirical valence bond description of protonatable amino acids," *J. Phys. Chem. A*, vol. 110, no. 2, pp. 631–639, 2006.
- [106] H. Chen, T. Yan, and G. A. Voth, "A computer simulation model for proton transport in liquid imidazole," J. Phys. Chem. A, vol. 113, no. 16, pp. 4507–4517, 2009.
- [107] L. Zhang, J. Han, H. Wang, R. Car, and E. Weinan, "Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics," *Phys. Rev. Lett.*, vol. 120, no. 14, p. 143 001, 2018.

- [108] S. Amabilino, L. A. Bratholm, S. J. Bennie, A. C. Vaucher, M. Reiher, and D. R. Glowacki, "Training Neural Nets to Learn Reactive Potential Energy Surfaces Using Interactive Quantum Chemistry in Virtual Reality," *J. Phys. Chem. A*, vol. 123, no. 20, pp. 4486– 4499, 2019.
- [109] J. Behler, "Neural network potential-energy surfaces in chemistry: A tool for large-scale simulations," vol. 13, no. 40, pp. 17930–17955, 2011.
- [110] H. Schreiber and O. Steinhauser, "Molecular dynamics studies of solvated polypeptides: Why the cut-off scheme does not work," *Chem. Phys.*, vol. 168, no. 1, pp. 75–89, 1992.
- [111] G. S. Del Buono, F. E. Figueirido, and R. M. Levy, "Dielectric response of solvent surrounding an ion pair: Ewald potential versus spherical truncation," *Chem. Phys. Lett.*, vol. 263, no. 3-4, pp. 521–529, 1996.
- [112] L. Perera, U. Essmann, and M. L. Berkowitz, "Effect of the treatment of long-range forces on the dynamics of ions in aqueous solutions," *J. Chem. Phys.*, vol. 102, no. 1, pp. 450– 456, 1995.
- [113] B. Parsaeifard, D. S. De, J. A. Finkler, and S. Goedecker, "Fingerprint-Based Detection of Non-Local Effects in the Electronic Structure of a Simple Single Component Covalent System," *Condens. Matter*, vol. 6, no. 1, p. 9, 2021.
- [114] D. P. Metcalf, A. Jiang, S. A. Spronk, D. L. Cheney, and C. D. Sherrill, "Electron-Passing Neural Networks for Atomic Charge Prediction in Systems with Arbitrary Molecular Charge," J. Chem. Inf. Model., vol. 61, no. 1, pp. 115–122, 2021.
- [115] Z. L. Glick, A. Koutsoukas, D. L. Cheney, and C. D. Sherrill, "Cartesian message passing neural networks for directional properties: Fast and transferable atomic multipoles," *J. Chem. Phys.*, vol. 154, no. 22, p. 224 103, 2021.
- [116] D. M. Flanigan, F. Romanov-Michailidis, N. A. White, and T. Rovis, "Organocatalytic Reactions Enabled by N-Heterocyclic Carbenes," *Chem. Rev.*, vol. 115, no. 17, pp. 9307– 9387, 2015.
- [117] M. N. Hopkinson, C. Richter, M. Schedler, and F. Glorius, "An overview of N-heterocyclic carbenes," *Nature*, vol. 510, no. 7506, pp. 485–496, 2014.
- [118] R. Ratti, "Ionic Liquids: Synthesis and Applications in Catalysis," *Adv. Chem.*, vol. 2014, pp. 1–16, 2014.
- [119] I. Chiarotto, L. Mattiello, F. Pandolfi, D. Rocco, and M. Feroci, "NHC in Imidazolium Acetate Ionic Liquids: Actual or Potential Presence?" *Front. Chem.*, vol. 6, p. 355, 2018.

- [120] I. Chiarotto, M. Feroci, G. Forte, M. Orsini, and A. Inesi, "Proton-Exchange Equilibrium between Bases and [BMIm][BF4]: An Electrochemical Procedure to Evaluate the Presence of Carbenes for Synthetic Applications," *ChemElectroChem*, vol. 1, no. 9, pp. 1525–1530, 2014.
- [121] M. Feroci, I. Chiarotto, and A. Inesi, "Advances in the knowledge of N-heterocyclic carbenes properties. The backing of the electrochemical investigation," *Catalysts*, vol. 6, no. 11, p. 178, 2016.
- [122] M. Thomas, M. Brehm, O. Hollóczki, and B. Kirchner, "How can a carbene be active in an ionic liquid?" *Chem. Eur. J.*, vol. 20, no. 6, pp. 1622–1629, 2014.
- [123] N. M. A. N. Daud, E. Bakis, J. P. Hallett, C. C. Weber, and T. Welton, "Evidence for the spontaneous formation of N-heterocyclic carbenes in imidazolium based ionic liquids," *Chem. Commun.*, vol. 53, no. 81, pp. 11 154–11 156, 2017.
- [124] S. Gehrke, W. Reckien, I. Palazzo, T. Welton, and O. Hollóczki, "On the Carbene-Like Reactions of Imidazolium Acetate Ionic Liquids: Can Theory and Experiments Agree?" *Eur. J. Org. Chem.*, vol. 2019, no. 2-3, pp. 504–511, 2019.
- [125] Z. Kelemen, O. Hollóczki, J. Nagy, and L. Nyulászi, "An organocatalytic ionic liquid," *Org. Biomol. Chem.*, vol. 9, no. 15, p. 5362, 2011.
- [126] A. C. Baumruck, D. Tietze, A. Stark, and A. A. Tietze, "Reactions of Sulfur-Containing Organic Compounds and Peptides in 1-Ethyl-3-methyl-imidazolium Acetate," *J. Org. Chem.*, vol. 82, no. 14, pp. 7538–7545, 2017.
- [127] F. Pandolfi, M. Feroci, and I. Chiarotto, "Role of Anion and Cation in the 1-Methyl-3-butyl Imidazolium Ionic Liquids BMImX: The Knoevenagel Condensation," *ChemistrySelect*, vol. 3, no. 17, pp. 4745–4749, 2018.
- [128] I. Chiarotto, M. Feroci, G. Forte, and A. Inesi, "Stability of electrogenerated 1-butyl-3methylimidazol-2-ylidene in DMF. Part 2. Role of acid substrates. [1]," *Electrochim. Acta*, vol. 176, pp. 627–635, 2015.
- [129] I. Chiarotto, M. Feroci, and A. Inesi, "First direct evidence of N-heterocyclic carbene in BMIm acetate ionic liquids. An electrochemical and chemical study on the role of temperature," *New J. Chem.*, vol. 41, no. 16, pp. 7840–7843, 2017.
- [130] P. Jain, V. R. Chaudhari, and A. Kumar, "Water-assisted stability of carbene: cyclic voltammetric investigation of 1-ethyl-3-methylimidazolium ethylsulfate ionic liquid," *Phys. Chem. Chem. Phys.*, vol. 21, no. 43, pp. 24126–24131, 2019.
- [131] J. F. Wishart, "Energy applications of ionic liquids," *Energy Environ. Sci.*, vol. 2, no. 9, pp. 956–961, 2009.

- [132] S. Schneider, T. Hawkins, M. Rosander, G. Vaghjiani, S. Chambreau, and G. Drake, "Ionic liquids as hypergolic fuels," *Energy Fuels*, vol. 22, no. 4, pp. 2871–2872, 2008.
- [133] S. D. Chambreau *et al.*, "Thermal decomposition mechanisms of alkylimidazolium ionic liquids with cyano-functionalized anions," *J. Phys. Chem. A*, vol. 118, no. 47, pp. 11119– 11132, 2014.
- [134] N. Artrith, T. Morawietz, and J. Behler, "High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide," *Phys. Rev. B*, vol. 83, no. 15, p. 153 101, 2011.
- [135] J. B. Schriber, D. R. Nascimento, A. Koutsoukas, S. A. Spronk, D. L. Cheney, and C. D. Sherrill, "CLIFF: A component-based, machine-learned, intermolecular force field," *J. Chem. Phys.*, vol. 154, no. 18, p. 184 110, 2021.
- [136] T. Bereau, R. A. DiStasioJr., A. Tkatchenko, and O. A. v. Lilienfeld, "Non-covalent interactions across organic and biological subsets of chemical space: Physics-based potentials parametrized from machine learning," *J. Chem. Phys.*, vol. 148, no. 24, p. 241 706, 2018.
- [137] J. L. Sonnenberg and H. B. Schlegel, "Empirical valence bond models for reactive potential energy surfaces. II. Intramolecular proton transfer in pyridone and the Claisen reaction of allyl vinyl ether," *Mol. Phys.*, vol. 105, no. 19-22, pp. 2719–2729, 2007.
- [138] Y. T. Chang and W. H. Miller, "An empirical valence bond model for constructing global potential energy surfaces for chemical reactions of polyatomic molecular systems," J. Phys. Chem., vol. 94, no. 15, pp. 5884–5888, 1990.
- [139] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.*, vol. 77, no. 18, pp. 3865–3868, 1996.
- [140] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," J. Chem. Phys., vol. 132, no. 15, p. 154 104, 2010.
- [141] S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," *J. Comput. Chem.*, vol. 32, no. 7, pp. 1456–1465, 2011.
- [142] S. Goedecker and M. Teter, "Separable dual-space Gaussian pseudopotentials," *Phys. Rev. B*, vol. 54, no. 3, pp. 1703–1710, 1996.
- [143] P. Pulay, "Convergence acceleration of iterative sequences. the case of scf iteration," *Chem. Phys. Lett.*, vol. 73, no. 2, pp. 393–398, 1980.

- [144] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, "An overview of the Amber biomolecular simulation package," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 3, no. 2, pp. 198– 210, 2013.
- [145] B. R. Brooks *et al.*, "CHARMM: The biomolecular simulation program," *J. Comput. Chem.*, vol. 30, no. 10, pp. 1545–1614, 2009.
- [146] Lindahl, Abraham, Hess, and v. d. Spoel, GROMACS 2021 Manual, 2021.
- [147] J. G. McDaniel, E. Choi, C. Y. Son, J. R. Schmidt, and A. Yethiraj, "Ab Initio Force Fields for Imidazolium-Based Ionic Liquids," *J. Phys. Chem. B*, vol. 120, no. 28, pp. 7024–7036, 2016.
- [148] Z. Liu, S. Huang, and W. Wang, "A refined force field for molecular simulation of imidazolium-based ionic liquids," J. Phys. Chem. B, vol. 108, no. 34, pp. 12978–12989, 2004.
- [149] J. G. McDaniel and J. R. Schmidt, "Next-Generation Force Fields from Symmetry-Adapted Perturbation Theory," *Annu. Rev. Phys. Chem*, vol. 67, no. 1, pp. 467–488, 2016.
- [150] J. G. McDaniel and J. R. Schmidt, "Physically-Motivated Force Fields from Symmetry-Adapted Perturbation Theory," *J. Phys. Chem. A*, vol. 117, no. 10, pp. 2053–2066, 2013.
- [151] E. Choi, J. G. McDaniel, J. R. Schmidt, and A. Yethiraj, "First-principles, physically motivated force field for the ionic liquid [BMIM][BF4]," *J. Phys. Chem. Lett.*, vol. 5, no. 15, pp. 2670–2674, 2014.
- [152] J. G. McDaniel and A. Yethiraj, "Influence of Electronic Polarization on the Structure of Ionic Liquids," *J. Phys. Chem. Lett.*, vol. 9, no. 16, pp. 4765–4770, 2018.
- [153] J. A. Lemkul, J. Huang, B. Roux, and J. Alexander D. MacKerell, "An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications," *Chem. Rev.*, vol. 116, no. 9, pp. 4983–5013, 2016.
- [154] J. Huang, J. A. Lemkul, P. K. Eastman, and A. D. MacKerell, "Molecular dynamics simulations using the drude polarizable force field on GPUs with OpenMM: Implementation, validation, and benchmarks," *J. Comput. Chem.*, vol. 39, no. 21, pp. 1682–1689, 2018.
- [155] M. J. V. Vleet, A. J. Misquitta, A. J. Stone, and J. R. Schmidt, "Beyond Born–Mayer: Improved Models for Short-Range Repulsion in ab Initio Force Fields," J. Chem. Theory Comput., vol. 12, no. 8, pp. 3851–3870, 2016.
- [156] K. Patkowski, "Recent developments in symmetry-adapted perturbation theory," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 10, no. 3, 2020.

- [157] B. Kaduk, T. Kowalczyk, and T. Van Voorhis, "Constrained density functional theory," *Chem. Rev.*, vol. 112, no. 1, pp. 321–370, 2012.
- [158] W. D. Derricotte, "Symmetry-adapted perturbation theory decomposition of the reaction force: Insights into substituent effects involved in hemiacetal formation mechanisms," J. Phys. Chem. A, vol. 123, no. 36, pp. 7881–7891, 2019.
- [159] T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno, and C. D. Sherrill, "Levels of symmetry adapted perturbation theory (sapt). i. efficiency and performance for interaction energies," J. Chem. Phys., vol. 140, no. 9, p. 94 106, 2014.
- [160] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "Multidimensional free-energy calculations using the weighted histogram analysis method," *J. Comput. Chem.*, vol. 16, no. 11, pp. 1339–1350, 1995.
- [161] A. Hjorth Larsen *et al.*, "The atomic simulation environment A Python library for working with atoms," *J. Phys. Condens. Matter*, vol. 29, no. 27, p. 30, 2017.
- [162] J. Stoppelman, *Ase-pbnn-calculator*, 2021.
- [163] P. Eastman *et al.*, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics," *PLOS Comput. Biol.*, vol. 13, no. 7, e1005659, 2017.
- [164] Y. Luo, W. Jiang, H. Yu, A. D. MacKerell, and B. Roux, "Simulation study of ion pairing in concentrated aqueous salt solutions with a polarizable force field," *Faraday Discuss.*, vol. 160, no. 0, pp. 135–149, 2013.
- [165] R. Jinnouchi, K. Miwa, F. Karsai, G. Kresse, and R. Asahi, "On-the-Fly Active Learning of Interatomic Potentials for Large-Scale Atomistic Simulations," *J. Phys. Chem. Lett.*, vol. 11, no. 17, pp. 6946–6955, 2020.
- [166] R. K. Lindsey, L. E. Fried, N. Goldman, and S. Bastea, "Active learning for robust, highcomplexity reactive atomistic simulations," *J. Chem. Phys.*, vol. 153, no. 13, p. 134 117, 2020.
- [167] G. Fonseca, I. Poltavsky, V. Vassilev-Galindo, and A. Tkatchenko, "Improving molecular force fields across configurational space by combining supervised and unsupervised machine learning," J. Chem. Phys., vol. 154, no. 12, p. 124 102, 2021.
- [168] J. P. Stoppelman and J. G. McDaniel, "N-Heterocyclic Carbene Formation in the Ionic Liquid [EMIM+][OAc–]: Elucidating Solvation Effects with Reactive Molecular Dynamics Simulations," J. Phys. Chem. B, vol. 127, no. 23, pp. 5317–5333, 2023.

- [169] M. C. Jahnke and F. Ekkehardt Hahn, "Introduction to N-Heterocyclic Carbenes: Synthesis and Stereoelectronic Parameters," *RSC Catalysis Series*, vol. 2017-January, no. 27, pp. 1– 45, 2016.
- [170] P. Bellotti, M. Koy, M. N. Hopkinson, and F. Glorius, "Recent advances in the chemistry and applications of N-heterocyclic carbenes," *Nat. Rev. Chem.*, vol. 5, no. 10, pp. 711–725, 2021.
- [171] A. J. Arduengo, R. L. Harlow, and M. Kline, "A Stable Crystalline Carbene," J. Am. Chem. Soc., vol. 113, no. 1, pp. 361–363, 1991.
- [172] A. J. I. Arduengo, H. V. R. Dias, R. L. Harlow, and M. Kline, "Electronic stabilization of nucleophilic carbenes," J. Am. Chem. Soc., vol. 114, no. 14, pp. 5530–5534, 1992.
- [173] A. J. Arduengo, "Looking for Stable Carbenes: The Difficulty in Starting Anew," Acc. Chem. Res., vol. 32, no. 11, pp. 913–921, 1999.
- [174] K. V. Ranganath, S. Onitsuka, A. K. Kumar, and J. Inanaga, "Recent progress of Nheterocyclic carbenes in heterogeneous catalysis," *Catal. Sci. Technol.*, vol. 3, no. 9, pp. 2161–2181, 2013.
- [175] J. B. Ernst, S. Muratsugu, F. Wang, M. Tada, and F. Glorius, "Tunable Heterogeneous Catalysis: N-Heterocyclic Carbenes as Ligands for Supported Heterogeneous Ru/K-Al<sub>2</sub>O<sub>3</sub> Catalysts to Tune Reactivity and Selectivity," *J. Am. Chem. Soc.*, vol. 138, no. 34, pp. 10718–10721, 2016.
- [176] M. Koy, P. Bellotti, M. Das, and F. Glorius, "N-Heterocyclic carbenes as tunable ligands for catalytic metal surfaces," *Nat. Catal.*, vol. 4, no. 5, pp. 352–363, 2021.
- [177] H. W. Wanzlick and H. J. Schönherr, "Direct Synthesis of a Mercury Salt-Carbene Complex," Angew. Chem., Int. Ed. Engl., vol. 7, no. 2, pp. 141–142, 1968.
- [178] V. Chernyshev, E. Denisova, D. Eremin, and V. Ananikov, "The key role of R-NHC couplings (R = C, H, heteroatom) and M-NHC bond cleavage in the evolution of M/NHC complexes and formation of catalytically active species," *Chem. Sci.*, 2020.
- [179] H. Jacobsen, A. Correa, A. Poater, C. Costabile, and L. Cavallo, "Understanding the M{single bond}(NHC) (NHC = N-heterocyclic carbene) bond," *Coord. Chem. Rev.*, vol. 253, no. 5-6, pp. 687–703, 2009.
- [180] H. Shen *et al.*, "N-heterocyclic carbene coordinated metal nanoparticles and nanoclusters," *Coord. Chem. Rev.*, vol. 458, p. 214 425, 2022.
- [181] F. E. Hahn, M. C. Jahnke, F. E. Hahn, and M. C. Jahnke, "Heterocyclic Carbenes: Synthesis and Coordination Chemistry," *Angew. Chem. Int. Ed.*, vol. 47, no. 17, pp. 3122–3172, 2008.

- [182] Y. Jiang *et al.*, "N-Heterocyclic Carbene-Stabilized Ultrasmall Gold Nanoclusters in a Metal-Organic Framework for Photocatalytic CO2 Reduction," *Angew. Chem. Int. Ed.*, vol. 60, no. 32, pp. 17 388–17 393, 2021.
- [183] M. Sirignano *et al.*, "Catalytic and biological activity of silver and gold complexes stabilized by nhc with hydroxy derivatives on nitrogen atoms," *Catalysts*, vol. 12, no. 1, 2022.
- [184] M. M. Li, X. Chen, Y. Deng, and J. Lu, "Recent advances of N-heterocyclic carbenes in the applications of constructing carbo- and heterocyclic frameworks with potential biological activity," *RSC Adv.*, vol. 11, no. 60, pp. 38 060–38 078, 2021.
- [185] J. Li *et al.*, "Iridium-Catalyzed B-H Bond Insertion Reactions Using Sulfoxonium Ylides as Carbene Precursors toward α-Boryl Carbonyls," *Org. Lett.*, vol. 21, no. 22, pp. 9005– 9008, 2019.
- [186] A. Patra, F. Gelat, X. Pannecoucke, T. Poisson, T. Besset, and A. T. Biju, "Synthesis of 4-Difluoromethylquinolines by NHC-Catalyzed Umpolung of Imines," *Org. Lett.*, vol. 20, no. 4, pp. 1086–1089, 2018.
- [187] D. Enders, O. Niemeier, and A. Henseler, "Organocatalysis by N-heterocyclic carbenes," *Chem. Rev.*, vol. 107, no. 12, pp. 5606–5655, 2007.
- [188] M. Fèvre, J. Pinaud, Y. Gnanou, J. Vignolle, and D. Taton, "N-Heterocyclic carbenes (NHCs) as organocatalysts and structural components in metal-free polymer synthesis," *Chem. Soc. Rev.*, vol. 42, no. 5, pp. 2142–2172, 2013.
- [189] Z. Fu, J. Xu, T. Zhu, W. W. Y. Leong, and Y. R. Chi, "β-Carbon activation of saturated carboxylic esters through N-heterocyclic carbene organocatalysis," *Nat. Chem.*, vol. 5, no. 10, pp. 835–839, 2013.
- [190] S. Barik and A. T. Biju, "N-Heterocyclic carbene (NHC) organocatalysis using aliphatic aldehydes," *Chem. Commun.*, vol. 56, no. 99, pp. 15484–15495, 2020.
- [191] T. Ishii, K. Nagao, and H. Ohmiya, "Recent advances in N-heterocyclic carbene-based radical catalysis," *Chem. Sci.*, vol. 11, no. 22, pp. 5630–5636, 2020.
- [192] A. Vellé, A. Cebollada, R. Macías, M. Iglesias, M. Gil-Moles, and P. J. Sanz Miguel, "From Imidazole toward Imidazolium Salts and N-Heterocyclic Carbene Ligands: Electronic and Geometrical Redistribution," ACS Omega, vol. 2, no. 4, pp. 1392–1399, 2017.
- [193] J. P. Canal, T. Ramnial, D. A. Dickie, and J. A. Clyburne, "From the reactivity of N-heterocyclic carbenes to new chemistry in ionic liquids," *Chem. Commun.*, no. 17, pp. 1809–1818, 2006.

- [194] M. D. Green and T. E. Long, "Designing Imidazole-Based Ionic Liquids and Ionic Liquid Monomers for Emerging Technologies," *Polym. Rev.*, vol. 49, no. 4, pp. 291–314, 2009.
- [195] C. P. Fredlake, J. M. Crosthwaite, D. G. Hert, S. N. Aki, and J. F. Brennecke, "Thermophysical properties of imidazolium-based ionic liquids," *J. Chem. Eng. Data*, vol. 49, no. 4, pp. 954–964, 2004.
- [196] M. Galiński, A. Lewandowski, and I. Stepniak, "Ionic liquids as electrolytes," *Electrochim. Acta*, vol. 51, no. 26, pp. 5567–5580, 2006.
- [197] K. N. Marsh, J. A. Boxall, and R. Lichtenthaler, "Room temperature ionic liquids and their mixtures - A review," *Fluid Ph. Equilibria*, vol. 219, no. 1, pp. 93–98, 2004.
- [198] T. L. Amyes, S. T. Diver, J. P. Richard, F. M. Rivas, and K. Toth, "Formation and Stability of N-Heterocyclic Carbenes in Water: The Carbon Acid p*Ka* of Imidazolium Cations in Aqueous Solution," *J. Am. Chem. Soc.*, vol. 126, no. 13, pp. 4366–4374, 2004.
- [199] T. Welton, "Ionic Liquids: A Brief History.," *Biophys. Rev.*, vol. 10, no. 3, pp. 691–706, 2018.
- [200] A. P. De Los Ríos, A. Irabien, F. Hollmann, and F. J. H. Fernández, "Ionic Liquids: Green Solvents for Chemical Processing," J. Chem., vol. 2013, 2013.
- [201] N. V. Ignat'ev *et al.*, "Ionic Liquids—Advanced Reaction Media for Organic Synthesis," *Phosphorus Sulfur Silicon Relat. Elem.*, vol. 186, no. 5, pp. 1205–1216, 2011.
- [202] L. Sun, G. K. Ramesha, P. V. Kamat, and J. F. Brennecke, "Switching the reaction course of electrochemical co2 reduction with ionic liquids," *Langmuir*, vol. 30, no. 21, pp. 6302– 6308, 2014.
- [203] J. Dupont and J. Spencer, "On the Noninnocent Nature of 1,3-Dialkylimidazolium Ionic Liquids," *Angew. Chem. Int. Ed.*, vol. 43, no. 40, pp. 5296–5297, 2004.
- [204] V. Nair, S. Bindu, and V. Sreekumar, "N-Heterocyclic Carbenes: Reagents, Not Just Ligands!" *Angew. Chem. Int. Ed.*, vol. 43, no. 39, pp. 5130–5135, 2004.
- [205] B. Gorodetsky, T. Ramnial, N. R. Branda, and J. A. Clyburne, "Electrochemical reduction of an imidazolium cation: a convenient preparation of imidazol-2-ylidenes and their observation in an ionic liquid," *Chem. Commun.*, no. 17, pp. 1972–1973, 2004.
- [206] J. Liu, S. D. Chambreau, and G. L. Vaghjiani, "Dynamics simulations and statistical modeling of thermal decomposition of 1-ethyl-3-methylimidazolium dicyanamide and 1-ethyl-2,3-dimethylimidazolium dicyanamide," *J. Phys. Chem. A*, vol. 118, no. 47, pp. 11133– 11144, 2014.

- [207] J. Liu, W. Zhou, S. D. Chambreau, and G. L. Vaghjiani, "Molecular dynamics simulations and product vibrational spectral analysis for the reactions of NO<sub>2</sub> with 1-ethyl-3-methylimidazolium dicyanamide (EMIM<sup>+</sup>DCA<sup>-</sup>), 1-butyl-3methylimidazolium dicyanamide (BMIM<sup>+</sup>DCA<sup>-</sup>), and 1-allyl-3-methylimidazolium dicyanamide (AMIM<sup>+</sup>DCA<sup>-</sup>)," J. Phys. Chem. B, vol. 124, no. 21, pp. 4303–4325, 2020.
- [208] H. Rodríguez, G. Gurau, J. D. Holbrey, and R. D. Rogers, "Reaction of elemental chalcogens with imidazolium acetates to yield imidazole-2-chalcogenones: Direct evidence for ionic liquids as proto-carbenes," *Chem. Commun.*, vol. 47, no. 11, pp. 3222–3224, 2011.
- [209] F. C. Binks, G. Cavalli, M. Henningsen, B. J. Howlin, and I. Hamerton, "Investigating the mechanism through which ionic liquids initiate the polymerisation of epoxy resins," *Polymer*, vol. 139, pp. 163–176, 2018.
- [210] R. Lambert, P. Coupillaud, A. L. Wirotius, J. Vignolle, and D. Taton, "Imidazolium-Based Poly(Ionic Liquid)s Featuring Acetate Counter Anions: Thermally Latent and Recyclable Precursors of Polymer-Supported N-Heterocyclic Carbenes for Organocatalysis," *Macromol. Rapid Commun.*, vol. 37, no. 14, pp. 1143–1149, 2016.
- [211] O. Hollóczki et al., "Carbenes in ionic liquids," New J. Chem., vol. 34, no. 12, p. 3004, 2010.
- [212] C. Salvitti *et al.*, "Free N-heterocyclic carbenes from Brønsted acidic ionic liquids: Direct detection by electrospray ionization mass spectrometry," *Rapid Commun. Mass Spectrom.*, vol. 36, no. 17, e9338, 2022.
- [213] S. Gehrke and O. Hollóczki, "Are There Carbenes in N-Heterocyclic Carbene Organocatalysis?" *Angew. Chem. Int. Ed.*, vol. 56, no. 51, pp. 16395–16398, 2017.
- [214] H. Du and X. Qian, "The effects of acetate anion on cellulose dissolution and reaction in imidazolium ionic liquids," *Carbohydr. Res.*, vol. 346, no. 13, pp. 1985–1990, 2011.
- [215] F. Yan, N. R. Dhumal, and H. J. Kim, "CO2 capture in ionic liquid 1-alkyl-3methylimidazolium acetate: a concerted mechanism without carbene," *Phys. Chem. Chem. Phys.*, vol. 19, no. 2, pp. 1361–1368, 2017.
- [216] M. Brehm, H. Weber, A. S. Pensado, A. Stark, and B. Kirchner, "Proton transfer and polarity changes in ionic liquid – water mixtures: a perspective on hydrogen bonds from ab initio molecular dynamics at the example of 1-ethyl-3-methylimidazolium acetate – water mixtures—Part 1," *Phys. Chem. Chem. Phys.*, vol. 14, no. 15, pp. 5030–5044, 2012.
- [217] O. Holloczki *et al.*, "Carbene formation in ionic liquids: Spontaneous, induced, or prohibited?" *J. Phys. Chem. B*, vol. 117, no. 19, pp. 5898–5907, 2013.

- [218] H. Wang, G. Gurau, and R. D. Rogers, "Ionic liquid processing of cellulose," *Chem. Soc. Rev.*, vol. 41, no. 4, pp. 1519–1537, 2012.
- [219] A. M. da Costa Lopes, K. G. João, A. R. C. Morais, E. Bogel-Łukasik, and R. Bogel-Łukasik, "Ionic liquids as a tool for lignocellulosic biomass fractionation," *Sustain. Chem. Process.*, vol. 1, no. 1, p. 3, 2013.
- [220] A. Parviainen *et al.*, "Predicting cellulose solvating capabilities of acid–base conjugate ionic liquids," *ChemSusChem*, vol. 6, no. 11, pp. 2161–2169, 2013.
- [221] A. J. Holding, M. Heikkilä, I. Kilpeläinen, and A. W. T. King, "Amphiphilic and phaseseparable ionic liquids for biomass processing," *ChemSusChem*, vol. 7, no. 5, pp. 1422– 1434, 2014.
- [222] H. Lin and D. G. Truhlar, "QM/MM: what have we learned, where are we, and where do we go from here?" *Theor. Chem. Acc.*, vol. 117, no. 2, pp. 185–199, 2006.
- [223] R. Jacobi, F. Joerg, O. Steinhauser, and C. Schröder, "Emulating proton transfer reactions in the pseudo-protic ionic liquid 1-methylimidazolium acetate," *Phys. Chem. Chem. Phys.*, vol. 24, no. 16, pp. 9277–9285, 2022.
- [224] J. Ingenmey, S. Gehrke, and B. Kirchner, "How to Harvest Grotthuss Diffusion in Protic Ionic Liquid Electrolyte Systems," *ChemSusChem*, vol. 11, no. 12, pp. 1900–1910, 2018.
- [225] Y. L. S. Tse, C. Chen, G. E. Lindberg, R. Kumar, and G. A. Voth, "Propensity of Hydrated Excess Protons and Hydroxide Anions for the Air-Water Interface," J. Am. Chem. Soc., vol. 137, no. 39, pp. 12610–12616, 2015.
- [226] M. Galib and D. T. Limmer, "Reactive uptake of N2O5 by atmospheric aerosol is dominated by interfacial processes," *Science*, vol. 371, no. 6532, pp. 921–925, 2021.
- [227] I. Benjamin, "Chemical Reactions and Solvation at Liquid Interfaces: A Microscopic Perspective," *Chem. Rev.*, vol. 96, no. 96, pp. 1449–1476, 1996.
- [228] P. Dubois *et al.*, "Ionic liquid droplet as e-microreactor," *Anal. Chem.*, vol. 78, no. 14, pp. 4909–4917, 2006.
- [229] M. López-Pastor, A. Domínguez-Vidal, M. J. Ayora-Cañada, T. Laurell, M. Valcárcel, and B. Lendl, "Containerless reaction monitoring in ionic liquids by means of Raman microspectroscopy," *Lab Chip*, vol. 7, no. 1, pp. 126–132, 2006.
- [230] M. Girod, E. Moyano, D. I. Campbell, and R. G. Cooks, "Accelerated bimolecular reactions in microdroplets studied by desorption electrospray ionization mass spectrometry," *Chem. Sci.*, vol. 2, no. 3, pp. 501–510, 2011.

- [231] J. K. Lee, S. Kim, H. G. Nam, and R. N. Zare, "Microdroplet fusion mass spectrometry for fast reaction kinetics," *Proc. Natl. Acad. Sci. U.S.A*, vol. 112, no. 13, pp. 3898–3903, 2015.
- [232] Y. Li, X. Yan, and R. G. Cooks, "The role of the interface in thin film and droplet accelerated reactions studied by competitive substituent effects," *Angew. Chem. Int. Ed.*, vol. 55, no. 10, pp. 3433–3437, 2016.
- [233] X. Yan, R. M. Bain, and R. G. Cooks, "Organic reactions in microdroplets: Reaction acceleration revealed by mass spectrometry," *Angew. Chem. Int. Ed.*, vol. 55, no. 42, pp. 12960– 12972, 2016.
- [234] R. M. Bain, S. Sathyamoorthi, and R. N. Zare, ""on-droplet" chemistry: The cycloaddition of diethyl azodicarboxylate and quadricyclane," *Angew. Chem. Int. Ed.*, vol. 56, no. 47, pp. 15083–15087, 2017.
- [235] S. Mondal, S. Acharya, R. Biswas, B. Bagchi, and R. N. Zare, "Enhancement of reaction rate in small-sized droplets: A combined analytical and simulation study," *J. Chem. Phys.*, vol. 148, no. 24, p. 244 704, 2018.
- [236] Z. Wei, Y. Li, R. G. Cooks, and X. Yan, "Accelerated reaction kinetics in microdroplets: Overview and recent developments," *Annu. Rev. Phys. Chem.*, vol. 71, no. 1, pp. 31–51, 2020.
- [237] J. M. Swanson *et al.*, "Proton solvation and transport in aqueous and biomolecular systems: Insights from computer simulations," *J. Phys. Chem. B*, vol. 111, no. 17, pp. 4300–4314, 2007.
- [238] R. Vuilleumier and D. Borgis, "An Extended Empirical Valence Bond Model for Describing Proton Mobility in Water," *Isr. J. Chem.*, vol. 39, no. 3-4, pp. 457–467, 1999.
- [239] C. Li and G. A. Voth, "Accurate and Transferable Reactive Molecular Dynamics Models from Constrained Density Functional Theory," J. Phys. Chem. B, vol. 125, no. 37, pp. 10471–10480, 2021.
- [240] G. A. Voth, "Computer simulation of proton solvation and transport in aqueous and biomolecular systems," *Acc. Chem. Res.*, vol. 39, no. 2, pp. 143–150, 2006.
- [241] J. Huang and A. D. Mackerell, "CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data," *J. Comput. Chem.*, vol. 34, no. 25, pp. 2135– 2145, 2013.
- [242] J. G. McDaniel and J. R. Schmidt, "First-Principles Many-Body Force Fields from the Gas Phase to Liquid: A "Universal" Approach," J. Phys. Chem. B, vol. 118, no. 28, pp. 8042– 8053, 2014.

- [243] K. Y. Baek, J. H. Jo, J. H. Moon, J. Yoon, and J. Y. Lee, "Systematic Strategy for Designing Immidazolium Containing Precursors To Produce N-Heterocyclic Carbenes: A DFT Study," J. Org. Chem., vol. 80, no. 3, pp. 1878–1886, 2015.
- [244] M. Bayat, A. Sedghi, L. Ebrahimkhani, and S. J. Sabounchei, "N-Heterocyclic carbene or phosphorus ylide: which one forms a stronger bond with group 11 metals? A theoretical study," *Dalton Trans.*, vol. 46, no. 1, pp. 207–220, 2017.
- [245] J. C. Bernhammer, G. Frison, and H. V. Huynh, "Electronic Structure Trends in N-Heterocyclic Carbenes (NHCs) with Varying Number of Nitrogen Atoms and NHC Transition-Metal Bond Properties," *Chem. Eur. J.*, vol. 19, no. 38, pp. 12892–12905, 2013.
- [246] M. Bonomi *et al.*, "Promoting transparency and reproducibility in enhanced molecular simulations," *Nat. Methods*, vol. 16, no. 8, pp. 670–673, 2019.
- [247] I. C. Yeh and M. L. Berkowitz, "Ewald summation for systems with slab geometry," J. *Chem. Phys.*, vol. 111, no. 7, pp. 3155–3162, 1999.
- [248] P. Jungwirth and D. J. Tobias, "Specific ion effects at the air/water interface," *Chem. Rev.*, vol. 106, no. 4, pp. 1259–1281, 2006.
- [249] M. Zheng and M. P. Waller, "Adaptive quantum mechanics/molecular mechanics methods," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 6, no. 4, pp. 369–385, 2016.
- [250] S. S. Bystrov *et al.*, "Translational Diffusion in a Set of Imidazolium-Based Ionic Liquids [bmim]+A- and Their Mixtures with Water," *J. Phys. Chem. B*, vol. 123, no. 43, pp. 9187– 9197, 2019.
- [251] M. Brehm, M. Thomas, S. Gehrke, and B. Kirchner, "TRAVIS—A free analyzer for trajectories from molecular simulation," *J. Chem. Phys.*, vol. 152, no. 16, p. 164 105, 2020.
- [252] E. F. Pettersen *et al.*, "UCSF ChimeraX: Structure visualization for researchers, educators, and developers," *Protein Science*, vol. 30, no. 1, pp. 70–82, 2021.
- [253] T. Yan, S. Li, W. Jiang, X. Gao, B. Xiang, and G. A. Voth, "Structure of the liquid-vacuum interface of room-temperature ionic liquids: A molecular dynamics study," *J. Phys. Chem. B*, vol. 110, no. 4, pp. 1800–1806, 2006.
- [254] J. K. Konieczny and B. Szefczyk, "Structure of alkylimidazolium-based ionic liquids at the interface with vacuum and water-a molecular dynamics study," *J. Phys. Chem. B*, vol. 119, no. 9, pp. 3795–3807, 2015.
- [255] M. Brehm and D. Sebastiani, "Simulating structure and dynamics in small droplets of 1ethyl-3-methylimidazolium acetate," *J. Chem. Phys.*, vol. 148, no. 19, p. 193 802, 2018.

- [256] J. G. McDaniel and A. Yethiraj, "Understanding the properties of ionic liquids: Electrostatics, structure factors, and their sum rules," *J. Phys. Chem. B*, vol. 123, no. 16, pp. 3499– 3512, 2019.
- [257] J. P. Stoppelman and J. G. McDaniel, "Proton Transport in [BMIM+][BF4-] / Water Mixtures Near the Percolation Threshold," *J. Phys. Chem. B*, vol. 2020, p. 3, 2020.
- [258] S. Peighambardoust, S. Rowshanzamir, and M. Amjadi, "Review of the Proton Exchange Membranes for Fuel Cell Applications," *Int. J. Hydrogen Energy*, vol. 35, no. 17, pp. 9349– 9384, 2010.
- [259] D. U. Eberle and D. R. von Helmolt, "Sustainable Transportation Based on Electric Vehicle Concepts: A Brief Overview," *Energy Environ. Sci.*, vol. 3, no. 6, p. 689, 2010.
- [260] U. Eberle, B. Müller, and R. von Helmolt, "Fuel Cell Electric Vehicles and Hydrogen Infrastructure: Status 2012," *Energy Environ. Sci.*, vol. 5, no. 10, p. 8780, 2012.
- [261] S. J. Hamrock and M. A. Yandrasits, "Proton Exchange Membranes for Fuel Cell Applications," J. Macromol. Sci. Polymer Rev., vol. 46, no. 3, pp. 219–244, 2006.
- [262] M. Casciola, G. Alberti, M. Sganappa, and R. Narducci, "On the Decay of Nafion Proton Conductivity at High Temperature and Relative Humidity," *J. Power Sources*, vol. 162, no. 1, pp. 141–145, 2006.
- [263] Y. Shao, G. Yin, Z. Wang, and Y. Gao, "Proton Exchange Membrane Fuel Cell from Low Temperature to High Temperature: Material Challenges," *J. Power Sources*, vol. 167, no. 2, pp. 235–242, 2007.
- [264] C. Schmidt, T. Glück, and G. Schmidt-Naake, "Modification of Nafion Membranes by Impregnation with Ionic Liquids," *Chem. Eng. Technol.*, vol. 31, no. 1, pp. 13–22, 2008.
- [265] S. Authayanun, K. Im-orb, and A. Arpornwichanop, "A review of the development of high temperature proton exchange membrane fuel cells," *Chinese J. Catal.*, vol. 36, no. 4, pp. 473–483, 2015.
- [266] A. K. Santra and D. W. Goodman, "Catalytic oxidation of CO by platinum group metals: From ultrahigh vacuum to elevated pressures," *Electrochim. Acta*, vol. 47, no. 22-23, pp. 3595–3609, 2002.
- [267] Q. Li, R. He, J. O. Jensen, and N. J. Bjerrum, "Approaches and Recent Development of Polymer Electrolyte Membranes for Fuel Cells Operating above 100 °C," *Chem. Mater.*, vol. 15, no. 26, pp. 4896–4915, 2003.

- [268] M. Armand, F. Endres, D. R. MacFarlane, H. Ohno, and B. Scrosati, "Ionic-Liquid Materials for the Electrochemical Challenges of the Future," *Nat. Mater.*, vol. 8, no. 8, pp. 621– 629, 2009.
- [269] T. Welton, "Room-Temperature Ionic Liquids. Solvents for Synthesis and Catalysis," *Chem. Rev.*, vol. 99, no. 8, pp. 2071–2083, 1999.
- [270] J. P. Hallett and T. Welton, "Room-Temperature Ionic Liquids: Solvents for Synthesis and Catalysis. 2," *Chem. Rev.*, vol. 111, no. 5, pp. 3508–3576, 2011.
- [271] R. Sood *et al.*, "Proton Conducting Ionic Liquid Doped Nafion Membranes: Nano-Structuration, Transport Properties and Water Sorption," *J. Phys. Chem. C*, vol. 116, no. 46, pp. 24413–24423, 2012.
- [272] J. Schauer *et al.*, "Ion-Conductive Polymer Membranes Containing 1-Butyl-3methylimidazolium Trifluoromethanesulfonate and 1-Ethylimidazolium Trifluoromethanesulfonate," J. Membr. Sci., vol. 367, no. 1-2, pp. 332–339, 2011.
- [273] L. Yu *et al.*, "Conductivity, Spectroscopic, and Computational Investigation of H<sub>3</sub>O<sup>+</sup> Solvation in Ionic Liquid BMIBF<sub>4</sub>," *J. Phys. Chem. B*, vol. 117, no. 23, pp. 7057–7064, 2013.
- [274] Y. Nakata, K. Kohara, K. Matsumoto, and R. Hagiwara, "Thermal Properties of Ionic Liquid + Water Binary Systems Applied to Heat Pipes," J. Chem. Eng. Data, vol. 56, no. 5, pp. 1840–1846, 2011.
- [275] N. Merkel, C. Weber, M. Faust, and K. Schaber, "Influence of Anion and Cation on the Vapor Pressure of Binary Mixtures of Water + Ionic Liquid and on the Thermal Stability of the Ionic Liquid," *Fluid Ph. Equilibria*, vol. 394, pp. 29–37, 2015.
- [276] A. Stoppa, J. Hunger, and R. Buchner, "Conductivities of Binary Mixtures of Ionic Liquids with Polar Solvents," *J. Chem. Eng. Data*, vol. 54, no. 2, pp. 472–479, 2009.
- [277] J. N. Canongia Lopes *et al.*, "Polarity, Viscosity, and Ionic Conductivity of Liquid Mixtures Containing [C<sub>4</sub>C<sub>1</sub>im][Ntf<sub>2</sub>] and a Molecular Component," *J. Phys. Chem. B*, vol. 115, no. 19, pp. 6088–6099, 2011.
- [278] J. G. McDaniel and C. Y. Son, "Ion Correlation and Collective Dynamics in BMIM/BF ¡sub¿4¡/sub¿ -Based Organic Electrolytes: From Dilute Solutions to the Ionic Liquid Limit," J. Phys. Chem. B, vol. 122, no. 28, pp. 7154–7169, 2018.
- [279] W. Choi, Z. W. Ulissi, S. F. Shimizu, D. O. Bellisario, M. D. Ellison, and M. S. Strano, "Diameter-Dependent Ion Transport Through the Interior of Isolated Single-Walled Carbon Nanotubes," *Nat. Commun.*, vol. 4, no. 1, p. 2397, 2013.

- [280] S. H. Lee and J. C. Rasaiah, "Proton Transfer and the Diffusion of H<sup>+</sup> and OH<sup>-</sup> Ions Along Water Wires," *J. Chem. Phys.*, vol. 139, no. 12, p. 124 507, 2013.
- [281] Z. Cao, Y. Peng, T. Yan, S. Li, A. Li, and G. A. Voth, "Mechanism of Fast Proton Transport along One-Dimensional Water Chains Confined in Carbon Nanotubes," J. Am. Chem. Soc., vol. 132, no. 33, pp. 11 395–11 397, 2010.
- [282] C. Dellago, M. M. Naor, and G. Hummer, "Proton Transport through Water-Filled Carbon Nanotubes," *Phys. Rev. Lett.*, vol. 90, no. 10, p. 105 902, 2003.
- [283] C. Dellago and G. Hummer, "Kinetics and Mechanism of Proton Transport across Membrane Nanopores," *Phys. Rev. Lett.*, vol. 97, no. 24, p. 245 901, 2006.
- [284] M. L. Brewer, U. W. Schmitt, and G. A. Voth, "The Formation and Dynamics of Proton Wires in Channel Environments.," *Biophys. J.*, vol. 80, no. 4, pp. 1691–702, 2001.
- [285] N. Agmon, "The Grotthuss Mechanism," *Chem. Phys. Lett.*, vol. 244, no. 5-6, pp. 456–462, 1995.
- [286] K. D. Kreuer, "Proton conductivity: Materials and applications," *Chem. Mater.*, vol. 8, no. 3, pp. 610–641, 1996.
- [287] K. D. Kreuer, "On the complexity of proton conduction phenomena," *Solid State Ion.*, vol. 136-137, pp. 149–160, 2000.
- [288] J. G. McDaniel and A. Verma, "On the Miscibility and Immiscibility of Ionic Liquids and Water," *J. Phys. Chem. B*, vol. 123, no. 25, pp. 5343–5356, 2019.
- [289] A. Verma, J. P. Stoppelman, and J. G. McDaniel, "Tuning Water Networks via Ionic Liquid/Water Mixtures," *Int. J. Mol. Sci.*, vol. 21, no. 2, p. 403, 2020.
- [290] J. G. Huddleston, A. E. Visser, W. M. Reichert, H. D. Willauer, G. A. Broker, and R. D. Rogers, "Characterization and comparison of hydrophilic and hydrophobic room temperature ionic liquids incorporating the imidazolium cation," *Green Chem.*, vol. 3, no. 4, pp. 156–164, 2001.
- [291] K. E. Gutowski *et al.*, "Controlling the aqueous miscibility of ionic liquids: Aqueous biphasic systems of water-miscible ionic liquids and water-structuring salts for recycle, metathesis, and separations," *J. Am. Chem. Soc.*, vol. 125, no. 22, pp. 6632–6633, 2003.
- [292] L. Cammarata, S. G. Kazarian, P. A. Salter, and T. Welton, "Molecular states of water in room temperature ionic liquids," *Phys. Chem. Chem. Phys.*, vol. 3, no. 23, pp. 5192–5200, 2001.

- [293] Y. Kohno and H. Ohno, "Ionic liquid/water mixtures: From hostility to conciliation," *ChemComm*, vol. 48, no. 57, pp. 7119–7130, 2012.
- [294] K. R. Seddon, A. Stark, and M. J. Torres, "Influence of chloride, water, and organic solvents on the physical properties of ionic liquids," *Pure Appl. Chem.*, vol. 72, no. 12, pp. 2275– 2287, 2000.
- [295] O. Markovitch, H. Chen, S. Izvekov, F. Paesani, G. A. Voth, and N. Agmon, "Special Pair Dance and Partner Selection: Elementary Steps in Proton Transport in Liquid Water," J. Phys. Chem. B, vol. 112, no. 31, pp. 9456–9466, 2008.
- [296] M. Tuckerman, K. Laasonen, M. Sprik, and M. Parrinello, "Ab initio molecular dynamics simulation of the solvation and transport of h<sub>3</sub>o<sup>+</sup> and oh<sup>-</sup> ions in water," J. Phys. Chem., vol. 99, no. 16, pp. 5749–5752, 1995.
- [297] R. Kumar, C. Knight, and G. Voth, "Exploring the behavior of the hydrated excess proton at hydrophobic interfaces," *Farad. Discuss.*, vol. 167, p. 263, 2013.
- [298] Z. Cao, R. Kumar, Y. Peng, and G. Voth, "Hydrated proton structure and diffusion at platinum surfaces," *J. Phys. Chem. C*, vol. 119, p. 14675, 2015.
- [299] J. M. J. Swanson *et al.*, "Proton Solvation and Transport in Aqueous and Biomolecular Systems: Insights from Computer Simulations," *J. Phys. Chem. B*, 2007.
- [300] Y. Wu and G. A. Voth, "A Computer Simulation Study of the Hydrated Proton in a Synthetic Proton Channel," *Biophys. J.*, vol. 85, no. 2, pp. 864–875, 2003.
- [301] J. G. McDaniel and A. Yethiraj, "Importance of hydrophobic traps for proton diffusion in lyotropic liquid crystals," *J. Chem. Phys.*, vol. 144, no. 9, p. 094 705, 2016.
- [302] J. G. McDaniel and A. Yethiraj, "Proton Diffusion through Bilayer Pores," J. Phys. Chem. B, vol. 121, no. 39, pp. 9247–9259, 2017.
- [303] K. D. Kreuer, S. J. Paddison, E. Spohr, and M. Schuster, "Transport in Proton Conductors for Fuel-Cell Applications: Simulations, Elementary Reactions, and Phenomenology," *Chem. Rev.*, vol. 104, no. 10, pp. 4637–4678, 2004.
- [304] K. Park, W. Lin, and F. Paesani, "A Refined MS-EVB Model for Proton Transport in Aqueous Environments," *J. Phys. Chem. B*, vol. 116, no. 1, pp. 343–352, 2012.
- [305] D. D. Borges *et al.*, "Proton Transport in a Highly Conductive Porous Zirconium-Based Metal-Organic Framework: Molecular Insight," *Angew. Chem. Int. Ed.*, vol. 55, no. 12, pp. 3919–3924, 2016.

- [306] M. K. Petersen, S. S. Iyengar, T. J. Day, and G. A. Voth, "The hydrated proton at the water liquid/vapor interface," *J. Phys. Chem. B*, vol. 108, no. 39, pp. 14804–14806, 2004.
- [307] D. Marx, M. E. Tuckerman, J. Hutter, and M. Parrinello, "The Nature of the Hydrated Excess Proton in Water," *Nature*, vol. 397, no. 6720, pp. 601–604, 1999.
- [308] R. Biswas and G. A. Voth, "Role of Solvation Structure in the Shuttling of the Hydrated Excess Proton," *J. Chem. Sci.*, vol. 129, no. 7, pp. 1045–1051, 2017.
- [309] M. K. Petersen, A. J. Hatt, and G. A. Voth, "Orientational Dynamics of Water in the Nafion Polymer Electrolyte Membrane and Its Relationship to Proton Transport," *J. Phys. Chem. B*, vol. 112, no. 26, pp. 7754–7761, 2008.
- [310] R. Kumar, R. Christie, and K. Jordan, "A modified msevb force field for protonated water clusters," *J. Phys. Chem. B*, vol. 113, p. 4111, 2009.
- [311] U. W. Schmitt and G. A. Voth, "The Computer Simulation of Proton Transport in Water," *J. Chem. Phys.*, vol. 111, no. 20, pp. 9361–9381, 1999.
- [312] J. Xu, S. Izvekov, and G. A. Voth, "Structure and Dynamics of Concentrated Hydrochloric Acid Solutions," *J. Phys. Chem. B*, vol. 114, no. 29, pp. 9555–9562, 2010.
- [313] S. Feng and G. A. Voth, "Proton Solvation and Transport in Hydrated Nafion," J. Phys. Chem. B, vol. 115, no. 19, pp. 5903–5912, 2011.
- [314] S. Feng, J. Savage, and G. A. Voth, "Effects of Polymer Morphology on Proton Solvation and Transport in Proton-Exchange Membranes," J. Phys. Chem. C, vol. 116, no. 36, pp. 19104–19116, 2012.
- [315] A. M. Smondyrev and G. A. Voth, "Molecular Dynamics Simulation of Proton Transport Througn the Influenza A Virus M2 Channel," *Biophys. J.*, vol. 83, no. 4, pp. 1987–1996, 2002.
- [316] H. Chen, B. Ilan, Y. Wu, F. Zhu, K. Schulten, and G. A. Voth, "Charge Delocalization in Proton Channels, I: The Aquaporin Channels and Proton Blockage," *Biophys. J.*, vol. 92, no. 1, pp. 46–60, 2007.
- [317] Y. Wu, B. Ilan, and G. A. Voth, "Charge Delocalization in Proton Channels, II: The Synthetic LS2 Channel and Proton Selectivity," *Biophys. J.*, vol. 92, no. 1, pp. 61–69, 2007.
- [318] F. Wang, S. Izvekov, and G. A. Voth, "Unusual "Amphiphilic" Association of Hydrated Protons in Strong Acid Solution," J. Am. Chem. Soc., vol. 130, no. 10, pp. 3120–3126, 2008.

- [319] W. Grot, "Commercial Membrane Types," in *Fluorinated Ionomers*, Elsevier, 2011, pp. 185–199.
- [320] M. K. Petersen, F. Wang, N. P. Blake, H. Metiu, and G. A. Voth, "Excess Proton Solvation and Delocalization in a Hydrophilic Pocket of the Proton Conducting Polymer Membrane Nafion," *J. Phys. Chem. B*, vol. 109, no. 9, pp. 3727–3730, 2005.
- [321] M. K. Petersen and G. A. Voth, "Characterization of the Solvation and Transport of the Hydrated Proton in the Perfluorosulfonic Acid Membrane nafion," *J. Phys. Chem. B*, vol. 110, no. 37, pp. 18594–18600, 2006.
- [322] R. Jorn, J. Savage, and G. A. Voth, "Proton Conduction in Exchange Membranes across Multiple Length Scales," *Acc. Chem. Res.*, vol. 45, no. 11, pp. 2002–2010, 2012.
- [323] J. A. Elliott and S. J. Paddison, "Modelling of morphology and proton transport in PFSA membranes," *Phys. Chem. Chem. Phys.*, vol. 9, no. 21, pp. 2602–2618, 2007.
- [324] M. Eikerling and A. A. Kornyshev, "Proton transfer in a single pore of a polymer electrolyte membrane," *J. Electroanal. Chem.*, vol. 502, no. 1-2, pp. 1–14, 2001.
- [325] C. Arntsen, J. Savage, Y.-L. S. Tse, and G. A. Voth, "Simulation of Proton Transport in Proton Exchange Membranes with Reactive Molecular Dynamics," *Fuel Cells*, vol. 16, no. 6, pp. 695–703, 2016.
- [326] Y. L. S. Tse, A. M. Herring, K. Kim, and G. A. Voth, "Molecular Dynamics Simulations of Proton Transport in 3M and Nafion Perfluorosulfonic Acid Membranes," *J. Phys. Chem. C*, vol. 117, no. 16, pp. 8079–8091, 2013.
- [327] J. Savage, Y. L. S. Tse, and G. A. Voth, "Proton Transport Mechanism of Perfluorosulfonic Acid Membranes," *J. Phys. Chem. C*, vol. 118, no. 31, pp. 17436–17445, 2014.
- [328] H. Liu, S. Cavaliere, D. J. Jones, J. Rozière, and S. J. Paddison, "Morphology of Hydrated Nafion through a Quantitative Cluster Analysis: A Case Study Based on Dissipative Particle Dynamics Simulations," J. Phys. Chem. C, vol. 122, no. 24, pp. 13 130–13 139, 2018.
- [329] D. Stauffer, "Scaling Theory of Percolation Clusters," *Phys. Rep.*, vol. 54, no. 1, pp. 1–74, 1979.
- [330] K. A. Zweig, "Graph Theory, Social Network Analysis, and Network Science," in *Network Analysis Literacy*, 2016, pp. 23–55.
- [331] C. J. Stam and J. C. Reijneveld, "Graph Theoretical Analysis of Complex Networks in the Brain," *Nonlinear Biomed. Phys.*, vol. 1, no. 1, p. 3, 2007.

- [332] L. Metcalf and W. Casey, "Chapter 5 Graph theory," in *Cybersecurity and Applied Mathematics*, 2016, pp. 67–94, ISBN: 9780128044520.
- [333] J.-H. Choi, H. Lee, H. R. Choi, and M. Cho, "Graph Theory and Ion and Molecular Aggregation in Aqueous Solutions," *Annu. Rev. Phys. Chem.*, vol. 69, no. 1, pp. 125–149, 2018.
- [334] S. Kim, H. Kim, J.-H. Choi, and M. Cho, "Ion Aggregation in High Salt Solutions: Ion Network Versus Ion Cluster," *J. Chem. Phys.*, vol. 141, no. 12, p. 124510, 2014.
- [335] H. Lee, J.-H. Choi, P. K. Verma, and M. Cho, "Spectral Graph Analyses of Water Hydrogen-Bonding Network and Osmolyte Aggregate Structures in Osmolyte–Water Solutions," J. Phys. Chem. B, vol. 119, no. 45, pp. 14402–14412, 2015.
- [336] P. Eastman *et al.*, "OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics," *PLOS Comput. Biol.*, pp. 1–17, 2017.
- [337] Y. Wu, H. L. Tepper, and G. A. Voth, "Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties," *J. Chem. Phys.*, vol. 124, no. 2, 2006.
- [338] C. Y. Son, J. G. McDaniel, J. R. Schmidt, Q. Cui, and A. Yethiraj, "First-Principles United Atom Force Field for the Ionic Liquid BMIM<sup>+</sup>BF4<sup>-</sup>: An Alternative to Charge Scaling," *J. Phys. Chem. B*, vol. 120, no. 14, pp. 3560–3568, 2016.
- [339] J. G. McDaniel, "Polarization Effects in Binary [BMIM+][BF4-]/1,2-Dichloroethane, Acetone, Acetonitrile, and Water Electrolytes," J. Phys. Chem. B, vol. 122, no. 15, pp. 4345– 4355, 2018.
- [340] A. Heßelmann, G. Jansen, and M. Schütz, "Density-Functional Theory-Symmetry-Adapted Intermolecular Perturbation Theory with Density Fitting: A New Efficient Method to Study Intermolecular Interaction Energies," J. Chem. Phys., vol. 122, no. 1, p. 014 103, 2005.
- [341] R. M. Parrish *et al.*, "Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability," *J. Chem. Theory Comput.*, vol. 13, no. 7, pp. 3185–3197, 2017.
- [342] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring Network Structure, Dynamics, and Function using NetworkX," in *Proceedings of the 7th Python in Science Conference*, G. Varoquaux, T. Vaught, and J. Millman, Eds., Pasadena, CA USA, 2008, pp. 11–15.
- [343] A. Luzar and D. Chandler, "Structure and Hydrogen Bond Dynamics of Water–Dimethyl Sulfoxide Mixtures by Computer Simulations," J. Chem. Phys., vol. 98, no. 10, pp. 8160– 8173, 1993.

- [344] A. Luzar, "Resolving the Hydrogen Bond Dynamics Conundrum," J. Chem. Phys., vol. 113, no. 23, pp. 10663–10675, 2000.
- [345] V. Marques Leite dos Santos, F. Brady Moreira, and R. L. Longo, "Topology of the Hydrogen Bond Networks in Liquid Water at Room and Supercritical Conditions: A Small-World Structure," *Chem. Phys. Lett.*, vol. 390, no. 1-3, pp. 157–161, 2004.
- [346] M. I. Oh, M. Gupta, C. I. Oh, and D. F. Weaver, "Understanding the Effect of Nanoconfinement on the Structure of Water Hydrogen Bond Networks," *Phys. Chem. Chem. Phys.*, vol. 21, no. 47, pp. 26237–26250, 2019.
- [347] A. Geiger, F. H. Stillinger, and A. Rahman, "Aspects of the Percolation Process for Hydrogen-Bond Networks in Water," *J. Chem. Phys.*, vol. 70, no. 9, pp. 4185–4193, 1979.
- [348] H. Neitsch and S. H. L. Klapp, "Percolation Transition of Short-Ranged Square Well Fluids in Bulk and Confinement," *J. Chem. Phys.*, vol. 138, no. 6, p. 064 904, 2013.
- [349] J.-P. Hovi and A. Aharony, "Scaling and universality in the spanning probability for percolation," *Phys. Rev. E*, vol. 53, no. 1, pp. 235–253, 1996.
- [350] T. Edvinsson, P. J. Råsmark, and C. Elvingson, "Cluster Identification and Percolation Analysis Using a Recursive Algorithm," *Mol. Simul.*, vol. 23, no. 3, pp. 169–190, 1999.
- [351] S. S. Zhang, "A Review on Electrolyte Additives for Lithium-Ion Batteries," J. Power Sources, vol. 162, no. 2 SPEC. ISS. Pp. 1379–1394, 2006.
- [352] K. L. Van Aken, M. Beidaghi, and Y. Gogotsi, "Formulation of Ionic-Liquid Electrolyte To Expand the Voltage Window of Supercapacitors," *Angew. Chem. Int. Ed.*, vol. 54, no. 16, pp. 4806–4809, 2015.
- [353] L. Cheng *et al.*, "Accelerating Electrolyte Discovery for Energy Storage with High-Throughput Screening," *J. Phys. Chem. Lett.*, vol. 6, no. 2, pp. 283–291, 2015.
- [354] K. A. Mauritz and R. B. Moore, "State of Understanding of Nafion," *Chem. Rev.*, vol. 104, no. 10, pp. 4535–4585, 2004.
- [355] O. N. Kalugin, I. V. Voroshylova, A. V. Riabchunova, E. V. Lukinova, and V. V. Chaban, "Conductometric Study of Binary Systems Based on Ionic Liquids and Acetonitrile in a Wide Concentration Range," *Electrochim. Acta*, vol. 105, pp. 188–199, 2013.
- [356] W. Li, Z. Zhang, B. Han, S. Hu, Y. Xie, and G. Yang, "Effect of Water and Organic Solvents on the Ionic Dissociation of Ionic Liquids," *J. Phys. Chem. B*, 2007.
- [357] J. Cascão, W. Silva, A. S. Ferreira, and E. J. Cabrita, "Ion pair and solvation dynamics of [Bmim][BF<sub>4</sub>] + water system," *Magn. Resn. Chem.*, vol. 56, no. 2, pp. 127–139, 2018.

- [358] M. R. Harpham, B. M. Ladanyi, N. E. Levinger, and K. W. Herwig, "Water motion in reverse micelles studied by quasielastic neutron scattering and molecular dynamics simulations," J. Chem. Phys., vol. 121, no. 16, pp. 7855–7868, 2004.
- [359] Y. Yoshimura *et al.*, "Confirmation of Local Water Structure Confined in Ionic Liquids Using H/D Exchange," J. Mol. Liq., vol. 286, 2019.
- [360] X. Zhong, Z. Fan, Z. Liu, and D. Cao, "Local Structure Evolution and its Connection to Thermodynamic and Transport Properties of 1-Butyl-3-methylimidazolium Tetrafluoroborate and Water Mixtures by Molecular Dynamics Simulations," *J. Phys. Chem. B*, vol. 116, no. 10, pp. 3249–3263, 2012.
- [361] Y. Peng, J. M. J. Swanson, S.-g. Kang, R. Zhou, and G. A. Voth, "Hydrated Excess Protons Can Create Their Own Water Wires," J. Phys. Chem. B, vol. 119, no. 29, pp. 9212–9218, 2015.
- [362] H. L. Tepper and G. A. Voth, "Protons May Leak through Pure Lipid Bilayers via a Concerted Mechanism," *Biophys. J.*, vol. 88, no. 5, pp. 3095–3108, 2005.
- [363] C. Knight and G. A. Voth, "The Curious Case of the Hydrated Proton," Acc. Chem. Res., vol. 45, no. 1, pp. 101–109, 2012.
- [364] M. K. Petersen and G. A. Voth, "Amphiphilic Character of the Hydrated Proton in Methanol-Water Solutions," *J. Phys. Chem. B*, vol. 110, no. 14, pp. 7085–7089, 2006.
- [365] A. A. Gurtovenko and I. Vattulainen, "Pore Formation Coupled to Ion Transport Through Lipid Membranes as Induced by Transmembrane Ionic Charge Imbalance: Atomistic Molecular Dynamics Study," J. Am. Chem. Soc., vol. 127, no. 50, pp. 17 570–17 571, 2005.
- [366] H. Leontiadou, A. E. Mark, and S. J. Marrink, "Ion Transport Across Transmembrane Pores," *Biophys. J.*, vol. 92, no. 12, pp. 4209–4215, 2007.
- [367] J. P. Stoppelman, J. G. McDaniel, and M. T. Cicerone, "Excitations follow (or lead?) density scaling in propylene carbonate," *J. Chem. Phys.*, vol. 157, no. 20, p. 204 506, 2022.
- [368] H. Miyagawa, Y. Hiwatari, B. Bernu, and J. P. Hansen, "Molecular dynamics study of binary soft-sphere mixtures: Jump motions of atoms in the glassy state," J. Chem. Phys., vol. 88, no. 6, pp. 3879–3886, 1988.
- [369] C. Bennemann, C. Donati, J. Baschnagel, and S. C. Glotzer, "Growing range of correlated motion in a polymer melt on cooling towards the glass transition," *Nature*, vol. 399, no. 6733, pp. 246–249, 1999.
- [370] T. F. Middleton and D. J. Wales, "Energy landscapes of some model glass formers," *Phys. Rev. B*, vol. 64, no. 2, p. 024 205, 2001.

- [371] D. Chandler and J. P. Garrahan, "Dynamics on the Way to Forming Glass: Bubbles in Space-Time," *Annu. Rev. Phys. Chem.*, vol. 61, no. 1, pp. 191–217, 2010.
- [372] A. S. Keys, L. O. Hedges, J. P. Garrahan, S. C. Glotzer, and D. Chandler, "Excitations Are Localized and Relaxation Is Hierarchical in Glass-Forming Liquids," *Phys. Rev. X*, vol. 1, no. 2, p. 021013, 2011.
- [373] Y. S. Elmatad, D. Chandler, and J. P. Garrahan, "Corresponding States of Structural Glass Formers," *J. Phys. Chem. B*, vol. 113, no. 16, pp. 5563–5567, 2009.
- [374] Y. Jung, J. P. Garrahan, and D. Chandler, "Excitation lines and the breakdown of Stokes-Einstein relations in supercooled liquids," *Phys. Rev. E*, vol. 69, no. 6, p. 061 205, 2004.
- [375] M. Merolle, J. P. Garrahan, and D. Chandler, "Space–time thermodynamics of the glass transition," *Proc Natl Acad Sci U S A*, vol. 102, no. 31, p. 10837, 2005.
- [376] L. O. Hedges, L. Maibaum, D. Chandler, and J. P. Garrahan, "Decoupling of exchange and persistence times in atomistic models of glass formers," *J. Chem. Phys.*, vol. 127, no. 21, p. 211 101, 2007.
- [377] A. S. Keys, J. P. Garrahan, and D. Chandler, "Calorimetric glass transition explained by hierarchical dynamic facilitation," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, no. 12, pp. 4482– 4487, 2013.
- [378] A. Tolle, "Neutron scattering studies of the model glass former ortho-terphenyl," *Rep. Prog. Phys.*, vol. 64, no. 11, p. 1473, 2001.
- [379] N. Gnan, T. B. Schrøder, U. R. Pedersen, N. P. Bailey, and J. C. Dyre, "Pressure-energy correlations in liquids. iv. "isomorphs" in liquid phase diagrams," *J. Chem. Phys.*, vol. 131, no. 23, p. 234 504, 2009.
- [380] M. Paluch *et al.*, "A relationship between intermolecular potential, thermodynamics, and dynamic scaling for a supercooled ionic liquid," *J. Phys. Chem. Lett.*, vol. 1, no. 6, pp. 987–992, 2010.
- [381] K. Adrjanowicz, J. Pionteck, and M. Paluch, "Isochronal superposition and density scaling of the intermolecular dynamics in glass-forming liquids with varying hydrogen bonding propensity," *RSC Adv.*, vol. 6, pp. 49 370–49 375, 55 2016.
- [382] H. W. Hansen, B. Frick, S. Capaccioli, A. Sanz, and K. Niss, "Isochronal superposition and density scaling of the alpha-relaxation from pico- to millisecond," *J. Chem. Phys.*, vol. 149, no. 21, p. 214 503, 2018.
- [383] F. Puosi, O. Chulkin, S. Bernini, S. Capaccioli, and D. Leporini, "Thermodynamic scaling of vibrational dynamics and relaxation," *J. Chem. Phys.*, vol. 145, no. 23, p. 234 904, 2016.

- [384] I. H. Bell, "Probing the link between residual entropy and viscosity of molecular fluids and model potentials," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 116, no. 10, p. 201 815 943, 2019.
- [385] R. Casalini and C. M. Roland, "Thermodynamical scaling of the glass transition dynamics," *Phys. Rev. E*, vol. 69, p. 062 501, 6 2004.
- [386] C. Dreyfus, A. Le Grand, J. Gapinski, W. Steffen, and A. Patkowski, "Scaling the  $\alpha$ -relaxation time of supercooled fragile organic liquids," *Eur. Phys. J. B*, vol. 42, no. 3, pp. 309–319, 2004.
- [387] C. M. Roland, S. Hensel-Bielowka, M. Paluch, and R. Casalini, "Supercooled dynamics of glass-forming liquids and polymers under hydrostatic pressure," *Rep. Prog. Phys.*, vol. 68, no. 6, pp. 1405–1478, 2005.
- [388] C. M. Roland, S. Bair, and R. Casalini, "Thermodynamic scaling of the viscosity of van der waals, h-bonded, and ionic liquids," *J. Chem. Phys.*, vol. 125, no. 12, p. 124 508, 2006.
- [389] R. Casalini, S. Capaccioli, and C. M. Roland, "What Can We Learn by Squeezing a Liquid?" J. Phys. Chem. B, vol. 110, no. 23, pp. 11491–11495, 2006.
- [390] N. P. Bailey, U. R. Pedersen, N. Gnan, T. B. Schrøder, and J. C. Dyre, "Pressure-energy correlations in liquids. i. results from computer simulations," *J. Chem. Phys.*, vol. 129, no. 18, p. 184 507, 2008.
- [391] N. P. Bailey, U. R. Pedersen, N. Gnan, T. B. Schrøder, and J. C. Dyre, "Pressure-energy correlations in liquids. ii. analysis and consequences," *J. Chem. Phys.*, vol. 129, no. 18, p. 184 508, 2008.
- [392] T. B. Schrøder, N. P. Bailey, U. R. Pedersen, N. Gnan, and J. C. Dyre, "Pressure-energy correlations in liquids. iii. statistical mechanics and thermodynamics of liquids with hidden scale invariance," *J. Chem. Phys.*, vol. 131, no. 23, p. 234 503, 2009.
- [393] U. R. Pedersen, N. P. Bailey, T. B. Schrøder, and J. C. Dyre, "Strong pressure-energy correlations in van der waals liquids," *Phys. Rev. Lett.*, vol. 100, no. 1, p. 015 701, 2008.
- [394] J. P. Gabriel, M. Tress, W. Kossack, L. Popp, and F. Kremer, "Molecular heterogeneities in the thermal expansivity of polyalcohols," *J. Chem. Phys.*, vol. 154, no. 2, p. 024 503, 2021.
- [395] C. M. Roland, R. Casalini, R. Bergman, and J. Mattsson, "Role of hydrogen bonds in the supercooled dynamics of glass-forming liquids at high pressures," *Phys. Rev. B*, vol. 77, no. 1, p. 012 201, 2008.
- [396] A. A. Veldhorst, J. C. Dyre, and T. B. Schrøder, "Scaling of the dynamics of flexible lennard-jones chains: Effects of harmonic bonds," *J. Chem. Phys.*, vol. 143, no. 19, p. 194 503, 2015.

- [397] C. M. Roland and R. Casalini, "Entropy basis for the thermodynamic scaling of the dynamics of o-terphenyl," *J. Phys. Condens. Matter.*, vol. 19, no. 20, p. 205 118, 2007.
- [398] S. Bernini, F. Puosi, and D. Leporini, "Thermodynamic scaling of relaxation: insights from anharmonic elasticity," *J. Phys. Condens. Matter*, vol. 29, no. 13, p. 135 101, 2017.
- [399] K. G. Honnell, C. K. Hall, and R. Dickman, "On the pressure equation for chain molecules," *J. Chem. Phys.*, vol. 87, no. 1, pp. 664–674, 1987.
- [400] A. E. Olsen, J. C. Dyre, and T. B. Schrøder, "Communication: Pseudoisomorphs in liquids with intramolecular degrees of freedom," J. Chem. Phys., vol. 145, no. 24, p. 241 103, 2016.
- [401] M. C. C. Ribeiro, T. Scopigno, and G. Ruocco, "Computer simulation study of thermodynamic scaling of dynamics of 2Ca(NO3)2·3KNO3," J. Chem. Phys., vol. 135, no. 16, p. 164 510, 2011.
- [402] M. T. Cicerone, Q. Zhong, and M. Tyagi, "Picosecond Dynamic Heterogeneity, Hopping, and Johari-Goldstein Relaxation in Glass-Forming Liquids," *Phys. Rev. Lett.*, vol. 113, no. 11, p. 117 801, 2014.
- [403] M. T. Cicerone and M. Tyagi, "Metabasin transitions are Johari-Goldstein relaxation events," J. Chem. Phys., vol. 146, no. 5, 2017.
- [404] F. H. Stillinger, "A Topographic View of Supercooled Liquids and Glass Formation," *Science*, vol. 267, no. 5206, p. 1935, 1995.
- [405] M. T. Cicerone, D. Averett, and J. J. d. Pablo, "The role of hopping on transport above Tc in glycerol," J. Non-Cryst. Solids, vol. 407, no. 0, pp. 118–125, 2015.
- [406] M. T. Cicerone and K. Badilla-Nunez, "Do crossovers in liquid transport mechanisms arise from percolation of discrete dynamic environments?" *arXiv*, 2022. arXiv: 2201.12593.
- [407] S. Pawlus, R. Casalini, C. M. Roland, M. Paluch, S. J. Rzoska, and J. Ziolo, "Temperature and volume effects on the change of dynamics in propylene carbonate," *Phys. Rev. E*, vol. 70, no. 6, pp. 171–4, 2004.
- [408] J.-C. Soetens, C. Millot, B. Maigret, and I. Bakó, "Molecular dynamics simulation and x—ray diffraction studies of ethylene carbonate, propylene carbonate and dimethyl carbonate in liquid phase," J. Mol. Liq., vol. 92, no. 3, pp. 201–216, 2001.
- [409] X. You, M. I. Chaudhari, S. B. Rempe, and L. R. Pratt, "Dielectric relaxation of ethylene carbonate and propylene carbonate from molecular dynamics simulations," *J. Phys. Chem. B*, vol. 120, no. 8, pp. 1849–1853, 2016.

- [410] V. A. Koverga *et al.*, "Local structure and hydrogen bonding in liquid gammabutyrolactone and propylene carbonate: A molecular dynamics simulation," *J. Mol. Liq.*, vol. 287, p. 110912, 2019.
- [411] G. Lamoureux and B. Roux, "Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm," *J. Chem. Phys.*, vol. 119, no. 6, pp. 3025–3039, 2003.
- [412] J. M. Martínez and L. Martínez, "Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking," J. Comput. Chem., vol. 24, no. 7, pp. 819–825, 2003.
- [413] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," *J. Chem. Theory Comput.*, vol. 4, no. 3, pp. 435–447, 2008.
- [414] S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras, and A. J. Liu, "A structural approach to relaxation in glassy liquids," *Nat. Phys.*, vol. 12, no. 5, pp. 469–471, 2016.
- [415] J. G. McDaniel and J. R. Schmidt, "First-principles many-body force fields from the gas phase to liquid: A "universal" approach," *J. Phys. Chem. B*, vol. 118, pp. 8042–8053, 2014.
- [416] M. Bonetti and A. Dubois, "Isochronal superpositioning in the equilibrium regime of superpressed propylene carbonate to 1.8 GPa: A study by diffusivity measurement of the fluorescent probe Coumarin 1," *Eur. Phys. J. E*, vol. 42, no. 8, pp. 1–14, 2019.
- [417] M. Vogel, B. Doliwa, A. Heuer, and S. C. Glotzer, "Particle rearrangements during transitions between local minima of the potential energy landscape of a binary Lennard-Jones liquid," J. Chem. Phys., vol. 120, no. 9, pp. 4404–4414, 2004.
- [418] J. S. Bender, M. Zhi, and M. T. Cicerone, "The polarizability response of a glass-forming liquid reveals intrabasin motion and interbasin transitions on a potential energy landscape," *Soft Matter*, vol. 16, no. 24, pp. 5588–5598, 2020.
- [419] K. Goossens, L. Smeller, and K. Heremans, "Pressure tuning spectroscopy of the lowfrequency Raman spectrum of liquid amides," J. Chem. Phys., vol. 99, no. 8, pp. 5736– 5741, 1993.
- [420] H. Shimizu, Y. Ikeda, and S. Sasaki, "High-pressure Raman study of liquid and crystalline carbonyl sulfide," *Chem. Phys. Lett.*, vol. 175, no. 4, pp. 349–353, 1990.
- [421] K. Aoki, Y. Kakudate, S. Usuba, M. Yoshida, K. Tanaka, and S. Fujiwara, "High-pressure Raman study of liquid and crystalline C 2 H 2," J. Chem. Phys., vol. 88, no. 8, pp. 4565– 4568, 1988.

- [422] B. Frick, D. Richter, W. Petry, and U. Buchenau, "Study of the glass transition order parameter in amorphous polybutadiene by incoherent neutron scattering," *Z. Phys. B*, vol. 70, no. 1, pp. 73–79, 1988.
- [423] M. T. Cicerone and C. L. Soles, "Fast Dynamics and Stabilization of Proteins: Binary Glasses of Trehalose and Glycerol," *Biophys. J.*, vol. 86, no. 6, pp. 3836–3845, 2004.
- [424] L. Larini, A. Ottochian, C. D. Michele, and D. Leporini, "Universal scaling between structural relaxation and vibrational dynamics in glass-forming liquids and polymers," *Nat. Phys.*, vol. 4, no. 1, pp. 42–45, 2008.
- [425] D. S. Simmons, M. T. Cicerone, Q. Zhong, M. Tyagi, and J. F. Douglas, "Generalized localization model of relaxation in glass-forming liquids," *Soft Matter*, vol. 8, no. 45, pp. 11455–11461, 2012.
- [426] F. J., *Kinetic theory of liquids* (International series of monographs on physics.), R. Fowler, P. Kapitza, and N. Mott, Eds. Oxford: Oxford University Press, 1946.
- [427] D. J. Tobias and C. L. Brooks, "Molecular dynamics with internal coordinate constraints," J. Chem. Phys., vol. 89, no. 8, pp. 5115–5127, 1988.
- [428] S. He and H. A. Scheraga, "Macromolecular conformational dynamics in torsional angle space," *J. Chem. Phys.*, vol. 108, no. 1, pp. 271–286, 1998.
- [429] N. Vaidehi and A. Jain, "Internal coordinate molecular dynamics: A foundation for multiscale dynamics," *J. Phys. Chem. B*, vol. 119, no. 4, pp. 1233–1242, 2015.
- [430] J. P. Stoppelman, A. P. Wilkinson, and J. G. McDaniel, "Equation of State Predictions for ScF<sub>3</sub> and CaZrF<sub>6</sub> with Neural Network-Driven Molecular Dynamics," *J. Chem. Phys. Submitted*, 2023.
- [431] Q. Li *et al.*, "Chemical Diversity for Tailoring Negative Thermal Expansion," *Chem. Rev.*, vol. 122, pp. 8438–8486, 9 2021.
- [432] G. Ventura and M. Perfetti, *Thermal Properties of Solids at Room and Cryogenic Temperatures* (International Cryogenics Monograph Series). Dordrecht: Springer Netherlands, 2014, ISBN: 978-94-017-8968-4.
- [433] G. Hartwig, *Polymer Properties at Room and Cryogenic Temperatures* (International Cryogenics Monograph Series). Springer US, 1994.
- [434] Y. Rong *et al.*, "Large negative thermal expansion in non-perovskite lead-free ferroelectric Sn<sub>2</sub>P<sub>2</sub>S<sub>6</sub>," *Phys. Chem. Chem. Phys.*, vol. 18, no. 8, pp. 6247–6251, 2016.

- [435] D. Dubbeldam *et al.*, "Exceptional Negative Thermal Expansion in Isoreticular Metal–Organic Frameworks," *Angew. Chem. Int. Ed.*, vol. 46, no. 24, pp. 4496–4499, 2007.
- [436] M. K. Gupta, B. Singh, R. Mittal, S. Rols, and S. L. Chaplot, "Lattice dynamics and thermal expansion behavior in the metal cyanides MCN (M=Cu, Ag, Au): Neutron inelastic scattering and first-principles calculations," *Phys. Rev. B*, vol. 93, no. 13, p. 134 307, 2016.
- [437] P. J. Attfield, "Mechanisms and materials for NTE," *Front. Chem.*, vol. 6, no. AUG, p. 371, 2018.
- [438] T. A. Mary, J. S. Evans, T. Vogt, and A. W. Sleight, "Negative Thermal Expansion from 0.3 to 1050 Kelvin in ZrW<sub>2</sub>O<sub>8</sub>," *Science*, vol. 272, no. 5258, pp. 90–92, 1996.
- [439] J. S. Evans, T. A. Mary, and A. W. Sleight, "Negative thermal expansion materials," *Physica B: Condens. Matter*, vol. 241-243, pp. 311–316, 1997.
- [440] T. Chatterji, T. C. Hansen, M. Brunelli, and P. F. Henry, "Negative thermal expansion of ReO<sub>3</sub> in the extended temperature range," *Appl. Phys. Lett.*, vol. 94, no. 24, p. 241 902, 2009.
- [441] M. Dapiaggi and A. N. Fitch, "Negative (and very low) thermal expansion in ReO<sub>3</sub> from 5 to 300 K," *J. Appl. Crystallogr.*, vol. 42, no. 2, pp. 253–258, 2009.
- [442] J. Z. Tao and A. W. Sleight, "The role of rigid unit modes in negative thermal expansion," *J. Solid State Chem.*, vol. 173, no. 2, pp. 442–448, 2003.
- [443] B. K. Greve, K. L. Martin, P. L. Lee, P. J. Chupas, K. W. Chapman, and A. P. Wilkinson, "Pronounced Negative Thermal Expansion from a Simple Structure: Cubic ScF<sub>3</sub>," *J. Am. Chem. Soc.*, vol. 132, no. 44, pp. 15 496–15 498, 2010.
- [444] B. J. Kennedy and T. Vogt, "Powder X-ray diffraction study of the rhombohedral to cubic phase transition in TiF<sub>3</sub>," *Mater. Res. Bull.*, vol. 37, no. 1, pp. 77–83, 2002.
- [445] S. Chaudhuri, P. J. Chupas, M. Wilson, P. Madden, and C. P. Grey, "Study of the nature and mechanism of the rhombohedral-to-cubic phase transition in  $\alpha$ -AlF<sub>3</sub> with molecular dynamics simulations," *J. Phys. Chem. B*, vol. 108, no. 11, pp. 3437–3445, 2004.
- [446] Z. Wei *et al.*, "Colossal Pressure-Induced Softening in Scandium Fluoride," *Phys. Rev. Lett.*, vol. 124, no. 25, p. 255 502, 2020.
- [447] C. Yang *et al.*, "Size effects on negative thermal expansion in cubic ScF<sub>3</sub>," *Appl. Phys. Lett.*, vol. 109, no. 2, p. 023 110, 2016.
- [448] L. Hu *et al.*, "Localized Symmetry Breaking for Tuning Thermal Expansion in ScF<sub>3</sub> Nanoscale Frameworks," *J. Am. Chem. Soc.*, vol. 140, no. 13, pp. 4477–4480, 2018.

- [449] C. R. Morelock, B. K. Greve, L. C. Gallington, K. W. Chapman, and A. P. Wilkinson, "Negative thermal expansion and compressibility of  $Sc_{1-x}Y_xF_3$  ( $x \le 0.25$ )," *J. Appl. Phys.*, vol. 114, no. 21, p. 213 501, 2013.
- [450] C. R. Morelock, L. C. Gallington, and A. P. Wilkinson, "Solid solubility, phase transitions, thermal expansion, and compressibility in  $Sc_{1x}$  Al<sub>x</sub> F<sub>3</sub>," *J. Solid State Chem.*, vol. 222, pp. 96–102, 2015.
- [451] C. R. Morelock, L. C. Gallington, and A. P. Wilkinson, "Evolution of negative thermal expansion and phase transitions in  $Sc_{1-x}Ti_xF_3$ ," *Chem. Mater.*, vol. 26, no. 5, pp. 1936–1940, 2014.
- [452] F. Qin *et al.*, "Isotropic Zero Thermal Expansion and Local Vibrational Dynamics in (Sc,Fe)F<sub>3</sub>," *Inorg. Chem.*, vol. 56, no. 18, pp. 10840–10843, 2017.
- [453] L. Hu *et al.*, "Zero thermal expansion and ferromagnetism in cubic  $Sc_{1-x}M_xF_3$  (M = Ga, Fe) over a wide temperature range," *J. Am. Chem. Soc.*, vol. 136, no. 39, pp. 13566–13569, 2014.
- [454] J. Chen *et al.*, "Tunable thermal expansion in framework materials through redox intercalation," *Nat. Commun.*, vol. 8, no. 1, pp. 1–7, 2017.
- [455] S. J. Baxter, K. V. Loske, A. J. Lloyd, and A. P. Wilkinson, "Controlling the Phase Behavior of Low and Negative Thermal Expansion  $\text{ReO}_3$ -Type Fluorides using Interstitial Anions:  $\text{Sc}_{1-x}\text{Zr}_x\text{F}_{3+x}$ ," *Inorg. Chem.*, vol. 59, no. 10, pp. 7188–7194, 2020.
- [456] T. Wang *et al.*, "Tunable thermal expansion and magnetism in Zr-doped ScF<sub>3</sub>," *Appl. Phys. Lett.*, vol. 109, no. 18, p. 181 901, 2016.
- [457] J. C. Hancock *et al.*, "Large Negative Thermal Expansion and Anomalous Behavior on Compression in Cubic ReO<sub>3</sub>-Type A<sup>I</sup>IB<sup>I</sup>VF<sub>6</sub>: CaZrF<sub>6</sub> and CaHfF<sub>6</sub>," *Chem. Mater.*, vol. 27, no. 11, pp. 3912–3918, 2015.
- [458] B. R. Hester, A. M. Dos Santos, J. J. Molaison, J. C. Hancock, and A. P. Wilkinson, "Synthesis of Defect Perovskites  $(\text{He}_{2-x}\Box_x)(\text{CaZr})F_6$  by Inserting Helium into the Negative Thermal Expansion Material CaZrF<sub>6</sub>," *J. Am. Chem. Soc.*, vol. 139, no. 38, pp. 13284–13287, 2017.
- [459] A. J. Lloyd *et al.*, "Hybrid Double Perovskite Containing Helium: [He<sub>2</sub>][CaZr]F<sub>6</sub>," *Chem. Mater.*, vol. 33, no. 9, pp. 3132–3138, 2021.
- [460] C. A. Occhialini, S. U. Handunkanda, E. B. Curry, and J. N. Hancock, "Classical, quantum, and thermodynamics of a lattice model exhibiting structural negative thermal expansion," *Phys. Rev. B*, vol. 95, no. 9, p. 094 106, 2017.

- [461] T. A. Bird *et al.*, "Anharmonicity and scissoring modes in the negative thermal expansion materials ScF<sub>3</sub> and CaZrF<sub>6</sub>," *Phys. Rev. B*, vol. 101, no. 6, p. 064 306, 2020.
- [462] F. Han *et al.*, "Local structure and controllable thermal expansion in the solid solution  $(Mn_{1x}Ni_x)ZrF_6$ ," *Inorg. Chem. Front.*, vol. 4, no. 2, pp. 343–347, 2017.
- [463] D. Wendt *et al.*, "Entropic elasticity and negative thermal expansion in a simple cubic crystal," *Sci. Adv.*, vol. 5, no. 11, 2019.
- [464] L. Wang *et al.*, "First-Principles Study of  $Sc_{1x}Ti_xF_3$  ( $x \le 0.375$ ): Negative Thermal Expansion, Phase Transition, and Compressibility," *J. Am. Ceram. Soc.*, vol. 98, no. 9, pp. 2852–2857, 2015.
- [465] V. Heine, P. R. Welche, and M. T. Dove, "Geometrical Origin and Theory of Negative Thermal Expansion in Framework Structures," J. Am. Ceram. Soc., vol. 82, no. 7, pp. 1793– 1802, 1999.
- [466] A. V. Tkachenko and I. A. Zaliznyak, "Empty perovskites as Coulomb floppy networks: Entropic elasticity and negative thermal expansion," *Phys. Rev. B*, vol. 103, no. 13, p. 134 106, 2021.
- [467] Y. Oba, T. Tadano, R. Akashi, and S. Tsuneyuki, "First-principles study of phonon anharmonicity and negative thermal expansion in ScF<sub>3</sub>," *Phys. Rev. Mater.*, vol. 3, no. 3, p. 033 601, 2019.
- [468] S. Baroni, P. Giannozzi, and E. Isaev, "Density-Functional Perturbation Theory for Quasi-Harmonic Calculations," *Rev. Mineral. Geochem.*, vol. 71, no. 1, pp. 39–57, 2010.
- [469] C. W. Li *et al.*, "Structural relationship between negative thermal expansion and quartic anharmonicity of cubic ScF3," *Phys. Rev. Lett.*, vol. 107, no. 19, p. 195 504, 2011.
- [470] T. Tadano and S. Tsuneyuki, "Self-consistent phonon calculations of lattice dynamical properties in cubic SrTiO<sub>3</sub> with first-principles anharmonic force constants," *Phys. Rev. B Condens. Matter Mater. Phys.*, vol. 92, no. 5, p. 054 301, 2015.
- [471] N. R. Werthamer, "Self-Consistent Phonon Formulation of Anharmonic Lattice Dynamics," *Phys. Rev. B*, vol. 1, no. 2, p. 572, 1970.
- [472] V. V. Goldman, G. K. Horton, and M. L. Klein, "An Improved Self-Consistent Phonon Approximation," *Phys. Rev. Lett.*, vol. 21, no. 22, p. 1527, 1968.
- [473] F. Zhou, W. Nielson, Y. Xia, and V. Ozoliņš, "Compressive sensing lattice dynamics. I. General formalism," *Phys. Rev. B*, vol. 100, no. 18, p. 184 308, 2019.

- [474] F. Zhou, B. Sadigh, D. Åberg, Y. Xia, and V. Ozoliņš, "Compressive sensing lattice dynamics. II. Efficient phonon calculations and long-range interactions," *Phys. Rev. B*, vol. 100, no. 18, p. 184 309, 2019.
- [475] P. Lazar, T. Bučko, and J. Hafner, "Negative thermal expansion of ScF<sub>3</sub>: Insights from density-functional molecular dynamics in the isothermal-isobaric ensemble," *Phys. Rev. B Condens. Matter Mater. Phys.*, vol. 92, no. 22, p. 224 302, 2015.
- [476] D. Bocharov, M. Krack, Y. Rafalskij, A. Kuzmin, and J. Purans, "Ab initio molecular dynamics simulations of negative thermal expansion in ScF<sub>3</sub>: The effect of the supercell size," *Comput. Mater. Sci.*, vol. 171, p. 109 198, 2020.
- [477] C. Wang *et al.*, "Negative thermal expansion of ScF<sub>3</sub>: first principles vs empirical molecular dynamics," in *IOP Conference Series: Materials Science and Engineering*, vol. 503, IOP Publishing, 2019, p. 012 001.
- [478] M. K. Gupta, B. Singh, R. Mittal, and S. L. Chaplot, "Negative thermal expansion behavior in MZrF<sub>6</sub> (M=Ca,Mg,Sr): Ab initio lattice dynamical studies," *Phys. Rev. B*, vol. 98, no. 1, p. 014 301, 2018.
- [479] N. Wang, J. Deng, J. Chen, and X. Xing, "Phonon spectrum attributes for the negative thermal expansion of  $MZrF_6$  (M = Ca, Mn-Ni, Zn)," *Inorg. Chem. Front.*, vol. 6, no. 4, pp. 1022–1028, 2019.
- [480] J. Schmidt, M. R. Marques, S. Botti, and M. A. Marques, "Recent advances and applications of machine learning in solid-state materials science," *npj Comput. Mater.*, vol. 5, no. 1, pp. 1–36, 2019.
- [481] F. Noé, A. Tkatchenko, K. R. Müller, and C. Clementi, "Machine Learning for Molecular Simulation," *Annu. Rev. Phys. Chem.*, vol. 71, pp. 361–390, 2020.
- [482] P. O. Dral, "Quantum Chemistry in the Age of Machine Learning," J. Phys. Chem. Lett., vol. 11, no. 6, pp. 2336–2347, 2020.
- [483] J. Behler, R. Martoňák, D. Donadio, and M. Parrinello, "Pressure-induced phase transitions in silicon studied by neural network-based metadynamics simulations," *Phys. Status Solidi B*, vol. 245, no. 12, pp. 2618–2629, 2008.
- [484] H. Eshet, R. Z. Khaliullin, T. D. Kühne, J. Behler, and M. Parrinello, "Ab initio quality neural-network potential for sodium," *Phys. Rev. B Condens. Matter Mater. Phys.*, vol. 81, no. 18, p. 184 107, 2010.
- [485] C. Chen, Z. Deng, R. Tran, H. Tang, I. H. Chu, and S. P. Ong, "Accurate force field for molybdenum by machine learning large materials data," *Phys. Rev. Matter.*, vol. 1, no. 4, p. 043 603, 2017.

- [486] X. Chen *et al.*, "Machine learning enhanced empirical potentials for metals and alloys," *Comput. Phys. Commun.*, vol. 269, p. 108 132, 2021.
- [487] S. Ma and Z. P. Liu, "Machine learning potential era of zeolite simulation," *Chem. Sci.*, vol. 13, no. 18, pp. 5055–5068, 2022.
- [488] O. Tayfuroglu, A. Kocak, and Y. Zorlu, "A neural network potential for the IRMOF series and its application for thermal and mechanical behaviors," *Phys. Chem. Chem. Phys.*, vol. 24, no. 19, pp. 11882–11897, 2022.
- [489] P. Friederich, F. Häse, J. Proppe, and A. Aspuru-Guzik, "Machine-learned potentials for next-generation matter simulations," *Nat. Mater.*, vol. 20, no. 6, pp. 750–761, 2021.
- [490] Y. Liu, T. Zhao, W. Ju, and S. Shi, "Materials discovery and design using machine learning," J. Materiomics, vol. 3, no. 3, pp. 159–177, 2017.
- [491] S. Lin, Y. Wang, Y. Zhao, L. R. Pericchi, A. J. Hernández-Maldonado, and Z. Chen, "Machine-learning-assisted screening of pure-silica zeolites for effective removal of linear siloxanes and derivatives," *J. Mater. Chem. A*, vol. 8, no. 6, pp. 3228–3237, 2020.
- [492] C. Altintas, O. F. Altundal, S. Keskin, and R. Yildirim, "Machine Learning Meets with Metal Organic Frameworks for Gas Storage and Separation," *J. Chem. Inf. Model.*, vol. 61, no. 5, pp. 2131–2146, 2021.
- [493] A. Jain *et al.*, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Mater.*, vol. 1, no. 1, p. 011 002, 2013.
- [494] A. Dal Corso, "Pseudopotentials periodic table: From H to Pu," *Comput. Mater. Sci.*, vol. 95, pp. 337–350, 2014.
- [495] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *J. Appl. Phys.*, vol. 52, no. 12, p. 7182, 1998.
- [496] L. Zhang, J. Han, H. Wang, W. Saidi, R. Car, and W. E, "End-to-end Symmetry Preserving Inter-atomic Potential Energy Model for Finite and Extended Systems," in Advances in Neural Information Processing Systems, vol. 31, 2018, pp. 8024–8035.
- [497] T. A. Young, T. Johnston-Wood, V. L. Deringer, and F. Duarte, "A transferable activelearning strategy for reactive molecular force fields," *Chem. Sci.*, vol. 12, no. 32, pp. 10944–10955, 2021.
- [498] H. J. Berendsen, J. P. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *J. Chem. Phys.*, vol. 81, no. 8, p. 3684, 1998.
- [499] G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant pressure molecular dynamics algorithms," *J. Chem. Phys.*, vol. 101, no. 5, p. 4177, 1998.
- [500] M. Ceriotti, J. More, and D. E. Manolopoulos, "i-PI: A Python interface for ab initio path integral molecular dynamics simulations," *Comput. Phys. Commun.*, vol. 185, no. 3, pp. 1019–1026, 2014.
- [501] V. Kapil *et al.*, "i-PI 2.0: A universal force engine for advanced molecular simulations," *Comput. Phys. Commun.*, vol. 236, pp. 214–223, 2019.
- [502] G. Bussi, T. Zykova-Timan, and M. Parrinello, "Isothermal-isobaric molecular dynamics using stochastic velocity rescaling," *J. Chem. Phys.*, vol. 130, no. 7, p. 074 101, 2009.
- [503] C. Wang *et al.*, "Size and crystal symmetry breaking effects on negative thermal expansion in ScF<sub>3</sub> nanostructures," *Phys. Chem. Chem. Phys.*, vol. 23, no. 43, pp. 24814–24822, 2021.
- [504] W. J. Kim, M. Kim, E. K. Lee, S. Lebègue, and H. Kim, "Failure of Density Functional Dispersion Correction in Metallic Systems and Its Possible Solution Using a Modified Many-Body Dispersion Correction," *J. Phys. Chem. Lett.*, vol. 7, no. 16, pp. 3278–3283, 2016.
- [505] A. Tkatchenko, R. A. Distasio, R. Car, and M. Scheffler, "Accurate and efficient method for many-body van der Waals interactions," *Phys. Rev. Lett.*, vol. 108, no. 23, p. 236402, 2012.
- [506] G. P. Chen, V. K. Voora, M. M. Agee, S. G. Balasubramani, and F. Furche, "Random-Phase Approximation Methods," *Annu. Rev. Phys. Chem.*, vol. 68, pp. 421–445, 2017.
- [507] C.-K. Lee *et al.*, "Transfer learning with graph neural networks for optoelectronic properties of conjugated oligomers," *J. Chem. Phys.*, vol. 154, no. 2, p. 024 906, 2021.
- [508] K. N. Trueblood *et al.*, "Atomic Displacement Parameter Nomenclature. Report of a Subcommittee on Atomic Displacement Parameter Nomenclature," *Acta. Cryst.*, vol. 52, no. 5, pp. 770–781, 1996.

## VITA

John Stoppelman was born on June 2, 1996 to David and Lisa Stoppelman in Elk Grove Village, IL. He grew up in the Chicagoland suburb of Darien. He attended high school at Benet Academy in Lisle, IL and graduated in 2014. After graduation, he attended Loras College in Dubuque, IA, where he ran Division III cross country and track and field for four years. He gained an interest in molecular simulation while performing undergraduate research with Dr. Adam Moser at Loras and Dr. Lee-Ping Wang at University of California, Davis. He graduated maxima cum laude with a Bachelor of Science degree in Chemistry in 2018. After this, he attended the Georgia Institute of Technology to pursue his PhD in Chemistry under Dr. Jesse McDaniel. He will begin a postdoc with Dr. Chao-Ping Hsu at Academia Sinica in Taipei, Taiwan in January 2024.