

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Modeling Chemical Reactivity for Battery Applications through Data-driven Methodologies

Permalink

<https://escholarship.org/uc/item/8516003k>

Author

Xie, Xiaowei

Publication Date

2022

Peer reviewed|Thesis/dissertation

Modeling Chemical Reactivity for Battery Applications through Data-driven Methodologies

by

Xiaowei Xie

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Kristin A. Persson, Co-chair
Associate Professor Eric Neuscamman, Co-chair
Assistant Professor David Limmer
Associate Professor Ali Mesbah

Fall 2022

Modeling Chemical Reactivity for Battery Applications through Data-driven Methodologies

Copyright 2022
by
Xiaowei Xie

Abstract

Modeling Chemical Reactivity for Battery Applications through Data-driven Methodologies

by

Xiaowei Xie

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Kristin A. Persson, Co-chair

Associate Professor Eric Neuscamman, Co-chair

The following dissertation discusses the usage of data-driven approaches, combined with density functional theory (DFT), to study the complex reactions involved in the battery systems, e.g., electrolyte degradation and solid electrolyte interphase (SEI) formation. Data-driven approaches include the development of machine learning (ML) interatomic potentials for the challenging charged/open-shell molecular species, reaction network construction based on high-throughput DFT, ML, automated fragmentation and recombination strategies to yield a thermodynamic landscape of millions of possible reactions for the SEI formation, and computational screening of novel electrolytes for multivalent ion batteries based on their kinetic barriers to reductive decomposition.

Chapter 1 provides a brief overview of battery chemistry and the computational modeling tools for investigating such systems, e.g., ab-initio methods or classical/ML force fields. Furthermore, this overview highlights the challenges and the strategies for the computational modeling of battery materials, bulk electrolyte properties, and electrochemical and interfacial reactions.

Chapter 2 describes the development of the Becke Population Neural Network (BpopNN) model as an ML interatomic potential that is capable of handling charged molecules, which are essential in electrochemistry and battery systems. This is accomplished by incorporating electronic information (electronic population on each atom) into the descriptor and minimizing the total energy self-consistently with the total population constrained to be the total

number of electrons in the system.

Chapter 3 extends from the work discussed in Chapter 2 and demonstrates the application of the BpopNN model to Li-organic systems, which have direct relevance to battery electrolytes. Spin populations are incorporated into the model, and possibilities of using BpopNN to model charge-transfer excited states are discussed.

Chapter 4 describes the study of using an automated reaction network and DFT to probe the formation mechanisms of Lithium Ethylene Monocarbonate (LEMC), a controversial organic component of the SEI in lithium-ion batteries (LIB). It manifests another application of data-driven methods and ML models to studying the complex reaction cascades or mechanisms in battery systems. It was found that water is essential in the formation of LEMC; therefore, LEMC is not likely to be the major component of the SEI if rigorously dried electrolytes are used. Several novel mechanisms for forming LEMC were also proposed in an automated fashion through a path-finding algorithm applied to the massive reaction network.

Chapter 5 describes a computational screening effort for novel electrolytes in the weakly-coordinating fluorinated alkoxyborate and alkoxyaluminate salt family for Mg/Ca-ion battery applications. The design metric employed is based on the redox potentials and the kinetics for reductive decomposition, covering various decomposition patterns, e.g., the breaking of Al/B-O, C-O, or C-F bonds. The kinetic barriers for such decomposition pathways are obtained and demonstrated to correlate well with their electrochemical performance for existing electrolytes. Novel electrolytes that are never synthesized before are also proposed based on this criterion.

Chapter 6 summarizes and concludes the thesis.

To my family and my boyfriend David W. Small

Contents

Contents	ii
List of Figures	iv
List of Tables	ix
1 Introduction	1
I Overview	2
II Chemical processes in battery systems	4
III Modeling techniques	8
Bibliography	16
2 Incorporating Electronic Information into Machine Learning Potential Energy Surfaces: Development of BpopNN	26
I Introduction	27
II Theory	30
III Results and Discussion	35
IV Conclusions	46
V Appendix	48
Bibliography	57
3 Extension of BpopNN to Li-organic systems	69
I Introduction	70
II Methods	72
III Results and Discussion	74
IV Conclusions	79
Bibliography	81

4	Data-driven Prediction of Formation Mechanisms of Lithium Ethylene Monocarbonate with an Automated Reaction Network	85
I	Introduction	86
II	Methods	88
III	Results and Discussion	93
IV	Conclusion	104
V	Appendix	105
	Bibliography	139
5	The reductive decomposition kinetics and thermodynamics that govern the design of fluorinated alkoxyaluminate/borate salts for Mg-ion and Ca-ion batteries	148
I	Introduction	149
II	Results and discussion	151
III	Conclusion	166
IV	Appendix	167
	Bibliography	226
6	Conclusions and Future Work	231
	Bibliography	236

List of Figures

1.1	(a) Typical organic electrolytes for LIB. (b) Example molecular cluster structures of Li^+ solvated in EC (left) and Mg^{2+} with $\text{Mg}[\text{TFSI}]_2$ salt solvated in DME (right).	3
1.2	Electrochemical stability window.	4
1.3	(a) The mechanism for reductive decomposition of the EC electrolyte in LIB. (b) The mechanism for reductive decomposition of the $\text{Mg}[\text{TFSI}]_2$ salt in Mg-ion batteries.	5
1.4	Schematic presentation of the mosaic structure of the SEI on Li/graphite anode (proposed by Peled [41]). An ideal SEI should be ionic-conducting but electron-insulating.	7
2.1	General concept for the ML potential	32
2.2	Performance on sparse $\text{Li}_{29}\text{H}_{29}$ and non-sparse $\text{Li}_{32}\text{H}_{32}$ test sets. All energies are in a.u.	38
2.3	Charge optimization result for a neutral $\text{Li}_{32}\text{H}_{32}$	40
2.4	Geometry optimization on a cubic $\text{Li}_{32}\text{H}_{32}$ structure	41
2.5	Energy profile for various structural transformations for neutral $\text{Li}_{12}\text{H}_{12}$. Superpositions and RMSD values were obtained with Maestro.[133]	42
2.6	Energy profile for various structural transformations for anion $\text{Li}_{12}\text{H}_{12}$. Superpositions and RMSD values were obtained with Maestro.[133]	43
2.7	A charge separation example for neutral $\text{Li}_{15}\text{H}_{15}$	44
2.8	Example geometries of the training data	50
2.9	DFT energy per atom plot for the neutral, cation, anion training set. To aid visibility, the data has been slightly shifted horizontally.	51
3.1	A representative set of training molecules.	75
3.2	BpopNN vs. DFT adiabatic IPs and EAs	76
3.3	(a) Types of Li interactions in the training set. (b) An example of a test molecule created through fusing two training molecules together. (c) Parity plot of DFT vs. BpopNN excitation gaps for a test set of anions.	78

3.4	More ES examples at different total charges. The orbital indices used for excitation are given in the purple text for cation examples in (a) and the carbon nanohoop example in (b). If the orbital indices are not labeled, it corresponds to HOMO to LUMO excitation. Charges or spins are transferred from the top orbital to the bottom orbital.	80
4.1	Workflow for generating relevant molecules	89
4.2	Overview of the reaction networks generated in this work	91
4.3	Direct paths to LEMC from EC, Li^+ and H_2O identified by the reaction network	95
4.4	LEMC paths from DFT	97
4.5	Paths to LEMC involving recombinant molecules from EC, Li^+ and H_2O identified by the reaction network	98
4.6	LEMC paths from DFT	100
4.7	Paths to LEMC with no water participation identified by the 2-bond-change reaction network	101
4.8	Radical paths, corresponding to key steps in path A in Figure 4.7	102
4.9	Transition state structure for water splitting on a Li cluster	107
4.10	One-step LEDC hydrolysis path	108
4.11	LEMC paths involving recombinant molecules	109
4.12	A LEMC path without water participation, corresponding to the key step in path B in Figure 4.7	110
4.13	Energy diagrams for the 3 major reactions involved in the 8 non-water LEMC paths in the 5-bond-change reaction network	129
4.14	The EC hydrolysis paths with one explicit solvent molecule (EC)	130
4.15	The LEDC hydrolysis path with one explicit solvent molecule (EC)	131
4.16	The LEDC hydrolysis path computed with implicit solvation model with different dielectric constants	132
4.17	Distribution of BonDNet prediction error by reactant charge for the old recombinant molecules	134
4.18	Distribution of BonDNet prediction error by bond dissociation type for the old recombinant molecules	135
4.19	Reactions with large prediction error for the old recombinant molecules from the current BonDNet model	136
4.20	Prediction errors for the new test set from the current model and the new model	137
4.21	Reactions with large prediction error for the new recombinant molecules from the new BonDNet model	138

5.1	(a) Concept of contact ion pairs (CIP) and bare anions. (b) Electrochemical windows for the anions. Bare anions (X^- = monoanionic ligands) were used for oxidation potentials (upper limit) and contact ion pairs ($[M^{2+}X^-]^+$, $M = Mg/Ca$) were used for reduction potentials (bottom limit). Two levels of theory were used for comparison: the redox potentials shown in the regular bars are based on $\omega B97X-D3/def2-TZVPPD/CPCM/RIJCOSX//\omega B97X-D3/def2-SVPD/CPCM/RIJCOSX$ level of theory, whereas the black dashed lines and the numbers in parenthesis are from $B3LYP-D3BJ/def2-TZVPPD/CPCM//B3LYP-D3BJ/def2-SVP/CPCM$ level of theory. The numbers in the middle of the bars are the energy gaps between the two redox potentials (stability windows).	152
5.2	^{11}B NMR experiments on the Mg borate salts.	154
5.3	Representative decomposition pathways through Al/B-O bond breakage. (a) $Ca^+[Al(tftb)_4]^-$ decomposition pathway. (b) $Ca^+[B(tftb)_4]^-$ decomposition pathway.	156
5.4	Representative decomposition pathways through C-O and C-F bond breakage. (a) $Ca^+[Al(hftb)_4]^-$ decomposition through C-F bond breakage. (b) $Ca^+[Al(hftb)_4]^-$ decomposition through C-O bond breakage.	157
5.5	Decomposition pathways and rate-limiting barrier heights. (a) Decomposition pathways for Al-O/B-O bond breakage. (b) Rate-limiting decomposition free energy barriers for Al-O/B-O, C-F, and C-O bond breakage. The numbers on each bar from the top to bottom are the rate-limiting barriers for Al/B-O bond breakage, C-F bond breakage, and C-O bond breakage. Note that the C-O bond breaking barrier is only shown when it is less than 1.0 eV.	158
5.6	NCI plot of the $Ca^+[Al(tftb)_4]^-$ and $Ca^+[B(tftb)_4]^-$ complexes	161
5.7	Al/B-O bond breaking barrier analysis for aluminates/borates. (a) Illustrative picture of how the complexes are divided into two regions for probing the origin of the barrier differences. (b) Natural charge analysis for the Ca and the CaOR (OR=tftb,hfip,hftb) fragment. (c) Electronic energy barriers for different fragments. (1) The electronic energy barriers for the CaOR and the $Al(OR)_3$ fragments for the TS1 geometry (see Figure 5.5(a)). Note that different total charges are used for the SM and TS in computing the barriers for each fragment. The CaOR barriers are calculated from the $[CaOR]^0 \rightarrow [CaOR]^+$ transition; the $Al(OR)_3$ barriers are calculated from the $[Al(OR)_3]^0 \rightarrow [Al(OR)_3]^-$ transition. Similarly, the binding energy barriers correspond to subtracting the binding energy between $[CaOR]^0 - [Al(OR)_3]^0$ in the SM from the binding energy between $[CaOR]^+ - [Al(OR)_3]^-$ in the TS. The CaOR barrier, the $Al(OR)_3$ barrier, and the binding energy barrier sum to the total energy barrier. (2) is the same as (1) but is for the TS2 geometry in aluminates. (3) Barriers for the borates calculated with the core part of the ligand part taken from the original SM and TS geometries.	163

5.8	Speculative scheme for the kinetic control of the decomposition pathways (Al-O/B-O bond breakage)	165
5.9	Electrochemical windows for CIPs. Two level of theories were used for comparison: the redox potentials shown in the regular bars are based on ω B97X-D3/def2-TZVPPD/CPCM/RIJCOSX// ω B97X-D3/def2-SVPD/CPCM/RIJCOSX level of theory, whereas the black dashed lines and the numbers in parenthesis are from B3LYP-D3BJ/def2-TZVPPD/CPCM//B3LYP-D3BJ/def2-SVP/CPCM level of theory. The numbers in the middle of the bars are the energy gaps between the two redox potentials (stability windows).	167
5.10	Electrochemical windows for bare anions. Two levels of theory were used for comparison: the redox potentials shown in the regular bars are based on ω B97X-D3/def2-TZVPPD/CPCM/RIJCOSX// ω B97X-D3/def2-SVPD/CPCM/RIJCOSX level of theory, whereas the black dashed lines and the numbers in parenthesis are from B3LYP-D3BJ/def2-TZVPPD/CPCM//B3LYP-D3BJ/def2-SVP/CPCM level of theory. The numbers in the middle of the bars are the energy gaps between the two redox potentials (stability windows).	168
5.11	$\text{Mg}^+[\text{Al}(\text{tfip})_4]^-$ decomposition pathway through Al-O bond breakage	169
5.12	$\text{Mg}^+[\text{Al}(\text{tftb})_4]^-$ decomposition pathway through Al-O bond breakage	170
5.13	$\text{Mg}^+[\text{Al}(\text{hfip})_4]^-$ decomposition pathway through Al-O bond breakage	171
5.14	$\text{Mg}^+[\text{Al}(\text{hftb})_4]^-$ decomposition pathway through Al-O bond breakage	172
5.15	$\text{Mg}^+[\text{Al}(\text{pftb})_4]^-$ decomposition pathway through Al-O bond breakage	173
5.16	$\text{Ca}^+[\text{Al}(\text{tfip})_4]^-$ decomposition pathway through Al-O bond breakage	174
5.17	$\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$ decomposition pathway through Al-O bond breakage	175
5.18	$\text{Ca}^+[\text{Al}(\text{hfip})_4]^-$ decomposition pathway through Al-O bond breakage	176
5.19	$\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$ decomposition pathway through Al-O bond breakage	177
5.20	$\text{Ca}^+[\text{Al}(\text{pftb})_4]^-$ decomposition pathway through Al-O bond breakage	178
5.21	$\text{Mg}^+[\text{B}(\text{tfip})_4]^-$ decomposition pathway through B-O bond breakage	179
5.22	$\text{Mg}^+[\text{B}(\text{tftb})_4]^-$ decomposition pathway through B-O bond breakage	180
5.23	$\text{Mg}^+[\text{B}(\text{hfip})_4]^-$ decomposition pathway through B-O bond breakage	181
5.24	$\text{Mg}^+[\text{B}(\text{hftb})_4]^-$ decomposition pathway through B-O bond breakage	182
5.25	$\text{Mg}^+[\text{B}(\text{pftb})_4]^-$ decomposition pathway through B-O bond breakage	183
5.26	$\text{Ca}^+[\text{B}(\text{tfip})_4]^-$ decomposition pathway through B-O bond breakage	184
5.27	$\text{Ca}^+[\text{B}(\text{tftb})_4]^-$ decomposition pathway through B-O bond breakage	185
5.28	$\text{Ca}^+[\text{B}(\text{hfip})_4]^-$ decomposition pathway through B-O bond breakage	186
5.29	$\text{Ca}^+[\text{B}(\text{hftb})_4]^-$ decomposition pathway through B-O bond breakage	187
5.30	$\text{Ca}^+[\text{B}(\text{pftb})_4]^-$ decomposition pathway through B-O bond breakage	188
5.31	$\text{Mg}^+[\text{Al}(\text{tfip})_4]^-$ decomposition pathway through C-O bond breakage	189
5.32	$\text{Mg}^+[\text{Al}(\text{hfip})_4]^-$ decomposition pathway through C-O bond breakage	190
5.33	$\text{Mg}^+[\text{Al}(\text{hftb})_4]^-$ decomposition pathway through C-O bond breakage	191
5.34	$\text{Mg}^+[\text{Al}(\text{pftb})_4]^-$ decomposition pathway through C-O bond breakage	192

5.35	$\text{Ca}^+[\text{Al}(\text{tfip})_4]^-$	decomposition pathway through C-O bond breakage	193
5.36	$\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$	decomposition pathway through C-O bond breakage	194
5.37	$\text{Ca}^+[\text{Al}(\text{hfip})_4]^-$	decomposition pathway through C-O bond breakage	195
5.38	$\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$	decomposition pathway through C-O bond breakage	196
5.39	$\text{Ca}^+[\text{Al}(\text{pftb})_4]^-$	decomposition pathway through C-O bond breakage	197
5.40	$\text{Mg}^+[\text{B}(\text{tfip})_4]^-$	decomposition pathway through C-O bond breakage	198
5.41	$\text{Mg}^+[\text{B}(\text{tftb})_4]^-$	decomposition pathway through C-O bond breakage	199
5.42	$\text{Mg}^+[\text{B}(\text{hfip})_4]^-$	decomposition pathway through C-O bond breakage	200
5.43	$\text{Mg}^+[\text{B}(\text{hftb})_4]^-$	decomposition pathway through C-O bond breakage	201
5.44	$\text{Mg}^+[\text{B}(\text{pftb})_4]^-$	decomposition pathway through C-O bond breakage	202
5.45	$\text{Ca}^+[\text{B}(\text{tfip})_4]^-$	decomposition pathway through C-O bond breakage	203
5.46	$\text{Ca}^+[\text{B}(\text{tftb})_4]^-$	decomposition pathway through C-O bond breakage	204
5.47	$\text{Ca}^+[\text{B}(\text{hfip})_4]^-$	decomposition pathway through C-O bond breakage	205
5.48	$\text{Ca}^+[\text{B}(\text{hftb})_4]^-$	decomposition pathway through C-O bond breakage	206
5.49	$\text{Ca}^+[\text{B}(\text{pftb})_4]^-$	decomposition pathway through C-O bond breakage	207
5.50	$\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$	decomposition pathway through C-F bond breakage	208
5.51	$\text{Ca}^+[\text{Al}(\text{pftb})_4]^-$	decomposition pathway through C-F bond breakage	209
5.52	$\text{Mg}^+[\text{B}(\text{hftb})_4]^-$	decomposition pathway through C-F bond breakage	210
5.53	$\text{Mg}^+[\text{B}(\text{pftb})_4]^-$	decomposition pathway through C-F bond breakage	211
5.54	$\text{Ca}^+[\text{B}(\text{hftb})_4]^-$	decomposition pathway through C-F bond breakage	212
5.55	$\text{Ca}^+[\text{B}(\text{pftb})_4]^-$	decomposition pathway through C-F bond breakage	213
5.56	Barrier analysis for borates from CaOR-B(OR)_3 separation		214
5.57	EDA analysis of the interaction energy barriers between the four ligands for the B-O bond breaking decomposition for borates		215
5.58	EDA analysis of the interaction energy barriers between the four ligands for the Al-O bond breaking decomposition (TS2) for aluminates		216
5.59	Various electron affinity values for Al(OR)_3 and B(OR)_3		217
5.60	Spin density plot for $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$		218
5.61	Spin density plot for $\text{Ca}^+[\text{Al}(\text{hfip})_4]^-$		219
5.62	Spin density plot for $\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$		220
5.63	Spin density plot for $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$		221
5.64	Spin density plot for $\text{Ca}^+[\text{B}(\text{hfip})_4]^-$		222
5.65	Spin density plot for $\text{Ca}^+[\text{B}(\text{hftb})_4]^-$		223

List of Tables

2.1	Energy and charge errors for $\text{Li}_{29}\text{H}_{29}$ sparse and $\text{Li}_{32}\text{H}_{32}$ non-sparse test sets . .	37
2.2	Energy statistics for a cubic $\text{Li}_{32}\text{H}_{32}$ after geometry optimization	41
2.3	A charge separation example for neutral $\text{Li}_{15}\text{H}_{15}$	45
2.4	Comparison across NN models with different radial cutoff distances for $\text{Li}_{29}\text{H}_{29}$ sparse test set	52
2.5	Comparison across NN models with different radial cutoff distances for $\text{Li}_{32}\text{H}_{32}$ non-sparse test set (Models are the same as those in Table 2.4)	53
2.6	Energy statistics for a cubic $\text{Li}_{32}\text{H}_{32}$ after geometry optimization (Models are the same as those in Table 2.4)	53
2.7	Performance of GAP models (trained with unique geometries with equilibrium charges, more details below) on the $\text{Li}_{29}\text{H}_{29}$ sparse and $\text{Li}_{32}\text{H}_{32}$ non-sparse test sets (energy MAE in kcal/mol/atom).	54
2.8	Performance of GAP models on $\text{Li}_{32}\text{H}_{32}$ geometry optimizations. Absolute energies are shown for the optimized geometries in each force field.	54
3.1	Energy, charge, and spin errors for the larger Li-organic test set	77
4.1	The reaction barrier for water splitting on a Li cluster	107
5.1	Selected parameters (The Laplacian of the density $\nabla^2\rho$, density at the Bond Critical Point $\rho @\text{BCP}(3,-1)$, Mayer bond order m , Bader atomic charges q) for the QTAIM analysis of the starting complexes)	161

Acknowledgments

Sometimes I wonder how my life would have been if I did not step into the PhD program at UC Berkeley. Life would be different, but I am confident that I would not have learned or grown this much as an independent researcher in these five and a half years. I have significantly benefited from the world-leading academic environment and the talented people around me. During PhD, I experienced one of the most challenging times in my life when I had to switch research groups. I was not sure what was waiting for me when my research direction turned 180 degrees around from experimental work in organic chemistry to pure computation and theory. Still, I was lucky that everything worked out.

First and foremost, I would like to thank my PhD advisor, Prof. Kristin Persson. She is a star in the materials science field. Every time I listen to her talking about materials discovery, I can strongly feel how modern data-driven technologies change how novel materials are discovered, from dumb trial-and-error approaches to fully automated pipelines. I was constantly amazed by her breadth of knowledge during any meeting. She leads a group of such diversity, and her students' projects include solid-state DFT, molecular DFT, MD, and ML in terms of methodologies, materials discovery, electrolyte design, and surface catalysis in terms of subjects. No need to mention that she also manages the Material's project, an open-source platform for materials calculated with first principles methods, which millions of scientists love! It is incredible how she could navigate the different fields and give constructive guidance.

I want to thank all the lab members in the Persson research group and all my friends in the Department of Chemistry. Working with such a group of talented and kind people who are enthusiastic about science and discussing potential solutions to complex problems together was a pleasure. In particular, I am thankful to Dr. Samuel Blau, Dr. Mingjian Wen, Evan Walter Clark Spotte-Smith, Hetal Patel, Dr. Tingzheng Hou, Dr. Hui Zheng, Dr. Ruoxi Yang, Dr. Matthew Horton, Alex Epstein, Oxana Andriuc, and Dr. Trevor Seguin for their helpful discussions that gave me valuable insights on how to approach my daily research problems. I am also grateful to the administrative assistants in the Persson group, Alice Muller and Sonia Dominguez, and my experimental collaborators at Argonne National Lab, Dr. Noel Leon and Dr. Chen Liao. It was wonderful to see how combined theory and experiments can provide valuable insights into electrolyte design problems. I want to express gratitude to my family for their love and unconditional support. Due to various reasons, e.g., visa issues and the pandemic, I have only seen them once during my entire PhD; this was the longest I have been away from home. My family provided numerous support over the years, and thanks to them, I could concentrate on my PhD study without worrying about many things. Finally, I have to thank my boyfriend, David Small, for his love and support during difficult times when I had to switch research groups. Without his support and encouragement, I could not have stepped into the computational research world.

Chapter 1

Introduction

I Overview

Since the birth of the industrial revolution, the CO₂ emissions from fossil fuel combustion and industrial processes have increased dramatically. In 2000, a total of 34.81 billion metric tons of CO₂ was released; from 2000-2019, it was further increased by 45%, reaching 34.81 billion metric tons. [1] As such, replacing fossil fuels with cleaner and renewable energy resources is urged. Rechargeable batteries play a vital role in a green society for energy storage, consumption, and transportation. Featured by the 2019 Nobel Prize in Chemistry, Li-ion battery (LIB) technology is one of the most revolutionary technologies that have completely reshaped people’s daily life. The global lithium-ion battery (LIB) market is estimated to be USD 44.5 billion in 2022 and is projected to reach USD 135.1 billion by 2031. [2] Despite the success of LIBs, larger and larger scale applications set higher requirements, such as higher volumetric and gravimetric energy density, higher rate capability, longer cycle life, safety, and lower cost. Therefore in the past decade, tremendous research efforts have been dedicated to improving the current Li-ion technology and developing new battery chemistries.

A battery is a complex multi-component system, encompassing cathode, anode, electrolyte, separator, collector, etc. The complicated electrochemically coupled transport processes across a wide range of time and length scales constitute a hurdle for a complete understanding of the structure-property relationship between battery materials, electrolytes, and the overall electrochemical performance. The traditional way of conducting battery research relies heavily on trial and error, requiring extensive research resources and sophisticated domain knowledge and experience. In recent years, data-driven techniques have emerged as the fourth paradigm of materials research in addition to the existing empirical, model-based, and computation-based science. [3–5] They open up new avenues for accelerated materials discovery and design through both forward models (prediction of molecular or materials properties) and inverse models (discovery and design of novel materials).

Although the introduction of data-driven methods to chemical/materials science is still recent, there already exist some successful applications: prediction of new stable materials, [6–8], prediction of various materials properties, [9–14] reducing the number of required first-principle calculations, [15] or the computational cost of such calculations.[16]

At the core of data-driven methods lies the need for high-quality data. In the case of computational science, this means we need an accurate description of the chemical system of interest. Especially in the search for methods to improve battery performance, theoretical studies addressing the structure-property relationship in batteries on an atomic level play an increasingly important role. Ab initio or quantum chemistry (QC) methods — either based on wavefunction descriptions or density functional theory (DFT) — provide a full quantum mechanical treatment of the atomic system within certain approximations. These methods are ideal for studying complicated electrochemical processes, such as electron transfer processes and the consecutive radical reactions where open-shell, charged species are typically

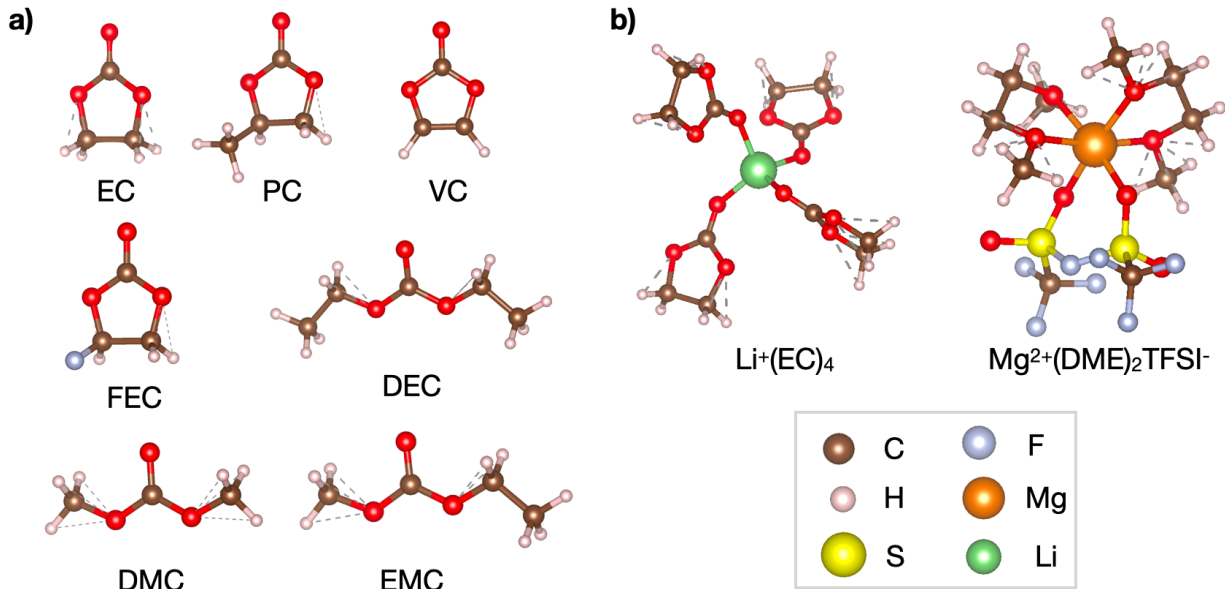


Figure 1.1: (a) Typical organic electrolytes for LIB. (b) Example molecular cluster structures of Li^+ solvated in EC (left) and Mg^{2+} with $\text{Mg}[\text{TFSI}]_2$ salt solvated in DME (right).

involved. There are two types of computational modeling techniques that are typically used for studying the battery system properties, (1) static calculations and (2) dynamical simulations. If DFT is employed, they would be static DFT calculations and ab initio molecular dynamics (AIMD). [17] The former method is suited for providing in-depth insights on the specific chemical process(es) of interest by evaluating the ground state energies or properties of molecules, fragments, clusters, and crystal structures, providing access to redox stabilities, electrolyte decomposition mechanisms, reaction or transport energies or barriers. Within the latter method, the system is propagated in time by numerically integrating Newton's equation of motion on the potential energy surface (PES) of typically an ensemble of molecules, affording statistical properties, such as structural and transport properties of the electrolyte.

The main goal of this thesis is to use data-driven approaches combined with DFT to study the complex reactions involved in the battery systems, e.g., electrolyte degradation and solid electrolyte interphase (SEI) formation.

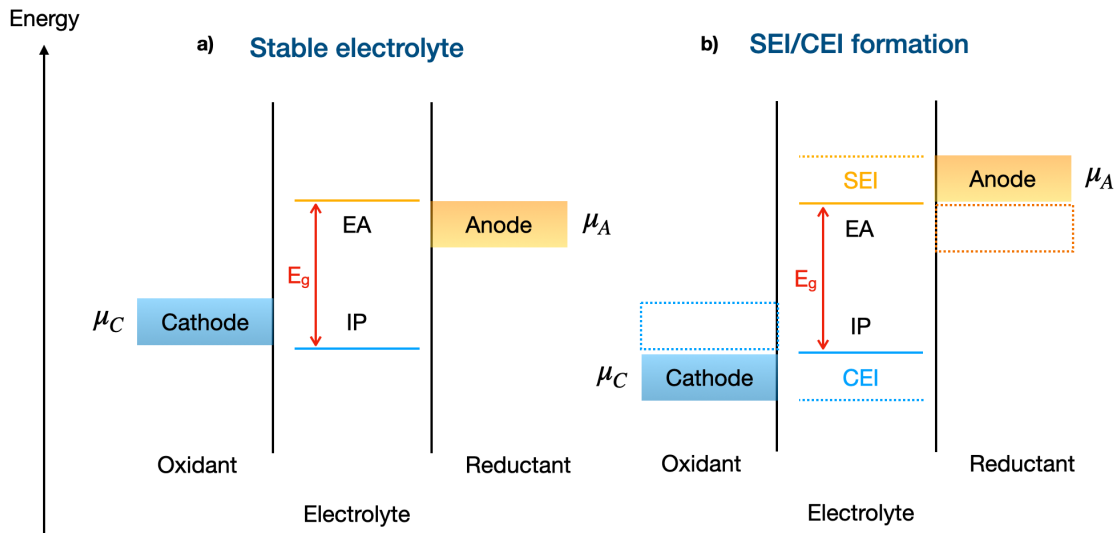


Figure 1.2: Electrochemical stability window.

II Chemical processes in battery systems

2.1 Battery operation mechanisms

A LIB typically contains a lithium or carbon-based anode, a lithium transition metal oxide cathode (e.g., lithium cobalt oxide (LiCoO_2)), and a polyolefin-based separator that is placed between anode and cathode and soaked with the electrolyte. In most cases, the electrolyte consists of an organic solvent and a Li^+ conducting salt. The solvent is usually a mixture of cyclic and linear organic carbonates, e.g., ethylene carbonate (EC), propylene carbonate (PC), dimethyl carbonate (DMC), diethyl carbonate (DEC), and ethyl methyl carbonate (EMC). Figure 1.1(a) shows some typical organic solvents used in LIB. As conducting salt, lithium hexafluorophosphate (LiPF_6) or lithium perchlorate (LiClO_4) can be used. During discharge, Li^+ moves through the electrolyte from the anode to the cathode. The corresponding electrons flow through an external electric circuit, and the electrical power is supplied to the connected device. The charging process is the reverse: Li^+ are incorporated into the anode from the cathode, and the electrical power is stored in the battery.

2.2 Electrolyte decomposition

Figure 1.2 is a schematic illustration of the relative electron energies in the electrodes and the electrolytes in a standard battery cell. The cathode is the oxidant, and the anode is the reductant. The energy separation E_g between the adiabatic ionization potential (IP)

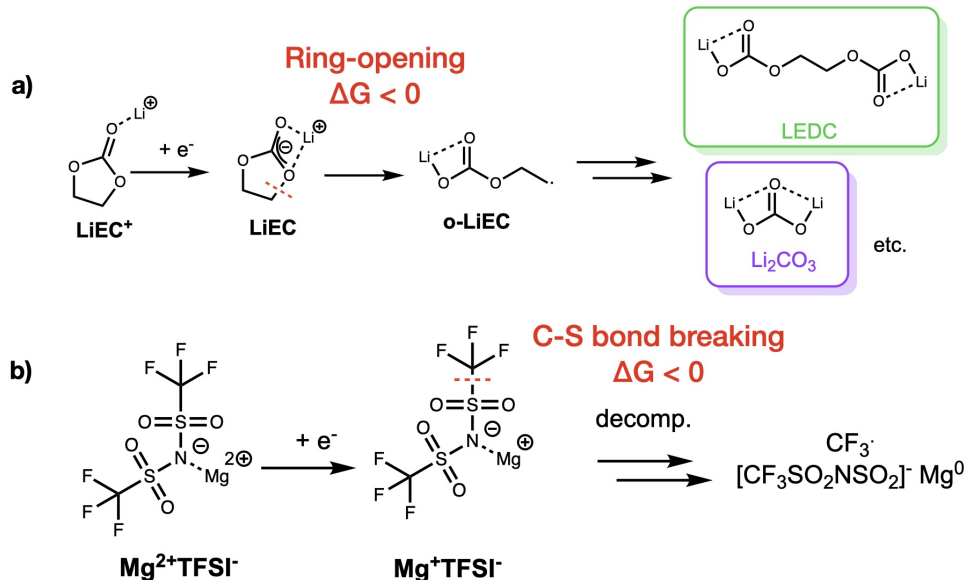


Figure 1.3: (a) The mechanism for reductive decomposition of the EC electrolyte in LIB. (b) The mechanism for reductive decomposition of the $\text{Mg}[\text{TFSI}]_2$ salt in Mg-ion batteries.

and electron affinity (EA) of the electrolyte dictates the electrochemical stability window of the electrolyte. Note that some literature uses the gap between the HOMO and LUMO of the electrolyte as the stability window, which is a similar concept but less rigorous than IP and EA. Figure 1.2 (a) shows the scenario when the electrolyte is thermodynamically stable, which means the electrochemical potentials of both electrodes (μ_A for the anode and μ_C for the cathode, referenced to e.g. Li/Li^+) locate within the E_g of the electrolyte. If this condition cannot be satisfied (Figure 1.2 (b)), reduction/oxidation of the electrolyte will happen, creating solid-electrolyte interphase (SEI) and cathode-electrolyte interphase (CEI) on the anode and cathode respectively. Such reduction/oxidation reactions are expected to carry on during battery cycling unless the as-formed passivation layer creates a barrier for electron transfer.

For elemental Li^0 and graphite, the two most common negative electrodes, the fermi energy of the anode, μ_A , lies above the EA(LUMO) of practical, known nonaqueous electrolytes, creating a thermodynamic driving force for the electrochemical reduction of electrolytes at the surface of these negative electrodes.[18, 19] Anions and anion radicals produced by the salt anion decomposition could further trigger a cascade of reactions that promote the decomposition of other molecules.

Two well-known reductive decomposition mechanisms, for the EC electrolyte in LIB and the $\text{Mg}[\text{TFSI}]_2$ salt in MIB, respectively, are shown in Figure 1.3. Some typical molecular

structures for the solvation shells of these two systems are given in Figure 1.1 (b). For Li^+EC , a shoulder bond in the ring (shown in red in Figure 1.3 (a)) is energetically favorable to break upon reduction, creating ring-opened LiEC , which further leads to the formation of lithium ethylene dicarbonate (LEDC), Li_2CO_3 etc. [20] In the $\text{Mg}[\text{TFSI}]_2$ case, one C-S bond in the reduced contact ion pair Mg^+TFSI^- is exergonic to break (Figure 1.3 (b)), likely decomposing into the fragments $\text{CF}_3\cdot$ and $[\text{CF}_3\text{SO}_2\text{NSO}_2]-\text{Mg}^0$. [21] The decomposition mechanisms for a novel family of salts, the weakly-coordinating fluorinated alkoxyborate and alkoxyaluminate salts for multivalent ion battery applications will be discussed in depth in chapter 5.

2.3 Solid-electrolyte interphase (SEI) formation

In the field of battery research, understanding electrolyte interfaces are of paramount importance because their properties crucially affect the overall performance and aging of the battery cell. A famous example is the solid-electrolyte interphase (SEI), a thin surface passivation film that grows in situ at the anode/liquid organic electrolyte interface due to the reductive decomposition of electrolytes (both solvents and salts). In the case of LIB, SEI typically forms onto the electrode surface at potentials below 1 V vs Li/Li^+ (0.7 V vs Li/Li^+ for most electrolytes).[22] An ideal SEI should be able to shut down electron transport and thus prevent continuous reductive degradation of the electrolytes while allowing the working ions to pass through.

These surface films usually form through out-of-equilibrium interfacial reactions. Such reactions are notoriously difficult to characterize, and robust prediction of the chemical evolution and associated functionality of the resulting surface film is one of the grand challenges of materials chemistry. Progress in understanding the SEI layers has been achieved using techniques such as X-ray photoelectron spectroscopy (XPS), [23, 24] X-Ray Diffraction (XRD),[25, 26] Fourier transform infrared (FTIR) spectroscopy, [27, 28] nuclear magnetic resonance spectroscopy (NMR), [29, 30] cryo-scanning transmission electron microscopy (cryo-STEM), [31–33] Atomic Force Microscopy (AFM), [34–36] etc. Despite significant research work, there is still limited understanding of what SEI layers consist of, by what mechanisms they are formed, and how they influence SEI properties and battery performance.

Our current understanding of the SEI comes from a progressive buildup of knowledge in the past decades. In 1983, Pelled proposed a double-layer SEI structure with a thin, compact layer close to the electrode and a thick porous secondary layer closer to the electrolyte, without mentioning any compositional information. [37] Nazri and Muller[25, 38] first showed the presence of Li_2CO_3 and oligomers (or polymers) in SEI on lithium surfaces using XPS, FTIR, and in situ XRD. Later Aurbach et al.[39] revealed through FTIR/XPS that lithium alkyl carbonates (dicarbonates or semicarbonates) instead of Li_2CO_3 are the main SEI components resulting from carbonate solvent decomposition. Kanamura et al.[40] reported by XPS that the compact layer near the lithium surface consists of LiF and Li_2O ,

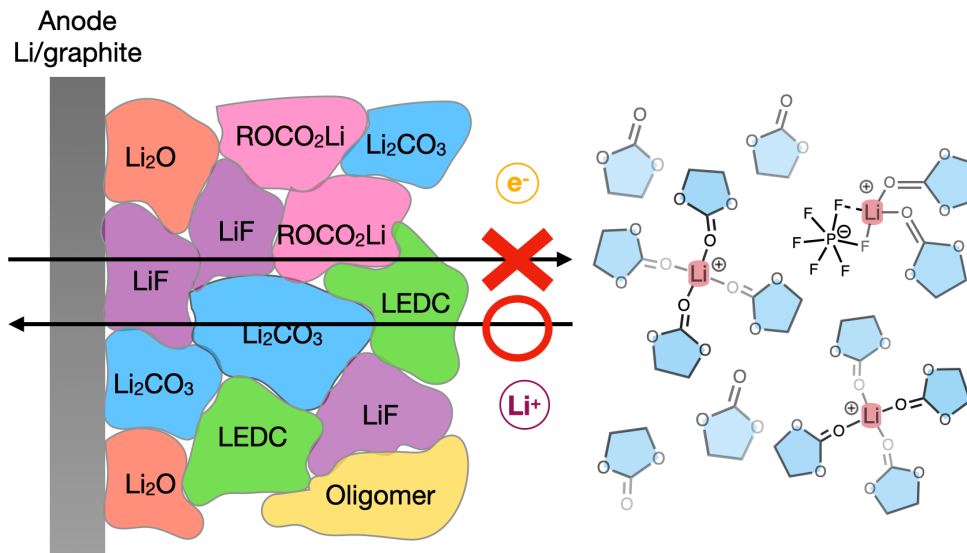


Figure 1.4: Schematic presentation of the mosaic structure of the SEI on Li/graphite anode (proposed by Peled [41]). An ideal SEI should be ionic-conducting but electron-insulating.

whereas the porous layer above contains LiF and organic compounds. Peled et al. further summarized previous findings and reported the mosaic structure of the SEI[41] (a schematic figure is shown in Figure 1.4), which is frequently mentioned in the literature today. In short, the SEI mosaic structure consists of multiple inorganic and organic products from electrolyte decomposition: compact layers are formed of inorganic species like Li_2O , Li_2CO_3 , and LiF near the anode surface, whereas the domains near the electrolyte consist mainly of oligomer species and semicarbonates. [27, 40, 41]

In the case of the popular carbonate solvents, Li_2CO_3 , lithium dicarbonates ($(\text{ROCO}_2\text{Li})_2$), and semicarbonates (ROCO_2Li) predominate in the SEI. Specifically, it has been long believed that EC forms a stable SEI composed mainly of LEDC.[26, 27, 29, 39, 42–50] In 2019, work by Wang et al. [51] cast doubt on the LEDC characterization: it was suggested that common LEDC standards[46, 48, 50] are comprised of lithium ethylene monocarbonate (LEMC). This controversy will be investigated and discussed thoroughly in Chapter 4.

The composition and the stability of the SEI depend primarily on the formulation of the electrolyte (solvent, additive, and salt) and the nature of the electrode surface, but also on some other factors such as the charge and discharge rate, etc.[22] In reality, a slight change in the electrolyte chemical structure could result in a dramatic change in battery performance. For instance, propylene carbonate (PC), which has a single methyl group difference from the commonly used electrolyte ethylene carbonate (EC), cannot form a stable SEI and leads to the exfoliation of the graphitic anode because of co-intercalation.[52]

Besides tuning the formulation of the electrolyte, a few strategies can be implemented to improve the cycling ability of LIB systems. For instance, the surface termination can be modified to reduce the material's reactivity towards the electrolyte: it has been shown that electrochemical performances can be improved through mild oxidation,[53] deposition of metals oxides,[54] various coatings techniques.[55] It is also possible to deposit an artificial passivation layer on top of the anode to protect the electrolytes from reacting with the surface.[56, 57]

III Modeling techniques

3.1 Density functional theory (DFT)

Density functional theory (DFT) is undoubtedly the most widely used and cost-effective electronic structure method for studying ground-state molecular and material properties. Within DFT, the high-dimensional many-body wavefunction $\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n)$ is replaced by a simple variable, the electron density $n(\mathbf{r})$ in the 3-dimensional space, and the total energy along with other ground-state properties are defined as functionals of the electron density. This has made the problem much simpler: as Kohn noted in his Nobel lecture in 1998, DFT “has been most useful for systems of very many electrons where wave function methods encounter and are stopped by the exponential wall”. [58] The basis of DFT lies in the Hohenberg-Kohn theorem,[59] stating that the external potential is a functional of the ground-state density and that the ground-state energy/density can be determined by variationally minimizing the total energy with respect to the density. If we define

$$F[n(\mathbf{r})] = \langle \Psi | T + U | \Psi \rangle \quad (1.1)$$

where T is the kinetic energy operator and U is the coulomb operator,

$$T = \frac{1}{2} \int \nabla \psi^*(\mathbf{r}) \nabla \psi(\mathbf{r}) d\mathbf{r} \quad (1.2)$$

$$U = \frac{1}{2} \int \frac{1}{|\mathbf{r} - \mathbf{r}'|} \psi^*(\mathbf{r}) \psi^*(\mathbf{r}') \psi(\mathbf{r}) \psi(\mathbf{r}') d\mathbf{r} d\mathbf{r}' \quad (1.3)$$

we can write the energy functional as

$$E_v[n] = \int n(\mathbf{r}) V_{ext}(\mathbf{r}, \mathbf{R}) d\mathbf{r} + F[n] \quad (1.4)$$

where $V_{ext}(\mathbf{r}, \mathbf{R}) = \sum_I Z_I e^2 / |\mathbf{r} - \mathbf{R}_I|$ is the total Coulomb attraction between one electron and the nuclei.

It has been proven that $E_v[n]$ is at the minimum value at the correct ground state density, under the constraint that the density integrates to the correct number of particles.

$$\int n(\mathbf{r})d\mathbf{r} = N \quad (1.5)$$

The most commonly used method within DFT is the Kohn-Sham formalism. In the Kohn-Sham treatment, the total energy is expressed as a functional of m mutually orthonormal single-particle electron orbitals $\psi_i(\mathbf{r})$, $i = 1, \dots, m$. The orbitals are related to the electronic density according to $n(\mathbf{r}) = \sum_i f_i |\psi_i(\mathbf{r})|^2$, where f_i is the occupation of the i th orbital. For the spin-unrestricted case, $m = N$ and $f_i = 1$, whereas for the spin-restricted case, $m = N/2$ and $f_i = 2$. The total energy is given by

$$\begin{aligned} F[\{\psi\}, \mathbf{R}] = & \frac{1}{2} \sum_i f_i \int \nabla \psi_i^*(\mathbf{r}) \nabla \psi_i(\mathbf{r}) d\mathbf{r} \\ & + \frac{1}{2} \int \frac{1}{|\mathbf{r} - \mathbf{r}'|} \psi^*(\mathbf{r}) \psi^*(\mathbf{r}') \psi(\mathbf{r}) \psi(\mathbf{r}') d\mathbf{r} d\mathbf{r}' + E_{xc}[n] \\ & + \int n(\mathbf{r}) V_{ext}(\mathbf{r}, \mathbf{R}) d\mathbf{r}. \end{aligned} \quad (1.6)$$

Minimizing this functional over orbitals subject to an orthonormality condition $\langle \psi_i | \psi_j \rangle = \delta_{ij}$ yields the ground state energy.

Although Eq. 1.6 is formally exact, the exact form of the exchange/correlation functional, which accounts for all the quantum-mechanical many-body effects, is unknown and this is where the approximations in DFT functionals are coming from. Therefore, the quality of the electronic structure calculation relies on the quality of the approximation used for E_{xc} . The most common classes of approximations are the local-density approximations (LDA), [60] generalized-gradient approximations (GGA), [61, 62] meta-generalized-gradient approximations (meta GGA), [63] as well as hybrid functionals [64] where a portion of exact exchange from Hartree-Fock theory is incorporated to alleviate the self-interaction error. [60]

Static DFT calculations and AIMD simulations are especially suitable for studying the electrode-electrolyte interfacial reactions, such as solvent/salt decomposition and the early-stage SEI formation, as discussed in Sec II. Early pioneering work from Balbuena and coworkers based on static DFT calculations clarified the reduction mechanisms of common organic solvents and additives in LIB. [20, 65, 66] Since then, DFT calculations have been extensively used to study the decomposition mechanisms of different electrolyte compositions and how other factors could change the decomposition pathways, such as additives, electrode surfaces, temperature, etc. To list a few examples, DFT studies revealed that the electrochemical stability of electrolyte molecules varies in different environments. [67–69] Clarke-Hannaford et al. adopted AIMD simulations to study the decomposition mechanism of TFSI[−] and FSI[−] anions on the Li(001) surface at different temperatures. [70] AIMD results

from Liu et al. showed that the structure of the exposed Li metal facet significantly affects the Li_2S_8 decomposition, nucleation, and growth. [71] We need to note that although AIMD simulations are ideal in terms of accuracy, they can only access very limited time regimes, at most 100 ps for typical AIMD simulations; thus understanding chemical reactivities through AIMD simulations is challenging without advanced sampling techniques.

3.2 Classical force fields

Classical force fields (or empirical interatomic potentials) treat the individual atoms as classical particles. Since no electronic degree of freedom is considered and relative simple functional forms are used, MD simulations using classical force fields are very cheap and thus suitable for modeling the nonreactive dynamics of a multicomponent system consisting of (1000 - 1M particles) for a large time scale (10 ps to 1 s depending on the model, but typically on the order of ns). Classical MD simulations can therefore provide access to many important properties such as diffusion constants, ionic conductivity, viscosity, etc. for battery electrolytes.

In classical force fields, potential energy functions usually follow a fixed pattern. The force field describes the intramolecular and intermolecular potential energy of atoms in the form of mathematical functions:

$$U = U_{nb} + U_{bond} + U_{angle} + U_{torsion} \quad (1.7)$$

where U_{nb} represents interactions between atoms, and U_{bond} , U_{angle} , $U_{dihedral}$ represents bond interactions between atom pairs, angle interactions between triplets of atoms, and torsion interactions between quadruplets of atoms, respectively.

Detailed expressions of some widely used force fields, such as OPLS (optimized potentials for liquid simulations), [72] AMBER (assisted model building with energy refinement), [73] and CHARMM (chemistry at Harvard macromolecular mechanics), [74, 75] follow the formula below with some variations:

$$U_{nb} = \sum_{i < j} \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \quad (1.8)$$

$$U_{bond} = \sum_{bonds} K_r (r_{ij} - r_0)^2 \quad (1.9)$$

$$U_{angle} = \sum_{angles} K_\theta (\theta_{ijk} - \theta_0)^2 \quad (1.10)$$

$$U_{torsion} = \sum_{torsions} \sum_n K_{\phi,n} [1 + \cos(n\phi_{ijkl} + \delta_n)] \quad (1.11)$$

where ϵ_{ij} is the Lennard-Jone (LJ) well depth, σ_{ij} is the LJ radius, r_{ij} is the distance between two atoms, q_i and q_j are the charges of two atoms, ϵ_0 is the dielectric constant of a vacuum, K_r , K_θ , and $K_{\phi,n}$ are the force constants, r_θ is the equilibrium bond value, θ_{ijk} is the bond angle, θ_0 is the equilibrium angle value, n is the multiplicity, ϕ_{ijkl} is the dihedral angle, and δ_n is the phase. The parameters needed to construct the force fields can either be derived from experiments or fitted by ab-initio data for different atom types and interaction modes.

For liquid electrolytes, the presence of ions leads to strong local electric fields and often necessitates the use of so-called polarizable force fields. [76] While simulations using the above nonpolarizable force fields can predict decent molecular structures (e.g., solvation structures) in ionic systems, [77, 78] the predicted thermodynamic, dielectric and transport properties are often inconsistent with experiments, especially at higher salt concentrations.[78, 79] As examples of polarizable force fields, APPLE&P[80] (atomistic polarizable potentials for liquids, electrolytes, and polymers) and AMOEBA[81] (atomic multipole optimized energetics for biomolecular simulation) attempt to include more accurate descriptions of electrostatic interactions between molecules and the polarizability.

For instance, for the APPLE&P force field, U_{nb} takes the following form to account for induced dipoles: [80]

$$U_{nb} = \sum_{i < j} \left[A_{\alpha\beta} \exp(-B_{\alpha\beta} r_{ij}) - C_{\alpha\beta} r_{ij}^{-6} + D \left(\frac{12}{B_{\alpha\beta} r_{ij}} \right)^{12} + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right] - \frac{1}{2} \sum_i \vec{\mu}_i \cdot \vec{E}_i^0 \quad (1.12)$$

where $\vec{\mu}_i = \alpha_i \vec{E}_i^{tot}$ is an induced dipole at force center i , α_i is the isotropic atomic polarizability, $\vec{E}_i^{tot}$ is the total electrostatic field at the atomic site i due to permanent charges q_j and induced dipoles $\vec{\mu}_j$, ϵ_0 is the dielectric permittivity of vacuum, $\vec{E}_i^0$ is the electric field due to fixed charges only, $A_{\alpha\beta}$ and $B_{\alpha\beta}$ are the repulsion parameters and $C_{\alpha\beta}$ is the dispersion parameter for interaction between atoms i and j with atom types α and β .

Indeed, there is a significantly enhanced correlation between simulation vs. experiment data for the self-diffusion coefficients[80, 82] and electrolyte conductivity,[83, 84] obtained from simulations using such polarizable force fields.

Because the bond interaction in classical force fields is usually in the form of harmonic oscillators, as shown in Eqn. 1.9, as we dissociate a bond, the corresponding interaction energy increases significantly. This precludes the usage of classical force fields in probing chemical reactivities, i.e., breaking or forming chemical bonds. To solve this issue, ReaxFF is developed for modeling the interactions on the basis of bond orders to allow for bond forming and breaking in MD simulations.[85, 86]

Classical MD simulations are widely used in studying the bulk solvation structures, ionic transport mechanisms, and physicochemical properties of battery electrolytes as described below.

Solvation structures

According to the coordination structure between cations and anions, solvated cations can be classified as either free ions, solvent-separated ion pairs (SSIPs), contact ion pairs (CIPs), and aggregates (AGGs), which are in the order of increasing number of anions participating in the cation solvation. Highly concentrated electrolytes have unique solution structures totally different from those of the corresponding dilute electrolytes. Generally, cations in free or SSIP states dominate in low-concentration electrolytes (LCEs),[87], and CIPs and AGGs take over in high-concentration electrolytes (HCEs). [88–91] MD simulations can provide insights into the fractions of various cation structures by counting the number of free ions, SSIPs, CIPs, and AGGs. This is relevant to the dissolution and dissociation degree of salts because anions compete for the coordination of cations with solvents. And a high dissolution and dissociation degree of salts are generally beneficial to implement ionic transport and desirable rate performances for rechargeable battery electrolytes.

Ion transport mechanisms

Generally, the transport mechanisms in liquid electrolytes can be classified into two categories: vehicular diffusion, where ions and species diffuse in their solvation shells in a concerted fashion,[92] and structural diffusion, where ionic transport is accompanied by desolvation and resolution processes (also referred to as the ion-hopping mechanism).[93, 94] The actual diffusion mechanism could also be mixed, where partial components depart from the solvation shell and the remaining migrates along with the working ion. In MD simulations, an understanding towards the ion transport mechanisms can be obtained through a comparison between the length scale of collective diffusion and that of the solvation shell.

Physicochemical properties

MD simulations can also deliver understandings to many physicochemical properties of the electrolytes at the atomic level. For example, the Green-Kubo relation derived from the linear response theory gives properties such as the diffusion coefficient and viscosity.[95] The Nernst-Einstein equation further relates the diffusion coefficient to ionic conductivity. The dielectric constants of liquids, which determines their capability of separating cations from anions, can be calculated from MD simulations based on the fluctuation in the total dipole moment of the box.[96]

3.3 Machine learning potential energy surface

As stated in the previous sections, one would prefer to use DFT whenever possible to study battery materials. However, there exist many situations where statistical sampling or dynamical behaviors are of interest, where MD simulations are preferable. MD simulations based on classical force fields are, however, limited in terms of the accuracy they can achieve and the capability of modeling chemical reactions (except for the reactive force fields). AIMDs can model chemical reactions but are usually too computationally expensive to reach the time scale to observe the desired reactions.

In the last few years, with the advent of machine learning techniques, a new class of potentials has emerged that rely on complex functions to describe the potential energy landscape, usually trained on data generated through DFT calculations. [97–100] Given some suitable input features/descriptors to encode the geometrical characters of the molecules, which usually also obey symmetries, a sophisticated non-linear function can be learned to match the energies, forces, and some other molecular properties obtained from first-principles calculations. If trained properly, this can, in principle, produce an ML force field with DFT accuracy but only at a cost similar to classical force fields. In contrast to empirical potentials (which are also often fitted to DFT data), ML models impose no particular functional form, thus they can adapt more easily and flexibly to the training data, although this could potentially cause overfitting. It is impressive that MD simulations of 100 million atoms have already been achieved with ML force fields for water and copper systems. [101, 102] Clearly, if ML potentials can describe the correct atomistic interactions, it opens up the possibility of highly accurate investigations of battery materials, electrolytes, and even electrolyte-electrode interphase on larger length and time scales.

In ML-based PESs, typically the total energy E of a chemical system is represented as a sum of atomic energies E_i : [98]

$$E = \sum_i^N E_i. \quad (1.13)$$

Atomic energies E_i are predicted through a regression model such as neural networks (NN) or gaussian process regression (GPR) based on descriptors that can represent the local geometrical environment around each atom. Decomposing the total energy into atomic energy contributions affords several important advantages, such as size-extensivity and near linear-scaling cost. Global descriptors are sometimes used, such as Coulomb matrix, [103, 104] in which case the total energy of the entire molecule is predicted, but those models are not designed to obey the energy behavior at the dissociation limit (if a sufficiently large distance separates two molecules, the energy of the complex should be the same as the sum of the energies of the two isolated molecules computed with the same method). Many molecular descriptors are designed up-to-date; prominent ones include the atom-centered symmetry functions (ACSF), [105, 106], smooth overlap of atomic positions (SOAP), [107] bispectrum, [107] the FCHL descriptor, [108] the Many Body Tensor Representation descriptor, [109] etc.

In recent years, it came to the realization that many traditional hand-crafted descriptors have some significant limitations because they often do not satisfy the one-to-one map, i.e., the representation does not uniquely determine the molecule. [110] The existence of indistinguishable configurations affects the expressive power of ML models because the model cannot give distinguishable predictions for such configurations by construction. Further, as an atom does not "see" the information from the atoms far away from itself (beyond the cut-off distance), long-range information is completely missed, and this would cause problems for incorporating electrostatic interactions, as they are inherently long-range. This aspect rep-

resents a challenge for modeling battery chemistry, as electrostatic interactions are essential in ionic systems. Structures that are open-shell or at a non-zero total charge could also contribute to the complex electrochemical reaction pathways, further necessitating an accurate description of long-range electronic structure, as the optimal electron density distribution will be different for molecules at different total charges.

One way of tackling the first problem is to construct higher-order descriptors that can capture beyond 3-body correlations.[110] Schemes such as MTP,[111] ACE,[112] and NICE[113] can in principle give a complete representation if expanded in a high enough order. Another method is to design "learnable" descriptors and to construct complicated architectures such as graph neural networks (or MPNN, message passing neural networks) where the descriptors are automatically determined/learned from the geometric structure of the entire molecule. A lot of the degeneracies (e.g., the degenerate pair discussed in ref. [110]) can be alleviated once the information about the whole molecule is redistributed onto each atom during training. In MPNN, each atom is initially described by a feature vector, which is iteratively updated in an interaction block using information about neighboring atoms (or even all atoms through an attention mechanism[114, 115]), such as distances and their feature vectors at the current step. This approach also provides a solution for the second long-range problem. It was later shown that global features such as total charge/spin could also be incorporated into MPNN models, where the appropriate distribution of atomic partial charges/spins can be predicted through multiple steps of iterative feature updates, and energies for molecules with different total charges/spins can be predicted. [115, 116] However, describing systems with variable electronic structure or where long-range interactions are essential is still an ongoing challenge in general. The proposed ML models need sufficient testing, not only on the well-known QM9 dataset,[117] but also on how well they can perform for realistic systems in terms of geometry optimizations, MD simulations, etc. Chapters 2 and 3 will discuss a different method for approaching the long-range problem: to train nonlinear energy functional of nuclear position and electron populations (electron density projected onto each atom).

Although a lot of success has been achieved with ML force fields, and it was claimed that ML potentials could show quantum mechanical accuracy without explicitly including electrons, [118]the majority of the actual applications of such models are still on systems that are more or less homogeneous such that the electronic information can be more likely captured by a descriptor that can sufficiently describe the local geometrical environment and short-range interactions. For the application of ML-based potential for the investigation of battery electrodes and solid electrolytes, there are a number of recent studies on systems that are semi-homogeneous, e.g. Li-C system [119], Li-S alloys, [120], $\text{Li}_{10}\text{GeP}_2\text{S}_{12}$ -type superionic conductors. [121] SEI has been typically modeled by force-field methods.[122] The complexity of the SEI requires ML potentials to reproduce polarization effects and long-range anisotropic electrostatic interactions and polarization. Furthermore, in the SEI formation, structures that are far from equilibrium geometry and structures that are open-shell or at a non-zero total charge could contribute to the complex reaction pathways. These not only

pose a challenge for data generation but also require the ML model to contain the correct physics to handle non-equilibrium, open-shell, and charged species simultaneously. Therefore the real application of ML potentials in SEI models has not been seen; only a few examples exist for simulating the solid-water interface where the reactions involved are simpler, i.e., proton transfer reactions. [123–125]

Bibliography

- [1] *Historical carbon dioxide emissions from global fossil fuel combustion and industrial processes from 1750 to 2020*. accessed 09/2022. URL: <https://www.statista.com/statistics/264699/worldwide-co2-emissions>.
- [2] *Lithium-ion Battery Market by type*. <https://www.marketsandmarkets.com/Market-Reports/lithium-ion-battery-market-49714593.html>. accessed 09/2022.
- [3] Ankit Agrawal and Alok Choudhary. “Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science”. In: *Appl Materials* 4.5 (2016), p. 053208.
- [4] Jonathan Schmidt et al. “Recent advances and applications of machine learning in solid-state materials science”. In: *Npj Comput. Mater.* 5.1 (2019), pp. 1–36.
- [5] Keith T Butler et al. “Machine learning for molecular and materials science”. In: *Nature* 559.7715 (2018), pp. 547–555.
- [6] Kevin Ryan, Jeff Lengyel, and Michael Shatruk. “Crystal structure prediction via deep learning”. In: *Journal of the American Chemical Society* 140.32 (2018), pp. 10158–10168.
- [7] Wei Li, Ryan Jacobs, and Dane Morgan. “Predicting the thermodynamic stability of perovskite oxides using machine learning models”. In: *Computational Materials Science* 150 (2018), pp. 454–463.
- [8] Felix A Faber et al. “Machine learning energies of 2 million elpasolite (A B C 2 D 6) crystals”. In: *Physical review letters* 117.13 (2016), p. 135502.
- [9] Jesús Carrete et al. “Finding unprecedentedly low-thermal-conductivity half-Heusler semiconductors via high-throughput materials modeling”. In: *Physical Review X* 4.1 (2014), p. 011019.
- [10] Tian Xie and Jeffrey C Grossman. “Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties”. In: *Physical review letters* 120.14 (2018), p. 145301.
- [11] Olexandr Isayev et al. “Universal fragment descriptors for predicting properties of inorganic crystals”. In: *Nature communications* 8.1 (2017), pp. 1–12.

- [12] Al’ona Furmanchuk, Ankit Agrawal, and Alok Choudhary. “Predictive analytics for crystalline materials: bulk modulus”. In: *RSC advances* 6.97 (2016), pp. 95246–95251.
- [13] Han Wei et al. “Predicting the effective thermal conductivities of composite materials and porous media by machine learning methods”. In: *International Journal of Heat and Mass Transfer* 127 (2018), pp. 908–916.
- [14] Shenghong Ju et al. “Designing nanostructures for phonon transport via Bayesian optimization”. In: *Physical Review X* 7.2 (2017), p. 021024.
- [15] Randy Jalem et al. “Bayesian-driven first-principles calculations for accelerating exploration of fast ion conductors for rechargeable battery application”. In: *Scientific reports* 8.1 (2018), pp. 1–10.
- [16] Haikuo Zhang et al. “Ultra-fast and accurate binding energy prediction of shuttle effect-suppressive sulfur hosts for lithium-sulfur batteries using machine learning”. In: *Energy Storage Materials* 35 (2021), pp. 88–98.
- [17] Richard Car and Mark Parrinello. “Unified approach for molecular dynamics and density-functional theory”. In: *Physical review letters* 55.22 (1985), p. 2471.
- [18] John B Goodenough and Youngsik Kim. “Challenges for rechargeable Li batteries”. In: *Chemistry of materials* 22.3 (2010), pp. 587–603.
- [19] Magali Gauthier et al. “Electrode–electrolyte interface in Li-ion batteries: current understanding and new insights”. In: *The journal of physical chemistry letters* 6.22 (2015), pp. 4653–4672.
- [20] Yixuan Wang et al. “Theoretical studies to understand surface chemistry on carbon anodes for lithium-ion batteries: reduction mechanisms of ethylene carbonate”. In: *J. Am. Chem. Soc.* 123.47 (2001), pp. 11708–11718.
- [21] Nav Nidhi Rajput et al. “The coupling between stability and ion pair formation in magnesium electrolytes from first-principles quantum mechanics and classical molecular dynamics”. In: *Journal of the American Chemical Society* 137.9 (2015), pp. 3411–3420.
- [22] Alexandre Chagnes and Jolanta Swiatowska. “Electrolyte and solid-electrolyte interphase layer in lithium-ion batteries”. In: *Lithium Ion Batteries-New Developments* (2012), pp. 145–172.
- [23] Fredrik Lindgren et al. “SEI formation and interfacial stability of a Si electrode in a LiTDI-salt based electrolyte with FEC and VC additives for Li-ion batteries”. In: *ACS applied materials & interfaces* 8.24 (2016), pp. 15758–15766.
- [24] Anna M Andersson et al. “Electrochemically lithiated graphite characterised by photoelectron spectroscopy”. In: *Journal of Power Sources* 119 (2003), pp. 522–527.

- [25] Gholamabbas Nazri and Rolf H Muller. “Composition of surface layers on Li electrodes in PC, LiClO₄ of very low water content”. In: *Journal of the Electrochemical Society* 132.9 (1985), p. 2050.
- [26] Feifei Shi et al. “A catalytic path for electrolyte reduction in lithium-ion cells revealed by in situ attenuated total reflection-Fourier transform infrared spectroscopy”. In: *Journal of the American Chemical Society* 137.9 (2015), pp. 3181–3184.
- [27] D Aurbach et al. “A comparative study of synthetic graphite and Li electrodes in electrolyte solutions based on ethylene carbonate-dimethyl carbonate mixtures”. In: *Journal of the Electrochemical Society* 143.12 (1996), p. 3809.
- [28] Guorong V Zhuang et al. “A study of electrochemical reduction of ethylene and propylene carbonate electrolytes on graphite using ATR-FTIR spectroscopy”. In: *Electrochemical and Solid State Letters* 8.9 (2005), A441.
- [29] Alison L Michan, Michal Leskes, and Clare P Grey. “Voltage dependent solid electrolyte interphase formation in silicon electrodes: monitoring the formation of organic decomposition products”. In: *Chemistry of Materials* 28.1 (2016), pp. 385–398.
- [30] Alison L Michan et al. “Solid electrolyte interphase growth and capacity loss in silicon electrodes”. In: *Journal of the American Chemical Society* 138.25 (2016), pp. 7918–7931.
- [31] AM Andersson and Kristina Edström. “Chemical composition and morphology of the elevated temperature SEI on graphite”. In: *Journal of the Electrochemical Society* 148.10 (2001), A1100.
- [32] Mengyun Nie et al. “Lithium ion battery graphite solid electrolyte interphase revealed by microscopy and spectroscopy”. In: *J. Phys. Chem. C* 117.3 (2013), pp. 1257–1267.
- [33] Michael J Zachman et al. “Cryo-STEM mapping of solid–liquid interfaces and dendrites in lithium-metal batteries”. In: *Nature* 560.7718 (2018), pp. 345–349.
- [34] Soon-Ki Jeong et al. “Surface film formation on graphite negative electrode in lithium-ion batteries: AFM study in an ethylene carbonate-based solution”. In: *Journal of The Electrochemical Society* 148.9 (2001), A989.
- [35] Yasuhiro Domi et al. “In situ AFM study of surface film formation on the edge plane of HOPG for lithium-ion batteries”. In: *The Journal of Physical Chemistry C* 115.51 (2011), pp. 25484–25489.
- [36] Arthur v Cresce et al. “In situ and quantitative characterization of solid electrolyte interphases”. In: *Nano letters* 14.3 (2014), pp. 1405–1412.
- [37] E Peled. “Film forming reaction at the lithium/electrolyte interface”. In: *Journal of Power Sources* 9.3 (1983), pp. 253–266.

- [38] Gholamabbas Nazri and Rolf H Muller. “In situ X-ray diffraction of surface layers on lithium in nonaqueous electrolyte”. In: (1984).
- [39] D Aurbach et al. “Identification of surface films formed on lithium in propylene carbonate solutions”. In: *Journal of The Electrochemical Society* 134.7 (1987), p. 1611.
- [40] Kiyoshi Kanamura, Hiroshi Tamura, and Zen-ichiro Takehara. “XPS analysis of a lithium surface immersed in propylene carbonate solution containing various salts”. In: *Journal of electroanalytical chemistry* 333.1-2 (1992), pp. 127–142.
- [41] E Peled, D Golodnitsky, and G Ardel. “Advanced model for solid electrolyte interphase electrodes in liquid and polymer electrolytes”. In: *Journal of the Electrochemical Society* 144.8 (1997), p. L208.
- [42] Doron Aurbach and Yosef Gofer. “The behavior of lithium electrodes in mixtures of alkyl carbonates and ethers”. In: *Journal of the Electrochemical Society* 138.12 (1991), p. 3529.
- [43] Doron Aurbach et al. “The behaviour of lithium electrodes in propylene and ethylene carbonate: The major factors that influence Li cycling efficiency”. In: *Journal of Electroanalytical Chemistry* 339.1-2 (1992), pp. 451–471.
- [44] Doron Aurbach et al. “The study of electrolyte solutions based on ethylene and diethyl carbonates for rechargeable Li batteries: II. Graphite electrodes”. In: *Journal of The Electrochemical Society* 142.9 (1995), p. 2882.
- [45] CR Yang, YY Wang, and CC Wan. “Composition analysis of the passive film on the carbon electrode of a lithium-ion battery with an EC-based electrolyte”. In: *Journal of power sources* 72.1 (1998), pp. 66–70.
- [46] Guorong V Zhuang et al. “Lithium ethylene dicarbonate identified as the primary product of chemical and electrochemical reduction of EC in 1.2 M LiPF₆/EC: EMC electrolyte”. In: *The Journal of Physical Chemistry B* 109.37 (2005), pp. 17567–17573.
- [47] R Dedryvère et al. “Characterization of lithium alkyl carbonates by X-ray photoelectron spectroscopy: experimental and theoretical study”. In: *The Journal of Physical Chemistry B* 109.33 (2005), pp. 15868–15875.
- [48] Kang Xu et al. “Syntheses and characterization of lithium alkyl mono-and dicarbonates as components of surface films in Li-ion batteries”. In: *The Journal of Physical Chemistry B* 110.15 (2006), pp. 7708–7719.
- [49] S-H Kang et al. “Investigating the solid electrolyte interphase using binder-free graphite electrodes”. In: *Journal of Power Sources* 175.1 (2008), pp. 526–532.
- [50] Daniel M Seo et al. “Reduction reactions of carbonate solvents for lithium ion batteries”. In: *ECS Electrochemistry Letters* 3.9 (2014), A91.

- [51] Luning Wang et al. “Identifying the components of the solid–electrolyte interphase in Li-ion batteries”. In: *Nature chemistry* 11.9 (2019), pp. 789–796.
- [52] Rosamaria Fong, Ulrich Von Sacken, and Jeff R Dahn. “Studies of lithium intercalation into carbons using nonaqueous electrochemical cells”. In: *Journal of The Electrochemical Society* 137.7 (1990), p. 2009.
- [53] Hilmi Buqa et al. “Modified carbons for improved anodes in lithium ion cells”. In: *Journal of Power Sources* 97 (2001), pp. 122–125.
- [54] Jong-Chan Lee and Morton H Litt. “Ring-opening polymerization of ethylene carbonate and depolymerization of poly (ethylene oxide-co-ethylene carbonate)”. In: *Macromolecules* 33.5 (2000), pp. 1618–1627.
- [55] Hongyu Wang et al. “Characterization of carbon-coated natural graphite as a lithium-ion battery anode material”. In: *Journal of The Electrochemical Society* 149.4 (2002), A499.
- [56] Jolanta Światowska et al. “XPS, XRD and SEM characterization of a thin ceria layer deposited onto graphite electrode for application in lithium-ion batteries”. In: *Applied Surface Science* 257.21 (2011), pp. 9110–9119.
- [57] Yujing Liu et al. “Self-assembled monolayers direct a LiF-rich interphase toward long-life lithium metal batteries”. In: *Science* 375.6582 (2022), pp. 739–745.
- [58] Walter Kohn. “Nobel Lecture: Electronic structure of matter—wave functions and density functionals”. In: *Reviews of Modern Physics* 71.5 (1999), p. 1253.
- [59] Pierre Hohenberg and Walter Kohn. “Inhomogeneous electron gas”. In: *Physical review* 136.3B (1964), B864.
- [60] Walter Kohn and Lu Jeu Sham. “Self-consistent equations including exchange and correlation effects”. In: *Physical review* 140.4A (1965), A1133.
- [61] John P Perdew et al. “Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation”. In: *Physical review B* 46.11 (1992), p. 6671.
- [62] Axel D Becke. “Density-functional exchange-energy approximation with correct asymptotic behavior”. In: *Physical review A* 38.6 (1988), p. 3098.
- [63] Jianmin Tao et al. “Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids”. In: *Physical Review Letters* 91.14 (2003), p. 146401.
- [64] Axel D Becke. “A new mixing of Hartree–Fock and local density-functional theories”. In: *The Journal of chemical physics* 98.2 (1993), pp. 1372–1377.

- [65] Yixuan Wang et al. “Theoretical studies to understand surface chemistry on carbon anodes for lithium-ion batteries: how does vinylene carbonate play its role as an electrolyte additive?” In: *Journal of the American Chemical Society* 124.16 (2002), pp. 4408–4421.
- [66] Jiamei Yu et al. “Hybrid DFT functional-based static and molecular dynamics studies of excess electron in liquid ethylene carbonate”. In: *Journal of The Electrochemical Society* 158.4 (2011), A400.
- [67] Tingzheng Hou et al. “The influence of FEC on the solvation structure and reduction reaction of LiPF₆/EC electrolytes and its implication for solid electrolyte interphase formation”. In: *Nano Energy* 64 (2019), p. 103881.
- [68] Jaebeom Han et al. “Calculated Reduction Potentials of Electrolyte Species in Lithium–Sulfur Batteries”. In: *The Journal of Physical Chemistry C* 124.38 (2020), pp. 20654–20670.
- [69] Jaebeom Han and Perla B Balbuena. “First-principles calculations of oxidation potentials of electrolytes in lithium–sulfur batteries and their variations with changes in environment”. In: *Physical Chemistry Chemical Physics* 20.27 (2018), pp. 18811–18827.
- [70] Jonathan Clarke-Hannaford et al. “Stability of boronium cation-based ionic liquid electrolytes on the Li metal anode surface”. In: *ACS Applied Energy Materials* 3.6 (2020), pp. 5497–5509.
- [71] Zhixiao Liu et al. “Li₂S film formation on lithium anode surface of Li–S batteries”. In: *ACS applied materials & interfaces* 8.7 (2016), pp. 4700–4708.
- [72] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. “Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids”. In: *Journal of the American Chemical Society* 118.45 (1996), pp. 11225–11236.
- [73] Wendy D Cornell et al. “A second generation force field for the simulation of proteins, nucleic acids, and organic molecules”. In: *Journal of the American Chemical Society* 117.19 (1995), pp. 5179–5197.
- [74] Alex D MacKerell Jr et al. “All-atom empirical potential for molecular modeling and dynamics studies of proteins”. In: *The journal of physical chemistry B* 102.18 (1998), pp. 3586–3616.
- [75] Kenno Vanommeslaeghe et al. “CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields”. In: *Journal of computational chemistry* 31.4 (2010), pp. 671–690.
- [76] Dmitry Bedrov et al. “Molecular dynamics simulations of ionic liquids and electrolytes using polarizable force fields”. In: *Chemical reviews* 119.13 (2019), pp. 7940–7995.

- [77] José NA Canongia Lopes and Agílio AH Pádua. “Nanostructural organization in ionic liquids”. In: *The Journal of Physical Chemistry B* 110.7 (2006), pp. 3330–3335.
- [78] Nav Nidhi Rajput et al. “Elucidating the solvation structure and dynamics of lithium polysulfides resulting from competitive salt and solvent interactions”. In: *Chemistry of Materials* 29.8 (2017), pp. 3375–3379.
- [79] Ildikó Pethes. “A comparison of classical interatomic potentials applied to highly concentrated aqueous lithium chloride solutions”. In: *Journal of Molecular Liquids* 242 (2017), pp. 845–858.
- [80] Oleg Borodin. “Polarizable force field development and molecular dynamics simulations of ionic liquids”. In: *The Journal of Physical Chemistry B* 113.33 (2009), pp. 11463–11478.
- [81] Pengyu Ren and Jay W Ponder. “Polarizable atomic multipole water model for molecular mechanics simulation”. In: *The Journal of Physical Chemistry B* 107.24 (2003), pp. 5933–5947.
- [82] Seiji Tsuzuki et al. “Molecular dynamics simulations of ionic liquids: cation and anion dependence of self-diffusion coefficients of ions”. In: *The Journal of Physical Chemistry B* 113.31 (2009), pp. 10641–10649.
- [83] Oleg Borodin and Grant D Smith. “Quantum chemistry and molecular dynamics simulation study of dimethyl carbonate: ethylene carbonate electrolytes doped with LiPF₆”. In: *The Journal of Physical Chemistry B* 113.6 (2009), pp. 1763–1776.
- [84] Oleg Borodin and GD Smith. “Li⁺ transport mechanism in oligo (ethylene oxide) s compared to carbonates”. In: *Journal of solution chemistry* 36.6 (2007), pp. 803–813.
- [85] Michael F Russo Jr and Adri CT Van Duin. “Atomistic-scale simulations of chemical reactions: Bridging from quantum chemistry to engineering”. In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 269.14 (2011), pp. 1549–1554.
- [86] Thomas P Senftle et al. “The ReaxFF reactive force-field: development, applications and future directions”. In: *npj Computational Materials* 2.1 (2016), pp. 1–14.
- [87] Liang Deng et al. “Self-optimizing weak solvation effects achieving faster low-temperature charge transfer kinetics for high-voltage Na₃V₂ (PO₄)₂F₃ cathode”. In: *Energy Storage Materials* 44 (2022), pp. 82–92.
- [88] Yuki Yamada and Atsuo Yamada. “Superconcentrated electrolytes for lithium batteries”. In: *Journal of The Electrochemical Society* 162.14 (2015), A2406.
- [89] Oleg Borodin et al. “Uncharted waters: super-concentrated electrolytes”. In: *Joule* 4.1 (2020), pp. 69–100.

- [90] Yuki Yamada et al. “Advances and issues in developing salt-concentrated battery electrolytes”. In: *Nature Energy* 4.4 (2019), pp. 269–280.
- [91] Xia Cao et al. “localized high-concentration electrolytes for lithium batteries”. In: *Journal of The Electrochemical Society* 168.1 (2021), p. 010522.
- [92] CJF Solano et al. “A joint theoretical/experimental study of the structure, dynamics, and Li⁺ transport in bis ([tri] fluoro [methane] sulfonyl) imide [T] FSI-based ionic liquids”. In: *The Journal of Chemical Physics* 139.3 (2013), p. 034502.
- [93] Oleg Borodin and Grant D Smith. “LiTFSI structure and transport in ethylene carbonate from molecular dynamics simulations”. In: *The Journal of Physical Chemistry B* 110.10 (2006), pp. 4971–4977.
- [94] Julian Self, Kara D Fong, and Kristin A Persson. “Transport in superconcentrated LiPF₆ and LiBF₄/propylene carbonate electrolytes”. In: *ACS Energy Letters* 4.12 (2019), pp. 2843–2849.
- [95] Melville S Green. “Brownian motion in a gas of noninteracting molecules”. In: *The Journal of Chemical Physics* 19.8 (1951), pp. 1036–1046.
- [96] M Neumann and O Steinhauser. “Computer simulation and the dielectric constant of polarizable polar systems”. In: *Chemical physics letters* 106.6 (1984), pp. 563–569.
- [97] Oliver T Unke et al. “Machine learning force fields”. In: *Chemical Reviews* 121.16 (2021), pp. 10142–10186.
- [98] Jörg Behler. “Four generations of high-dimensional neural network potentials”. In: *Chemical Reviews* 121.16 (2021), pp. 10037–10072.
- [99] Volker L Deringer et al. “Gaussian process regression for materials and molecules”. In: *Chemical Reviews* 121.16 (2021), pp. 10073–10141.
- [100] Sergei Manzhos and Tucker Carrington Jr. “Neural network potential energy surfaces for small molecules and reactions”. In: *Chemical Reviews* 121.16 (2020), pp. 10187–10217.
- [101] Han Wang et al. “DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics”. In: *Comput. Phys. Commun.* 228 (2018), pp. 178–184.
- [102] Denghui Lu et al. “86 PFLOPS Deep Potential Molecular Dynamics simulation of 100 million atoms with ab initio accuracy”. In: *Computer Physics Communications* 259 (2021), p. 107624.
- [103] Matthias Rupp et al. “Fast and accurate modeling of molecular atomization energies with machine learning”. In: *Phys. Rev. Lett.* 108.5 (2012), p. 058301.

- [104] Matthias Rupp, Raghunathan Ramakrishnan, and O Anatole Von Lilienfeld. “Machine learning for quantum mechanical properties of atoms in molecules”. In: *The Journal of Physical Chemistry Letters* 6.16 (2015), pp. 3309–3313.
- [105] Jörg Behler and Michele Parrinello. “Generalized neural-network representation of high-dimensional potential-energy surfaces”. In: *Phys. Rev. Lett.* 98.14 (2007), p. 146401.
- [106] Jörg Behler. “Atom-centered symmetry functions for constructing high-dimensional neural network potentials”. In: *J. Chem. Phys.* 134.7 (2011), p. 074106.
- [107] Albert P Bartók, Risi Kondor, and Gábor Csányi. “On representing chemical environments”. In: *Phys. Rev. B* 87.18 (2013), p. 184115.
- [108] Felix A Faber et al. “Alchemical and structural distribution based representation for universal quantum machine learning”. In: *The Journal of chemical physics* 148.24 (2018), p. 241717.
- [109] Haoyan Huo and Matthias Rupp. “Unified representation of molecules and crystals for machine learning”. In: *arXiv preprint arXiv:1704.06439* (2017).
- [110] Sergey N Pozdnyakov et al. “Incompleteness of atomic structure representations”. In: *Physical Review Letters* 125.16 (2020), p. 166001.
- [111] Alexander V Shapeev. “Moment tensor potentials: A class of systematically improvable interatomic potentials”. In: *Multiscale Modeling & Simulation* 14.3 (2016), pp. 1153–1173.
- [112] Ralf Drautz. “Atomic cluster expansion for accurate and transferable interatomic potentials”. In: *Physical Review B* 99.1 (2019), p. 014104.
- [113] Jigyasa Nigam, Sergey Pozdnyakov, and Michele Ceriotti. “Recursive evaluation and iterative contraction of N-body equivariant features”. In: *The Journal of Chemical Physics* 153.12 (2020), p. 121101.
- [114] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).
- [115] Oliver T Unke et al. “SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects”. In: *Nature communications* 12.1 (2021), pp. 1–14.
- [116] Roman Zubatyuk et al. “Teaching a neural network to attach and detach electrons from molecules”. In: *Nature communications* 12.1 (2021), pp. 1–11.
- [117] Raghunathan Ramakrishnan et al. “Quantum chemistry structures and properties of 134 kilo molecules”. In: *Scientific data* 1.1 (2014), pp. 1–7.
- [118] Albert P Bartók et al. “Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons”. In: *Physical review letters* 104.13 (2010), p. 136403.

- [119] So Fujikake et al. “Gaussian approximation potential modeling of lithium intercalation in carbon nanostructures”. In: *The Journal of chemical physics* 148.24 (2018), p. 241714.
- [120] Berk Onat et al. “Implanted neural network potentials: Application to Li-Si alloys”. In: *Physical Review B* 97.9 (2018), p. 094106.
- [121] Jianxing Huang et al. “Deep potential generation scheme and simulation protocol for the Li₁₀GeP₂S₁₂-type superionic conductors”. In: *The Journal of Chemical Physics* 154.9 (2021), p. 094703.
- [122] Oleg Borodin and Dmitry Bedrov. “Interfacial structure and dynamics of the lithium alkyl dicarbonate SEI components in contact with the lithium battery electrolyte”. In: *The Journal of Physical Chemistry C* 118.32 (2014), pp. 18362–18371.
- [123] Vanessa Quaranta, Jörg Behler, and Matti Hellström. “Structure and dynamics of the liquid–water/zinc-oxide interface from machine learning potential simulations”. In: *The Journal of Physical Chemistry C* 123.2 (2018), pp. 1293–1304.
- [124] Marcos F Calegari Andrade et al. “Free energy of proton transfer at the water–TiO₂ interface from ab initio deep potential molecular dynamics”. In: *Chemical Science* 11.9 (2020), pp. 2335–2341.
- [125] Christoph Schran et al. “Machine learning potentials for complex aqueous systems made simple”. In: *Proceedings of the National Academy of Sciences* 118.38 (2021), e2110077118.

Chapter 2

Incorporating Electronic Information into Machine Learning Potential Energy Surfaces: Development of BpopNN

Parts of this chapter were reprinted with permission from: “Incorporating Electronic Information into Machine Learning Potential Energy Surfaces via Approaching the Ground-State Electronic Energy as a Function of Atom-Based Electronic Populations” Xiaowei Xie, Kristin A. Persson, David W. Small. *J. Chem. Theory Comput.* **2020**, 16, 7, 4256–4270.

I Introduction

DFT computations continue to be the predominant approach to first-principles modeling of chemical systems,[1–3] while empirical force-field based approximations continue to dominate large scale simulations.[4, 5] As one category of approaches intent on bridging this divide, ML-based methods for approximating first-principles PESs are drawing much attention. These methods largely reduce the computational cost while remaining remarkably near to the accuracy of quantum chemistry calculations, and practical applications in chemistry, physics and material science have been demonstrated[6–9]. So far, predictions of the electronic energy for neutral, closed-shell molecules (equilibrium or off-equilibrium) have been successfully demonstrated. A ubiquitous theme in such methodologies is the development of translationally, rotationally, and permutationally invariant descriptors based on nuclear positions. Prominent examples include the atom-centered symmetry functions[10, 11] of Behler and Parrinello and the smooth overlap of atomic positions (SOAP)[12] by Bartók et al. Some specific ML models that fit the general positions-to-energy characterization are RuNNer,[13] Tensormol,[14–18] SchNet,[19, 20] ANI-1,[21–23] AMP,[24] DeepMD,[25] LASP,[26] QML,[27, 28] PhysNet, [29] GAP,[30] models based on atom-density representations by Ceriotti and coworkers,[31–35] the spectral neighbor analysis potential (SNAP), [36] graph kernel methods,[37, 38] graph neural networks, [39, 40] methods based upon eigenvalues of the Coulomb matrix, [41, 42] Bag-of-Bonds, [43] permutation invariant polynomials,[44] and others. Impressive examples of large-scale applications of these methods include simulating proteins, [14] amorphous carbon, [45–47] constructing phase diagrams of amorphous Li_xSi , [48] etc.

However, descriptors based on atom positions alone exhibit some important limitations. For example, they cannot be applied to more than one overall charge or spin state. Furthermore, far-away changes in the chemical environment may induce changes in the local charge or spin state of a region of a molecule even if the local geometry is not changed appreciably. We illustrate this with an example in section 3.2.3 below. This general problem applies to many molecules and materials. Examples relating to conjugation, like the emergence of polyradicality with increasing length in acenes,[49–51] armchair versus zigzag edge effects in graphene nanoribbons,[52–54] and the gradual ascension to the effective conjugation length in a variety of oligomers,[55–57] come to mind. Even more marked are examples involving explicit ionization and charge transfer (CT). These include outer-sphere electron transfer processes such as those occurring in Ferredoxin protein cores and other proteins[58, 59] and large-scale organic donor-acceptor complexes with appealing electronic properties.[60–62]

We stress the need for advanced ML-based methods to describe complicated heterogeneous systems such as found in electrode surfaces like the solid electrolyte interface (SEI) in lithium-ion batteries (LIB).[63–67] As a small glance into the complexities involved here, consider that the ground state of neutral lithium ethylene carbonate (LiEC), a key intermediate in the reductive decomposition process to form the LIB SEI, varies in character between neutral and CT, i.e. unpaired electron on Li or on the rest of the molecule. This variation is subtly dependent on where the Li atom locates around the ethylene carbonate (EC) molecule;[68] the transition between the two characters can be fairly rapid, and so for position-based descriptors, whose terms don’t change rapidly in such transitions, this situation is clearly a practical difficulty. This is just one part of the SEI’s multiplex of charged and uncharged species and the reactions between them, underscoring significant challenges for ML descriptors. In principle, these general effects can be captured by a position based descriptor along with the total charge and spin multiplicity of the molecule, as after all these are the only inputs needed for a ground-state calculation. However, size extensibility, the crucial property that a ML model may be uniformly and consistently applied to systems of different sizes, favors the use of local, typically atom-centered, descriptors, which are difficult to reconcile with global parameters like total charge and spin. And, stretching these to include the (possibly very) long range information required for these cases is challenging in practice in terms of computational efficiency and numerical stability.

The above issues stem from a lack of explicit electronic information in the descriptors, which suggests a general remedial approach. To successfully incorporate electronic information, the basic scheme of the parent DFT model should be followed such that electronic terms can adapt according to changes in neighboring electronic information, preferably in a self-consistent or equilibrated way.

In this context, we would like to briefly reassess some of the models mentioned above. The SOAP based descriptors make use of smeared atom density distributions, and together these bear a rough resemblance to electron density. A recent variant[33] uses the electrostatic potential (ESP) associated with such distributions as the basis for the SOAP expansion. This approach provides enhanced long-range information, and we will comment more on this important development at the end of this chapter. Nevertheless, these density distributions are fixed and do not adapt to the electronic environment.

Several of the above ML approaches include a separate long-range, pairwise electrostatic contribution to the energy. This inclusion has been shown to be very useful for describing large Zn_nO_n clusters, [69], water clusters, [14, 70–72] proteins,[14] and other examples [73, 74]. Popelier and coworkers first proposed to employ NNs to construct environment dependent multipoles.[75] Artrith et. al.,[69] and later on Yao et. al. [14] adopted a scheme where they pretrain a neural network to predict atomic point charges, using the same descriptors (i.e. symmetry functions) as they used to train energies. These charges were subsequently used to compute the electrostatic term through Ewald summation [76] for periodic systems, or Coulomb’s law for molecules. Another approach fits molecular dipoles to infer partial charges.

[77] We note that there are also independent ML models specialized for predicting partial charges, [78, 79] multipole moments,[32, 80, 81] and even the full electron density. [31, 82]. The latter method has been used e.g. to guide the inverse design of chemical materials based on electron density information, to serve as an initial guess for SCF convergence, and more. The above-mentioned methods all predict the electronic terms directly from local position-based descriptors, and thus are not properly adaptable to the electronic environment.

In another direction, the various charge equilibration schemes, such as EEM[83] and QEq[84] and several newer examples,[85–91] compute atomic charges self-consistently. These models therefore exhibit some basic flexibility in the context of the above issues. However, they also employ very simple forms; this is advantageous for application to very large molecular systems, but it limits the accuracy as compared to the above ML-based models. Ghasemi, et al[92] developed a charge equilibration scheme where the intra-atomic energy function comes from a deep NN that is parametrized by DFT data, and they used this to predict the energies for both neutral and ionized NaCl clusters. To our knowledge, this important development is the only extant positions-to-energy model that employs a high-level ML approximation and that is also able to accommodate varying electronic structure. However, the model is exclusively trained on self-consistent DFT energies and its charges do not enter the descriptor. Hence the charges are effectively auxiliary parameters as opposed to being more directly associated with the electron density. This precludes, for example, obtaining more than one SCF solution at a given geometry.

An open question remains as to how much electronic information should be incorporated. At one extreme, it is possible to utilize the full electron density and create ML models that map this directly to the energy. Some of these efforts are concerned with finding new DFT approaches with unprecedented accuracy.[93–96] Others pertain more to the fact that commonly used functionals do not provide a direct non-iterative link between density and energy, so a ML approximation for this, as with other orbital-free approaches, can greatly reduce the associated burden.[20, 96–101] Effective self-consistency is then obtained by optimizing the energy with respect to the density. This level of electronic information is ideal in terms of accuracy, and it easily addresses the above issues, but it entails a significant increase in computational requirements compared to the above models.

In the present chapter, we attempt to identify the simplest amount of electronic information that is sufficient to solve the above issues, and how to incorporate it effectively. In a sense, we are seeking a good balance between the above position-only approaches and the full-density based approach. We focus on approximating regular DFT functionals, and we explore the incorporation of atomic electron populations and associated electrostatic interaction terms into the descriptor. Effectively, this divides the global charge and spin parameters into local terms that may be naturally incorporated into local descriptors. In the following sections, we exploit CDFT to realize a natural map from intrinsic populations (self-consistently optimal ones or otherwise) to energies, which may be approximated with deep neural networks, leading to a simple self-consistent approach that is qualitatively true to the parent

DFT functional.

II Theory

2.1 Target Energy Function

2.1.1 Formal viewpoint

In the Levy constrained search approach in DFT,[102] each density ρ maps to the lowest possible energy $E[\rho]$ obtainable from a many-electron wave function associated with that density. Thinking of the density as being a reduction of the many-electron density matrix, the constrained-search idea may be generalized to any collection of reduced variables $\{v_i\}$ (which may be functions, etc.):

$$E[\{v_i\}] = \min_{\Psi \rightarrow \{v_i\}} \langle \Psi | \mathbf{H} | \Psi \rangle, \quad (2.1)$$

where $\mathbf{H}$ is the many-electron Hamiltonian. Of course, in practice, the search over wave functions and the energy expectation value on the right hand side will be replaced by a search over orbitals and an approximate functional, respectively.

For $\{v_i\}$ in this chapter, we are primarily concerned with reducing the density to atom populations, in particular Becke populations.[103] For these, the density is multiplied by a weight function and integrated to produce a population value. For atom i and spin σ we have

$$p_{i,\sigma} = \int w_i(\mathbf{r}) \rho_\sigma(\mathbf{r}). \quad (2.2)$$

The weight functions w_i are localized to their respective atoms, they take on values between 0 and 1, maximizing on the pertinent atom, and they sum over all atoms to 1.

Our target function for machine learning is essentially $E[\mathbf{p}]$, where we have placed the populations into a vector $\mathbf{p}$. Technically, this function is defined only for population values that are non-negative and sum to appropriate values for the numbers of electrons of each spin, or equivalently, the total charge and spin multiplicity. We can extend this function to the larger domain of non-negative population vectors with no sum conditions by first applying P_{N_α, N_β} , which “projects” the population vector to a vector obeying the sum conditions:

$$[P_{N_\alpha, N_\beta}(\mathbf{p})]_{i,\sigma} = p_{i,\sigma} + \frac{N_\sigma - \sum_j p_{j,\sigma}}{N_\sigma}, \quad (2.3)$$

where N_α and N_β are the numbers of α and β electrons, respectively. Hence, our formal target function is $E[P_{N_\alpha, N_\beta}(\mathbf{p})]$.

The target function may be optimized with respect to the population values, with the latter considered as input variables. This is much the same as optimizing DFT energies with respect to the density or orbitals. Optimizing with respect to the populations thus defines an SCF procedure. As to the issue of representability here, we only need to ensure that the populations (after projection) are non-negative.

At this point, we need a way to obtain practical approximations to this target function, which is the subject of the next subsection.

2.1.2 The target function in practice: CDFT

The above ideas may be applied to any of the functionals used in common practice, e.g. B3LYP as we use below. That is, they each may be formally reduced to a functional of populations.

Training data for the target function must be obtained with non-standard DFT calculations. Fully self-consistent DFT solutions provide data for certain population choices, however, by itself this would constitute a very limited training set. Energies for other population choices may be obtained by using constrained DFT (CDFT),[104–106] which uses Lagrange multipliers to optimize the DFT energy under various constraints. The use of constraints on Becke-weight based populations has been implemented in Q-Chem,[107] and allows us to, in principle, obtain the energy for any given set of populations. Much as how there are generally numerous wave functions (and thereby energies) associated with a given density, there will overall be many energies associated with a specific population vector. In fact, the energy spread is much wider for populations than for the density, because there will generally be many densities associated with a given population vector. As with standard DFT calculations, care must be exercised to attempt to obtain the lowest energy CDFT solution.

The accuracy of ML models of the DFT energy is significantly enhanced by incorporating energy derivatives into the loss function, i.e. it is desirable to train the model towards both the direct value of the target function and its first derivatives. The xyz position derivatives for CDFT have been previously derived.[108] For this present chapter, we also need a way to compute the derivatives of the CDFT energy with respect to populations. This is described in the Appendix.

2.2 Model Energy Function

We adopted a widely accepted scheme in the ML PES field, where the total energy is decomposed into atom-centered contributions to ensure transferability across systems with different sizes and compositions. A deep neural network (NN) framework was used to map an atomistic descriptor $D_i[\mathbf{Z}, \mathbf{p}, \mathbf{r}]$ to the atomic energy. The model energy function can be written

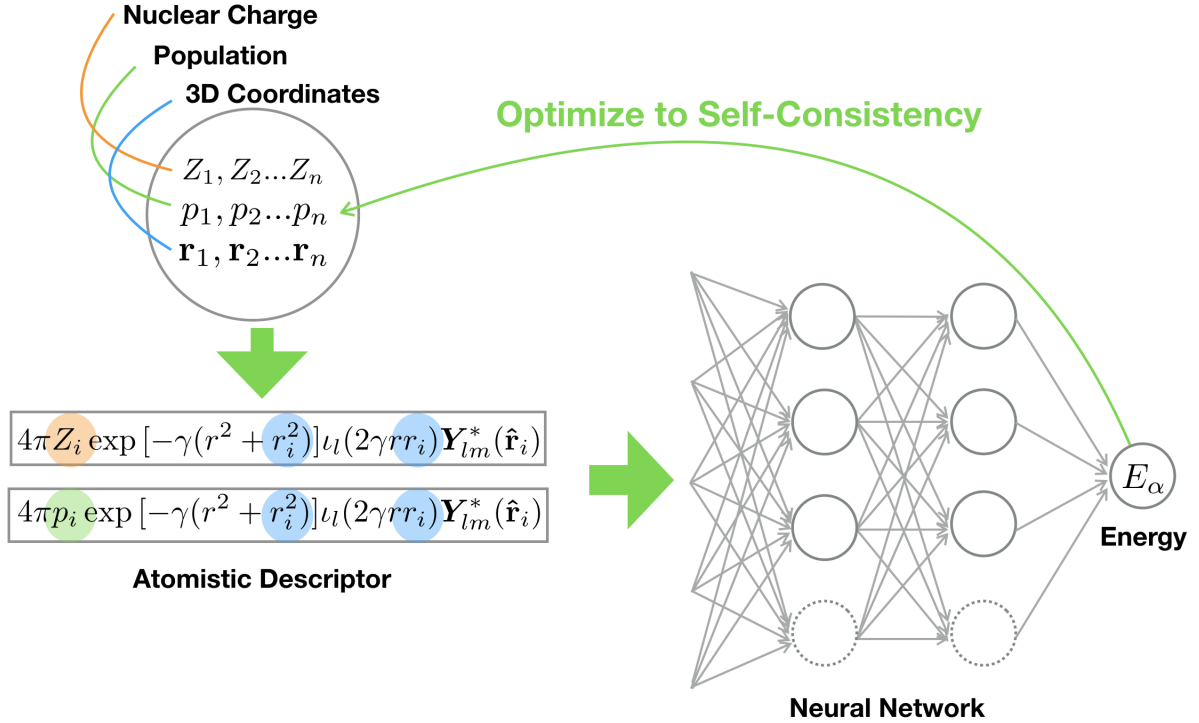


Figure 2.1: General concept for the ML potential

as

$$E_{\text{ML}} = \sum_{\mu} \sum_{i \in A_{\mu}} \epsilon_{\mu}[D_i[\mathbf{Z}, \mathbf{p}, \mathbf{r}]] + E_{\text{intra}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}] + E_{\text{el}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}], \quad (2.4)$$

where ϵ_{μ} represents the output from the atomic neural network for the element μ , and this function is applied to the descriptor for atom i ; A_{μ} is the set of atom indices corresponding to element μ ; $\mathbf{Z}$ and $\mathbf{r}$ contain the atomic numbers and coordinates, respectively, for the entire molecule. E_{intra} and E_{el} are simple intra-atomic and pairwise electrostatic terms, respectively, and will be described below. An overview of our model, which we will call “BpopNN” for “Becke Population Neural Network”, is shown in Figure 2.1. More details are given in the following subsections.

2.2.1 Atomistic Descriptor

The descriptor we use in this chapter is essentially a variant of SOAP,[12] modified to describe the electronic as well as nuclear environments. The regular SOAP descriptor is based on an atomic power spectrum of a basic density distribution that represents the nuclei: for each

atom, the latter is expanded in a set of radial basis functions (RBF) and spherical harmonics that are symmetric about that atom, and the coefficients of this expansion directly produce the spectrum. In our case, we want to adapt the density distribution to represent the varying electrostatic environment.

The basic idea is to use the ESPs of the nuclei, and the ESPs of the α populations treated as point charges, and likewise for the β populations. In fact, as in SOAP, we use separate distributions for each element type, since this makes for a more refined descriptor. In our lithium hydride applications below, we simplified the descriptor to be based only on the nuclear charges and the total electron population on each atom. In the present subsection, we will continue the development using spin populations. This is more general, and reducing this descriptor to the total-population one is straightforward.

To simplify the computation of the descriptor elements, we approximate the $1/r$ term in the ESPs with a single Gaussian function. Hence our density distributions take the form

$$\rho_{\mu,t}(\mathbf{r}) = \sum_{i \in A_\mu} \chi_i^t B \exp(-\gamma |\mathbf{r} - \mathbf{r}_i|^2), \quad (2.5)$$

where t is either ν (for nuclear), α , or β corresponding to the type of point population, and thus $\chi_i^\nu = Z_i$, $\chi_i^\alpha = p_{i,\alpha}$, and $\chi_i^\beta = p_{i,\beta}$. Here we used $B = 1.128$ and $\gamma = 0.171$, and r is in the unit Bohr radius.

To build the descriptor for any atom in the molecule, we set the origin to that position and project the radial slices of $\rho_{\mu,t}(\mathbf{r})$ onto the pertinent set of spherical harmonics $\mathbf{Y}_{lm}(\hat{\mathbf{r}})$:

$$\rho_{\mu,t}(\mathbf{r}) = \sum_{i \in A_\mu} \sum_{lm} c_{ilm}^t(r) \mathbf{Y}_{lm}(\hat{\mathbf{r}}). \quad (2.6)$$

Note that the i sum includes the central atom (if it is of the element type μ). In practice, the l sum must be limited to a range. In the applications below, we use a maximum l value of 6. Since χ_i^t does not depend directly on positions, the analytical form of the radial terms can be derived following previous works [12, 109] as

$$c_{ilm}^t(r) \equiv 4\pi\chi_i^t B \exp[-\gamma(r^2 + r_i^2)] \iota_l(2\gamma r r_i) \mathbf{Y}_{lm}^*(\hat{\mathbf{r}}_i), \quad (2.7)$$

where $*$ denotes complex conjugation and ι_l are the modified spherical Bessel functions of the first kind. Summing over pertinent atoms gives $\rho_{\mu,t}$ and projecting these onto RBFs gives a sequence of expansion coefficients $c_{nlm}^{\mu,t}$ with an extra index n . Underlying the projections is the evaluation of the integrals $\langle g_n \mathbf{Y}_{lm} | \rho \rangle$, with g_n denoting an RBF. In general, this can be done by either analytical or numerical integration (e.g. Gauss-Legendre quadrature), but in our case, we bypass these complications by simply evaluating $c_{ilm}^{\mu,t}(r)$ on a grid of r values. The resulting coefficients are smooth with respect to nuclear positions, and technically, one could choose RBFs whose projection coefficients match the evaluated grid values. Hence the

grid spacings can in principle be selected to produce a result that is as satisfactory as the explicit RBF/integration approach. A simple linear grid between 0.5 and 13.2 Å was used in this work; further refinement of the grid will be addressed in the future.

At this point, one can construct the power spectrum by contracting over m ,

$$p_{nl}^{\mu,t} = \sum_{m=-l}^l \|c_{nlm}^{\mu,t}\|^2 \quad (2.8)$$

It is easy to prove the rotational invariance of $p_{nl}^{\mu,t}$, as 3D rotations correspond to transformations of blocks of the $c_{nlm}^{\mu,t}$ by Wigner matrices, which are unitary.

For each atom in the molecule, we employ a separate power spectrum for each element-type in the molecule, i.e. the power spectrum is split into different element channels. There is active research aimed at reducing the cost associated with adding more atomic species,[18, 28, 35, 110] which is also of future interest for us. For now, the above treatment is sufficient to prove the principle.

A radial cutoff of 13.2 Å was used below in computing the descriptor, i.e. atom pairs beyond this cutoff were neglected. The reason for choosing this relatively long cutoff is that, as described below, in addition to the NN energies, we include a simple long-range pairwise electrostatic term, as has been shown to be very useful in previous papers.[14, 69] The influence of the switch in dominance between short-range NN and explicit electrostatic terms is a general issue, but because this question is separate from the main theme of this work, we chose the long cutoff to minimize this influence.

2.2.2 Neural Network Architecture

We employ separate neural networks for different element types, and for the applications below, each element network contains two hidden layers with 30 neurons in each layer. This size of neural network was used to best accommodate the number of training data and prevent overfitting. Nonlinearities were introduced using a modified softplus activation function $\ln(1 + \exp(100x))/100$. This form resembles the RELU activation function, which itself is highly popular because it overcomes the vanishing gradient problem,[111] yet this softplus variant has the advantage of being twice continuously differentiable.

In previous works, the ML model is usually trained towards the DFT atomization energy, as opposed to total energy. This has been shown to accelerate and balance training.[14] The definition of atomization energy is ambiguous in the case of charged molecules, hence we select a quadratic function of atomic charge that best fits the DFT energies for the pertinent charge states of the isolated atom. $E_{\text{intra}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}]$ is then the sum of these energies for each atom. For lithium and hydrogen atoms found in the molecular applications below, we used the following quadratic functions as baseline for atomic energies

$$E_{\text{intra,Li},i} = 0.0901q_i^2 + 0.1088q_i - 7.4787 \quad (2.9)$$

$$E_{\text{intra,H},i} = 0.2471q_i^2 + 0.2403q_i - 0.4805, \quad (2.10)$$

where q_i is the atomic partial charge for atom i , i.e. $q_i = Z_i - p_{i,\alpha} - p_{i,\beta}$. This resembles the intra-atomic term in charge equilibration models. [92]

We also used a simple pairwise Coulomb term to incorporate long-range electrostatics, as shown in the following equation.

$$E_{\text{el}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}] = \sum_{j>i} \tanh(\kappa r_{ij}) \frac{q_i q_j}{r_{ij}} \quad (2.11)$$

Parameters κ (one each for different element pair types) were trained together with other parameters in the neural network.

The loss function for the training is

$$\mathcal{L} = \sum_M \frac{(E_M^{\text{CDFT}} - E_M^{\text{ML}})^2}{N_{\text{atom},M}} + \gamma_1 \sum_M \frac{\|\mathbf{F}_M^{\text{CDFT}} - \mathbf{F}_M^{\text{ML}}\|^2}{N_{\text{atom},M}} + \gamma_2 \sum_M \frac{\|\mathbf{f}_M^{\text{CDFT}} - \mathbf{f}_M^{\text{ML}}\|^2}{N_{\text{atom},M}}. \quad (2.12)$$

For the model presented below, both γ_1 and γ_2 were set to 1.5. The above sums go over the training molecules, and $\mathbf{F}$ is the position gradient, $\mathbf{f}$ is the population gradient, and $N_{\text{atom},M}$ represents the number of atoms in molecule M . Adaptive moment solver (Adam)[112] was used to update the NN weights during training. Our implementation is built up from the open-source package TensorMol,[14] which takes advantage of the automatic differentiation scheme of Tensorflow.[113]

III Results and Discussion

We here test the performance of our method on Li_nH_n clusters of varying size and geometry, and overall charges of +1, 0, and -1. We chose this for the simplicity of having only two element types. Lithium hydride clusters have been studied in several theoretical[114–119] and experimental[120, 121] works, and have implications in various applications including hydrogen storage.[122–126]

In the calculations below, we employ the descriptor described above, but, as mentioned above, we modified it so that the total electron population on each atom, rather than the 2 spin populations, is used for the electronic terms.

3.1 Training-data generation and model training

In the training dataset, the cluster sizes vary from from $n = 1$ to 24, with increments of 3 or 4. We use two basic types of structures from which we obtain more samples: (1) collections of

relatively widely separated LiH molecules (2-5 Å between molecules), and (2) denser Li_nH_n systems with roughly cubic structures with alternating Li and H. We denote the first dataset as “sparse”, and the second as “non-sparse” from here on.

For the sparse data, structures from DFT geometry-optimization trajectories were used in the training dataset. For the non-sparse data, structures both from geometry-optimization trajectories and Ab Initio Molecular Dynamics (AIMD)[127] trajectories were used. The AIMD simulations used a temperature of 500 K and a time step of 20 a.u. (≈ 0.5 fs).

All structures were generated using overall charge 0, and calculations for the ionized molecules used the same geometries. During the geometry optimizations, the LiH units in the sparse structures cluster together in various ways and become less sparse, although they remain less dense than the non-sparse data. All of these geometries exhibit alternating Li and H atoms, i.e. no structures with Li clusters nor H_2 molecules were obtained. This is because Li clusters often entail spin polarization, and therefore we reserve the generation of such structures for future work in which spin populations are used. More details of the geometries used in the training dataset are included in the Appendix.

All DFT computations were conducted using the QChem[107] program. The geometry-optimization trajectories for the sparse data used the B3LYP[128, 129] functional with D3(BJ) dispersion[130], the AIMD trajectories for the non-sparse data used the $\omega\text{B97X-D}$ [131] functional, and the geometry-optimization trajectories for the non-sparse data used B3LYP (without dispersion). The reason for the overall use of different functionals is that we began with $\omega\text{B97X-D}$, then switched to using B3LYP-D3(BJ) for computational efficiency, and then encountered some implementational problems with D3(BJ) for Li-Li interactions, so we finally switched to using B3LYP alone. These variations only occurred for the generation of structures; for all geometries used in the training and test sets (the latter is described below), we (re-)computed the DFT energy using B3LYP. All DFT calculations in this chapter used the def2-SVPD basis set.[132]

For the final training set, we selected one per every several geometries from each trajectory, with an interval of 2 to 7 depending on the trajectory’s length. On each of these geometries, we computed unconstrained B3LYP solutions for the considered overall charges (+1,0,-1), and for each of these three, we performed 10 separate CDFT calculations. For each of these, we added a random (uniform distribution from -0.05 to 0.05) charge fluctuation to each atom’s charge value from the pertinent unconstrained SCF solution, and then computed the associated CDFT energy, the energy derivatives with respect to atomic positions and atomic charges (see Appendix for the latter).

The training data (~ 37000 points) were randomly separated to a training (95%) and a validation set (5%) at the outset, and the validation error provides an estimate on how well the model is trained. Typically the model reaches its best performance on test sets after ~ 500 epochs of training. The mean absolute error (MAE) on the independent validation set of the energy is 1.523 kcal/mol per molecule.

Table 2.1: Energy and charge errors for $\text{Li}_{29}\text{H}_{29}$ sparse and $\text{Li}_{32}\text{H}_{32}$ non-sparse test sets

	MAE charge (e/atom)	MAE energy (kcal/mol/atom)		MAE charge (e/atom)	MAE energy (kcal/mol/atom)
$\text{Li}_{29}\text{H}_{29}$ sparse neutral	0.009	0.188	$\text{Li}_{32}\text{H}_{32}$ non-sparse neutral	0.007	0.235
$\text{Li}_{29}\text{H}_{29}$ sparse cation	0.011	0.136	$\text{Li}_{32}\text{H}_{32}$ non-sparse cation	0.010	0.281
$\text{Li}_{29}\text{H}_{29}$ sparse anion	0.013	0.121	$\text{Li}_{32}\text{H}_{32}$ non-sparse anion	0.013	0.168

3.2 Testing the model

3.2.1 $\text{Li}_{29}\text{H}_{29}$ and $\text{Li}_{32}\text{H}_{32}$ clusters

We generated two test sets with larger molecules to test the transferability of our model: non-sparse $\text{Li}_{32}\text{H}_{32}$ (+1,0,-1 charge) from an AIMD trajectory and sparse $\text{Li}_{29}\text{H}_{29}$ (+1,0,-1 charge) from a geometry-optimization trajectory. For those test sets, CDFT calculations for off-equilibrium charges were not conducted as we focus on comparing the self-consistent solutions from DFT and the BpopNN model.

As discussed earlier, one advantage of our model is that the populations (i.e. partial charges at present) can be optimized self-consistently. In other words, the ML-model energies are optimized with respect to the partial charges. This process is reminiscent of the SCF procedure in DFT calculations, and we denote this charge optimization procedure as “SCF-q” in the remaining text. In general, each of these types of SCF can produce more than one stationary point, i.e. solution. Since our objective is to find counterpart solutions between DFT and ML and then compare them, we used the actual DFT partial charges as initial guesses for SCF-q calculations for each geometry in the $\text{Li}_{29}\text{H}_{29}$ and $\text{Li}_{32}\text{H}_{32}$ test sets. Other initial guesses can be used but this increases the odds of obtaining an SCF-q solution that does not correspond to the SCF one even when a pair of corresponding solutions exist. Of course, general and effective initial guess schemes will eventually need to be developed for the partial charges. The Li partial charges in our systems usually exhibit values between 0.5 and 0.8 (although the actual range covered throughout the whole dataset is -0.31 to 0.99). It turned out that we obtained the same SCF-q solutions when starting with uniform guesses of 0.4 for Li charges and -0.4 for H charges as when starting from DFT partial charges, demonstrating the model’s capability of dealing with general reasonable initial guesses for partial charges.

In Figure 2.2, we plot a comparison between the SCF-q BpopNN energies (converged to threshold 10^{-7} Hartree/e) and the DFT energies for neutral and ionized molecules. The statistics for energy and charge errors are shown in Table 2.1. For the sparse $\text{Li}_{29}\text{H}_{29}$

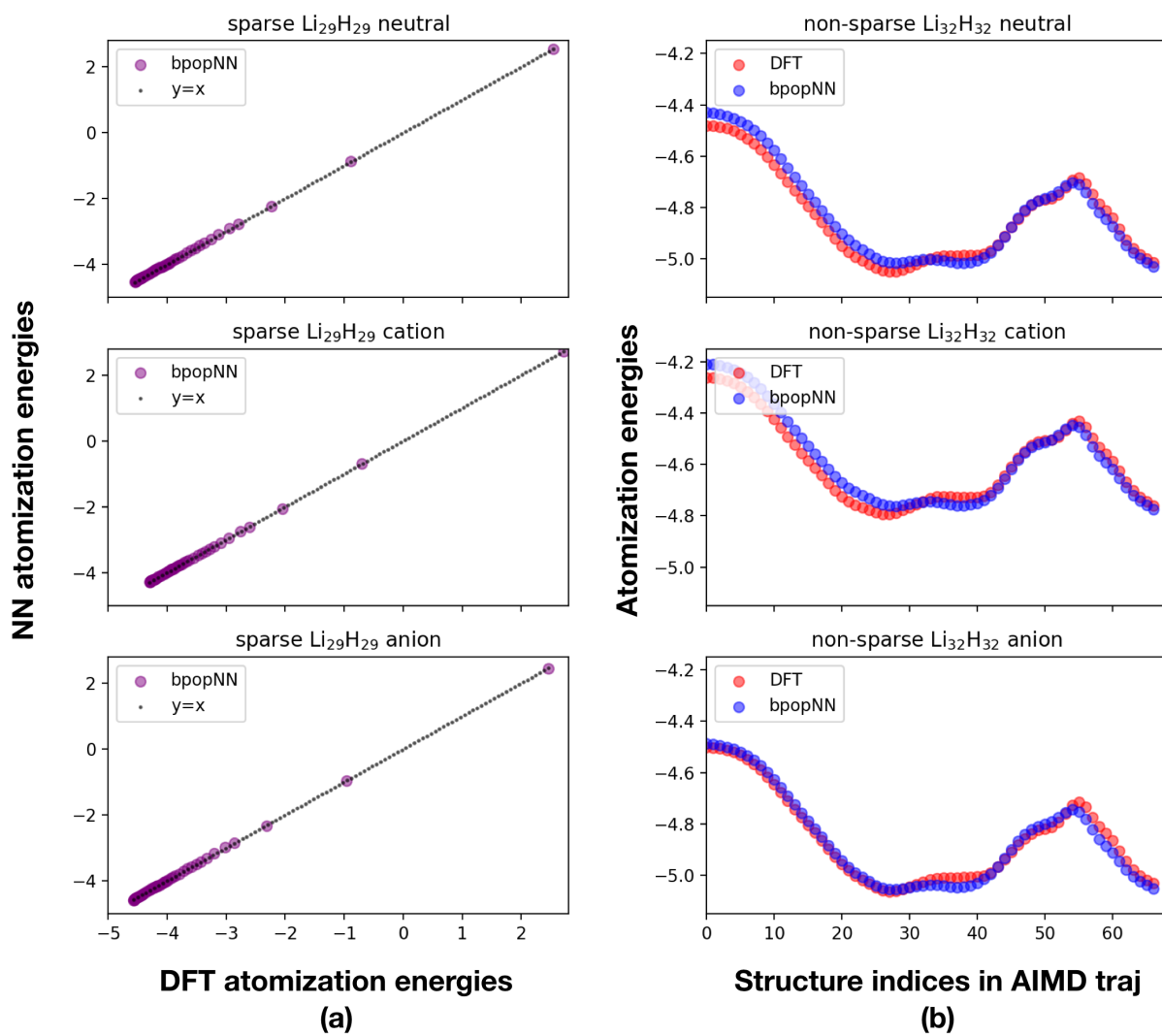


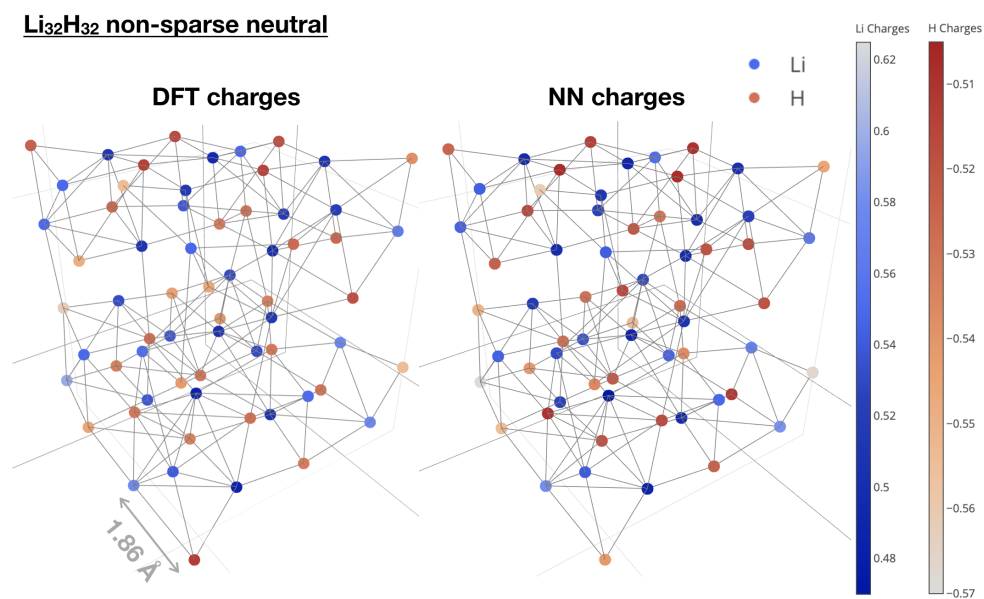
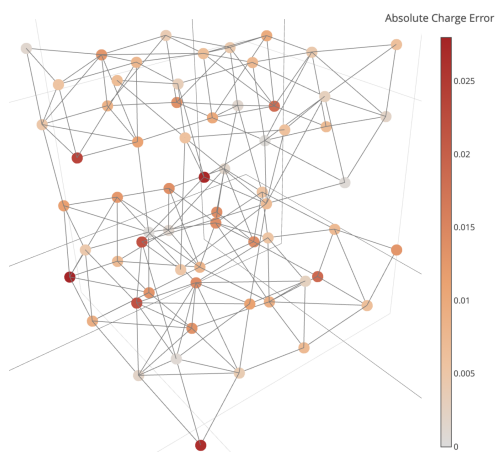
Figure 2.2: Performance on sparse $\text{Li}_{29}\text{H}_{29}$ and non-sparse $\text{Li}_{32}\text{H}_{32}$ test sets. All energies are in a.u.

test set (Figure 2.2 (a)), the BpopNN energies are showing close agreement with the DFT reference energies, with a charge MAE of ~ 0.01 per atom, an energy MAE of ~ 0.15 kcal/mol per atom. Non-sparse $\text{Li}_{32}\text{H}_{32}$ (Figure 2.2 (b)) provides a harder test set - there are more neighbors around each center atom and the Li-H bonds are actively forming and breaking in the AIMD trajectory, which makes it more challenging for the descriptor. We still observe a charge MAE of ~ 0.01 and energy MAE of ~ 0.2 kcal/mol per atom for those cases. There is no obvious deterioration in performance on ionized molecules compared with neutral molecules.

Figure 2.3 (a) compares the partial charges obtained from DFT and the BpopNN SCF-q procedure for a random test neutral $\text{Li}_{32}\text{H}_{32}$ structure; Figure 2.3 (b) depicts the absolute partial charge error for the same molecule. There is a good agreement between the two methods. The atoms on the corner of the molecule are showing slightly larger charge errors than the atoms on the edge or in the middle.

In addition to the SCF-q procedure, geometries and partial charges can be optimized simultaneously to a tight convergence (10^{-7} Hartree/Å for geometries and 10^{-7} Hartree/e for charges) using a conjugate gradient algorithm for both neutral and ionized molecules. We adopted the geometry-optimization procedure that DFT uses, i.e. converging SCF-q at each geometry cycle. In theory there is no need to follow this order, but this method turned out to be more stable and efficient in practice. The cubic $\text{Li}_{32}\text{H}_{32}$ structures were optimized using BpopNN separately for cation, neutral, anion, and compared with the DFT optimized geometries. Figure 2.4 shows the superposition of BpopNN and DFT optimized neutral $\text{Li}_{32}\text{H}_{32}$ structures. The equilibrium geometries were successfully predicted with a root mean square displacement (RMSD) of 0.0522 Å, 0.0958 Å, 0.0516 Å for neutral, cation, and anion, respectively, compared to the true DFT geometries. The superpositions and RMSD values were obtained with Maestro.[133] The final absolute DFT and BpopNN energies are compared in Table 2.2, along with the DFT and BpopNN predicted adiabatic ionization potential (IP) and electron affinity (EA) values. The predicted IP and EA are showing qualitative agreement with DFT (0.167 eV and 0.278 eV error for IP and EA respectively), although they are not highly accurate. These errors are fairly reasonable considering the size of the trial molecule (64 atoms) we are testing on, and the fact that B3LYP functional has 0.15 to 0.18 eV and 0.11 to 0.16 eV mean absolute deviations for IP and EA respectively for atoms and small main group molecules. [134]

To make a preliminary comparison to existing ML methods, we used our training set and the QUIP software package[135] to obtain a few variations of Gaussian approximation potential (GAP) models,[30] differing e.g. in the cutoffs and auxiliary potentials employed (see Appendix Sec. 5.4). It is necessary to train separate GAP models for cation, anion, and neutral. For this, only our training data for equilibrium (i.e., regular SCF) energies are applicable, which amounts to ~ 1100 data points (for each overall charge). For neutral and anion, the GAP energy errors for the sparse ($\text{Li}_{29}\text{H}_{29}$) and nonsparse ($\text{Li}_{32}\text{H}_{32}$) geometries are comparable to those of our model, while for cation, the GAP energy errors are significantly

(a) DFT and BpopNN partial charges for a neutral Li₃₂H₃₂(b) Charge error for a neutral Li₃₂H₃₂Figure 2.3: Charge optimization result for a neutral Li₃₂H₃₂

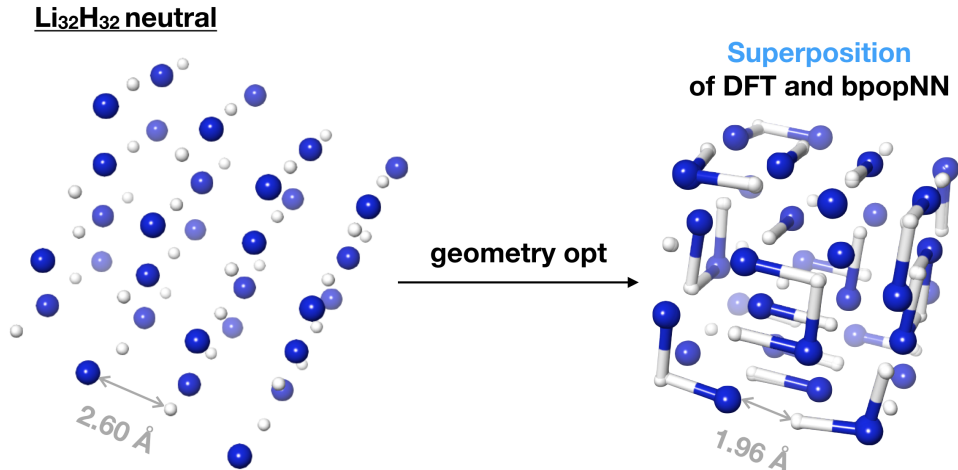


Figure 2.4: Geometry optimization on a cubic $\text{Li}_{32}\text{H}_{32}$ structure

Table 2.2: Energy statistics for a cubic $\text{Li}_{32}\text{H}_{32}$ after geometry optimization

	DFT energy (Hartree)	BpopNN energy (Hartree)	Energy Error (kcal/mol/atom)
Neutral	-260.956	-260.967	0.106
Cation	-206.711	-260.716	0.046
Anion	-260.962	-260.983	0.206
	DFT (eV)	BpopNN (eV)	
IP	6.656	6.823	
EA	0.172	0.450	

larger. As to the latter result, we note that the cation-neutral energy gaps (from DFT or ML) are generally significantly larger than the corresponding neutral-anion gaps, and as shown in Figure 2.9, the distribution of DFT cation energies in the training set is noticeably less uniform than in the other cases. The GAP optimized geometries for the 3 charge states of $\text{Li}_{32}\text{H}_{32}$ are less accurate than those of BpopNN in both energy and structure. Details for the above results are given in the Appendix Sec. 5.4.

3.2.2 Reaction Pathways

As another practical application of our model, we sketched the energy landscape of various structural transformations for neutral and anionic $\text{Li}_{12}\text{H}_{12}$. Diverse stationary-point structures were found for this system, ranging from cuboid structures to hexagonal structures

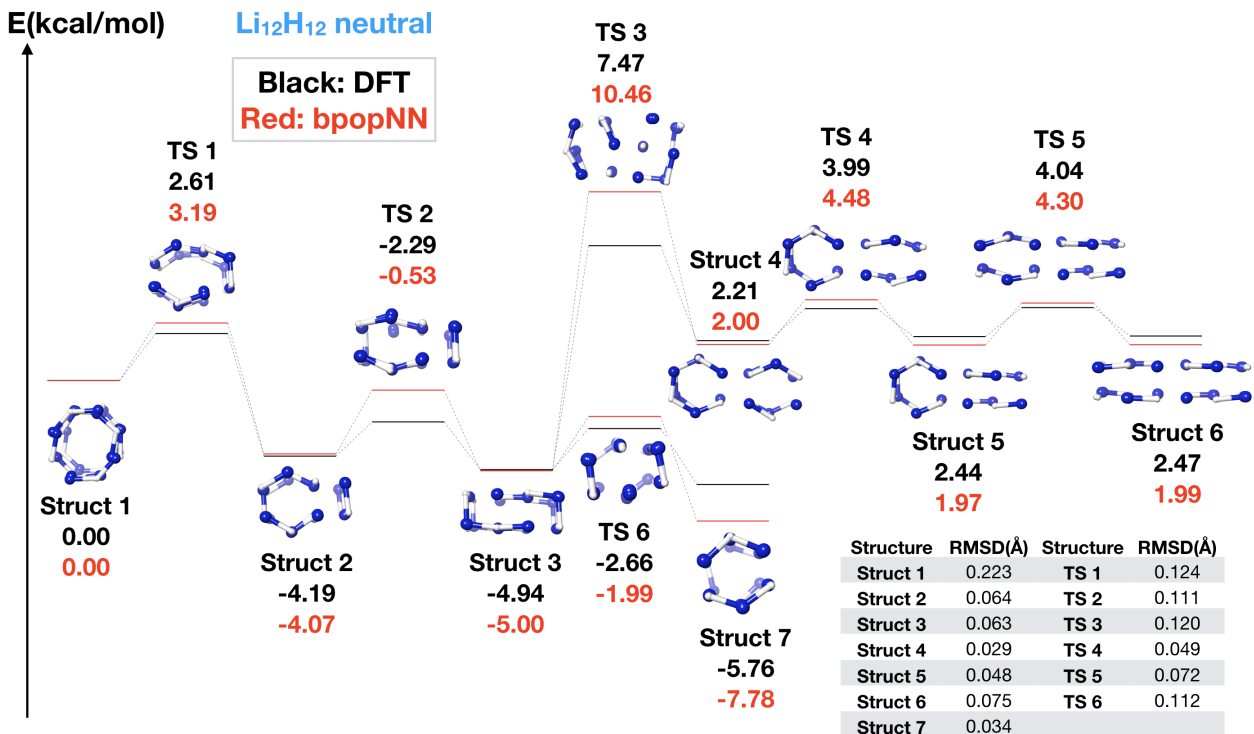


Figure 2.5: Energy profile for various structural transformations for neutral Li₁₂H₁₂. Superpositions and RMSD values were obtained with Maestro.[133]

to structures with fused 6-4-6 membered rings, etc. Although some motifs can be seen in a couple of optimized geometries in the training, these structures are overall distinct from the structures found in the training set. This provides an assessment of how well the model can perform in mapping out energetics for realistic reaction pathways. The resulting energy profile for neutral Li₁₂H₁₂ structures is shown in Figure 2.5 and for anion Li₁₂H₁₂ in Figure 2.6. All structures shown were individually optimized, e.g. the DFT neutral and anionic structures are distinct.

The energies of **Struct 1** from DFT and BpopNN were separately used as baselines, so that only the energy difference between structures is shown. The difference in energy between baseline structures for neutral relative to anion is 9.17 kcal/mol and 10.20 kcal/mol, respectively for DFT and BpopNN. All the minimum structures and the transition states (TSs) connecting them were obtained independently for DFT and BpopNN. Newton’s method was used in the BpopNN to converge the TSs (to geometry gradient threshold 10⁻⁷ Hartree/Å and charge gradient threshold 10⁻⁷ Hartree/e) and the presence of one imaginary frequency was confirmed for all cases. The pathways from TSs to minimum structures in BpopNN were

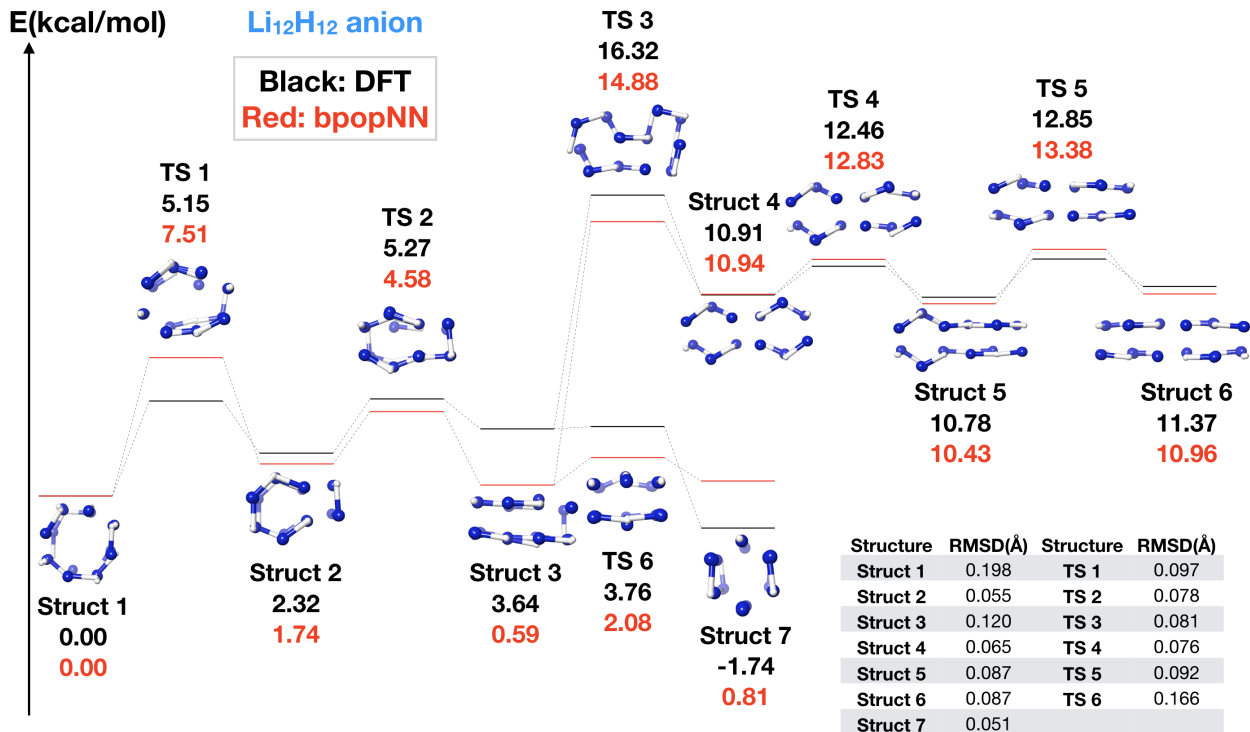


Figure 2.6: Energy profile for various structural transformations for anion Li₁₂H₁₂. Superpositions and RMSD values were obtained with Maestro.[133]

confirmed by a steepest descent algorithm with a small step size. All the DFT TSs were confirmed with one imaginary mode as well and pathways confirmed by Intrinsic Reaction Coordinate (IRC) calculations. In general, the BpopNN geometries show close agreement with DFT geometries (RMSD values are shown in the table inside each figure). Energies for equilibrium structures are qualitatively good, although some deviations (2 to 3 kcal/mol) can be seen for **Structure 7** and anion **Structure 3**. Reaction barriers are showing qualitative agreement as well, although there are cases with noticeable quantitative error (neutral **Structure 3** ↔ **TS 3** ↔ **Structure 4**, neutral **Structure 3** ↔ **TS 6** ↔ **Structure 7**, anion **Structure 2** ↔ **TS 2** ↔ **Structure 3**). Some caution needs to be exercised if one were to predict the lowest energy structure from ML models, especially when the energy range is extremely small (See anion **Structure 1** vs **Structure 7**). At the very least, however, it is helpful to take advantage of the computational speed of the NN and map out the whole transformation picture prior to refinement with full electronic structure calculations.

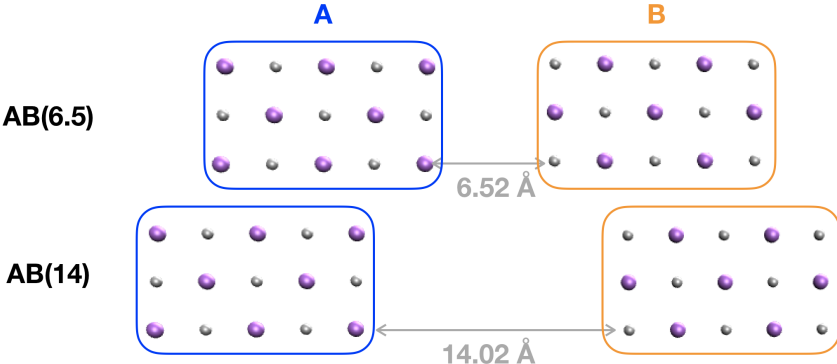


Figure 2.7: A charge separation example for neutral $\text{Li}_{15}\text{H}_{15}$

3.2.3 Long-range environmental effects

Finally, we would like to demonstrate the importance of optimizing the electronic components self-consistently. Namely, we consider a model system that exhibits significant long-range charge effects. Consider a neutral $\text{Li}_{15}\text{H}_{15}$ cluster, which is the side-by-side interaction of two sheets, one Li_8H_7 (“A”) and one Li_7H_8 (“B”) as shown in Figure 2.7. The two sheets are separated by 6.52 Å (“AB(6.5)”) or 14.02 Å (“AB(14)”). The first distance is beyond the cutoff used by other common local position-only descriptors and the second one is even beyond the long cutoff distance used in this work. For B3LYP, both inter-sheet distances exhibit nearly full charge separation: the left sheet (A) bearing +1 charge and the right sheet (B) bearing -1 charge. We note that a ML model based merely on local position-only descriptors will not succeed for this system. In that case, for AB(14), monomer A doesn’t “see” monomer B, so its predicted energy would mirror that for an isolated *neutral* A monomer, and likewise for B. The predicted atom charges and the intra-monomer energies would be incorrect, and the inter-monomer energy, which is largely electrostatic, would nearly be missed entirely.

Table 2.3: A charge separation example for neutral $\text{Li}_{15}\text{H}_{15}$

	DFT energy (Hartree)	BpopNN energy (Hartree)	Absolute error (kcal/mol)
AB(6.5)	-121.524 ^a	-121.534	6.232
AB(14)	-121.506 ^a	-121.525	11.801
AB(1000)	-121.489 ^a	-121.515	16.286
A	-64.213	-64.224	7.213
B	-57.241	-57.254	8.403
A ⁺	-64.110	-64.112	1.210
B ⁻	-57.372	-57.360	7.584
	DFT (kcal/mol)	BpopNN (kcal/mol)	
AB(6.5) BE ^b	-44.143	-34.760	
AB(14) BE	-33.168	-29.345	
	DFT (e)	BpopNN (e)	
$q(\text{A})^c$ in AB(6.5)	0.993	0.729	
$q(\text{A})$ in AB(14)	0.979	0.681	
$q(\text{A})$ in AB(1000)	0.683	0.609	

^a For CT solutions. There exist DFT “neutral-neutral” solutions, with energies of -121.4542 and -121.4536 respectively for AB(6.5) and AB(14). This is very close to the sum of the energies of A and B.

^b Binding Energy.

^c sum of partial charges on A atoms.

As shown in Table 2.3, this problem can be solved qualitatively by incorporating self-consistent electronic information: our model can predict the charge separation and qualitatively correct binding energies, although there were no direct scenarios of this kind in our training data.

At extremely large A-B separation, AB(1000), both DFT and BpopNN exhibit a partially ionized state. For DFT, this stems from the well-known “self interaction” error (also called “delocalization error”), and this kind of result is also expected of BpopNN. For AB(6.5) and AB(14), BpopNN shows increased A-B charge separation relative to infinite separation. The principal driving force for this, and the only one for AB(14), is the electrostatic attraction between A^+ and B^- . In other words, the long-range environment is the sole reason for increased charge separation in BpopNN in this case, which explicitly establishes that the model is adaptable to the environment.

The sum of BpopNN energies for A^+ and B^- is -121.4718 a.u., which actually is slightly above the sum of BpopNN energies for A and B, -121.4785 a.u. The situation is opposite for DFT, which may partially explain the increased A-B charge separation for DFT for AB(6.5) and AB(14).

IV Conclusions

In this chapter, we have highlighted the importance of incorporating adaptable electronic information into ML models of PESs. For this, we presented a new general approach for building ML models whose target function is a DFT populations-to-energy map, afforded in practice by CDFT. Such models are functions of atom positions and atom-based electron populations; optimization with respect to the latter variables allows for self-consistent adjustment to the electronic environment, akin to optimizing the density in DFT. This leads to features analogous to regular orbital or density based SCF: solutions for different overall charge states; solutions with different character (e.g. CT and non CT); spin polarized versus not, etc. In this sense, we view this as a reduced or simplified electronic structure approach. As a byproduct, this model can be used as a high-quality charge equilibration scheme, i.e. partial charges of this model can also be used by themselves to infer molecular properties, e.g. determining nucleophilic and electrophilic sites.

As a proof of concept for this approach, we trained a BpopNN model for Li_nH_n systems. For this, we simplified the target-energy map to use only total electron populations on each atom, as opposed to the two spin populations. Subsequent testing on a variety of Li_nH_n examples confirmed that the model exhibits qualitative accuracy for the energy of different overall charge states, a flexible reduced electronic structure whose populations are generally close to those of the parent DFT functional and that can adapt to the (particularly medium to long-range) environment, and transferability to clusters of different sizes. We conclude that BpopNN is a positive first step towards addressing some outstanding challenges for ML

models, namely the handling of systems with electronic structure that is sensitive to the surrounding environment.

This first application was designed for qualitative testing and it has several aspects that can explain various inaccuracies and that are ready for refinement in future work. Although partial charges are in principle adequate as a basis for the BpopNN approach, spin populations will need to be incorporated in practice to effectively model the intricacies of the CDFT map. In this context, it should then be noted that the reduction of the density to atom-based populations is the furthest one can reasonably go; clearly less drastic reductions should also be explored, and perhaps a useful hierarchy for this can be established.

Improvements in the descriptor can be made. Our descriptor is based on the ESP generated by the model’s populations as point charges, and it uses simple Gaussian functions to approximate the ESP for each point. We could better approximate these ESPs or incorporate them exactly. It would be particularly interesting to combine the ideas of this chapter with those of LODE:[33] instead of using point charges to generate the ESP, proxy densities, now integrating to the current populations on each atom, could be used to generate ESPs.

The fact that BpopNN models attempt to model DFT partial charges is advantageous for the incorporation of long-range electrostatic interactions. The simple approach used here, based on attenuated pairwise point-charge Coulomb interactions, would clearly benefit from refinements to more reasonably approximate the local morphology of the interacting density pieces.

Experimenting with the choice of ML architecture would be worthwhile, and especially experimenting with the NN depth. Of course, this would require a larger training dataset. The dataset used here is on the smaller side, so enlarging this is a focal point for future work. There is also the plausibility that our dataset is insufficiently balanced. More specifically, it only contains alternating Li-H clusters and cannot be reasonably applied to clusters with distinguishable Li_n and $(\text{H}_2)_n$ substructures, its geometries were obtained from trajectories using only the neutral charge state, and generally our sampling method of particular AIMD and optimization trajectories might not be varied enough. In the future we will explore more efficient and balanced sampling methods, such as normal mode sampling.

In the future, we are primarily interested in moving beyond the LiH-cluster scope to more general Li-organic systems. This significant increase in complexity should entail many adjustments to the general approach presented above, and will serve as a firm test of its efficacy.

V Appendix

5.1 Population gradient of CDFT energy

In the present context, the CDFT Lagrangian takes the following form:

$$L_{\text{CDFT}} = E[\gamma] + \sum_{\sigma} \sum_{i=1}^{N_{\text{atom}}-1} \lambda_{i,\sigma} (p_{i,\sigma}(\rho) - c_{i,\sigma}), \quad (2.13)$$

where γ is the density matrix, the $\lambda_{i,\sigma}$ are Lagrange multipliers, the $p_{i,\sigma}(\rho)$ are the functions defined in eqn. (2.2), and the $c_{i,\sigma}$ are the target values of the constraints. For each spin, the sum excludes one atom because the value of this last variable is implied by the usual condition that the density integrates to the total number of electrons of that spin. An equivalent result is obtained if a different atom is the one excluded from the sum, although of course the λ_i and c_i will adjust accordingly. This Lagrangian also formally applies to cases in which fewer population constraints are used. In that case, we simply set the c_i for any technically unconstrained population to its relaxed value, and the associated λ_i is 0.

Optimizing L_{CDFT} with respect to the orbitals (i.e. γ) and the Lagrange multipliers gives the CDFT energy, $E_{\text{CDFT}}[\{c_{i,\sigma}\}]$. At a stationary point of L_{CDFT} , we may apply the “envelope” theorem to obtain

$$\frac{\partial E_{\text{CDFT}}}{\partial c_{i,\sigma}} = \frac{\partial L_{\text{CDFT}}}{\partial c_{i,\sigma}} = -\lambda_{i,\sigma}. \quad (2.14)$$

The target energy function $E_t[\mathbf{p}]$ is essentially the same as $E_{\text{CDFT}}[\{c_{i,\sigma}\}]$ except that it is a function of all $2N_{\text{atom}}$ population variables. So we have to do a transformation between variables. Again excluding the last atom, we can fix all $c_{i,\sigma}$ but one, $c_{j,\sigma}$, and then vary $c_{j,\sigma}$. Upon (re-)optimizing L_{CDFT} for this new set of constraints, $p_{j,\sigma}$ changes identically as $c_{j,\sigma}$ does, and $p_{N_{\text{atom}},\sigma}$ changes by the opposite amount, while all other populations are unchanged. Denoting the direction corresponding to this overall change in populations by $\mathbf{e}_{j,\sigma}$, we thus have

$$\nabla E_t[\mathbf{p}] \cdot \mathbf{e}_{i,\sigma} = \frac{\partial E_{\text{CDFT}}}{\partial c_{i,\sigma}} = -\lambda_{i,\sigma}. \quad (2.15)$$

The vectors $\mathbf{e}_{i,\sigma}$ are (most of) the columns of a block diagonal matrix W with two N_{atom} by N_{atom} blocks, one for each spin type. That is, $W = W_s \oplus W_s$ with

$$W_s = \begin{bmatrix} 1 & 0 & \dots & 0 & 1 \\ 0 & 1 & \dots & 0 & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 1 \\ -1 & -1 & \dots & -1 & 1 \end{bmatrix}. \quad (2.16)$$

The last column in W_s corresponds to the fully symmetric direction of changing all populations (of one spin type) by the same amount. Because the total numbers of each spin are constant, the component of ∇E_t along this direction is 0. We thus have

$$\nabla E_t \cdot W = \boldsymbol{\nu}, \quad (2.17)$$

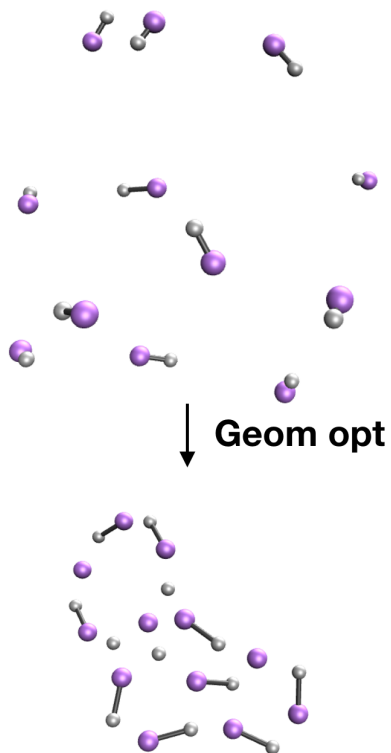
where the latter vector contains the $-\lambda_i$ along with two 0's for the symmetric directions:

$$\nu_i = \begin{cases} -\lambda_{i,\alpha} & \text{if } i < N_{\text{atom}} \\ 0 & \text{if } i = N_{\text{atom}} \\ -\lambda_{i-N_{\text{atom}},\beta} & \text{if } N_{\text{atom}} < i < 2N_{\text{atom}} \\ 0 & \text{if } i = 2N_{\text{atom}} \end{cases}. \quad (2.18)$$

By inverting W we readily obtain $\nabla E_t[\boldsymbol{p}]$ (in the regular population basis).

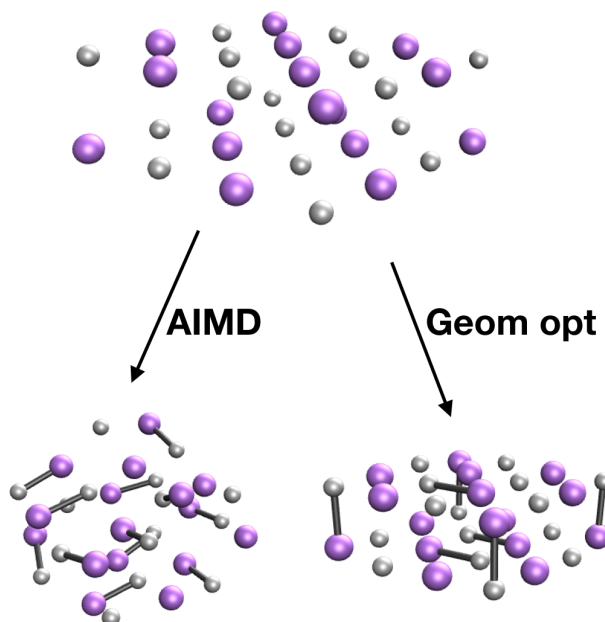
5.2 Details on training data

1. Sparse LinHn clusters from geom opt trajectories



**Li_4H_4 , Li_8H_8 , $\text{Li}_{12}\text{H}_{12}$,
 $\text{Li}_{16}\text{H}_{16}$, $\text{Li}_{20}\text{H}_{20}$, $\text{Li}_{24}\text{H}_{24}$**

2. Non-sparse LinHn clusters from aimd and geom opt trajectories



**LiH , Li_4H_4 , Li_8H_8 , $\text{Li}_{13}\text{H}_{13}$,
 $\text{Li}_{16}\text{H}_{16}$, $\text{Li}_{20}\text{H}_{20}$, $\text{Li}_{24}\text{H}_{24}$**

Figure 2.8: Example geometries of the training data

In Figure 2.9, we show a plot of the distribution of DFT energies for the training data.

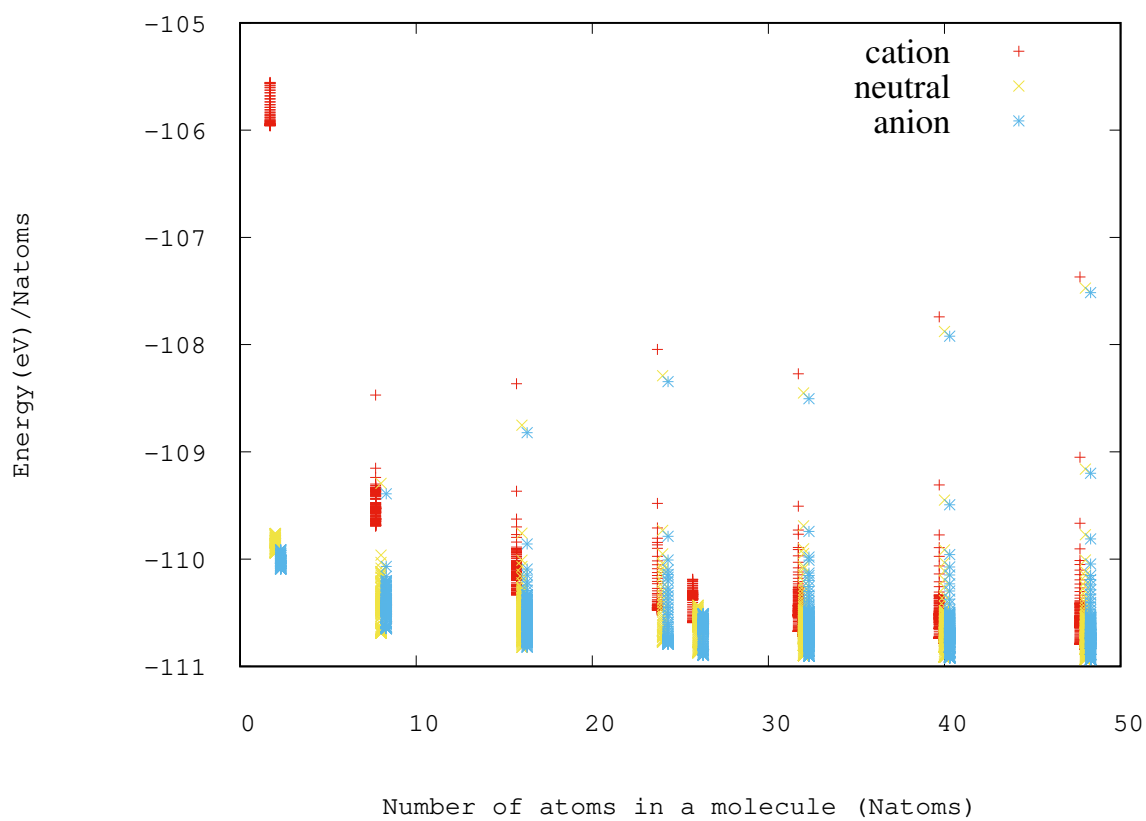


Figure 2.9: DFT energy per atom plot for the neutral, cation, anion training set. To aid visibility, the data has been slightly shifted horizontally.

5.3 Statistics for NN models with shorter cutoff

Table 2.4: Comparison across NN models with different radial cutoff distances for $\text{Li}_{29}\text{H}_{29}$ sparse test set

	MAE charge (e/atom)	MAE energy (kcal/mol/atom)	MAE charge (e/atom)	MAE energy (kcal/mol/atom)	MAE charge (e/atom)	MAE energy (kcal/mol/atom)
	NN-1 ^a		NN-2 ^b		NN-3 ^c	
$\text{Li}_{29}\text{H}_{29}$ neutral	0.009	0.112	0.009	0.181	0.011	0.129
$\text{Li}_{29}\text{H}_{29}$ cation	0.012	0.100	0.013	0.148	0.012	0.086
$\text{Li}_{29}\text{H}_{29}$ anion	0.018	0.148	0.017	0.138	0.017	0.075
	NN-4 ^d		NN-5 ^e		NN-6 ^f	
$\text{Li}_{29}\text{H}_{29}$ neutral	0.009	0.172	0.009	0.144	0.011	0.207
$\text{Li}_{29}\text{H}_{29}$ cation	0.011	0.144	0.011	0.103	0.013	0.181
$\text{Li}_{29}\text{H}_{29}$ anion	0.030	0.136	0.016	0.094	0.018	0.094
	NN-7 ^g		NN-8 ^h		NN-9 ⁱ	
$\text{Li}_{29}\text{H}_{29}$ neutral	0.009	0.083	0.010	0.251	0.009	0.105
$\text{Li}_{29}\text{H}_{29}$ cation	0.011	0.151	0.011	0.177	0.011	0.084
$\text{Li}_{29}\text{H}_{29}$ anion	0.016	0.161	0.050	0.156	0.017	0.148

^a NN-1: Cutoff 5 Å, evenly spacing radial grid of 32 grid points between 0.5 Å and 5 Å.

^b NN-2: Cutoff 6 Å, evenly spacing radial grid of 32 grid points between 0.5 Å and 6 Å.

^c NN-3: Cutoff 8 Å, evenly spacing radial grid of 32 grid points between 0.5 Å and 8 Å.

^d NN-4: Cutoff 13.2 Å, evenly spacing radial grid of 32 grid points between 0.5 Å and 5 Å.

^e NN-5: Cutoff 13.2 Å, evenly spacing radial grid of 32 grid points between 0.5 Å and 6 Å.

^f NN-6: Cutoff 13.2 Å, evenly spacing radial grid of 32 grid points between 0.5 Å and 7 Å.

^g NN-7: Cutoff 13.2 Å, evenly spacing radial grid of 25 grid points between 0.5 Å and 5 Å.

^h NN-8: Cutoff 13.2 Å, evenly spacing radial grid of 25 grid points between 0.5 Å and 6 Å.

ⁱ NN-9: Cutoff 13.2 Å, evenly spacing radial grid of 25 grid points between 0.5 Å and 7 Å.

Table 2.5: Comparison across NN models with different radial cutoff distances for $\text{Li}_{32}\text{H}_{32}$ non-sparse test set (Models are the same as those in Table 2.4)

	MAE charge (e/atom)	MAE energy (kcal/mol/atom)	MAE charge (e/atom)	MAE energy (kcal/mol/atom)	MAE charge (e/atom)	MAE energy (kcal/mol/atom)
	NN-1		NN-2		NN-3	
$\text{Li}_{32}\text{H}_{32}$ neutral	0.006	0.152	0.007	0.118	0.008	0.186
$\text{Li}_{32}\text{H}_{32}$ cation	0.010	0.105	0.010	0.145	0.011	0.149
$\text{Li}_{32}\text{H}_{32}$ anion	0.012	0.210	0.014	0.221	0.014	0.173
	NN-4		NN-5		NN-6	
$\text{Li}_{32}\text{H}_{32}$ neutral	0.007	0.218	0.007	0.163	0.008	0.299
$\text{Li}_{32}\text{H}_{32}$ cation	0.010	0.124	0.011	0.170	0.011	0.209
$\text{Li}_{32}\text{H}_{32}$ anion	0.014	0.173	0.013	0.210	0.043	0.203
	NN-7		NN-8		NN-9	
$\text{Li}_{32}\text{H}_{32}$ neutral	0.006	0.099	0.007	0.244	0.007	0.162
$\text{Li}_{32}\text{H}_{32}$ cation	0.010	0.160	0.011	0.134	0.011	0.153
$\text{Li}_{32}\text{H}_{32}$ anion	0.013	0.156	0.013	0.213	0.014	0.166

Table 2.6: Energy statistics for a cubic $\text{Li}_{32}\text{H}_{32}$ after geometry optimization (Models are the same as those in Table 2.4)

	DFT (a.u.)	NN-1 (a.u.)	NN-2 (a.u.)	NN-3 (a.u.)	NN-4 (a.u.)	NN-5 (a.u.)
Neutral	-260.956	-260.964	-260.967	-260.924	-260.948	-260.960
Cation	-206.711	-260.728	-260.782	-260.806	-260.721	-260.697
Anion	-260.962	-260.991	-260.983	-260.984	-260.983	-260.998
	NN-6 (a.u.)	NN-7 (a.u.)	NN-8 (a.u.)	NN-9 (a.u.)	Our best model (a.u.) ^a	
Neutral	-260.946	-260.946	-260.929	-260.949	-260.967	
Cation	-260.683	-260.713	-260.699	-260.691	-260.716	
Anion	-260.975	-260.963	-260.930	-260.967	-260.983	
	DFT (eV)	NN-1 (eV)	NN-2 (eV)	NN-3 (eV)	NN-4 (eV)	NN-5 (eV)
IP	6.656	6.427	5.038	3.226	6.181	7.177
EA	0.172	0.732	0.431	1.615	0.937	1.011
	NN-6 (eV)	NN-7 (eV)	NN-8 (eV)	NN-9 (eV)	Our best model (eV) ^a	
IP	7.172	6.324	6.262	7.000	6.823	
EA	0.788	0.477	0.009	0.501	0.450	

^a The BpopNN model used in the main text.

Note that significant deviations from the DFT geometries were observed for the following combinations: NN-2 $\text{Li}_{32}\text{H}_{32}$ cation, NN-3 $\text{Li}_{32}\text{H}_{32}$ cation and neutral .

5.4 Performance of the GAP models on the current test set

Table 2.7: Performance of GAP models (trained with unique geometries with equilibrium charges, more details below) on the $\text{Li}_{29}\text{H}_{29}$ sparse and $\text{Li}_{32}\text{H}_{32}$ non-sparse test sets (energy MAE in kcal/mol/atom).

	GAP-1 ^a	GAP-2 ^b	GAP-3 ^c	GAP-4 ^d	GAP-5 ^e
$\text{Li}_{29}\text{H}_{29}$ sparse neutral	0.277	0.253	0.467	0.323	1.864
$\text{Li}_{29}\text{H}_{29}$ sparse cation	2.777	2.778	2.909	2.910	2.929
$\text{Li}_{29}\text{H}_{29}$ sparse anion	0.473	0.439	0.466	0.479	1.934
$\text{Li}_{32}\text{H}_{32}$ non-sparse neutral	0.254	0.438	0.315	0.272	1.438
$\text{Li}_{32}\text{H}_{32}$ non-sparse cation	4.963	4.266	1.117	1.014	1.150
$\text{Li}_{32}\text{H}_{32}$ non-sparse anion	0.400	0.430	0.285	0.284	1.516

^a GAP-1: Using 2 and 3-body and SOAP potential.[30] This is showing much better result than training with SOAP alone.

^b GAP-2: Using 2 and 3-body and SOAP potential trained with both energy and nuclear position forces.

^c GAP-3: Using 2-body and SOAP potential.

^d GAP-4: Using 2-body and SOAP potential trained with both energy and nuclear position forces.

^e GAP-5: Using SOAP potential with long cutoff (13.2 Å).

Table 2.8: Performance of GAP models on $\text{Li}_{32}\text{H}_{32}$ geometry optimizations. Absolute energies are shown for the optimized geometries in each force field.

	DFT	Our best model ^a	GAP-1	GAP-2	GAP-3	GAP-4	GAP-5
Neutral Absolute Energy (Hartree)	-260.956	-260.967	-261.131	-261.103	-260.992	-261.284	N/A ^b
Cation Absolute Energy (Hartree)	-260.711	-260.716	N/A ^b	N/A ^b	-276.864	N/A ^b	N/A ^b
Anion Absolute Energy (Hartree)	-260.962	-260.983	-261.446	-261.078	-260.980	-260.980	N/A ^b

^a The BpopNN model used in the main text.

^b Unable to converge using BFGS line search algorithm in the ase package.[136]

The GAP models were trained with the QUIP code. [135] Details of the GAP model can be found in the literature.[30] Other parameters we used to train the GAP models are listed in the following command lines. The value of those parameters were determined by a rough

grid search, and the ones showing the best performance on the $\text{Li}_{29}\text{H}_{29}$ and $\text{Li}_{32}\text{H}_{32}$ test set were used. Note that for the geometry optimizations on the $\text{Li}_{32}\text{H}_{32}$ molecules, we found it necessary to use the DFT optimized geometries as the initial guess (unlike in the case of our model). Out of all the GAP models, GAP-3 gave reasonable geometries for $\text{Li}_{32}\text{H}_{32}$ neutral, $\text{Li}_{32}\text{H}_{32}$ anion compared with the DFT optimized geometries, with RMSD values of 0.0804 Å and 0.0798 Å respectively.

Below are the command lines we used to train the GAP models for the neutral training set. For training the cation and anion sets, a simple change to the corresponding xyz files can be made.

GAP-1:

```
$ gap_fit at_file=LiH_all_training_data_neutral_equilibrium.xyz gap={
  distance_Nb order=2 cutoff=5.0 covariance_type=ARD_SE theta_uniform=1.0
  n_sparse=15 delta=1.0: distance_Nb order=3 cutoff=4.0 covariance_type=
  ARD_SE theta_uniform=1.0 n_sparse=50 delta=0.004: soap cutoff=4.0
  covariance_type=dot_product zeta=2 delta=0.016 atom_sigma=0.7 l_max=6
  n_max=6 n_sparse=200 sparse_method=cur_points} e0={Li:-203.67672091316:
  H:-13.63995354104} default_sigma={0.001 0.5 0.0 0.0} do_copy_at_file=F
  sparse_separate_file=F gp_file=LiH_neutral_equi_gap_2b3bsoap.xml 2>&1 |
  grep -v FoX
```

GAP-2:

```
$ gap_fit at_file=LiH_all_training_data_neutral_equilibrium.xyz gap={
  distance_Nb order=2 cutoff=5.0 covariance_type=ARD_SE theta_uniform=1.0
  n_sparse=15 delta=1.0: distance_Nb order=3 cutoff=4.0 covariance_type=
  ARD_SE theta_uniform=1.0 n_sparse=50 delta=0.004: soap cutoff=4.0
  covariance_type=dot_product zeta=2 delta=0.016 atom_sigma=0.7 l_max=6
  n_max=6 n_sparse=200 sparse_method=cur_points} e0={Li:-203.67672091316:
  H:-13.63995354104} default_sigma={0.001 0.5 0.0 0.0} do_copy_at_file=F
  sparse_separate_file=F force_parameter_name=forces gp_file=
  LiH_neutral_equi_gap_2b3bsoap_w_forces.xml 2>&1 | grep -v FoX
```

GAP-3:

```
$ gap_fit at_file=LiH_all_training_data_neutral_equilibrium.xyz gap={
  distance_Nb order=2 cutoff=5.0 covariance_type=ARD_SE theta_uniform=1.0
  n_sparse=15 delta=1.0: soap cutoff=4.0 covariance_type=dot_product
  zeta=2 delta=0.016 atom_sigma=0.7 l_max=6 n_max=6 n_sparse=200
  sparse_method=cur_points} e0={Li:-203.67672091316:H:-13.63995354104}
  default_sigma={0.001 0.5 0.0 0.0} do_copy_at_file=F
  sparse_separate_file=F gp_file=LiH_neutral_equi_gap_2bsoap.xml 2>&1 |
  grep -v FoX &
```

GAP-4:

```
$ gap_fit at_file=LiH_all_training_data_neutral_equilibrium.xyz gap={
  distance_Nb order=2 cutoff=5.0 covariance_type=ARD_SE theta_uniform=1.0
  n_sparse=15 delta=1.0: soap cutoff=4.0 covariance_type=dot_product
  zeta=2 delta=0.016 atom_sigma=0.7 l_max=6 n_max=6 n_sparse=200
  sparse_method=cur_points} e0={Li:-203.67672091316:H:-13.63995354104}
  default_sigma={0.001 0.5 0.0 0.0} do_copy_at_file=F
  sparse_separate_file=F force_parameter_name=forces gp_file=
  LiH_neutral_equi_gap_2bsoap_w_forces.xml 2>&1 | grep -v FoX &
```

GAP-5:

```
$ gap_fit at_file=LiH_all_training_data_neutral_equilibrium.xyz gap={ soap
  cutoff=13.2 covariance_type=dot_product zeta=2 delta=0.001 atom_sigma
  =0.7 l_max=6 n_max=30 n_sparse=400 sparse_method=cur_points} e0={Li
  :-203.67672091316:H:-13.63995354104} default_sigma={0.001 0.5 0.0 0.0}
  do_copy_at_file=F sparse_separate_file=F gp_file=
  LiH_neutral_equi_gap_soap_cutoff_13dot2.xml 2>&1 | grep -v FoX &
```

Bibliography

- [1] Robert G Parr. “Density functional theory of atoms and molecules”. In: *Horizons of Quantum Chemistry*. Springer, 1980, pp. 5–15.
- [2] R. Car and M. Parrinello. “Unified approach for molecular dynamics and density-functional theory”. In: *Phys. Rev. Lett.* 55.22 (1985), p. 2471.
- [3] Dominik Marx and Jürg Hutter. *Ab initio molecular dynamics: basic theory and advanced methods*. Cambridge University Press, 2009.
- [4] Tomas Hansson, Chris Oostenbrink, and WilfredF van Gunsteren. “Molecular dynamics simulations”. In: *Curr. Opin. Struc. Biol.* 12.2 (2002), pp. 190–196.
- [5] Martin Karplus and J Andrew McCammon. “Molecular dynamics simulations of biomolecules”. In: *Nat. Struct. Mol. Biol.* 9.9 (2002), p. 646.
- [6] Jörg Behler. “First principles neural network potentials for reactive simulations of large molecular and condensed systems”. In: *Angew. Chem. Int. Ed.* 56.42 (2017), pp. 12828–12840.
- [7] Jonathan Schmidt et al. “Recent advances and applications of machine learning in solid-state materials science”. In: *Npj Comput. Mater.* 5.1 (2019), pp. 1–36.
- [8] Volker L Deringer, Miguel A Caro, and Gábor Csányi. “Machine Learning Interatomic Potentials as Emerging Tools for Materials Science”. In: *Adv. Mater.* 31.46 (2019), p. 1902765.
- [9] Gabriel R Schleder et al. “From DFT to machine learning: recent approaches to materials science—a review”. In: *J. Phys.: Mater.* 2.3 (2019), p. 032001.
- [10] Jörg Behler and Michele Parrinello. “Generalized neural-network representation of high-dimensional potential-energy surfaces”. In: *Phys. Rev. Lett.* 98.14 (2007), p. 146401.
- [11] Jörg Behler. “Atom-centered symmetry functions for constructing high-dimensional neural network potentials”. In: *J. Chem. Phys.* 134.7 (2011), p. 074106.
- [12] Albert P Bartók, Risi Kondor, and Gábor Csányi. “On representing chemical environments”. In: *Phys. Rev. B* 87.18 (2013), p. 184115.

-
- [13] J Behler. *RuNNer-A Neural Network Code for High-Dimensional Potential-Energy Surfaces*; Universität Göttingen, 2018.
- [14] Kun Yao et al. “The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics”. In: *Chem. Sci.* 9.8 (2018), pp. 2261–2269.
- [15] Kun Yao, John E Herr, and John Parkhill. “The many-body expansion combined with neural networks”. In: *J. Chem. Phys.* 146.1 (2017), p. 014106.
- [16] John E Herr et al. “Metadynamics for training neural network model chemistries: A competitive assessment”. In: *J. Chem. Phys.* 148.24 (2018), p. 241710.
- [17] Kun Yao et al. “Intrinsic bond energies from a bonds-in-molecules neural network”. In: *J. Phys. Chem. Lett.* 8.12 (2017), pp. 2689–2694.
- [18] John E Herr et al. “Compressing physics with an autoencoder: Creating an atomic species representation to improve machine learning models in the chemical sciences”. In: *J. Chem. Phys.* 151.8 (2019), p. 084103.
- [19] Kristof T Schütt et al. “SchNet—A deep learning architecture for molecules and materials”. In: *J. Chem. Phys.* 148.24 (2018), p. 241722.
- [20] KT For an orbital-based ML approach whose goal is related, see Schütt et al. “Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions”. In: *Nat. Commun.* 10.1 (2019), pp. 1–10.
- [21] Justin S Smith, Olexandr Isayev, and Adrian E Roitberg. “ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost”. In: *Chem. Sci.* 8.4 (2017), pp. 3192–3203.
- [22] Justin S Smith et al. “Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning”. In: *Nat. Commun.* 10.1 (2019), pp. 1–8.
- [23] Justin S Smith et al. “Less is more: Sampling chemical space with active learning”. In: *J. Chem. Phys.* 148.24 (2018), p. 241733.
- [24] Alireza Khorshidi and Andrew A Peterson. “Amp: A modular approach to machine learning in atomistic simulations”. In: *Comput. Phys. Commun.* 207 (2016), pp. 310–324.
- [25] Han Wang et al. “DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics”. In: *Comput. Phys. Commun.* 228 (2018), pp. 178–184.
- [26] Si-Da Huang et al. “LASP: Fast global potential energy surface exploration”. In: *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 9.6 (2019), e1415.
- [27] AS Christensen et al. “QML: A Python toolkit for quantum machine learning”. In: URL <https://github.com/qmlcode/qml> (2017).

-
- [28] Felix A Faber et al. “Alchemical and structural distribution based representation for universal quantum machine learning”. In: *J. Chem. Phys.* 148.24 (2018), p. 241717.
- [29] Oliver Thorsten Unke and Markus Meuwly. “PhysNet: a neural network for predicting energies, forces, dipole moments and partial charges”. In: *J. Chem. Theory Comput.* (2019).
- [30] Albert P Bartók and Gábor Csányi. “Gaussian approximation potentials: A brief tutorial introduction”. In: *Int. J. Quantum Chem.* 115.16 (2015), pp. 1051–1057.
- [31] Andrea Grisafi et al. “Transferable machine-learning model of the electron density”. In: *ACS Cent. Sci.* 5.1 (2018), pp. 57–64.
- [32] Andrea Grisafi et al. “Symmetry-adapted machine learning for tensorial properties of atomistic systems”. In: *Phys. Rev. Lett.* 120.3 (2018), p. 036002.
- [33] Andrea Grisafi and Michele Ceriotti. “Incorporating long-range physics in atomic-scale machine learning”. In: *J. Chem. Phys.* 151.20 (2019), p. 204105.
- [34] Sandip De et al. “Comparing molecules and solids across structural and alchemical space”. In: *Phys. Chem. Chem. Phys.* 18.20 (2016), pp. 13754–13769.
- [35] Michael J Willatt, Félix Musil, and Michele Ceriotti. “Feature optimization for atomistic machine learning yields a data-driven construction of the periodic table of the elements”. In: *Phys. Chem. Chem. Phys.* 20.47 (2018), pp. 29661–29668.
- [36] Aidan P Thompson et al. “Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials”. In: *J. Comput. Phys.* 285 (2015), pp. 316–330.
- [37] Yu-Hang Tang and Wibe A de Jong. “Prediction of atomization energy using graph kernel and active learning”. In: *J. Chem. Phys.* 150.4 (2019), p. 044107.
- [38] Grégoire Ferré, Terry Haut, and Kipton Barros. “Learning molecular energies using localized graph kernels”. In: *J. Chem. Phys.* 146.11 (2017), p. 114107.
- [39] Justin Gilmer et al. “Neural message passing for quantum chemistry”. In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org. 2017, pp. 1263–1272.
- [40] Chi Chen et al. “Graph networks as a universal machine learning framework for molecules and crystals”. In: *Chem. Mater.* 31.9 (2019), pp. 3564–3572.
- [41] Matthias Rupp et al. “Fast and accurate modeling of molecular atomization energies with machine learning”. In: *Phys. Rev. Lett.* 108.5 (2012), p. 058301.
- [42] Katja Hansen et al. “Assessment and validation of machine learning methods for predicting molecular atomization energies”. In: *J. Chem. Theory Comput.* 9.8 (2013), pp. 3404–3419.

-
- [43] Katja Hansen et al. “Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space”. In: *J. Phys. Chem. Lett.* 6.12 (2015), pp. 2326–2331.
- [44] Kejie Shao et al. “Communication: Fitting potential energy surfaces with fundamental invariant neural network”. In: *J. Chem. Phys.* 145.7 (2016), p. 071101. DOI: 10.1063/1.4961454. eprint: <https://doi.org/10.1063/1.4961454>. URL: <https://doi.org/10.1063/1.4961454>.
- [45] Volker L Deringer and Gábor Csányi. “Machine learning based interatomic potential for amorphous carbon”. In: *Phys. Rev. B* 95.9 (2017), p. 094203.
- [46] Gabriele C Sosso et al. “Understanding the thermal properties of amorphous solids using machine-learning-based interatomic potentials”. In: *Mol. Simulat.* 44.11 (2018), pp. 866–880.
- [47] Miguel A Caro et al. “Growth Mechanism and Origin of High s p 3 Content in Tetrahedral Amorphous Carbon”. In: *Phys. Rev. Lett.* 120.16 (2018), p. 166101.
- [48] Nongnuch Artrith, Alexander Urban, and Gerbrand Ceder. “Constructing first-principles phase diagrams of amorphous Li x Si using machine-learning-assisted sampling with an evolutionary algorithm”. In: *J. Chem. Phys.* 148.24 (2018), p. 241711.
- [49] Michael Bendikov et al. “Oligoacenes: theoretical prediction of open-shell singlet diradical ground states.” In: *J. Am. Chem. Soc.* 126 (2004), pp. 7416–7417. ISSN: 0002-7863. DOI: 10.1021/ja048919w.
- [50] Johannes Hachmann et al. “The radical character of the acenes: a density matrix renormalization group study.” In: *J. Chem. Phys.* 127 (2007), p. 134309. ISSN: 0021-9606. DOI: 10.1063/1.2768362.
- [51] De-en Jiang and Sheng Dai. “Electronic Ground State of Higher Acenes”. In: *J. Phys. Chem. A* 112.2 (2008), pp. 332–335. DOI: 10.1021/jp0765087. eprint: <https://doi.org/10.1021/jp0765087>. URL: <https://doi.org/10.1021/jp0765087>.
- [52] Sudipta Dutta and Swapan K. Pati. “Novel properties of graphene nanoribbons: a review”. In: *J. Mater. Chem.* 20 (38 2010), pp. 8207–8223. DOI: 10.1039/C0JM00261E. URL: <http://dx.doi.org/10.1039/C0JM00261E>.
- [53] F. Cervantes-Sodi et al. “Edge-functionalized and substitutionally doped graphene nanoribbons: Electronic and spin properties”. In: *Phys. Rev. B* 77 (16 Apr. 2008), p. 165427. DOI: 10.1103/PhysRevB.77.165427. URL: <https://link.aps.org/doi/10.1103/PhysRevB.77.165427>.
- [54] Veronica Barone, Oded Hod, and Gustavo E. Scuseria. “Electronic Structure and Stability of Semiconducting Graphene Nanoribbons”. In: *Nano Lett.* 6.12 (2006), pp. 2748–2754. DOI: 10.1021/nl0617033.

- [55] H. Meier, U. Stalmach, and H. Kolshorn. “Effective conjugation length and UV/vis spectra of oligomers”. In: *Acta Polym.* 48.9 (1997), pp. 379–384. DOI: 10.1002/actp.1997.010480905. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/actp.1997.010480905>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/actp.1997.010480905>.
- [56] Tsuyoshi Izumi et al. “Synthesis and Spectroscopic Properties of a Series of β -Blocked Long Oligothiophenes up to the 96-mer: Revaluation of Effective Conjugation Length”. In: *J. Am. Chem. Soc.* 125.18 (2003), pp. 5286–5287. DOI: 10.1021/ja034333i. eprint: <https://doi.org/10.1021/ja034333i>.
- [57] S. Kishino et al. “Estimate of the effective conjugation length of polythiophene from its $|\chi^{(3)}(\omega; \omega, \omega, -\omega)|$ spectrum at excitonic resonance”. In: *Phys. Rev. B* 58 (20 Nov. 1998), R13430–R13433. DOI: 10.1103/PhysRevB.58.R13430. URL: <https://link.aps.org/doi/10.1103/PhysRevB.58.R13430>.
- [58] Harry B. Gray and W.R. Ellis, Jr. “Chapter 6: Electron Transfer”. In: ed. by Ivano Bertini et al. University Science Books, Mill Valley, CA, 1994.
- [59] R.A. Marcus and Norman Sutin. “Electron transfers in chemistry and biology”. In: *Biochim. Biophys. Acta* 811.3 (1985), pp. 265–322. ISSN: 0304-4173. DOI: [https://doi.org/10.1016/0304-4173\(85\)90014-X](https://doi.org/10.1016/0304-4173(85)90014-X). URL: <http://www.sciencedirect.com/science/article/pii/030441738590014X>.
- [60] Denis Jerome. “Organic Conductors: From Charge Density Wave TTF-TCNQ to Superconducting (TMTSF)₂PF₆”. In: *Chem. Rev.* 104.11 (2004), pp. 5565–5592. DOI: 10.1021/cr030652g. eprint: <https://doi.org/10.1021/cr030652g>. URL: <https://doi.org/10.1021/cr030652g>.
- [61] László Forró and László Mihály. “Electronic properties of doped fullerenes”. In: *Rep. Prog. Phys.* 64.5 (May 2001), pp. 649–699. DOI: 10.1088/0034-4885/64/5/202. URL: <https://doi.org/10.1088/0034-4885/64/5/202>.
- [62] Yukihiro Takahashi et al. “What Happens at the Interface between TTF and TCNQ Crystals (TTF = Tetrathiafulvalene and TCNQ = 7,7,8,8-Tetracyanoquinodimethane)?”. In: *J. Phys. Chem. C* 116.1 (2012), pp. 700–703. DOI: 10.1021/jp2074368. eprint: <https://doi.org/10.1021/jp2074368>. URL: <https://doi.org/10.1021/jp2074368>.
- [63] Emanuel Peled. “The electrochemical behavior of alkali and alkaline earth metals in nonaqueous battery systems—the solid electrolyte interphase model”. In: *J. Electrochem. Soc.* 126.12 (1979), p. 2047.
- [64] Mengyun Nie et al. “Lithium ion battery graphite solid electrolyte interphase revealed by microscopy and spectroscopy”. In: *J. Phys. Chem. C* 117.3 (2013), pp. 1257–1267.

- [65] Martin Winter. “The solid electrolyte interphase—the most important and the least understood solid electrolyte in rechargeable Li batteries”. In: *Z. Phys. Chem.* 223.10-11 (2009), pp. 1395–1406.
- [66] Aiping Wang et al. “Review on modeling of the anode solid electrolyte interphase (SEI) for lithium-ion batteries”. In: *Npj Comput. Mater.* 4.1 (2018), pp. 1–26.
- [67] E Peled and S Menkin. “SEI: past, present and future”. In: *J. Electrochem. Soc.* 164.7 (2017), A1703–A1719.
- [68] Yixuan Wang et al. “Theoretical studies to understand surface chemistry on carbon anodes for lithium-ion batteries: reduction mechanisms of ethylene carbonate”. In: *J. Am. Chem. Soc.* 123.47 (2001), pp. 11708–11718.
- [69] Nongnuch Artrith, Tobias Morawietz, and Jörg Behler. “High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide”. In: *Phys. Rev. B* 83.15 (2011), p. 153101.
- [70] Hao Wang and Weitao Yang. “Force Field for Water Based on Neural Network”. In: *J. Phys. Chem. Lett.* 9.12 (2018), pp. 3232–3240.
- [71] Tobias Morawietz and Jörg Behler. “A full-dimensional neural network potential-energy surface for water clusters up to the hexamer”. In: *Z. Phys. Chem.* 227.9-11 (2013), pp. 1559–1581.
- [72] Tobias Morawietz and Jörg Behler. “A density-functional theory-based neural network potential for water clusters including van der waals corrections”. In: *J. Phys. Chem. A* 117.32 (2013), pp. 7356–7366.
- [73] Zhi Deng et al. “An electrostatic spectral neighbor analysis potential for lithium nitride”. In: *Npj Comput. Mater.* 5.1 (2019), pp. 1–8.
- [74] Lin Shen, Jingheng Wu, and Weitao Yang. “Multiscale quantum mechanics/molecular mechanics simulations with neural networks”. In: *J. Chem. Theory Comput.* 12.10 (2016), pp. 4934–4946.
- [75] S Houlding, SY Liem, and PLA Popelier. “A polarizable high-rank quantum topological electrostatic potential developed using neural networks: Molecular dynamics simulations on the hydrogen fluoride dimer”. In: *Int. J. Quantum Chem.* 107.14 (2007), pp. 2817–2827.
- [76] Paul P Ewald. “Die Berechnung optischer und elektrostatischer Gitterpotentiale”. In: *Ann. Phys. (Berl.)* 369.3 (1921), pp. 253–287.
- [77] Michael Gastegger, Jörg Behler, and Philipp Marquetand. “Machine learning molecular dynamics for the simulation of infrared spectra”. In: *Chem. Sci.* 8.10 (2017), pp. 6924–6935.

- [78] Patrick Bleiziffer, Kay Schaller, and Sereina Riniker. “Machine learning of partial charges derived from high-quality quantum-mechanical calculations”. In: *J. Chem. Inf. Model.* 58.3 (2018), pp. 579–590.
- [79] Yuanqing Wang et al. “Graph Nets for Partial Charge Prediction”. In: *arXiv preprint arXiv:1909.07903* (2019).
- [80] Tristan Bereau, Denis Andrienko, and O Anatole Von Lilienfeld. “Transferable atomic multipole machine learning models for small organic molecules”. In: *J. Chem. Theory Comput.* 11.7 (2015), pp. 3225–3233.
- [81] Tristan Bereau et al. “Non-covalent interactions across organic and biological subsets of chemical space: Physics-based potentials parametrized from machine learning”. In: *J. Chem. Phys.* 148.24 (2018), p. 241706.
- [82] Anand Chandrasekaran et al. “Solving the electronic structure problem with machine learning”. In: *Npj Comput. Mater.* 5.1 (2019), pp. 1–7.
- [83] Wilfried J. Mortier, Karin Van Genechten, and Johann Gasteiger. “Electronegativity equalization: application and parametrization”. In: *J. Am. Chem. Soc.* 107.4 (1985), pp. 829–835. DOI: 10.1021/ja00290a017. eprint: <https://doi.org/10.1021/ja00290a017>. URL: <https://doi.org/10.1021/ja00290a017>.
- [84] Anthony K. Rappe and William A. Goddard. “Charge equilibration for molecular dynamics simulations”. In: *J. Phys. Chem.* 95.8 (1991), pp. 3358–3363. DOI: 10.1021/j100161a070. eprint: <https://doi.org/10.1021/j100161a070>. URL: <https://doi.org/10.1021/j100161a070>.
- [85] Razvan A. Nistor et al. “A generalization of the charge equilibration method for nonmetallic materials”. In: *J. Chem. Phys.* 125.9 (2006), p. 094108. DOI: 10.1063/1.2346671. eprint: <https://doi.org/10.1063/1.2346671>. URL: <https://doi.org/10.1063/1.2346671>.
- [86] Jiahao Chen and Todd J. Martinez. “QTPIE: Charge transfer with polarization current equalization. A fluctuating charge model with correct asymptotics”. In: *Chem. Phys. Lett.* 438.4 (2007), pp. 315–320. ISSN: 0009-2614. DOI: <https://doi.org/10.1016/j.cplett.2007.02.065>. URL: <http://www.sciencedirect.com/science/article/pii/S0009261407002618>.
- [87] Min Zhang and Rene Fournier. “Self-Consistent Charge Equilibration Method and Its Application to $\text{Au}_{13}\text{Na}_n$ ($n = 1, 10$) Clusters”. In: *J. Phys. Chem. A* 113.13 (2009), pp. 3162–3170. DOI: 10.1021/jp8063273. eprint: <https://doi.org/10.1021/jp8063273>. URL: <https://doi.org/10.1021/jp8063273>.
- [88] Christopher E. Wilmer, Ki Chul Kim, and Randall Q. Snurr. “An Extended Charge Equilibration Method”. In: *J. Phys. Chem. Lett.* 3.17 (2012), pp. 2506–2511. DOI:

- 10.1021/jz3008485. eprint: <https://doi.org/10.1021/jz3008485>. URL: <https://doi.org/10.1021/jz3008485>.
- [89] T. Verstraelen et al. “ACKS2: Atom-condensed Kohn-Sham DFT approximated to second order”. In: *J. Chem. Phys.* 138.7 (2013), p. 074108. DOI: 10.1063/1.4791569. eprint: <https://doi.org/10.1063/1.4791569>.
- [90] Brad A. Wells, Caspar De Bruin-Dickason, and Alan L. Chaffee. “Charge Equilibration Based on Atomic Ionization in Metal-Organic Frameworks”. In: *J. Phys. Chem. C* 119.1 (2015), pp. 456–466. DOI: 10.1021/jp510415h. eprint: <https://doi.org/10.1021/jp510415h>. URL: <https://doi.org/10.1021/jp510415h>.
- [91] Saber Naserifar et al. “Polarizable charge equilibration model for predicting accurate electrostatic interactions in molecules and solids”. In: *J. Chem. Phys.* 146.12 (2017), p. 124117. DOI: 10.1063/1.4978891. eprint: <https://doi.org/10.1063/1.4978891>. URL: <https://doi.org/10.1063/1.4978891>.
- [92] S Alireza Ghasemi et al. “Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network”. In: *Phys. Rev. B* 92.4 (2015), p. 045131.
- [93] Sebastian Dick and Marivi Fernandez-Serra. “Machine learning accurate exchange and correlation functionals of the electronic density”. In: *Nature communications* 11.1 (2020), pp. 1–10.
- [94] Jonathan Schmidt, Carlos L Benavides-Riveros, and Miguel AL Marques. “Machine Learning the Physical Nonlocal Exchange–Correlation Functional of Density-Functional Theory”. In: *J. phys. chem. lett.* 10.20 (2019), pp. 6425–6431.
- [95] Ryo Nagai, Ryosuke Akashi, and Osamu Sugino. “Completing density functional theory by machine-learning hidden messages from molecules”. In: *arXiv preprint arXiv:1903.00238* (2019).
- [96] Brian Kolb, Levi C Lentz, and Alexie M Kolpak. “Discovering charge density functionals and structure-property relationships with PROPhet: A general framework for coupling machine learning and first-principles methods”. In: *Sci. Rep.* 7.1 (2017), pp. 1–9.
- [97] John C Snyder et al. “Orbital-free bond breaking via machine learning”. In: *J. chem. phys.* 139.22 (2013), p. 224104.
- [98] John C Snyder et al. “Finding density functionals with machine learning”. In: *Phys. rev. lett.* 108.25 (2012), p. 253002.
- [99] Li Li et al. “Understanding machine-learned density functionals”. In: *Int. J. Quantum Chem.* 116.11 (2016), pp. 819–833.

-
- [100] Kun Yao and John Parkhill. “Kinetic energy of hydrocarbons as a function of electron density and convolutional neural networks”. In: *J. Chem. Theory. Comput.* 12.3 (2016), pp. 1139–1147.
- [101] Felix Brockherde et al. “Bypassing the Kohn-Sham equations with machine learning”. In: *Nat. Commun.* 8.1 (2017), pp. 1–10.
- [102] Mel Levy. “Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem”. In: *Proc. Natl. Acad. Sci. U. S. A.* 76.12 (1979), pp. 6062–6065. ISSN: 0027-8424. DOI: 10.1073/pnas.76.12.6062. eprint: <https://www.pnas.org/content/76/12/6062.full.pdf>. URL: <https://www.pnas.org/content/76/12/6062>.
- [103] A.D. Becke. “A multicenter numerical integration scheme for polyatomic molecules”. In: *J. Chem. Phys.* 88.4 (1988), pp. 2547–2553. DOI: 10.1063/1.454033. eprint: <https://doi.org/10.1063/1.454033>. URL: <https://doi.org/10.1063/1.454033>.
- [104] P. H. Dederichs et al. “Ground States of Constrained Systems: Application to Cerium Impurities”. In: *Phys. Rev. Lett.* 53 (26 Dec. 1984), pp. 2512–2515. DOI: 10.1103/PhysRevLett.53.2512.
- [105] Qin Wu and Troy Van Voorhis. “Direct optimization method to study constrained systems within density-functional theory”. In: *Phys. Rev. A* 72 (2 Aug. 2005), p. 024502. DOI: 10.1103/PhysRevA.72.024502. URL: <https://link.aps.org/doi/10.1103/PhysRevA.72.024502>.
- [106] Benjamin Kaduk, Tim Kowalczyk, and Troy Van Voorhis. “Constrained Density Functional Theory”. In: *Chem. Rev.* 112.1 (2012), pp. 321–370. DOI: 10.1021/cr200148b. eprint: <https://doi.org/10.1021/cr200148b>. URL: <https://doi.org/10.1021/cr200148b>.
- [107] Yihan Shao et al. “Advances in molecular quantum chemistry contained in the Q-Chem 4 program package”. In: *Mol. Phys.* 113 (2015), pp. 184–215. DOI: 10.1080/00268976.2014.952696. eprint: <http://dx.doi.org/10.1080/00268976.2014.952696>. URL: <http://dx.doi.org/10.1080/00268976.2014.952696>.
- [108] Qin Wu and Troy Van Voorhis. “Direct Calculation of Electron Transfer Parameters through Constrained Density Functional Theory”. In: *J. Phys. Chem. A* 110.29 (2006), pp. 9212–9218. DOI: 10.1021/jp061848y. eprint: <https://doi.org/10.1021/jp061848y>. URL: <https://doi.org/10.1021/jp061848y>.
- [109] K Kaufmann and W Baumeister. “Single-centre expansion of Gaussian basis functions and the angular decomposition of their overlap integrals”. In: *J. Phys. B* 22.1 (1989), p. 1.

-
- [110] Nongnuch Artrith, Alexander Urban, and Gerbrand Ceder. “Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species”. In: *Phys. Rev. B* 96.1 (2017), p. 014112.
- [111] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. “Deep sparse rectifier neural networks”. In: *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. 2011, pp. 315–323.
- [112] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [113] Martin Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: <http://tensorflow.org/>.
- [114] B.K. Rao and P. Jena. “Molecular cluster calculations of the electronic structure of lithium hydride”. In: *J. Phys. C: Solid State Phys.* 19.26 (Sept. 1986), pp. 5167–5172. DOI: 10.1088/0022-3719/19/26/016. URL: <https://doi.org/10.1088/0022-3719/19/26/016>.
- [115] Anne-Marie Sapse, Duli C. Jain, and Krishnan Raghavachari. “Theoretical Studies of Aggregates of Lithium Compounds”. In: *Lithium Chemistry: A Theoretical and Experimental Overview*. Ed. by Anne-Marie Sapse and Paul von R. Schleyer. John Wiley & Sons, 1995. Chap. 2, p. 45.
- [116] Yung-Lung Chen, Chun-Huei Huang, and Wei-Ping Hu. “Theoretical Study on the Small Clusters of LiH, NaH, BeH₂, and MgH₂”. In: *J. Phys. Chem. A* 109.42 (2005), pp. 9627–9636. DOI: 10.1021/jp051978r. eprint: <https://doi.org/10.1021/jp051978r>. URL: <https://doi.org/10.1021/jp051978r>.
- [117] S. J. Nolan et al. “Calculation of properties of crystalline lithium hydride using correlated wave function theory”. In: *Phys. Rev. B* 80 (16 Oct. 2009), p. 165109. DOI: 10.1103/PhysRevB.80.165109.
- [118] Takeshi Sato and Kenichi L. Ishikawa. “Time-dependent multiconfiguration self-consistent-field method based on the occupation-restricted multiple-active-space model for multielectron dynamics in intense laser fields”. In: *Phys. Rev. A* 91 (2 Feb. 2015), p. 023417. DOI: 10.1103/PhysRevA.91.023417. URL: <https://link.aps.org/doi/10.1103/PhysRevA.91.023417>.
- [119] Zhi-Quan Huang et al. “First-principles calculated decomposition pathways for LiBH₄ nanoclusters”. In: *Sci. Rep.* 6 (2016), p. 26056.
- [120] R. Antoine et al. “Dissociation pathways and binding energies of (LiH)_nLi⁺ and (LiH)_nLi⁺+3 clusters”. In: *J. Chem. Phys.* 104.1 (1996), pp. 110–119. DOI: 10.1063/1.470880. eprint: <https://doi.org/10.1063/1.470880>. URL: <https://doi.org/10.1063/1.470880>.

- [121] Xuefeng Wang and Lester Andrews. “Infrared Spectra and Theoretical Calculations of Lithium Hydride Clusters in Solid Hydrogen, Neon, and Argon”. In: *J. Phys. Chem. A* 111.27 (2007), pp. 6008–6019. DOI: 10.1021/jp071251y. eprint: <https://doi.org/10.1021/jp071251y>. URL: <https://doi.org/10.1021/jp071251y>.
- [122] Lucas K. Wagner et al. “Tuning metal hydride thermodynamics via size and composition: Li-H, Mg-H, Al-H, and Mg-Al-H nanoclusters for hydrogen storage”. In: *Phys. Chem. Chem. Phys.* 14 (18 2012), pp. 6611–6616. DOI: 10.1039/C2CP24063G. URL: <http://dx.doi.org/10.1039/C2CP24063G>.
- [123] Sjoerd Harder. “Molecular early main group metal hydrides: synthetic challenge, structures and applications”. In: *Chem. Commun.* 48 (91 2012), pp. 11165–11177. DOI: 10.1039/C2CC33478J. URL: <http://dx.doi.org/10.1039/C2CC33478J>.
- [124] Andreas Stasch. “Well-Defined, Nanometer-Sized LiH Cluster Compounds Stabilized by Pyrazolate Ligands”. In: *Angew. Chem. Int. Ed.* 53.5 (2014), pp. 1338–1341. DOI: 10.1002/anie.201308935. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201308935>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201308935>.
- [125] Liwu Huang et al. “Synthesis of destabilized nanostructured lithium hydride via hydrogenation of lithium electrochemically inserted into graphite”. In: *Int. J. Hydrogen Energ.* 40.40 (2015), pp. 13936–13941. ISSN: 0360-3199. DOI: <https://doi.org/10.1016/j.ijhydene.2015.07.130>. URL: <http://www.sciencedirect.com/science/article/pii/S0360319915019333>.
- [126] Lei Wang, Md Zakaria Quadir, and Kondo-Francois Aguey-Zinsou. “Direct and reversible hydrogen storage of lithium hydride (LiH) nanoconfined in high surface area graphite”. In: *Int. J. Hydrogen Energ.* 41.40 (2016), pp. 18088–18094. ISSN: 0360-3199. DOI: <https://doi.org/10.1016/j.ijhydene.2016.07.073>. URL: <http://www.sciencedirect.com/science/article/pii/S0360319916320602>.
- [127] Dominik Marx and Jurg Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009. DOI: 10.1017/CB09780511609633.
- [128] Axel D. Becke. “Density-functional thermochemistry. III. The role of exact exchange”. In: *J. Chem. Phys.* 98.7 (1993), pp. 5648–5652. DOI: <http://dx.doi.org/10.1063/1.464913>. URL: <http://scitation.aip.org/content/aip/journal/jcp/98/7/10.1063/1.464913>.
- [129] P. J. Stephens et al. “Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields”. In: *J. Phys. Chem.* 98.45 (1994), pp. 11623–11627. DOI: 10.1021/j100096a001. eprint: <http://pubs.acs.org/doi/pdf/10.1021/j100096a001>.

- [130] Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. “Effect of the damping function in dispersion corrected density functional theory”. In: *J. Comput. Chem.* 32.7 (2011), pp. 1456–1465.
- [131] Jeng-Da Chai and Martin Head-Gordon. “Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections”. In: *Phys. Chem. Chem. Phys.* 10.44 (2008), pp. 6615–6620.
- [132] Dmitrij Rappoport and Filipp Furche. “Property-optimized Gaussian basis sets for molecular response calculations”. In: *J. Chem. Phys.* 133.13 (2010), p. 134105. DOI: 10.1063/1.3484283. eprint: <https://doi.org/10.1063/1.3484283>. URL: <https://doi.org/10.1063/1.3484283>.
- [133] Schrödinger Release 2019-4: Maestro, Schrödinger, LLC, New York, NY, 2019.
- [134] Wolfram Koch and Max C Holthausen. *A chemist’s guide to density functional theory*. John Wiley & Sons, 2015.
- [135] *QUIP*. (accessed 05/2020). URL: <https://github.com/libAtoms/QUIP>.
- [136] *ase*. (accessed 05/2020). URL: <https://wiki.fysik.dtu.dk/ase/index.html>.

Chapter 3

Extension of BpopNN to Li-organic systems

I Introduction

In recent years, there has been a rise in interest in extending ML potentials beyond closed-shell, well-behaved organic molecules or homogeneous materials. One important realization is the limitations of local descriptors based on atomic positions: they are not directly suitable for describing long-range electronic interactions because only local geometrical information is encoded.

Resolutions to this problem more or less rely on deriving the appropriate partial charges or spins for a given total charge or spin state for a molecule, thereafter long-range electrostatics can be computed through Coulomb’s law and added on to the ML-predicted short-range energies. Prominent examples include the charge equilibration via neural network technique (CENT) method[1–3], where the optimal charge distribution is determined by minimizing a quadratic energy expression of atomic charges, parametrized by the atomic electronegativities and chemical hardness. The intuition is, compared with atomic charges, atomic electronegativities are “local” properties that can be predicted from local descriptors. SpookyNet[4] and a variation of AIMNet[5] directly predict the atomic partial charges and partial spins for a given total charge/spin multiplicity through an attention mechanism[6] or a message passing procedure. All the above methods can be expected to reliably reach the desired ground state (GS) when the electronic character of the GS within each total charge/spin multiplicity is similar across the chemical compound space of the training set and when the GS is well separated in energy from other SCF solutions or excited states (ES). Reasonable accuracy has been achieved for charged organic molecules,[2, 5] ionic clusters (NaCl clusters), [1, 2]metal clusters (Ag_3),[2, 4] singlet and triplet carbenes[4] and one example of periodic system (Au_2 cluster on the $\text{MgO}(001)$ surface).[2, 4]

However, complications arise when one wishes to apply such models to molecular systems with multiple low-lying SCF solutions, e.g., metal-containing organic complexes. For this kind of system, two or more low-lying SCF solutions are bundled in narrow spectral ranges, mixing their electronic characters and exchanging their order in response to small nuclear displacements. In these scenarios, standard SCF algorithms in DFT calculations, e.g., DIIS,[7, 8] GDM,[9] do not guarantee to converge to the GS (it also depends on the initial guesses); instead, several SCF solutions need to be manually obtained to determine which one is the true GS. Suppose one were to train an ML potential on this type of system, i.e. rapid change in the electronic structure for the GS when the geometry does not change appreciably. In that case, it is unclear how the ML model can bypass all the complexities that DFT faces (especially near the state-crossing region), and “directly” map the nuclear positions, total charge, and spin to the true GS.

Such systems are ubiquitous, especially seen in the electrochemical reaction cascades in battery environments, responsible for the ion-pair formation, the salt decomposition in the electrolyte solutions, and the formation of the solid-electrolyte interphase (SEI) on the

anode surface. For instance, one would be interested to know whether metal ions or organic electrolytes/anions preferentially uptake an electron under reductive conditions. In the case of LIB, the former is relevant to the formation of Li dendrite and the latter is responsible for the formation of a variety of organic SEI components, e.g. LEDC,[10–14] LEMC,[15] etc.; in Mg-ion batteries (MIB), it is shown that Mg-salt is prone to the ion-pair formation and Mg^{2+} , but not TFSI^- , is more likely to reduce first in the $\text{Mg}^{2+}\text{TFSI}^-$ complex to facilitate subsequent salt decomposition reactions. [16]

Both the CENT method and methods based on graph neural networks (GNN) can only obtain one “SCF” solution: the former approach relies on minimizing a convex target energy function of atomic charges which only has one minimum solution, and the latter method only outputs one set of atomic charges/spins by construction. The BpopNN method (chapter 2), on the other hand, encodes the atomic populations into the descriptor, and a full self-consistent procedure (through the variational principle, i.e., minimizing the total energy) within the ML framework is used to derive the total energy and the atomic populations. In this approach, unlike the CENT method, not only the long-range electrostatics, but the short-range atomic energies are also dependent on the atomic populations, i.e., the whole procedure is self-consistent. In GNN-based approaches, since different independent spin/charge embeddings are employed for different total charges/spins, little physical information is exchanged between molecules with different total charges/spins. Theoretically, there is no problem with constructing a separate ML PES for each electronic state, given that the electronic character of each state is known beforehand; this is the case for many existing ML models for ES[17–25]. Still, these ML models for ES are usually restricted to certain molecules or materials; it is challenging to transfer the knowledge about different states across chemical compound space. Ideally, geometrical and electronic information should be unified seamlessly and one ML model capable of describing multiple electronic states across chemical compound space should be trained.

BpopNN can model multiple SCF solutions by construction; thus it is capable of modeling excited states where the electronic character or the distribution of the electron density undergoes significant change upon excitation, e.g., charge transfer (CT) excitations. CT excited states are relevant in many photophysical phenomena observed in organic photovoltaics, such as aggregation-induced emission (AIE) [26] and thermally activated delayed fluorescence (TADF). [27, 28] Although time-dependent DFT (TDDFT) is a popular choice for modeling ES energies and properties, it has severe shortcomings in modeling ES whose density is significantly different from that of the GS:[29] this includes long-range CT, core-valence excitations[30] and double excitations. [31] Orbital optimized DFT, i.e. ΔSCF , [29, 32] is a complementary method for TDDFT that yields more accurate excitation energies for CT excitations. BpopNN can be naturally applied for ΔSCF calculations, representing a novel approach for modeling ES with ML potentials.

In this chapter, we build on the BpopNN method developed in chapter 2 and extend it to include spin populations. We test the model on the challenging Li-organic systems

and show that the model can predict adiabatic IPs and EAs in reasonable agreement with DFT. Adiabatic IPs and EAs represent useful design metrics for developing novel, redox-stable electrolytes. We further demonstrate that BpopNN can model CT excited states by obtaining multiple equilibrium charge/spin solutions within the model.

II Methods

We adopted the same method as described in the BpopNN work (chapter 2), [33] except for some slight modifications. In a nutshell, inspired by the Levy constrained search approach in DFT, we use ML to model a well-defined function, the lowest energy achievable with the atomic populations constrained to specific values. In practice, the training data can be obtained from constrained DFT (CDFT), [34–36] which allows us to compute the energy for any given set of population values. Here we used the Becke population scheme.[37] Once the model is trained, a self-consistent procedure can be used (as in DFT) to derive the optimal charge and spin populations for each atom in the molecule.

Our model energy function can be written as

$$E_{\text{ML}} = \sum_{\mu} \sum_{i \in A_{\mu}} \epsilon_{\mu}[D_i[\mathbf{Z}, \mathbf{p}, \mathbf{r}]] + E_{\text{intra}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}] + E_{\text{el}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}], \quad (3.1)$$

where ϵ_{μ} represents the output from the atomic neural network for the element μ , and this function is applied to the descriptor for atom i ; A_{μ} is the set of atom indices corresponding to element μ ; $\mathbf{Z}$, $\mathbf{r}$, and $\mathbf{p}$ contain the atomic numbers, coordinates, and atomic populations, respectively, for the entire molecule.

Here we used the following baseline terms for atomic energies. These are quadratic functions of atomic charge and spin that best fit the DFT energies for the pertinent charge and spin states of the isolated atom. And $E_{\text{intra}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}]$ is then the sum of these energies for each atom. For C, H, O, and Li atoms found in the molecular applications below, we used the following functions

$$E_{\text{intra,C},i} = 0.1982q_i^2 + 0.1982s_i^2 + 0.2243q_i - 37.7984 \quad (3.2)$$

$$E_{\text{intra,H},i} = 0.2586q_i^2 + 0.2586s_i^2 + 0.2433q_i - 0.5018 \quad (3.3)$$

$$E_{\text{intra,O},i} = 0.2271q_i^2 + 0.2271s_i^2 + 0.2841q_i - 74.9828 \quad (3.4)$$

$$E_{\text{intra,Li},i} = 0.0886q_i^2 + 0.0886s_i^2 + 0.1053q_i - 7.4832, \quad (3.5)$$

where q_i is the atomic partial charge for atom i , i.e. $q_i = Z_i - p_{i,\alpha} - p_{i,\beta}$ and s_i is the atomic partial spin for atom i , i.e. $s_i = p_{i,\alpha} - p_{i,\beta}$. This resembles the intra-atomic term in charge equilibration models. [1]

The $E_{\text{el}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}]$ term represents a simple pairwise Coulomb term for incorporating long-range electrostatics, as shown in the following equation.

$$E_{\text{el}}[\mathbf{Z}, \mathbf{p}, \mathbf{r}] = \sum_{j>i} \tanh(\kappa r_{ij}) \frac{q_i q_j}{r_{ij}} \quad (3.6)$$

Parameters κ (one each for different element pair types) were trained together with other parameters in the neural network.

The atomistic descriptor employed is similar to the one used in the previous BpopNN model, [33], but spin populations are also incorporated as input variables besides the total populations. The descriptor is a variant of SOAP, [38] where the proxy Gaussian densities on each atom are modified to describe the ESPs from the nuclei, total populations, and spin populations. Specifically, the proxy density takes the following form

$$\rho_{\mu,t}(\mathbf{r}) = \sum_{i \in A_\mu} \chi_i^t B \exp(-\gamma |\mathbf{r} - \mathbf{r}_i|^2), \quad (3.7)$$

where t is either ν (for nuclear), p , or s corresponding to the total population or spin population, and thus $\chi_i^\nu = Z_i$, $\chi_i^p = p_i$, and $\chi_i^s = s_i$. Here we used $B = 1.128$ and $\gamma = 0.171$, and r is in the unit Bohr radius, and these values were determined by fitting a single Gaussian function to the $1/r$ term in the ESPs. Technically we could choose to use the α populations and β populations instead of the total populations and spin populations; however, the latter performed better in practice.

Projecting this set of proxy densities onto spherical harmonics, we have

$$\rho_{\mu,t}(\mathbf{r}) = \sum_{i \in A_\mu} \sum_{lm} c_{ilm}^t(r) \mathbf{Y}_{lm}(\hat{\mathbf{r}}) \quad (3.8)$$

which has an analytical solution following previous works:[38, 39]

$$c_{ilm}^t(r) \equiv 4\pi \chi_i^t B \exp[-\gamma(r^2 + r_i^2)] \iota_l(2\gamma r r_i) \mathbf{Y}_{lm}^*(\hat{\mathbf{r}}_i), \quad (3.9)$$

where $*$ denotes complex conjugation and ι_l are the modified spherical Bessel functions of the first kind. Summing over pertinent atoms and self-contracting over m of the resulting sum gives the atomistic descriptor $p_{nl}^{\mu,t}$, which ensures the rotational invariance.

$$c_{lm}^{\mu,t}(r) = \sum_{i \in A_\mu} c_{ilm}^t(r), \quad (3.10)$$

$$p_{nl}^{\mu,t} = \sum_{m=-l}^l \|c_{nlm}^{\mu,t}\|^2 \quad (3.11)$$

We note that instead of projecting $c_{nlm}^{\mu,t}$ onto some radial basis functions, we choose to evaluate it on a simple linear radial grid between 0.5 and 6.0 Å.

The power spectrum is then split into different element channels, and a different neural network is employed to infer the atomic energies for each element type. A radial cutoff of 6.0 Å was used below in computing the descriptor, i.e., atom pairs beyond this cutoff were neglected. We employ separate neural networks for different element types, and for the applications below, each element network contains two hidden layers with 100 neurons in each layer. Nonlinearities were introduced using a modified softplus activation function $\ln(1 + \exp(100x))/100$.

The loss function for the training is

$$\begin{aligned} \mathcal{L} = & \sum_M \frac{(E_M^{\text{CDFT}} - E_M^{\text{ML}})^2}{N_{\text{atom},M}} + \gamma_1 \sum_M \frac{\|\mathbf{F}_M^{\text{CDFT}} - \mathbf{F}_M^{\text{ML}}\|^2}{N_{\text{atom},M}} + \gamma_2 \sum_M \frac{\|\mathbf{f1}_M^{\text{CDFT}} - \mathbf{f1}_M^{\text{ML}}\|^2}{N_{\text{atom},M}} \\ & + \gamma_3 \sum_M \frac{\|\mathbf{f2}_M^{\text{CDFT}} - \mathbf{f2}_M^{\text{ML}}\|^2}{N_{\text{atom},M}}. \end{aligned} \quad (3.12)$$

For the model presented below, both γ_1 and γ_2 were set to 5.0. The above sums go over the training molecules, and $\mathbf{F}$ is the position gradient, $\mathbf{f1}$ is the total population gradient, $\mathbf{f2}$ is the spin population gradient, and $N_{\text{atom},M}$ represents the number of atoms in molecule M . Adaptive moment solver (Adam)[40] was used to update the NN weights during training.

III Results and Discussion

3.1 Training-data generation and model training

We apply the BpopNN model to Li organic systems, consisting of element types C, H, O, and Li. The interaction between lithium and organic molecules is of interest, especially in the context of LIB, where organic electrolytes are typically employed. Figure 3.1 shows a representative set of training data.

The organic molecules used in training are taken from the ANI-1 dataset,[41] containing small organic molecules of up to 8 heavy atoms. Molecules with elements C, H, and O are selected, and either one or two Li are “placed” around the molecule, deleting H atoms as needed. Specifically, if the organic molecule consists of any oxygen atoms, the Li atoms are more likely to stay around those atoms as they have higher electronegativities; thus a random O is chosen, around which Li is placed. If the oxygen belongs to an ethereal or carbonyl oxygen type, Li is placed in a straightforward way; however, if the oxygen is part of a hydroxy group, a hydrogen atom that is bound to the oxygen is replaced with Li. If no oxygen atom is present in the molecule, a random carbon is selected, and one H atom that is attached to the C atom is replaced with a Li. Although not strictly thorough, this procedure

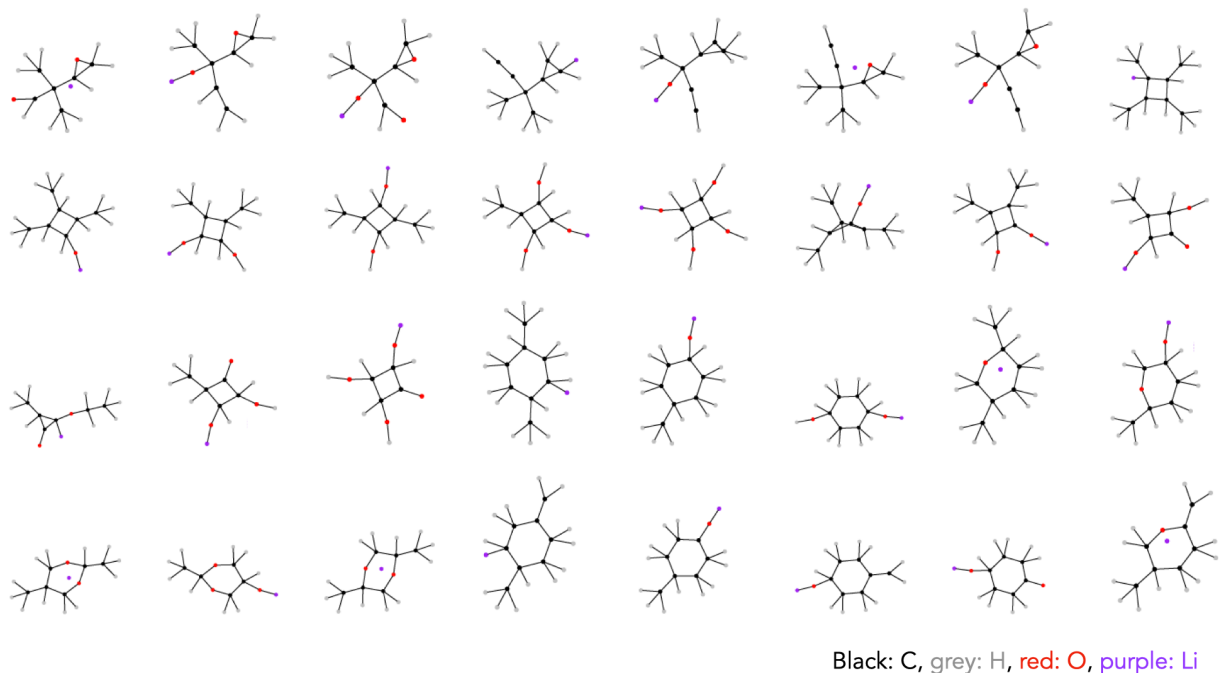


Figure 3.1: A representative set of training molecules.

covers various interactions between organic molecules and Li, as illustrated in Figure 3.3(a). Note that in the case of Li interacting with a carbonyl oxygen, if the total charge of the molecule is 0, two types of scenarios typically exist: (1) Li is Li^+ in character, and the organic part is bearing a minus charge, (2) Li is $\text{Li}^\cdot$ in character and the organic molecule is neutral. We obtained 13224 unique molecules in total through this procedure, which were then geometry optimized at total charges +1, 0, -1 at the $\omega\text{B97X-D}$ [42]/def2-SVPD[43] level of theory using the QChem[44] electronic structure code. After geometry optimization, we selected 3-4 geometries for each molecule along the optimization trajectory, and for each of those, we performed 3-4 CDFT calculations by adding a random (uniform distribution from -0.05 to 0.05) charge and spin fluctuation to each atom’s Becke charge/spin value from the pertinent unconstrained SCF solution. The CDFT energies and energy derivatives with respect to atomic positions, atomic charges, and atomic spins were used in the model training.

The training data (~ 1 million points) were randomly separated into a training (95%) and a validation set (5%) at the outset, and the validation error provides an estimate on how well the model is trained and can serve as an early-stopping criterion. The model reaches its best performance on test sets after ~ 250 epochs of training. The mean absolute error (MAE) on the independent validation set of the energy is 3.59 kcal/mol per molecule.

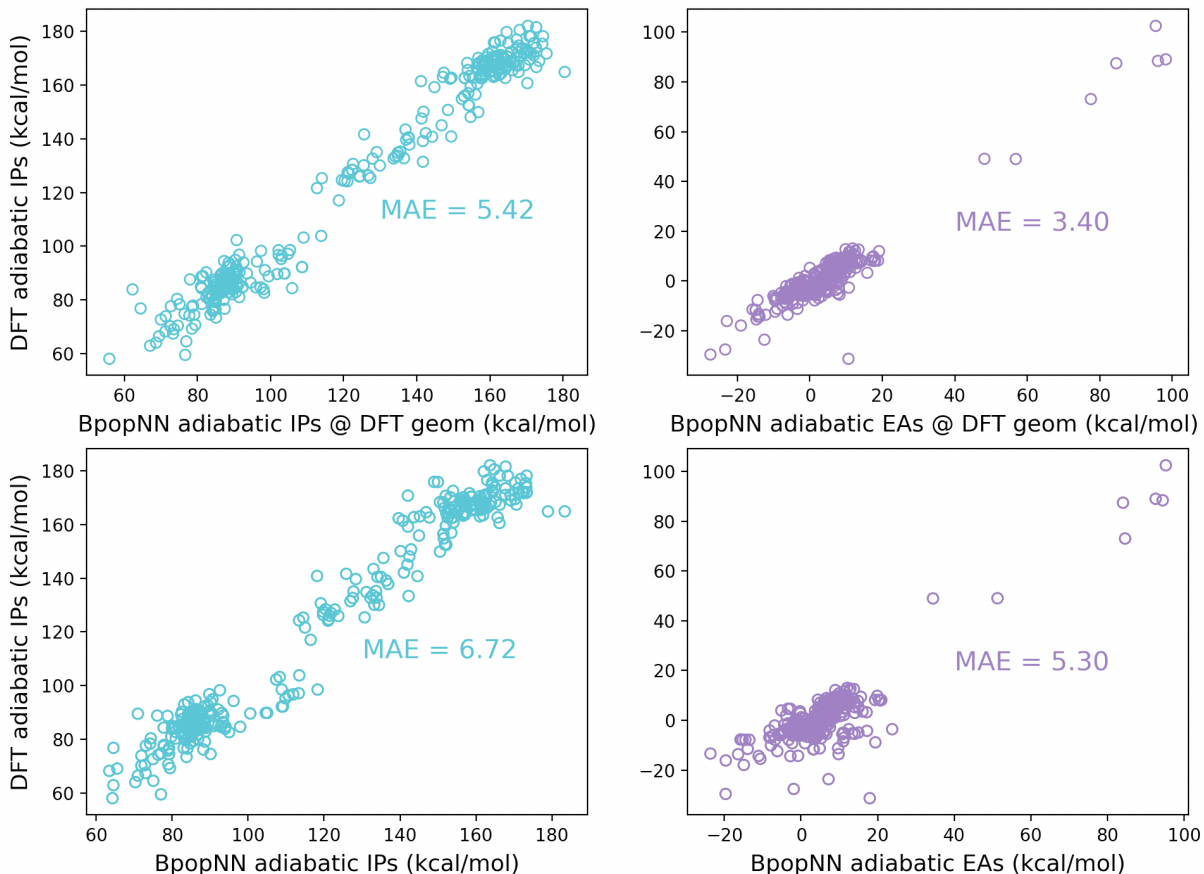


Figure 3.2: BpopNN vs. DFT adiabatic IPs and EAs

3.2 Testing the model

3.2.1 Adiabatic IP and EA on larger Li-organic molecules

To test the general applicability of the model, we generated an independent test set from 317 organic molecules with 9-12 heavy atoms (number of C atoms and O atoms, excluding Li) that were randomly taken from the ChempSpider database.[45] We show the performance of our model in predicting the adiabatic ionization potential (IP) and electron affinity (EA) values.

Similar to the SCF-q procedure defined in chapter 2, total populations and spin populations can be optimized together self-consistently; we will call this the SCF-q_s procedure in the remaining text. On the top panel of Figure 3.2, we plot the comparison between

Table 3.1: Energy, charge, and spin errors for the larger Li-organic test set

	MAE energy (kcal/mol/molecule)	MAE charge (e/atom)	MAE spin (e/atom)
Cation (scf-qs)	6.398	0.006	0.006
Neutral (scf-qs)	5.341	0.006	0.004
Anion (scf-qs)	4.336	0.009	0.003
Cation (@BpopNN geometry)	6.835	0.009	0.004
Neutral (@BpopNN geometry)	4.806	0.007	0.003
Anion (@BpopNN geometry)	6.608	0.010	0.002
	MAE energy (eV) @ DFT geometry	MAE energy (eV) @ BpopNN geometry	
IP	0.235	0.291	
EA	0.147	0.230	

the adiabatic IPs and EAs obtained from BpopNN (from the SCF-qs procedure converged to threshold 10^{-3} Hartree/e) and DFT at the optimal DFT geometries at different total charges. The bottom two plots in Figure 3.2 show the parity plots between the adiabatic IPs and EAs obtained from BpopNN at its optimized geometries and those obtained from DFT at the DFT optimized geometries. The statistics for energy, charge, and spin errors are shown in Table 3.1. The BpopNN charges and spins show close agreement with the DFT ones, with a charge and spin MAE of less than 0.01 e/atom. Energy errors are ~ 5 kcal/mol per molecule, which is not very accurate but reasonable considering how challenging this type of system is.

3.2.2 Charge transfer excited states

In this section, we show that it is possible to compute certain types of excited states with BpopNN, through a particular approach for computing them in DFT, i.e., Δ SCF.[29, 32] In Δ SCF, we make the orbital substitution(s) consistent with the excited state of interest, and seek a solution to the Kohn-Sham equations corresponding to this electronic configuration, i.e., orbitals are allowed to relax while maintaining the basic substituted configuration. It has been shown that every stationary point of the exact ground state energy functional $E_v(\rho)$ corresponds to an exact energy eigenstate, although the converse is not necessarily true. [46] Similar to how $E_v[\rho_i(\mathbf{r})]$ corresponds to the energy of the lowest energy state with density $\rho_i(\mathbf{r})$, CDFT yields the lowest-energy state under the constraint(s) that the electron population(s) on atoms is equal to the desired value(s). The target function map for BpopNN thus follows the definition of CDFT, where a set of electron populations on atoms are mapped onto the lowest possible energy with those populations. BpopNN is effective for

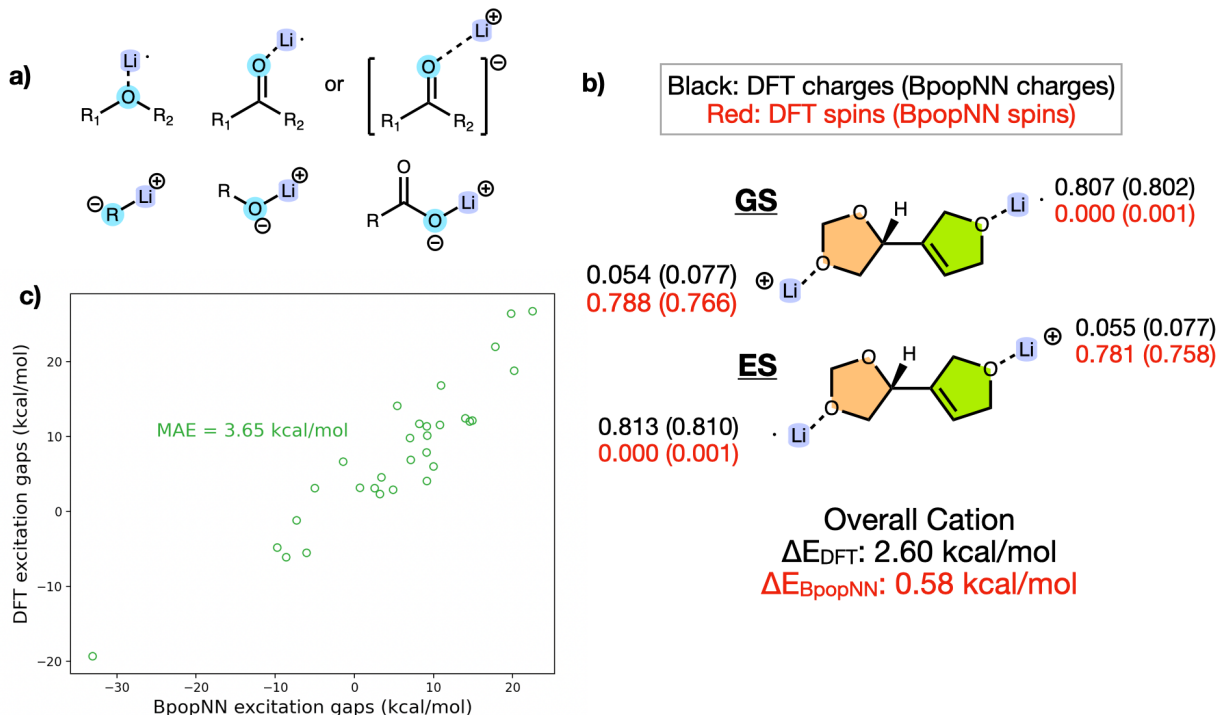


Figure 3.3: (a) Types of Li interactions in the training set. (b) An example of a test molecule created through fusing two training molecules together. (c) Parity plot of DFT vs. BpopNN excitation gaps for a test set of anions.

modeling an excited state when no state with the same set of electron populations is lower in energy. In other words, BpopNN is more suitable for modeling the type of excited states where the atomic electron populations are noticeably different from that of the ground state, e.g., charge-transfer excited states. There could be a problem with modeling, for example, $\pi \rightarrow \pi^*$ excitations when the molecular orbitals involved in the excitation differ mostly in phases and do not result in a noticeable difference in the atomic electron populations.

Since we have all the training molecules computed at total charge +1, 0, and -1, we can imagine making a new molecule by fusing two training molecules together (making a covalent bond) and studying the charge transfer across them. One example molecule is given in Figure 3.3 (b), which has an overall charge +1. One could roughly consider this molecule as a fused molecule between the orange molecule (left) at total charge +1 and the green molecule (right) at total charge 0, or it could also be a fusion between the orange molecule at total charge 0 and the green molecule at total charge +1. With DFT, we can indeed obtain two SCF solutions: at the ground state, the left Li is roughly neutral in character, and the right Li is cation in character; at the excited state, it is the opposite, as indicated by the

Becke charges and spins. Not surprisingly, these two solutions are also found as stationary points in BpopNN (converged to threshold 10^{-7} Hartree/e), with the equilibrium charges and spins showing close agreement with DFT. The excitation gap predicted by BpopNN is 0.48 kcal/mol, compared with 2.60 kcal/mol in DFT. This demonstrates that BpopNN is capable of obtaining multiple SCF solutions in practice.

Figure 3.3 (c) shows the DFT vs. BpopNN predicted excitation gaps for a benchmark set of molecules with total charge -1, created through fusing two training molecules together as described above (geometries of the fused molecules were not optimized). Anionic molecules were selected because in most cases CT excited states can be easily obtained through a HOMO $\rightarrow$ LUMO excitation. The mean absolute error (MAE) for the excitation gap of this test set is 3.65 kcal/mol. We note that if the excitation gap is negative, it means in the initial DFT calculation it did not converge to the ground state, and this only becomes apparent when a state with lower energy is also obtained. Some other examples of CT excited states at different total charges are given in Figure 3.4. For the cation examples in Figure 3.4 (a) and the carbon nanohoop example in Figure 3.4 (b), we needed to select the appropriate MOs that correspond to CT from one side of the molecule to another where the change of atomic electron populations between states is obvious. The performance of BpopNN on estimating the excitation gaps appears to be better for anionic molecules than for neutral or cationic molecules. One possible explanation is that the optimal cation geometries are usually very different from the optimal neutral geometries, so our test data are outside of the training regime in terms of the geometries. Further, the electrostatics term needs to be improved to produce the correct charge distributions.

IV Conclusions

In summary, we have incorporated the spin populations, in addition to the total electron populations, into the BpopNN model, expanding the scope of the open-shell molecules it can handle. We trained and tested the model on a variety of Li-organic molecules, and the model performed reasonably well in predicting the adiabatic IPs and EAs; these values bear close relevance to the redox stabilities of electrolytes in LIBs.

We have also provided a proof-of-concept illustration of modeling CT excited states with BpopNN. It is possible to obtain such excited states, even though we did not intentionally include any excited state data in the training set. For future work, we will further improve the BpopNN model towards conducting dynamical simulations of the charge transfer processes in the formation of the SEI.

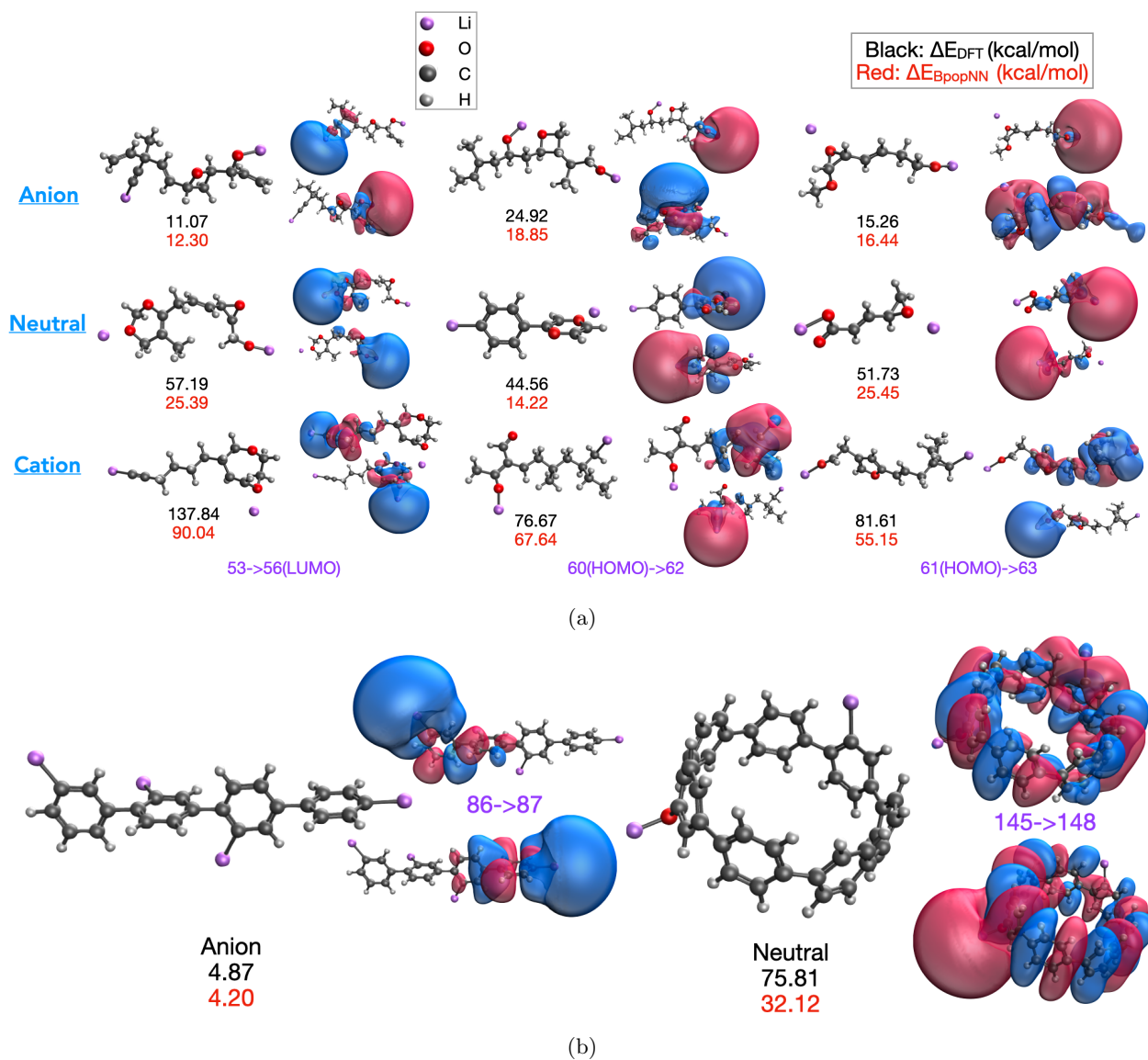


Figure 3.4: More ES examples at different total charges. The orbital indices used for excitation are given in the purple text for cation examples in (a) and the carbon nanohoop example in (b). If the orbital indices are not labeled, it corresponds to HOMO to LUMO excitation. Charges or spins are transferred from the top orbital to the bottom orbital.

Bibliography

- [1] S Alireza Ghasemi et al. “Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network”. In: *Phys. Rev. B* 92.4 (2015), p. 045131.
- [2] Tsz Wai Ko et al. “A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer”. In: *Nature communications* 12.1 (2021), pp. 1–11.
- [3] Tsz Wai Ko et al. “General-Purpose Machine Learning Potentials Capturing Nonlocal Charge Transfer”. In: *Accounts of Chemical Research* 54.4 (2021), pp. 808–817.
- [4] Oliver T Unke et al. “SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects”. In: *Nature communications* 12.1 (2021), pp. 1–14.
- [5] Roman Zubatyuk et al. “Teaching a neural network to attach and detach electrons from molecules”. In: *Nature communications* 12.1 (2021), pp. 1–11.
- [6] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).
- [7] Péter Pulay. “Convergence acceleration of iterative sequences. The case of SCF iteration”. In: *Chemical Physics Letters* 73.2 (1980), pp. 393–398.
- [8] Peter Pulay. “Improved SCF convergence acceleration”. In: *Journal of Computational Chemistry* 3.4 (1982), pp. 556–560.
- [9] Troy Van Voorhis and Martin Head-Gordon. “A geometric approach to direct minimization”. In: *Molecular Physics* 100.11 (2002), pp. 1713–1721.
- [10] Doron Aurbach and Yosef Gofer. “The behavior of lithium electrodes in mixtures of alkyl carbonates and ethers”. In: *Journal of the Electrochemical Society* 138.12 (1991), p. 3529.
- [11] Doron Aurbach et al. “The behaviour of lithium electrodes in propylene and ethylene carbonate: The major factors that influence Li cycling efficiency”. In: *Journal of Electroanalytical Chemistry* 339.1-2 (1992), pp. 451–471.

- [12] D Aurbach et al. “A comparative study of synthetic graphite and Li electrodes in electrolyte solutions based on ethylene carbonate-dimethyl carbonate mixtures”. In: *Journal of the Electrochemical Society* 143.12 (1996), p. 3809.
- [13] Doron Aurbach et al. “The study of electrolyte solutions based on ethylene and diethyl carbonates for rechargeable Li batteries: II. Graphite electrodes”. In: *Journal of The Electrochemical Society* 142.9 (1995), p. 2882.
- [14] D Aurbach et al. “Identification of surface films formed on lithium in propylene carbonate solutions”. In: *Journal of The Electrochemical Society* 134.7 (1987), p. 1611.
- [15] Luning Wang et al. “Identifying the components of the solid–electrolyte interphase in Li-ion batteries”. In: *Nature chemistry* 11.9 (2019), pp. 789–796.
- [16] Nav Nidhi Rajput et al. “The coupling between stability and ion pair formation in magnesium electrolytes from first-principles quantum mechanics and classical molecular dynamics”. In: *Journal of the American Chemical Society* 137.9 (2015), pp. 3411–3420.
- [17] RV Krems. “Bayesian machine learning for quantum molecular dynamics”. In: *Physical Chemistry Chemical Physics* 21.25 (2019), pp. 13392–13410.
- [18] Jörg Behler, Karsten Reuter, and Matthias Scheffler. “Nonadiabatic effects in the dissociation of oxygen molecules at the Al (111) surface”. In: *Physical Review B* 77.11 (2008), p. 115421.
- [19] Deping Hu et al. “Inclusion of machine learning kernel ridge regression potential energy surfaces in on-the-fly nonadiabatic molecular dynamics simulation”. In: *The journal of physical chemistry letters* 9.11 (2018), pp. 2725–2732.
- [20] Pavlo O Dral, Mario Barbatti, and Walter Thiel. “Nonadiabatic excited-state dynamics with machine learning”. In: *The journal of physical chemistry letters* 9.19 (2018), pp. 5660–5663.
- [21] David MG Williams and Wolfgang Eisfeld. “Neural network diabaticization: A new ansatz for accurate high-dimensional coupled potential energy surfaces”. In: *The Journal of chemical physics* 149.20 (2018), p. 204106.
- [22] Yafu Guan, Hua Guo, and David R Yarkony. “Extending the representation of multistate coupled potential energy surfaces to include properties operators using neural networks: Application to the 1, 21A states of ammonia”. In: *Journal of Chemical Theory and Computation* 16.1 (2019), pp. 302–313.
- [23] Jonathan P Alborzpour, David P Tew, and Scott Habershon. “Efficient and accurate evaluation of potential energy matrix elements for quantum dynamics using Gaussian process regression”. In: *The Journal of Chemical Physics* 145.17 (2016), p. 174112.

-
- [24] Yafu Guan, Hua Guo, and David R Yarkony. “Neural network based quasi-diabatic Hamiltonians with symmetry adaptation and a correct description of conical intersections”. In: *The Journal of Chemical Physics* 150.21 (2019), p. 214101.
- [25] Yafu Guan and David R Yarkony. “Accurate neural network representation of the ab initio determined spin-orbit interaction in the diabatic representation including the effects of conical intersections”. In: *The Journal of Physical Chemistry Letters* 11.5 (2020), pp. 1848–1858.
- [26] Wang Zhang Yuan et al. “Efficient solid emitters with aggregation-induced emission and intramolecular charge transfer characteristics: molecular design, synthesis, photo-physical behaviors, and OLED application”. In: *Chemistry of Materials* 24.8 (2012), pp. 1518–1528.
- [27] Hiroyuki Tanaka et al. “Twisted intramolecular charge transfer state for long-wavelength thermally activated delayed fluorescence”. In: *Chemistry of Materials* 25.18 (2013), pp. 3766–3771.
- [28] Qisheng Zhang et al. “Anthraquinone-based intramolecular charge-transfer compounds: computational molecular design, thermally activated delayed fluorescence, and highly efficient red electroluminescence”. In: *Journal of the American Chemical Society* 136.52 (2014), pp. 18070–18081.
- [29] Diptarka Hait and Martin Head-Gordon. “Orbital optimized density functional theory for electronic excited states”. In: *The Journal of Physical Chemistry Letters* 12.19 (2021), pp. 4517–4529.
- [30] Nicholas A Besley and Frans A Asmuruf. “Time-dependent density functional theory calculations of the spectroscopy of core electrons”. In: *Physical Chemistry Chemical Physics* 12.38 (2010), pp. 12024–12039.
- [31] David J Tozer and Nicholas C Handy. “On the determination of excitation energies using density functional theory”. In: *Physical Chemistry Chemical Physics* 2.10 (2000), pp. 2117–2121.
- [32] Tom Ziegler, Arvi Rauk, and Evert J Baerends. “On the calculation of multiplet energies by the Hartree-Fock-Slater method”. In: *Theoretica chimica acta* 43.3 (1977), pp. 261–271.
- [33] Xiaowei Xie, Kristin A Persson, and David W Small. “Incorporating electronic information into machine learning potential energy surfaces via approaching the ground-state electronic energy as a function of atom-based electronic populations”. In: *Journal of chemical theory and computation* 16.7 (2020), pp. 4256–4270.
- [34] P. H. Dederichs et al. “Ground States of Constrained Systems: Application to Cerium Impurities”. In: *Phys. Rev. Lett.* 53 (26 Dec. 1984), pp. 2512–2515. DOI: 10.1103/PhysRevLett.53.2512.

-
- [35] Qin Wu and Troy Van Voorhis. “Direct optimization method to study constrained systems within density-functional theory”. In: *Phys. Rev. A* 72 (2 Aug. 2005), p. 024502. DOI: 10.1103/PhysRevA.72.024502. URL: <https://link.aps.org/doi/10.1103/PhysRevA.72.024502>.
- [36] Benjamin Kaduk, Tim Kowalczyk, and Troy Van Voorhis. “Constrained Density Functional Theory”. In: *Chem. Rev.* 112.1 (2012), pp. 321–370. DOI: 10.1021/cr200148b. eprint: <https://doi.org/10.1021/cr200148b>. URL: <https://doi.org/10.1021/cr200148b>.
- [37] A.D. Becke. “A multicenter numerical integration scheme for polyatomic molecules”. In: *J. Chem. Phys.* 88.4 (1988), pp. 2547–2553. DOI: 10.1063/1.454033. eprint: <https://doi.org/10.1063/1.454033>. URL: <https://doi.org/10.1063/1.454033>.
- [38] Albert P Bartók, Risi Kondor, and Gábor Csányi. “On representing chemical environments”. In: *Phys. Rev. B* 87.18 (2013), p. 184115.
- [39] K Kaufmann and W Baumeister. “Single-centre expansion of Gaussian basis functions and the angular decomposition of their overlap integrals”. In: *J. Phys. B* 22.1 (1989), p. 1.
- [40] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [41] Justin S Smith, Olexandr Isayev, and Adrian E Roitberg. “ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost”. In: *Chem. Sci.* 8.4 (2017), pp. 3192–3203.
- [42] Jeng-Da Chai and Martin Head-Gordon. “Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections”. In: *Phys. Chem. Chem. Phys.* 10.44 (2008), pp. 6615–6620.
- [43] Dmitriy Rappoport and Filipp Furche. “Property-optimized Gaussian basis sets for molecular response calculations”. In: *The Journal of chemical physics* 133.13 (2010), p. 134105.
- [44] Yihan Shao et al. “Advances in molecular quantum chemistry contained in the Q-Chem 4 program package”. In: *Mol. Phys.* 113 (2015), pp. 184–215. DOI: 10.1080/00268976.2014.952696. eprint: <http://dx.doi.org/10.1080/00268976.2014.952696>. URL: <http://dx.doi.org/10.1080/00268976.2014.952696>.
- [45] Harry E Pence and Antony Williams. *ChemSpider: an online chemical information resource*. 2010.
- [46] John P Perdew and Mel Levy. “Extrema of the density functional for the energy: Excited states from the ground-state theory”. In: *Physical Review B* 31.10 (1985), p. 6264.

Chapter 4

Data-driven Prediction of Formation Mechanisms of Lithium Ethylene Monocarbonate with an Automated Reaction Network

Parts of this chapter were reprinted with permission from: “Data-Driven Prediction of Formation Mechanisms of Lithium Ethylene Monocarbonate with an Automated Reaction Network” Xiaowei Xie, Evan Walter Clark Spotte-Smith, Mingjian Wen, Hetal D. Patel, Samuel M. Blau, Kristin A. Persson. *J. Am. Chem. Soc.* **2021**, 143, 33, 13245–13258.

I Introduction

Out-of-equilibrium interfacial reactions which result in functional, transport-selective surface films play a crucial role in materials-based technologies, enabling applications from batteries, capacitors, water splitting, protective coatings, and alloy design. An exemplar in the family of spontaneously-formed, enabling interfaces is the passivation layer which results from the reductive decomposition of electrolytes on the anode of lithium-ion batteries (LIB) during the first few charging cycles. This solid electrolyte interphase (SEI) is considered “the most important but the least understood in rechargeable Li-ion batteries”, [1] due to the complex relationships between battery conditions, SEI composition/morphology, and SEI properties.[2–5] It is generally believed that an ideal SEI should be selectively permeable to the electrochemically active charge carrier, i.e. Li^+ , but also electron-insulating to prevent continuous reduction of the electrolyte species.[6]

Over the last decades, careful bulk and surface characterization techniques, such as X-ray photoelectron spectroscopy (XPS), [7, 8] Fourier transform infrared (FTIR) spectroscopy, [9, 10] nuclear magnetic resonance spectroscopy (NMR), [11, 12] cryo-scanning transmission electron microscopy (cryo-STEM) [13–15], etc., have been employed to elucidate the chemical composition of the SEI on both Li metal and graphite. Critical understanding has been gained as to early SEI formation as well as its aging. However, there are still numerous uncertainties as to specific reaction mechanisms, influence of the electrode chemistry and potential on competing reactions, and even the characterization of fundamental SEI components such as e.g. lithium ethylene dicarbonate (LEDC). LEDC was first proposed to be the major organic constituent of the SEI from the EC electrolyte by Aurbach et al. [9, 16–19], and later investigated extensively by numerous studies including synthetic efforts of the LEDC standards and comparisons with those synthetic standards based on FTIR, XPS, and NMR techniques.[11, 20–26]. Recent work by Wang et al. [27] cast doubt on the LEDC characterization: it was suggested that common LEDC standards[21, 23, 25] are comprised of lithium ethylene monocarbonate (LEMC). The authors further suggested that LEMC degrades to ethylene glycol (EG) under D_2O extraction, thus rendering the ^1H NMR spectra for LEDC and EG indistinguishable. However, while characterization of LEDC is under debate, questions also remain as to how LEMC forms under typical battery operation conditions and if it truly replaces LEDC as a major early SEI component.

Computational modeling of SEI-relevant reactions has also provided crucial insights into specific reaction mechanisms, including short-lived intermediate products.[28, 29] Prominent

examples include the early pioneering work from Balbuena and coworkers based on static DFT calculations to clarify the reduction mechanisms of common organic solvents and additives, [30–32] surface calculations [33–35], *ab initio* molecular dynamics (AIMD) simulations [36, 37], and others. However, currently employed methodologies such as static DFT, *ab initio* and classical molecular dynamics are inherently limited either by chemical intuition, simulation time scales and/or sampling. In particular, more complex, rare, co-operative and combined reaction pathways as well as the competition between different reaction mechanisms are difficult to capture. However, progress has been made in the automated searching for reaction pathways, traversing the potential energy surface (PES) using combinations of weighted sampling, heuristic rules and graph theory.[38–57] Recently, a new data-driven methodology was developed to analyze complex electrochemical reactive systems with minimal chemical intuitive bias through the use of computational reaction networks.[58] Using this methodology, pathways from Li^+ and EC to LEDC under reductive conditions corresponding to Li metal were explored. It was demonstrated that the reaction network automatically recovers previously suggested reaction mechanisms for the formation of LEDC as well as proposed novel, thermodynamically feasible mechanisms that had not previously been reported. Here, we utilize the same framework to identify possible reaction pathways leading to LEMC based on automated pathfinding in a landscape built on *ab initio* thermochemistry. We compare the proposed pathways on the basis of their thermodynamics, kinetics, and byproducts. We note that thermodynamic analysis alone is inadequate for assessing the energetic landscape, as kinetics also play a pivotal role in the SEI formation.[36, 59] In the present study, we also conduct static transition state (TS) calculations for the promising paths from the pathfinding algorithm using a total thermodynamic cost cutoff, so as to refine the reaction network predicted paths and assess their kinetic feasibility. The purpose of our work is two-fold:

(1) We implement and demonstrate a systematic and minimally biased methodology for the generation of relevant species and reactions. Subsequently, reaction pathways to a certain product of interest can be obtained in an automated fashion through pathfinding from a reaction network containing these species and reactions. All relevant molecules were generated through a systematic fragmentation/recombination procedure such that all reactions that satisfy a certain constraint (involving ≤ 5 bond changes here) were included. The effectiveness of this approach is tested on the formation of the new SEI component of interest, LEMC. We demonstrate that our data-driven methodology recovers not only previously proposed mechanisms, but also discover novel, competitive reaction pathways.

(2) We assess the reaction network-predicted pathways both in the context of experimental findings and from a comparison with DFT-refined pathways. We note that the advantage of the automatically generated pathways lies in the rapid evaluation of thermodynamically plausible reactions covering a vast electrochemical space and provides a thermochemically informed starting point for more bespoke computational and experimental explorations.

II Methods

2.1 Generation of Candidate Molecules and Molecular Intermediates through Fragmentation and Recombination

To generate the constituents of the reaction network, we employ heuristic rules based on thermodynamic data and associated machine-learned bond dissociation and formation, which we denote as the fragmentation and recombination procedure. This approach enables the generation of a large number of candidate molecular structures that represent the out-of-equilibrium, transient landscape of possible molecular fragments that form under electrochemical electrolyte decomposition. These candidate structures are subsequently optimized by electronic-structure methods to obtain minimum-energy structures. Molecular species are treated as graphs based on the connectivity of atoms. Fragmentation is defined as breaking one edge in a molecular graph and generating one or two molecular fragments; recombination is defined as forming one edge between two specified atoms on two molecular graphs and forming one recombinant molecular graph. In a one-step fragmentation, fragment molecular graphs are generated from breaking any one bond in the molecule. An n -step fragmentation involves multiple sequential one-step fragmentation processes; at each stage, the fragments from the previous stage undergo a one-step fragmentation, and the fragments from each such one-step procedure are collected. Similar to the fragmentation procedure, a general one-step recombination is conducted by iterating over all the molecular pairs in a given set of species, iterating over all atoms in both molecules, and generating all recombinant molecules by connecting a bond. During the process, a graph isomorphism comparison is applied at each step to check for structural duplicates, which are then removed. To compare and contrast LEDC with LEMC formation pathways as a result of electrochemical decomposition of an open system of Li^+ , EC, H_2O and electrons, we adopt the following procedure to generate the relevant species (an overview is shown in Figure 4.1).

(1) All possible unique fragments were generated from 2-step fragmentation of Li^+EC^- and one-step fragmentation of H_2O . An initial filtering of the resulting fragments is conducted to prevent the combinatorial explosion of the number of species. The filtering is done by first constructing a small reaction network with only the original molecules and the fragments, considering one-electron redox and one-bond change reactions, according to the previously reported procedure [58]. Pathfinding to each species mentioned above enable us to identify which fragments are thermodynamically accessible in the present model conditions by overall exergonic reaction pathways (overall $\Delta G < 0$). However, an exception is made for H_2O to include the fragments OH and H in all charge states, as breaking a bond in water is endergonic in molecular calculations, but will occur favorably in the presence of Li metal (as will be discussed in Sec 2.3 and Sec 3.1).

(2) All hypothetical recombinant structures were created by connecting one bond between

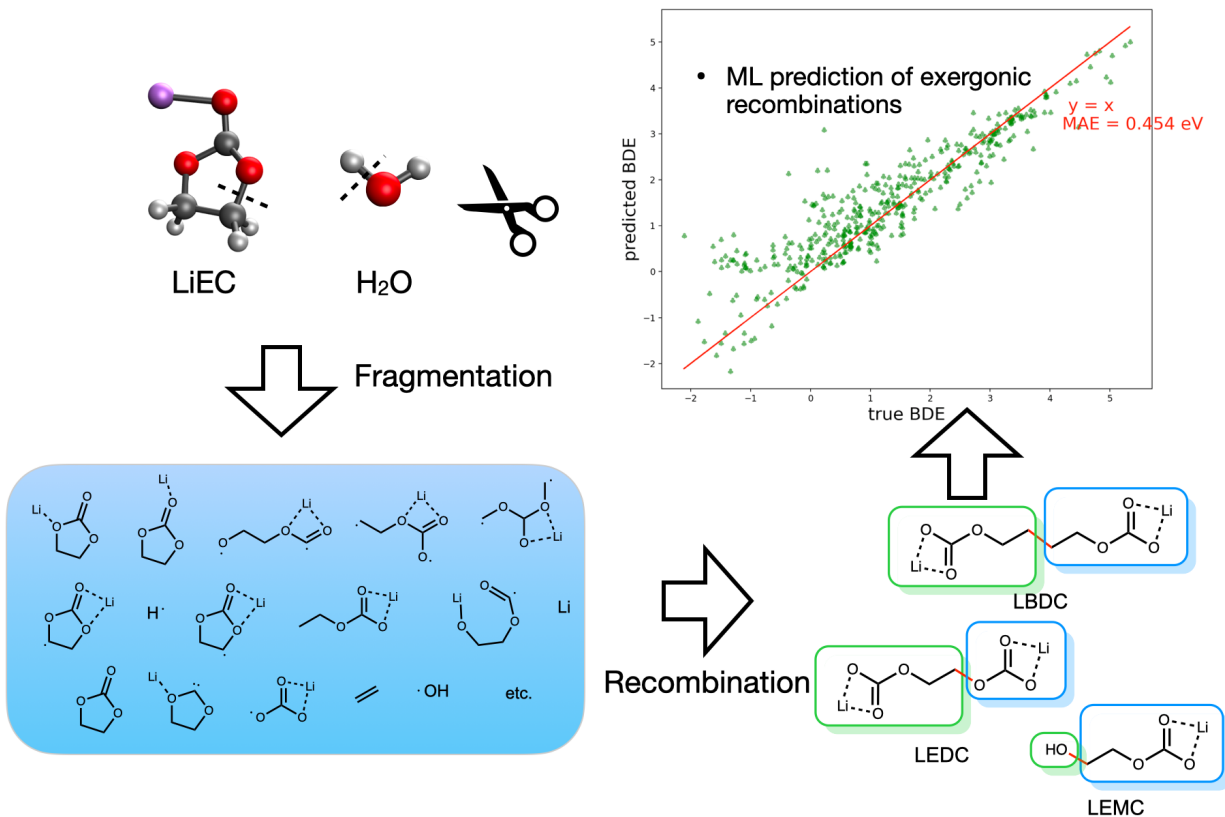


Figure 4.1: Workflow for generating relevant molecules

the generated fragment molecules. There were no other criteria imposed except to ensure chemically reasonable valences of the final recombinant products. Recombinant structures that do not satisfy the valence (e.g. a carbon with more than 4 bonds or an oxygen with more than 2 bonds) were discarded.

(3) A recent graph neural network BonDNet [60] was employed to predict the recombination reaction free energy and only those with negative ΔG upon recombination were retained. We note that the mean absolute error (MAE) on the prediction of reaction free energies for recombination is 0.45 eV with a high coefficient of determination (R^2) value of 0.84 for 449 unique recombination reactions. The figure on the top right corner of Figure 4.1 plots the BonDNet predicted vs DFT computed bond dissociation energies (BDEs; the recombination free energy would be the negative number of the corresponding BDE). The higher MAE than that reported in the BonDNet paper is likely to originate from (1) examples far away from the regime of the training set, e.g. forming a bond between sterically hindered oxygens and (or) carbons and (2) the larger conformational space of large flexible recombinant molecules,

as currently BonDNet is not taking the energies of different conformers into consideration. The performance of BonDNet can be iteratively improved by adding more representative data into the training set: by adding these recombinants back into the training, we observed a reduction of MAE from 0.496 eV to 0.355 eV when testing on a new test set consisting of a subset of the recombinants we previously eliminated (details given in Appendix Sec 5.6). Out of 382 newly calculated recombinants, we only excluded 7 molecules (less than 2% error) due to the BonDNet prediction error, which represents a low percentage loss and is unlikely to change our conclusions significantly. Detailed error analysis on both models is given in Appendix Sec 5.6.

(4) The geometries of the resulting molecules were optimized in DFT. The initial guesses for the 3D geometries of the recombinant molecules were generated based only on 2D molecular connectivities from the OPLS_2005 molecular force field [61] as implemented in the Schrödinger Python Suite. [62] To reduce the computational burden of generating the recombinant molecules, no conformational sampling was conducted. During geometry optimization, a portion of the recombinant molecules changed their connectivities; those molecules were removed from consideration as the results of BonDNet were no longer applicable and they were no longer in the regime of one-step recombination.

We obtained 570 candidate molecules in total through this procedure (45 from fragmentation and 525 from recombination). We note that the generation of the network species is self-contained, such that all fragments and recombinant molecules were generated from the procedure (1)-(4). Specifically we emphasize that no molecules were added from previous knowledge of Li electrolyte decomposition and that the target molecules LEDC and LEMC were obtained in this automated fashion. All the molecules used in this work are given in the Appendix; a more comprehensive dataset which contains, for example, fragments for various salts and other electrolytes, is released in another work.[63]

2.2 Construction of Reaction Network and Identification of Reaction Pathways

The reaction network was constructed as a connected graph using the same graph representation as described in previous work. [58] Allowing a broad range of reactions between the molecular species, we chose to include “concerted” reactions (reactions that involve multiple, simultaneous bonds breaking or forming) with up to 5 bond changes. Similar constraints (usually ≤ 2 bonds breaking and ≤ 2 bonds forming) were imposed in other works. [47, 51, 53] We emphasize that metal coordination reactions, such as Li bonding to a heteroatom, is treated in the same manner regardless of their ionic/dative nature, which motivates an inclusive, upper limit of allowed concerted bond changes. We note, however, the possibility that reactions considered here as concerted, may in reality proceed through multiple elementary steps. Hence, we consider reaction pathways as reasonable guesses for a collection of events

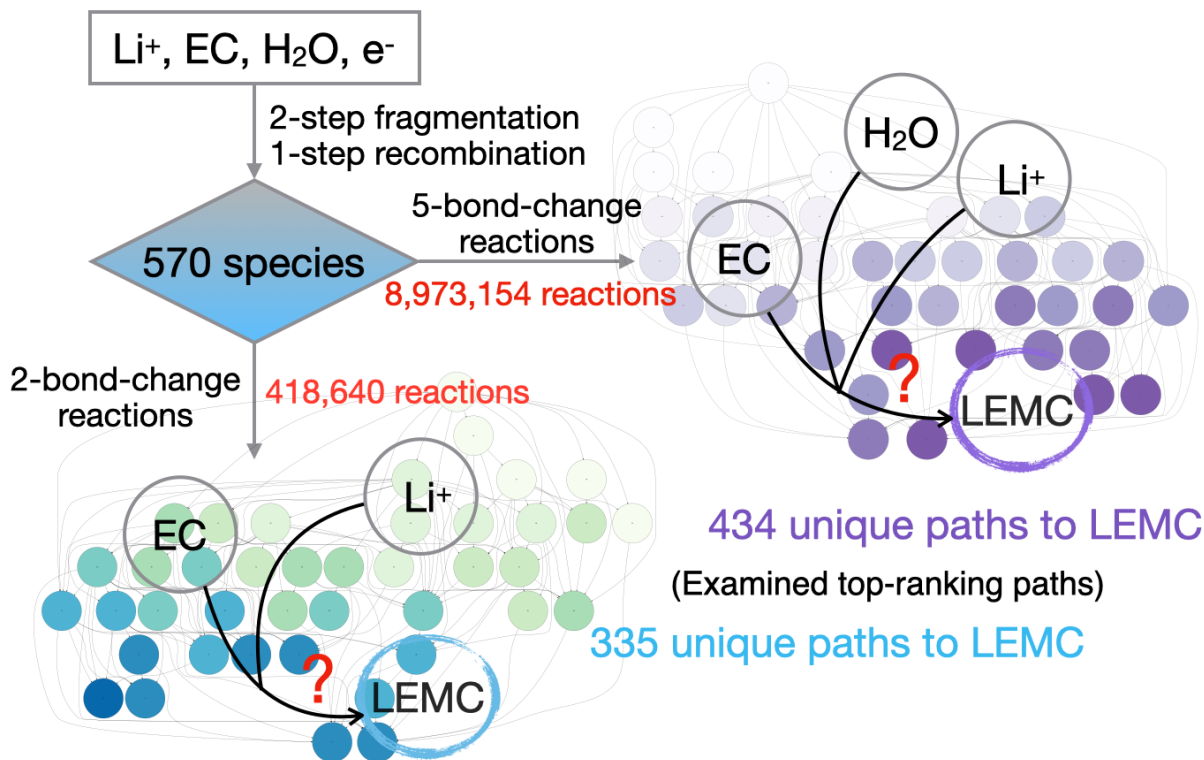


Figure 4.2: Overview of the reaction networks generated in this work

that should be refined using quantum chemical calculations.

To efficiently include concerted reactions, a mixed-integer linear programming (MILP) scheme modified from a previous work for atom-mapping[64] was utilized to calculate the minimum chemical distance (CD), i.e. the minimum number of bond changes needed to transform reactants to products, for all possible combinations of (up to two) reactants and products (for details, see Appendix Sec 5.1). All molecular pairs that satisfy a given stoichiometry were subsequently evaluated based on the minimum number of bond changes required to convert one molecule pair to another. This procedure resulted in a collection of all reactions that can be reached within 5 bond changes, totalling 570 species and 8,973,154 reactions.

To calculate the free energy for redox reactions relevant in Li-ion batteries, we employed a potential of $U = 0$ vs Li/Li⁺. Hence, to represent the reaction free energy for an arbitrary redox reaction, we employ the full-cell reaction $A + \text{Li(s)} \rightarrow A^- + \text{Li}^+$. The absolute electrode potential for the lithium electrode is calculated to be ~ 1.40 V by adding the Li/Li⁺ potential relative to standard hydrogen electrode (SHE) to the absolute electrode potential of hydrogen

electrode (4.44 V) [65–68]. We note that this is equivalent to a chemical potential of -1.40 eV for a Li metal electron. The effect of the electrode potential (or any applied external potential) can be easily accounted for, by shifting the free energy of the electron by $-eU$, where U is the potential relative to Li/Li⁺. For instance, the graphite anode in LIBs operates at a potential close to $U = 0.2$ V vs Li/Li⁺ which is equivalent to an electron chemical potential of -1.60 eV.

To identify promising thermodynamic pathways from the nearly 9 million hypothetical reactions, we adopted the same procedure for pathfinding as described in the previous work. [58] This approach provides us with a ranked list of reaction sequences and each reaction sequence connects the starting molecules to the target molecule (LEMC), within a grand canonical thermodynamic ensemble, allowing for a variable number of each specie. A soft-plus function, i.e. $\ln(1 + \exp(x))$, of reaction free energy is used to determine the cost for traversing a reaction in the reaction network based on computational efficiency and the tolerance for slightly endergonic reactions. However, we note that the resulting ranking from the cost function is not used to exclude pathways from consideration. Indeed, after obtaining the representative reaction pathways from the reaction network, we subsequently subject the proposed pathways to a detailed quantum mechanical study to obtain the transition state barriers ($\Delta G^\ddagger$) as well as the reaction energies of the complexes (ΔG) to verify their kinetic/thermodynamic feasibility. This information is intended to refine the pathways obtained from the reaction network which currently does not include any explicit kinetic information. Furthermore, concerted mechanisms are refined in this more thorough exploration.

2.3 Computational Methods

DFT calculations were performed using the QChem[69] electronic structure code. The geometry optimizations for the species included in the reaction network used the ω B97X-V[70] functional with the def2-TZVPPD basis set. [71, 72] Solvent effects were taken into account by SMD implicit solvation models[73], with user-defined parameters representing the Li-ion electrolyte. Specifically, an experimentally measured dielectric constant (for 3:7 EC/EMC binary mixtures[74]) of 18.5 was used; for all other parameters we used the values for pure EC.[75] For all stationary point geometries, frequency calculations were conducted to ensure they are true local minima on the PES for minimum structures or first-order saddle points for TS structures. Gibbs free energies based on harmonic approximations were used throughout to determine the energy landscape in the reaction pathways.

Geometry optimization and reaction path analysis for the predominant paths were performed at the ω B97X-D[76]/def2-SVPD[72] level of theory with the C-PCM implicit solvation model [77] (dielectric constant 18.5, optical dielectric 2.0, which is consistent with the SMD parameters), while single point energies at the stationary points were obtained using the ω B97X-V/def2-TZVPPD/SMD level of theory using the same SMD parameters as above. These two levels of theory have been shown to perform well for metal organic systems. [78,

79] We used the lower level of theory for computational efficiency in exploring the PES; in our experience, these two levels of theories give good agreements in both molecular geometries and relative energies. Vibrational frequencies were computed using the smaller basis, yielding zero-point energies and thermal corrections. A procedure based on taking small fixed-length steps in the steepest descent direction (the Euler’s method[80]) was performed to verify the TSs connect to the expected reactant(s) and product(s) through a smooth energy-descent path. For this verification in general, it is not necessary to obtain an *optimal* minimum-energy path (MEP), through e.g. Intrinsic Reaction Coordinate (IRC) calculations, which are more time-consuming.

Evaluating possible reaction paths to LEMC with and without water is a critical component of this study, however we note that a single water molecule does not spontaneously reduce in the presence of a single Li atom; it requires the presence of a metallic Li surface. Hence, the network, being ignorant of extended surface effects, will extend an endothermic (potentially artificial) cost to any water splitting reactions. As a consistency check, we verified that our level of theory does reproduce spontaneous water splitting reaction on a Li metal surface from a cluster-based DFT model (see Appendix 5.2 for computational details).

Besides surface effects, explicit solvation may exhibit a non-trivial impact on the reaction pathways. Although calculating every path with explicit solvation would be beyond the scope of the current work, we have obtained the reaction paths with one explicit EC molecule for two representative cases, the EC and LEDC hydrolysis paths as shown below. The results are included in Appendix Sec 5.5. In these two examples, the presence of one EC can lower the rate-limiting barrier of a path by 0.1-0.4 eV if the solvent molecule is near the reaction center. Nevertheless, in our case explicit solvation does not qualitatively affect our assessment of the kinetic feasibilities of the reaction paths.

III Results and Discussion

Herein we present and analyze the formation pathways to a recently identified, potential key component of the Li-ion SEI, LEMC.[27] Analyzing the top-ranked 30,000 pathways, first of all, we found that many of them were in fact the same, except for a difference in the order of reactions. These “duplicate” paths exhibit the same cost, and eliminating the paths with identical costs to threshold 10^{-3} resulted in 434 unique paths (the unique paths are included in the Appendix). We further grouped the paths roughly based on the main molecules involved in the paths and report the representative paths below. While this elimination may exclude subtle mechanistic differences in the paths, we believe that the main features are preserved.

Analyzing the participant molecules in the reaction pathways leading to LEMC formation, interestingly, we found that only 8 paths (among the 434 unique ones) proceeded without water. Furthermore, DFT refinement of these paths suggests none of these pathways are

kinetically feasible, with the rate-limiting barriers for the key reactions in the paths exceeding 2 eV (details shown in Appendix Sec 5.4). This important result indicates that water is likely to be critically important in the formation of LEMC. We note that the presence of kinetically infeasible paths is inevitable: this is a direct consequence of including a broader scope of reaction possibilities. Indeed, the identified infeasible mechanisms were found to include over-complicated concerted mechanisms connecting unstable radical and species, e.g. carbene. Since this artifact may result in missing important non-water mechanisms, we constructed another reaction network with more limited concerted reactions, and the details will be given in Sec 3.3 below. An overview of the two reaction networks considered in this work is given in Figure 4.2.

The two major pathways that have been proposed in the literature, namely the EC hydrolysis and LEDC hydrolysis reactions, were successfully identified by the reaction network, validating our data-driven approach. Additionally, other possibly interesting pathways were discovered and will be discussed below, such as a pathway that generates lithium hydride, pathways involving new, recombinant compounds that have not been conventionally considered to be important in the SEI formation, and a pathway involving the breakage of the $(\text{CH}_2)\text{O}-\text{C}(=\text{O})\text{OLi}$ bond in LEDC.

In the following sections, we divide the reaction pathways to LEMC into three categories: (i) direct paths from EC, Li^+ and H_2O , where only the fragments of LiEC serve as intermediates in the paths, (ii) paths involving recombinant molecules, and (iii) paths without water participation. We present, for each category, both the representative paths predicted from the reaction network and the DFT validated/refined free energy diagrams, with some discussion about the similarities/differences between them. We also discuss the physical implications of those paths in the context of experimental findings.

3.1 Direct reaction paths to LEMC from EC, Li^+ and H_2O

Figure 4.3 shows a representative set of “direct” paths leading to the formation of LEMC, utilizing EC, Li^+ and H_2O as the starting molecules. Although differing in the detailed steps, the two major mechanisms identified were (i) LiOH or OH^- attacking an EC or LiEC^+ to form LEMC (path A, B, C, E) and (ii) LiEC , water and a second electron react to form LEMC and hydride (path D). Figure 4.4 presents the energy profiles from DFT (ignoring explicit solvation and surface effects) corresponding to these two mechanisms. We note that the generation of LiOH and/or OH^- species can be corroborated by the known instability of water against Li metal,[81] which is also captured within our level of theory, when an explicit surface is included (see Appendix Sec 5.2). Since the reaction network does not have any knowledge of water splitting on the Li surface, it used indirect mechanisms for generating LiOH and/or OH^- .

The reaction of LiOH attacking EC to form LEMC was proposed by Wang et al. and also experimentally confirmed by adding anhydrous LiOH to an EC/ LiPF_6 solution. [27]

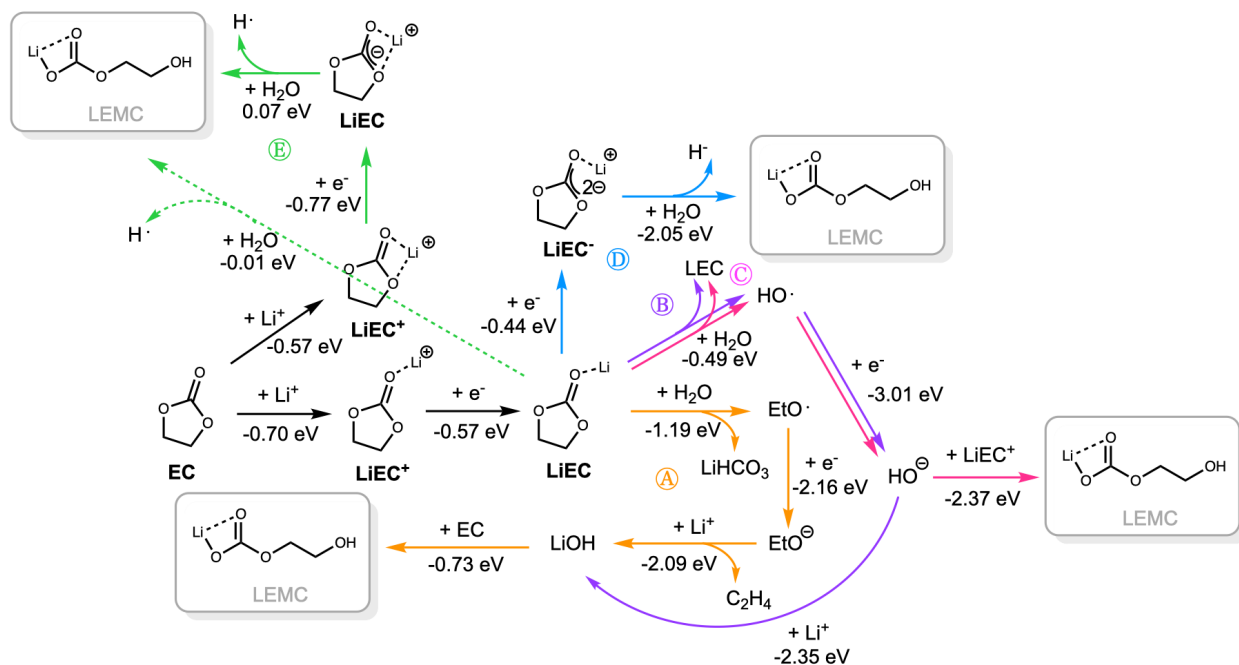


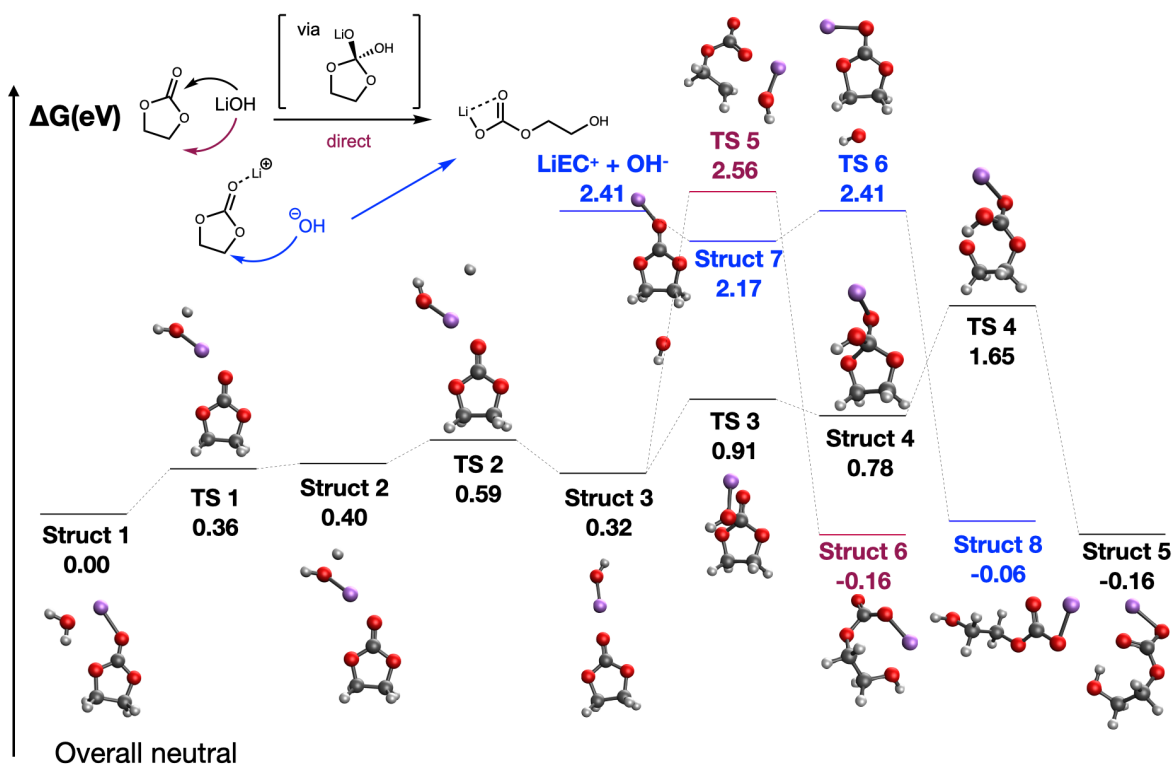
Figure 4.3: Direct paths to LEMC from EC, Li⁺ and H₂O identified by the reaction network

In the EC + LiOH → LEMC reaction, we identified two mechanisms through DFT (purple and black paths in Figure 4.4 (a)): one is a single-step process where LiOH directly attacks the ethylene carbon in the EC; another one is a step-wise process where LiOH attacks the carbonyl carbon to form a tetrahedral intermediate, and a subsequent shoulder bond breaking and proton transfer creates LEMC. The former one-step process has a high barrier of 2.24 eV (Struct 3 → TS 5; 1.95 eV with an explicit EC molecule, as shown in Figure 4.14, Struct 1 → TS 3), whereas the step-wise mechanism exhibits an effective overall barrier of 1.33 eV (Struct 3 → TS 4; 0.95 eV with an explicit EC molecule, as shown in Figure 4.14, Struct 1 → TS 2). The reaction barrier can be significantly lowered if the hydroxide ion is free and not bound to lithium (0.25 eV, Struct 7 → TS 6; 0.27 eV with an explicit EC molecule, as shown in Figure 4.14, Struct 6 → TS 4). This is consistent with the fact that Li⁺ is a strong Lewis acid, and the nucleophilicity of bare hydroxide anions is expected to be greatly reduced in the presence of Li⁺.^[82, 83] Similarly, the electrophilicity of EC would be enhanced by Li⁺, thus also increasing the reactivity between the two species. We consider OH⁻ + LiEC⁺ → LEMC a more realistic scenario for EC hydrolysis, as (i) LiOH has extremely low solubility in organic solvents, (ii) OH⁻ should be expected to be present because the electron-abundant anode represents a highly basic environment, ^[84] and (iii) LiEC⁺ is considered to be abundant in battery electrolytes and the Li⁺-coordinated EC is reduced at a higher potential than the

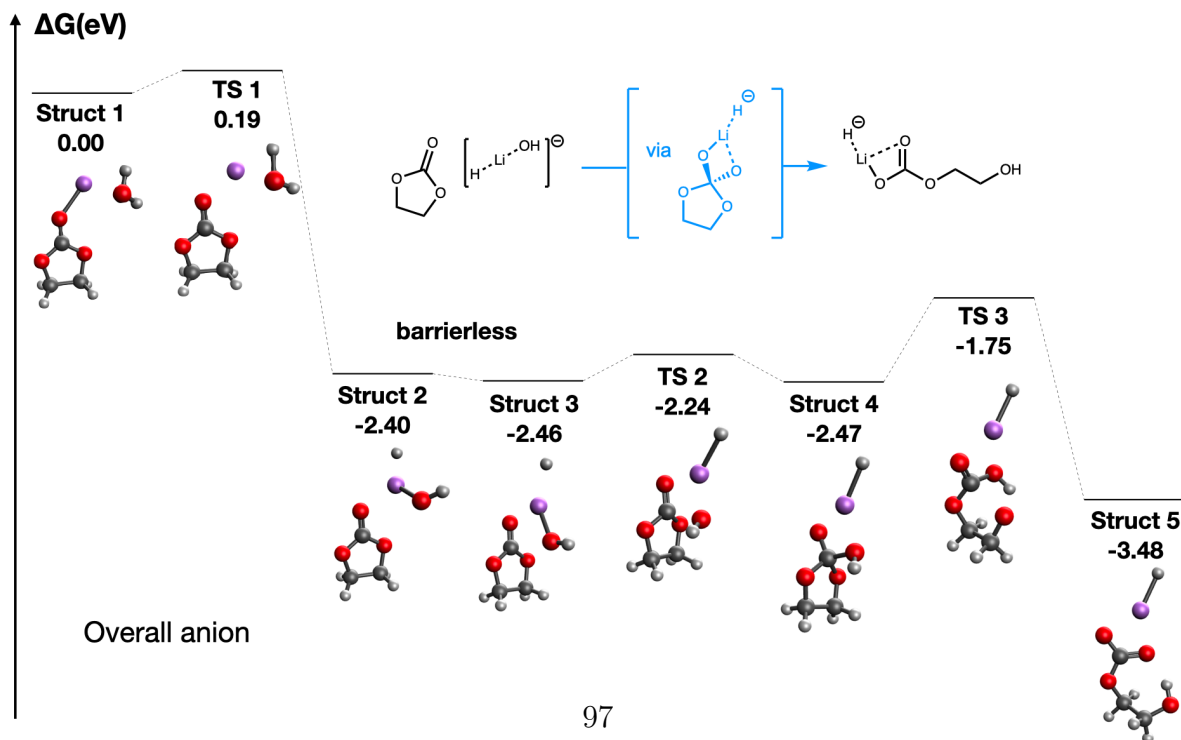
pure solvent.[30]

Path D in Figure 4.3 as proposed from the reaction network represents another possibility of LEMC formation. In path D, a neutral LiEC is further reduced to LiEC^- and reacts with water to form LEMC and a hydride ion. $(\text{c-EC}^{2-})\text{Li}^+$, where c-EC^{2-} represents the cyclic EC dianion, was proposed previously by Leung et al., [36] and the reduction rate of c-EC^- was estimated based on Marcus theory to be significantly higher (at $\sim 10^4$ order) than that of EC, but lower than that of o-EC^- , which is the ring-opened EC anion. Leung et al. also suggested that the doubly reduced $(\text{c-EC}^{2-})\text{Li}^+$ favorably ring opens at the shoulder bond and release CO. Regardless of the existence of c-EC^{2-} (or its life time), the reaction network suggests a reaction path according to $\text{EC} + \text{Li}^+ + 2\text{e}^- + \text{H}_2\text{O} \rightarrow \text{LEMC} + \text{H}^-$. This poses a possibility of a simultaneous (presumably asynchronous) formation of hydride and LEMC. We anticipate that the hydride will react rapidly with any available Li^+ ion to form LiH, which is likely to - in the presence of water - further react to LiOH on its surface. Experimentally, LiH was first characterized by IR spectroscopy on Li metal with wet electrolyte solutions, [85] and later observed as a main dendritic product in the Li-anode-SEI by cryo-STEM mapping [15] and XRD experiments. [86]

Examining this hypothetical path more carefully, we obtain a mechanism that resembles the black path in Figure 4.4 (a). Note that the overall charge state for the structures in the energy profile in Figure 4.4 (a) is neutral, whereas in Figure 4.4 (b), the overall charge is -1. We were unable to identify a path that proceeds through $(\text{c-EC}^{2-})\text{Li}^+$, presumably because water is more readily reduced than c-EC^- (the hydrogen evolution potential occurs at 2.21 V relative to Li/Li^+ , [87] whereas the reduction potential of c-EC^- is calculated to be 1.16 V relative to Li/Li^+ . [36]). Since the reaction network assigns a potentially overestimated endothermic penalty (depending on the distance to the fresh Li metal surface) to water reduction (see Sec 2.3), it proposes reduction of LiEC rather than H_2O . In the DFT refinement, we see a favorable water splitting in the anion state and identify a path corresponding to $\text{H-Li-OH}^- + \text{EC} \rightarrow \text{LEMC} + \text{H}^-$ (Figure 4.4 (b)). We include a water splitting step with LiEC^- at the beginning only for the sake of completing the energy diagram, however the LiEC^- in Struct 1 is Li^-EC in character, as revealed by the Mulliken[88] atomic charge (-1.17 e) and spin (0.0 a.u.), and the Li in the structure is considered a Li surface proxy, coordinating to EC. A similar argument can be made for the water splitting step in Figure 4.4 (a), where the Li is neutral in character (Struct 1). As mentioned in Sec. 2.3, metallic surface effects are neglected in the reaction network, and to confirm that water splitting is indeed spontaneous within our level of theory, we used a cluster DFT model consisting of 32 Li atoms and found that the reaction is almost barrierless (see Appendix Sec 5.2) and consistent with previous work (0.12 eV).[89] We note that although water reduction reactions are usually considered to be equivalent to hydrogen evolution reaction (HER), this is a simplified picture as the detailed mechanisms involve adsorbed species H_{ads} and OH_{ads} . [89, 90] Hence, the as-formed hydride ion, most likely created through water splitting on the Li surface, is presumably stabilizing the TS structures along the path and the overall barrier



(a) EC hydrolysis paths, corresponding to path A, B, C, E in Figure 4.3. Note that the free energy of a hydrogen radical was added for the structures to the right of TS 2, such that the energies are comparable.



(b) Hydride generating path, corresponding to path D in Figure 4.3

Figure 4.4: LEMC paths from DFT

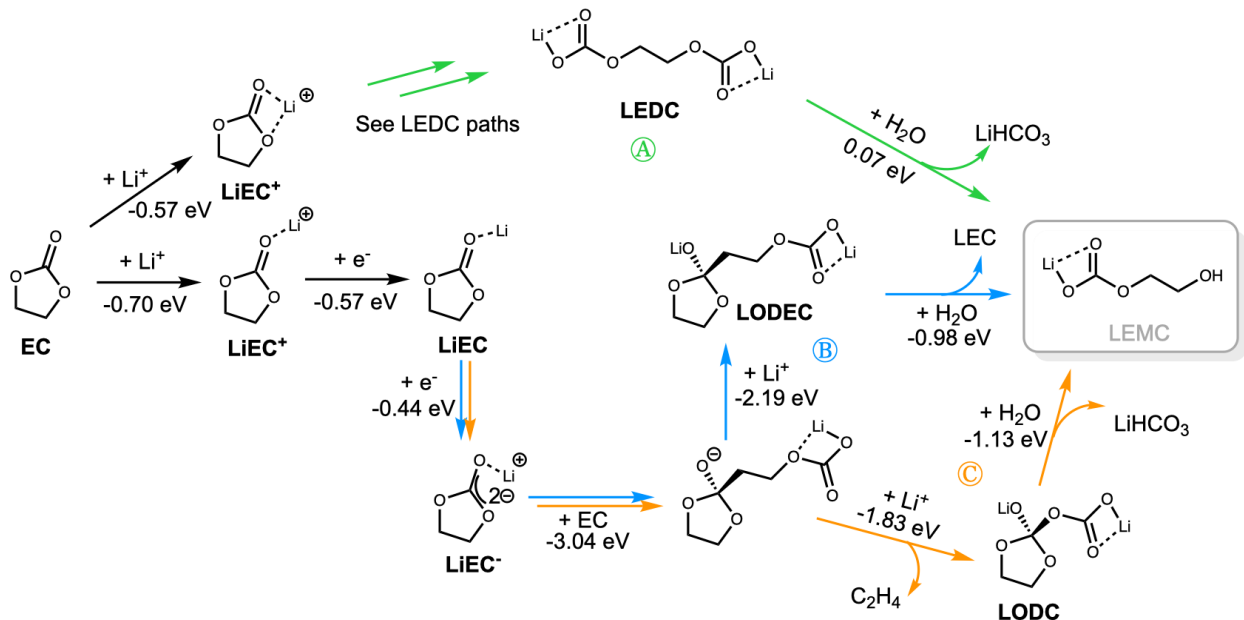


Figure 4.5: Paths to LEMC involving recombinant molecules from EC, Li^+ and H_2O identified by the reaction network

(0.72 eV as the rate-limiting barrier, Struct 4 $\rightarrow$ TS 3) is much lower than that for the black path in Figure 4.4 (a). Further investigation is necessary to determine the plausibility of this path and whether cooperative effects could further lower the barrier in this pathway.

Generally comparing our results with previous investigations of water reactivity with Li metal surface, our results are consistent with Shang et al.[89] who found the final products of H_2O reacting with a Li surface to include LiOH , a small amount of LiH , H_2 gas, and various water complexes.

3.2 Reaction paths to LEMC from EC, Li^+ and H_2O involving recombinant molecules

By including recombinant molecules into the reaction network, we were able to identify some pathways to LEMC involving new intermediate molecules en route to the formation of LEMC, as shown in Figure 4.5. Note that the inclusion of LEMC and LEDC, along with other recombinants is a direct consequence of graph enumeration. In other words, those molecules are naturally considered irrespective of any prior knowledge of the SEI components previously studied in the literature.

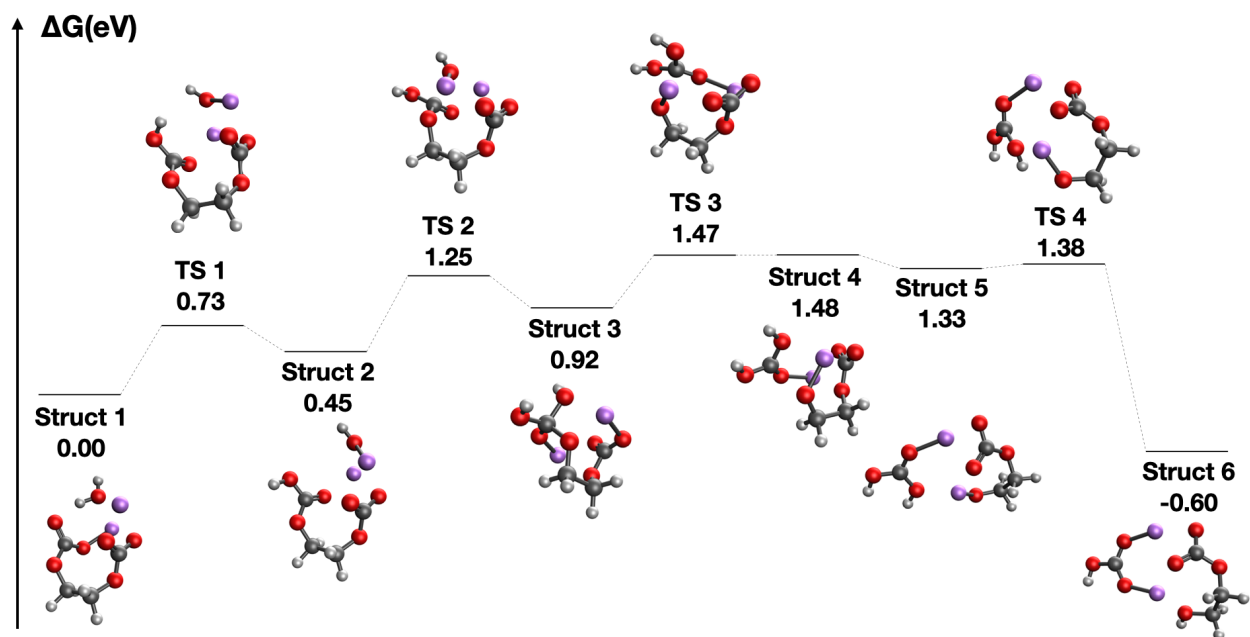
The hydrolysis reaction of LEDC (path A in Figure 4.5) was both proposed by Wang

et al. [27] and Rinkel et al. [84]. This pathway was successfully identified by our reaction network, albeit with a low ranking (ranking 352 among the 434 unique paths), presumably owing to the endergonicity of the last $\text{LEDC} + \text{H}_2\text{O} \rightarrow \text{LEMC} + \text{LiHCO}_3$ reaction step (0.07 eV). We also note that the LEDC and LEMC conformers used in the reaction network are linear configurations, due to inadequate sampling of the conformational space. Applying a conformational search under implicit solvation identified a puckered LEDC structure (see Figure 4.6 (a)), which is lower in energy than the linear conformer by 0.6 eV. While this puckered conformer obtained here may or may not be representative under explicit solvation (to our knowledge, there is no known crystal structure for LEDC[23, 27, 91]), the significant energy difference highlights the impact of LEDC conformation to the reaction energetics.

Examining more closely path A and its kinetics, we found that the single-step LEDC hydrolysis step exhibits a high barrier (3.05 eV, Figure 4.10, Struct 1 $\rightarrow$ TS 1). However, hydrolysis reactions do not usually proceed in a single step,[92–94] and therefore we attempt to obtain the barrier for a step-wise hydrolysis of LEDC passing through a tetrahedral intermediate, either with neutral water or OH^- . The rate-limiting barrier of the water version of this path is 1.48 eV (Figure 4.6 (a), Struct 1 $\rightarrow$ Struct 4; 1.37 eV with an explicit EC molecule, as shown in Figure 4.15, Struct 1 $\rightarrow$ TS 3), which is lowered significantly compared to the single-step path. Under basic conditions (starting from OH^-), the barrier was calculated to be 1.76 eV from the entrance complex (Struct 1 $\rightarrow$ TS 1), 0.62 eV from the starting materials ($\text{LEDC} + \text{OH}^- \rightarrow \text{TS 1}$) for the black path and 1.09 eV ($\text{LEDC} + \text{OH}^- \rightarrow \text{TS 4}$) for the green path (Figure 4.6 (b)).

In polymolecular reactions, such as the hydrolysis of LEDC, there is a possibility that the entrance complex before entering the transition state region is lower in energy than the sum of the starting materials, in which case the rate-limiting step may be determined by the energy difference between the TS and the entrance complex. Hence, whenever there is a non-negligible discrepancy between using the sum of the free energies of the individual reactants and the free energy of the entrance complex as the baseline for the barrier height, we included the results for both scenarios.

We observe that the LEDC hydrolysis reactions are less kinetically favorable compared to the EC hydrolysis pathway and the hydride pathway based on our calculations; however, LEDC hydrolysis was found to occur both in DMSO-d_6 solutions (with pure LEDC plus 1.5 equiv. of water) and in the solid state on a time scale of minutes, as evidenced by ^1H NMR and FTIR. [27] Explicit solvation effects, especially in the case of water, are known to significantly impact kinetics [42, 95, 96] and further investigation is needed to elucidate the fast kinetics of this reaction. Indeed, if the surrounding water plays a pivotal, catalytic role in the hydrolysis of LEDC to LEMC, it may be less favorable in the battery environments, where water is assumed to be at impurity concentrations. The presence of the DMSO solvent could also be of relevance (pure LEDC can only be synthesized as $\text{LEDC} \cdot 2 \text{DMSO}$ [27]) as LEDC is highly soluble in DMSO and solvation effects are known to impact reaction thermodynamics as well as kinetics.



(a) LEDC hydrolysis path from water, corresponding to path A in Figure 4.5

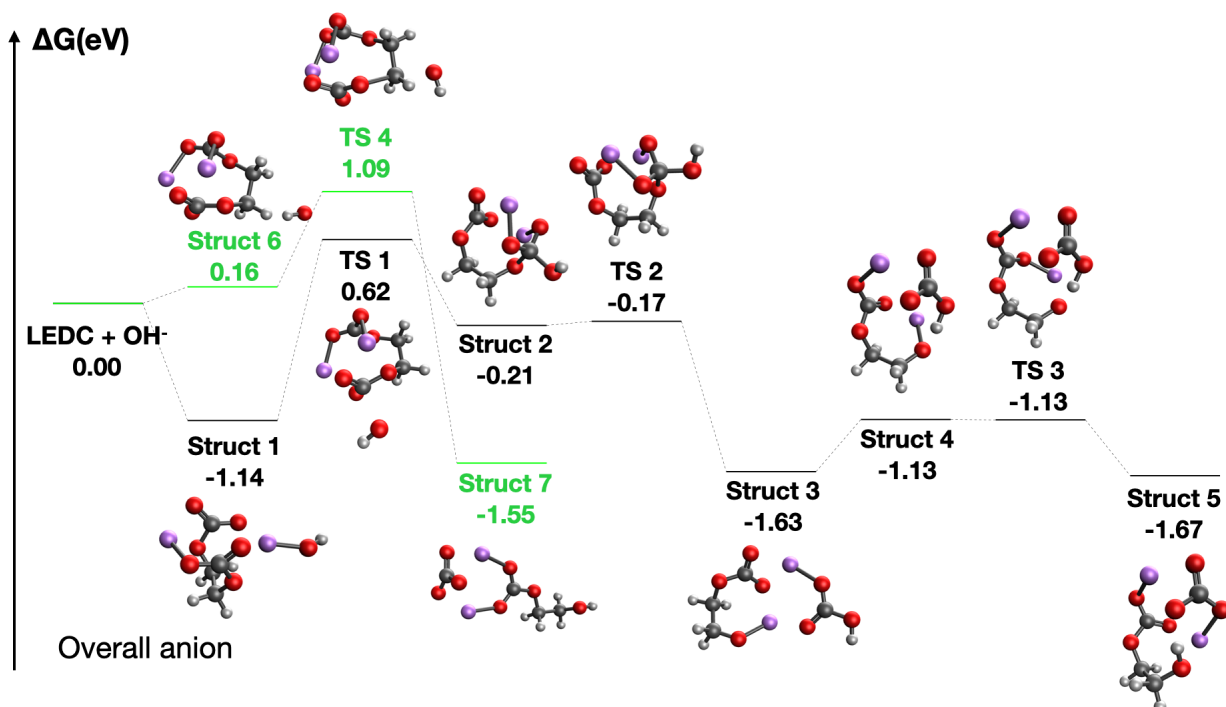
(b) LEDC hydrolysis path from OH^- , corresponding to path A in Figure 4.5

Figure 4.6: LEMC paths from DFT

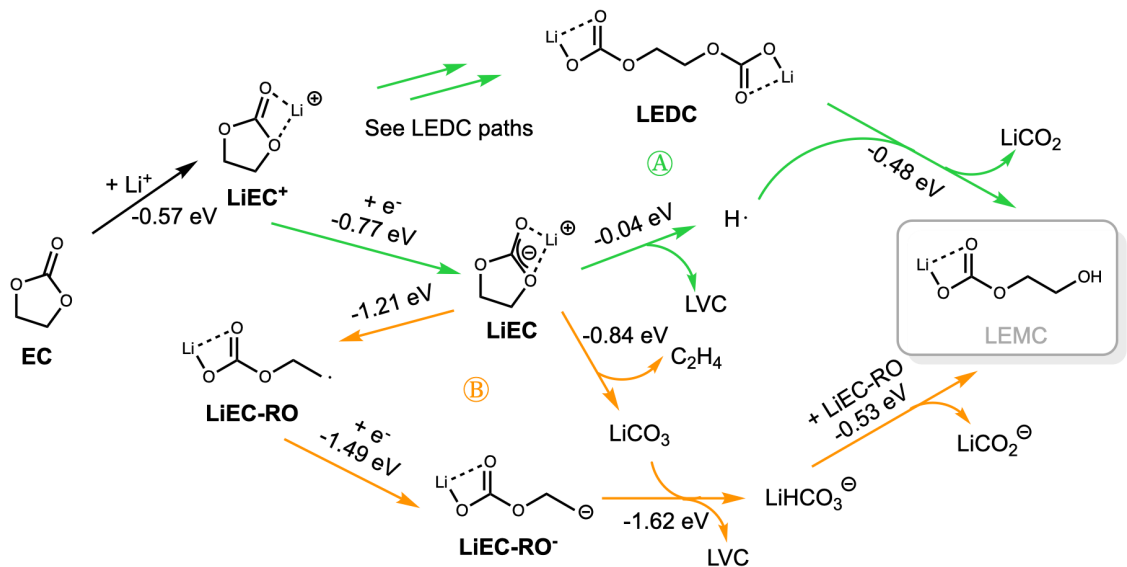


Figure 4.7: Paths to LEMC with no water participation identified by the 2-bond-change reaction network

Besides the LEDC mediated pathways, two additional representative recombinant molecules were identified in the formation of LEMC: LODEC (lithium 2-(2-oxido-1,3-dioxolan-2-yl)ethyl carbonate) in path B and LODC (lithium 2-oxido-1,3-dioxolan-2-yl carbonate) in path C in Figure 4.5. These molecules have not been proposed, to our knowledge, in the SEI literature, although LODEC was observed in an AIMD simulation of EC based electrolytes.[97] The reaction network suggest both recombinants to undergo hydrolysis and form LEMC. Examining the mechanistics of both reactions (Figure 4.11), we are unable to find a single-step hydrolysis reaction for both recombinants. Allowing for multistep reactions, in path B, LODEC and water transform to EC, LEC and LiOH crossing a 1.62 eV barrier and we surmise that EC and LiOH subsequently react according to the path in Figure 4.4 (a) to create LEMC. For path C, LODC initially decomposes to EC and Li_2CO_3 (which may exist as $\text{LiEC}^+ - \text{LiCO}_3^-$) and LiCO_3^- may facilitate water splitting and lead to the formation of a tetrahedral intermediate (Struct 4 in Figure 4.11 (b)) with the subsequent formation of LEMC. However, this is equivalent to the $\text{EC} + \text{LiOH} \rightarrow \text{LEMC}$ path (Figure 4.4 (a)), black path).

3.3 Reaction paths to LEMC without water

There are 8 reaction-network-identified paths starting from EC, Li^+ that do not involve water (see Appendix Sec 5.4 for the paths). However, these paths exhibit highly complex

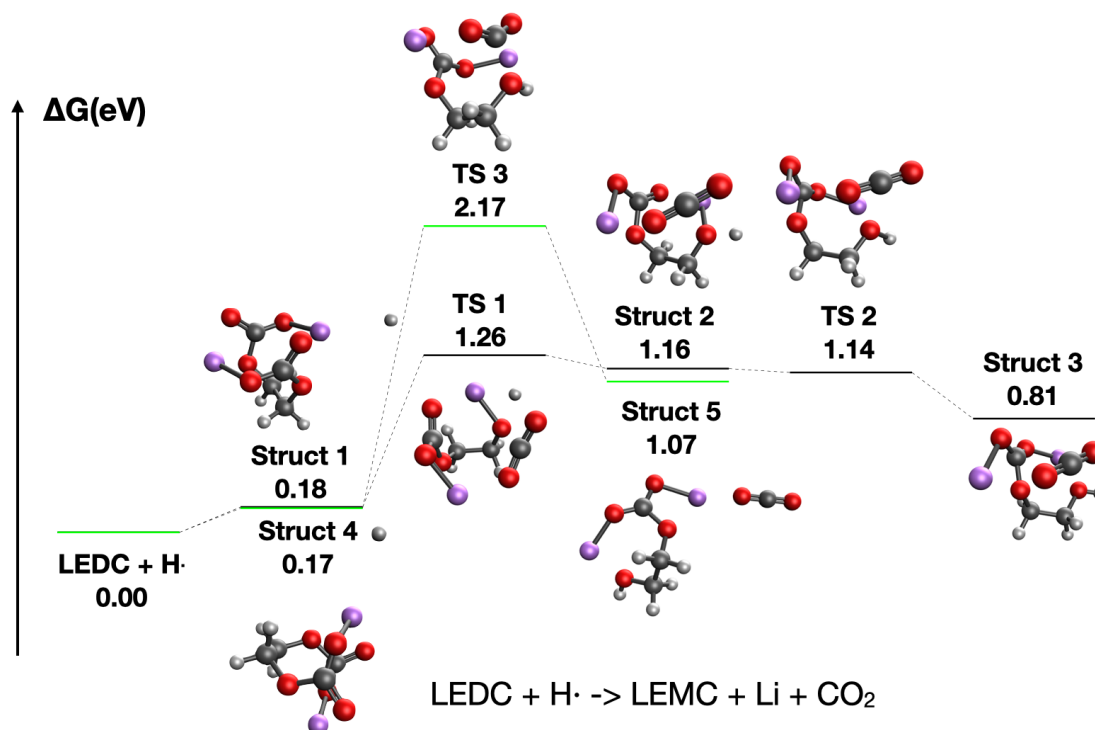


Figure 4.8: Radical paths, corresponding to key steps in path A in Figure 4.7

concerted reactions (involving 4 or 5 bond changes) connecting unstable open-shell species, with likely short lifetimes; those reactions were then verified to exhibit high barriers by DFT (see Appendix Sec 5.4). We note that reactions involving up to 5 bond changes were included in the reaction network to accommodate hydrolysis reaction mechanisms. Hence, to explore non-water containing reactions, another reaction network was constructed, limiting the concerted reactions to ≤ 2 bond changes and this reaction network consisted of 418,640 reactions. This procedure reveals some potentially feasible paths to LEMC starting with EC, Li^+ only (two representative paths are shown in Figure 4.7) and are discussed below. We note that these two paths are the two shortest paths from the 2-bond-change reaction network. The range of costs for the paths (where the cost of a path is defined as the sum of the softplus function values of reaction free energies for all the reactions involved in the path) in this smaller network is 2.39-3.12, which is significantly higher than that for the previous 5-bond-change reaction network (0.88-1.57), thus we considered it unnecessary to analyze the paths beyond the several top-ranking ones.

Path A in Figure 4.7 proceeds through an H-abstraction from LiEC, resulting in lithium vinyl carbonate (LVC) and a subsequent reaction between the H radical and LEDC creating

LEMC and LiCO_2 (neutral). This pathway is supported by previous findings by Shkrob et al., [98, 99] where VC was observed by NMR as a major product when EC undergoes electron radiolysis, and the formation mechanism was suggested to be an H-abstraction of EC by a secondary radical (e.g. ring-opened LiEC) creating $\text{EC}(-\text{H})\cdot$ (which is a H loss radical from EC), followed by a rapid radical disproportionation reaction. The existence of $\text{EC}(-\text{H})\cdot$ was confirmed by electron paramagnetic resonance (EPR) spectroscopy. Although radiolytic ionization creates both oxidation and reduction products, a detailed analysis of the relative yield between for the analogous radical and anion of propylene carbonate, $\text{PC}(-\text{H})\cdot$ and $\text{PC}\cdot^-$ and laser photoionization experiments (excluding the possibility of oxidation) enabled the authors in that work to conclude that $\text{PC}(-\text{H})\cdot$ radicals can indeed form through reduction chemistry. Thus a similar scenario should be expected for $\text{EC}(-\text{H})\cdot$ radicals. We note that, while the reaction network nicely captures the essence of this reaction series, it oversimplifies the first step to be $\text{LiEC} \rightarrow \text{LVC} + \text{H}\cdot$, as the process (according to ref. [98]) is presumably $\text{LiEC} \rightarrow \text{o-LiEC}$, then $\text{o-LiEC} + \text{EC} \rightarrow \text{LEC} + \text{EC}(-\text{H})\cdot$ and finally $\text{EC}(-\text{H})\cdot + \text{Li}^+ \rightarrow \text{LVC}$. The reaction barrier for the second step ($\text{LEDC} + \text{H}\cdot \rightarrow \text{LEMC} + \text{Li} + \text{CO}_2$), is calculated to be 1.26 eV (Figure 4.8). However, $\text{H}\cdot$ is not necessarily involved; we expect $\text{H}\cdot$ to be unstable and will react rapidly with its neighboring environment. The favorable thermodynamics of this path suggests that other radical species and/or the Li surface could facilitate a similar bond breakage (a homolytic cleavage of the $(\text{CH}_2)\text{O}-\text{C}(=\text{O})\text{OLi}$ bond) in LEDC, resulting in a $(\text{CH}_2)\text{O}\cdot$ moiety which would immediately abstract another $\text{H}\cdot$ (from for example EC) to form LEMC. Soto et al. reported LEDC decomposition in AIMD simulations, including CO_2^{2-} and CO_3^{2-} detachment from LEDC, in the absence or presence of radicals in the solution. [100] In fact, the C-O bond breakage of LEDC on the Li (100) surface, creating CO_2^- bound to the surface, was calculated from DFT (PBE functional) as $\Delta E^\ddagger = 0.22$ eV and $\Delta E = -1.94$ eV. [68] The resulting fragment may be further decomposed by breaking the $-\text{CH}_2-\text{O}$ bond, however with a high associated barrier (1.8 eV). Hence it is reasonable to assume the possibility of $\text{LiOC}(\text{O})\text{OCH}_2\text{OLi}$ undergoing H-abstraction reactions to create LEMC, although further investigations are recommended regarding the $\text{H}\cdot$ or H^+ source and the associated reaction barriers.

For path B in Figure 4.7, we were unable to locate the TS for the last step ($\text{LiHCO}_3^- + \text{o-LiEC} \rightarrow \text{LEMC} + \text{LiCO}_2^-$) with DFT. An overall neutral TS was identified, corresponding to $\text{LiHCO}_3 + \text{o-LiEC} \rightarrow \text{LEMC} + \text{LiCO}_2$ (Figure 4.12), however the step exhibits a high barrier of 2.75 eV and hence, under the current modeling conditions, presumably not kinetically feasible.

Finally, we comment on what the presented work means in the likelihood of LEMC being present in the early stage Li metal or graphite SEI, instead of LEDC. All identified, feasible paths to LEMC (the EC hydrolysis path, the hydride path, the LEDC hydrolysis path, the path involving $(\text{CH}_2)\text{O}-\text{C}(=\text{O})\text{OLi}$ bond breakage in LEDC) either required water and/or LEDC. Indeed, the currently proposed reaction mechanisms for forming LEMC in the literature all require water participation. [27, 84] Hence, assuming trace water impurities

in commercial electrolytes (~ 10 ppm of water in rigorously dried electrolyte) would not produce detectable amount of LEMC. However, it is possible that some water may originate from the cross-talk between the positive and negative electrodes, e.g. from the oxidation of EC at the cathode to produce water and CO_2 . [84, 101, 102] Our only non-water containing path is effectively a conversion of LEDC into LEMC, and as such requires LEDC to form first.

IV Conclusion

In this work, we developed a graph reaction network based on a database of systematically generated DFT-calculated molecular fragments and recombinants, specifically to explore reaction mechanisms for the formation of LEMC, relevant for the interface formation in lithium-ion batteries. The network allows for reactivity under a grand canonical formalism, treating all species as open. Our constructed reaction network is very dense, consisting of over 570 species and ~ 9 M reactions; this rich data allows us to embrace an unprecedented number of electrochemical reaction possibilities and predict reaction mechanisms without explicit knowledge of intermediates. A cost function based on the Gibbs free energy of the pathways was employed to rank the pathways, which were also manually inspected. Predominant pathways were further evaluated for mechanistic and kinetic feasibility using DFT. All relevant molecules, including our target molecule LEMC, were generated from fragmentation/recombination of the starting materials (EC, Li^+ and H_2O) with no imposed bias towards known SEI components. We note that our methodology can in principle be applied to any electrochemical, reactive system, although caution should be applied to avoid the rapid growth of combinatorial complexity. To prune the combinatorial complexity of possible recombinant molecules, machine learning was employed to predict reaction free energies to determine the feasibility of recombinant molecules. However, we emphasize that, in the present work, no other limitations, e.g. *a priori* identification of reactive sites, are imposed except for the elimination of recombinants with irregular valence.

Exploring the formation of LEMC with any amount of EC, H_2O , Li^+ , and electrons at the Li metal chemical potential, we successfully recovered - automatically - the two major literature-proposed mechanisms, namely EC and LEDC hydrolysis. Based on detailed DFT calculations of the reaction barriers, the direct EC hydrolysis mechanism was identified as the most kinetically feasible under the current model conditions, with the further suggestion of being significantly accelerated under basic conditions. In addition to these known paths, we also identify a hydride-generating mechanism and a radical mechanism through the $(\text{CH}_2)\text{O}-\text{C}(=\text{O})\text{OLi}$ bond breakage in LEDC, although those two mechanisms are more likely to be relevant on/near the Li metal surface.

Importantly, we find that nearly all feasible routes to LEMC, identified by the reaction network, require the participation of water, which is consistent with previously proposed

mechanisms.[27, 84] Since the formation of LEDC does not require water, our results suggest that decreasing or increasing the water concentration at the anode offer control of the SEI LEMC/LEDC ratio. Indeed, trace water impurities in rigorously dried electrolytes (~ 10 ppm) should not be able to produce detectable amounts of LEMC. However, recent works, [84, 101] show that water can originate from EC oxidation at the cathode to produce water and CO_2 .

We note that our predicted pathways to LEMC formation assume model conditions, such as the absence of an explicit surface as well as explicit solvation effects; both of which can contribute significantly to the mechanistics of the reactions. Despite these caveats, this work represents the first exploration of reaction pathways to LEMC based on a general, graph network of first-principles thermodynamics. The predicted pathways reproduce both existing, literature-suggested pathways as well as novel ones. Such predictions are critical for understanding the early-stage formation of the SEI and may be leveraged to develop next-generation electrolytes/additives for Li-ion batteries. Our work illustrates that our methodology can be used for discovering novel reaction mechanisms and demonstrate the potential of the chemically-consistent reaction network formalism for automated probing of complex, out-of-equilibrium electrochemical reaction mechanisms.

V Appendix

5.1 Details on the MILP algorithm

The MILP algorithm is slightly modified from a previous work on reaction atom-mapping (notations were kept the same). [64] Stereochemistry and bond orders are not considered because there are no stereocenters in the molecules in the database and bond orders cannot be unambiguously determined for most of the molecules.

$$\begin{aligned}
 \zeta &= \min \sum_{(i,j) \in B_R} (1 - \sum_{(k,l) \in B_P} \alpha_{ijkl}) + \sum_{(k,l) \in B_P} (1 - \sum_{(i,j) \in B_R} \alpha_{ijkl}) \\
 s.t. \quad &\sum_{k \in A} y_{ik} = 1 \quad \forall i \in A \\
 &\sum_{i \in A} y_{ik} = 1 \quad \forall k \in A \\
 &y_{ik} = 0 \quad \forall i, k \in A : T_R^i \neq T_P^k \\
 &\alpha_{ijkl} \leq y_{ik} + y_{il} \quad \forall (i,j) \in B_R \quad \forall (k,l) \in B_P \\
 &\alpha_{ijkl} \leq y_{jk} + y_{jl} \quad \forall (i,j) \in B_R \quad \forall (k,l) \in B_P
 \end{aligned}$$

where

$A = \{1, 2, \dots, n\}$: atom indices involved in the reaction

$B_R = \{(i, j) : i, j \in A, i < j, i \text{ and } j \text{ bonded in reactants}\}$

$B_P = \{(k, l) : k, l \in A, k < l, k \text{ and } l \text{ bonded in products}\}$

$T_R^i \forall i \in A$: type of atom i in reactants

$T_P^k \forall i \in A$: type of atom k in products

$$y_{ik} = \begin{cases} 1, & \text{if atom } i \text{ in reactants is mapped to atom } k \text{ in products} \\ 0, & \text{otherwise} \end{cases}$$

$\forall i, k \in A$

$$\alpha_{ijkl} = \begin{cases} 1, & \text{if bond } (i, j) \text{ is mapped to bond } (k, l) \\ 0, & \text{otherwise} \end{cases}$$

$\forall (i, j) \in B_R$;

$\forall (k, l) \in B_P$

5.2 Details on the water splitting reaction on the Li surface

The Li surface was derived from a body-centered cubic (bcc) Li-bulk structure taken from the Materials Project database[103, 104] with lattice parameter $a = 3.51 \text{ \AA}$, which shows good agreement with previous experimental ($3.49 \text{ \AA}$), [105] and theoretical results. [106] The lowest surface energy facet (100) was used. [33, 107] The cluster consists of 2 layers of Li with 16 atoms in each layer (see Figure 4.9). The bottom layer and the edge of the top layer were frozen during the geometry optimization and TS calculations to help maintain the bulk bcc structure. One imaginary frequency was confirmed for the block of the hessian matrix corresponding to the active atoms (Li atoms that are not frozen, water and the EC molecule) in the TS structure. The Gibbs free energies of the stationary structures were obtained by adding the vibrational free energy to the electronic energy (there are no translational or rotational degrees of freedom).

The TS structure for the water splitting reaction on an explicit Li surface is shown in Figure 4.9, and the barriers at various levels of theory are given in Table 4.1).

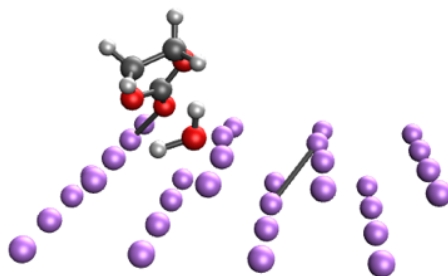


Figure 4.9: Transition state structure for water splitting on a Li cluster

Table 4.1: The reaction barrier for water splitting on a Li cluster

	free energy barrier (eV)
ω b97x-D/6-31G/gas phase	-0.067
ω b97x-D/def2-SVPD/gas phase	-0.016 ^a
ω b97x-D/def2-SVPD/PCM ^b	0.027 ^a
ω b97x-V/def2-TZVPPD/SMD ^b	0.041 ^a

^a Single point correction on the optimized geometry from ω b97x-D/6-31G/gas phase level of theory.

^b The EC parameters for implicit solvation was used, as described in the computational methods in Sec. 2.3

5.3 Supplementary free energy diagrams

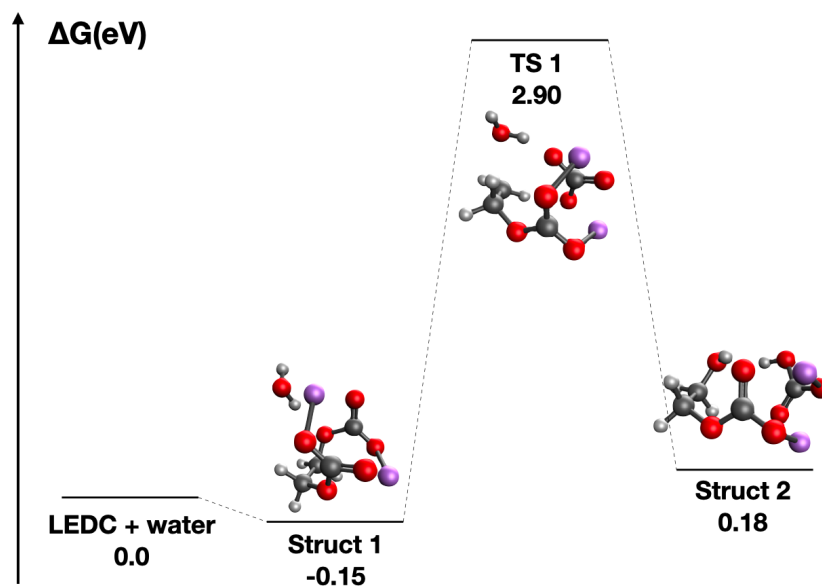
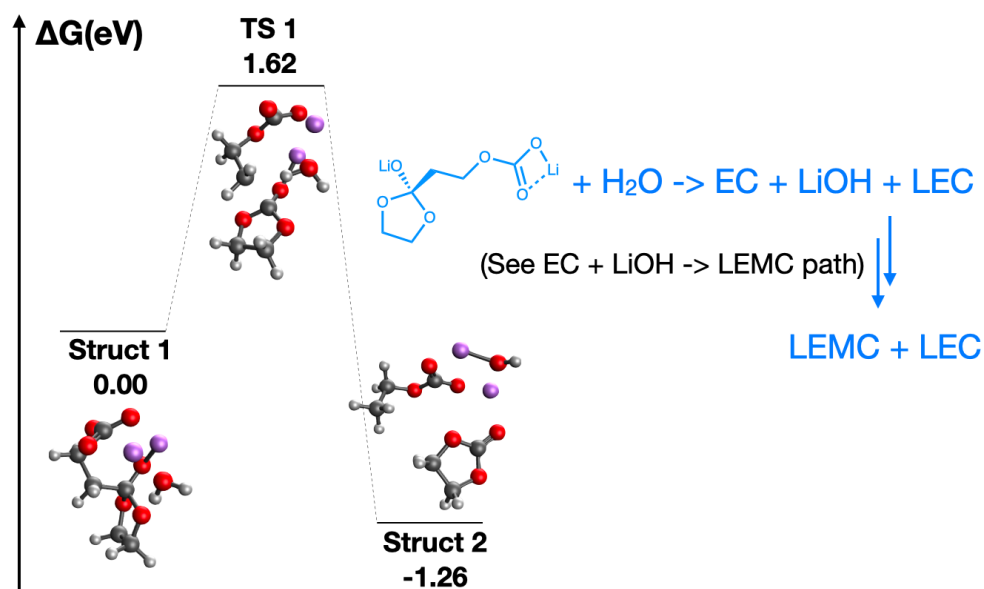
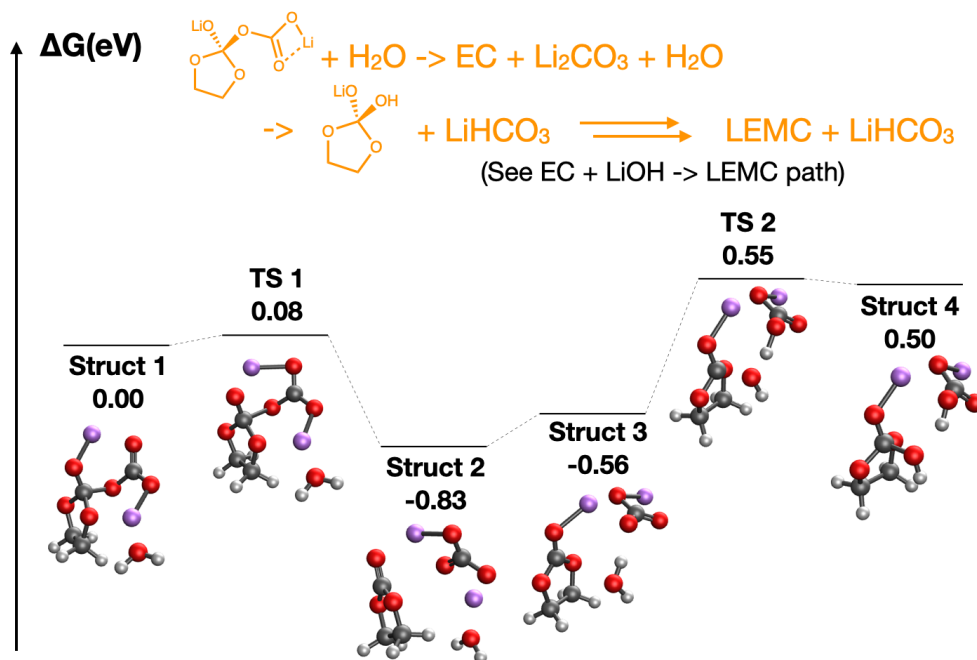


Figure 4.10: One-step LEDC hydrolysis path



(a) LEMC paths involving LODEC, corresponding to path B in Figure 4.5



(b) LEMC paths involving LODC, corresponding to path C in Figure 4.5

Figure 4.11: LEMC paths involving recombinant molecules

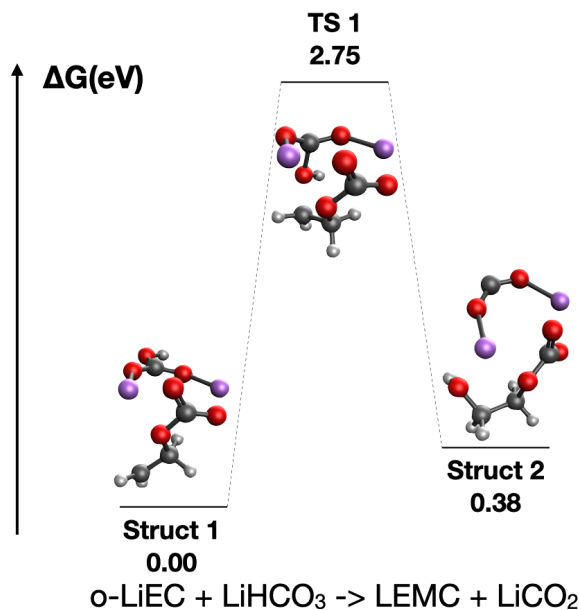


Figure 4.12: A LEMC path without water participation, corresponding to the key step in path B in Figure 4.7

5.4 Kinetically infeasible LEMC paths without water participation

Below we show 8 non-water paths identified by the 5-bond-change reaction network, and the DFT energy diagrams for the 3 major reactions involved in those paths, i.e. each path contains at least one of the 3 reactions. All 3 reactions exhibit high rate-limiting barriers (> 2 eV) from the paths we were able to find at a lower level of theory. Although we cannot prove these paths are the lowest-barrier ones connecting the starting materials and the products, this result strongly suggest they are unlikely to be kinetically feasible.

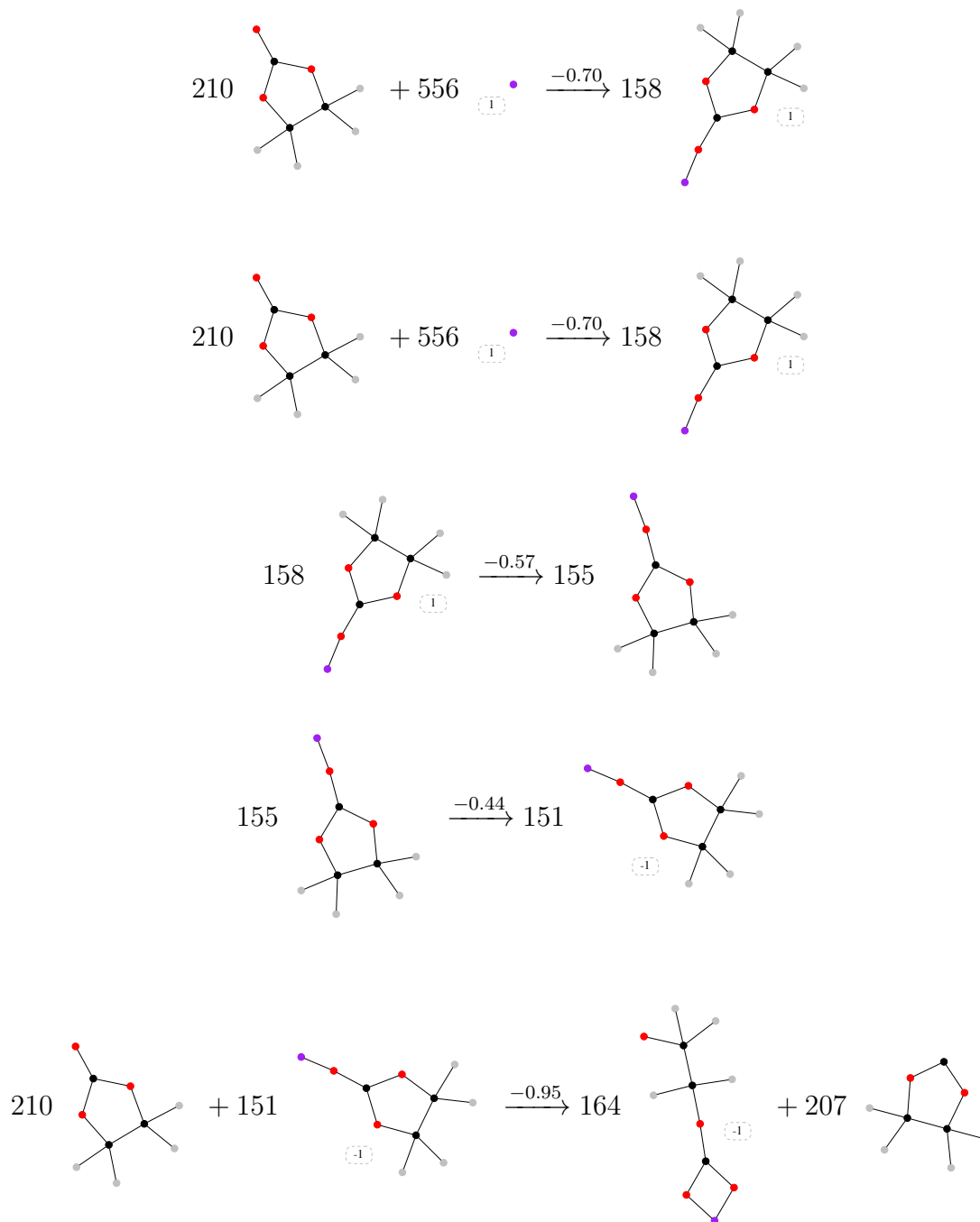
For the energy diagrams for reaction 1 (Figure 4.13 (a)) and reaction 2 (Figure 4.13 (b)), we used $\omega\text{B97X-D}[76]/\text{def2-SVP}[72]$ level of theory with the C-PCM implicit solvation model [77]. For reaction 3 (Figure 4.13 (c)), we encountered some numerical issues with the frequency calculations when optimizing the structures in C-PCM (some imaginary frequencies are artificial but hard to cancel), and thus switched to single point implicit solvation corrections on the structures optimized in gas phase at the same level of theory.

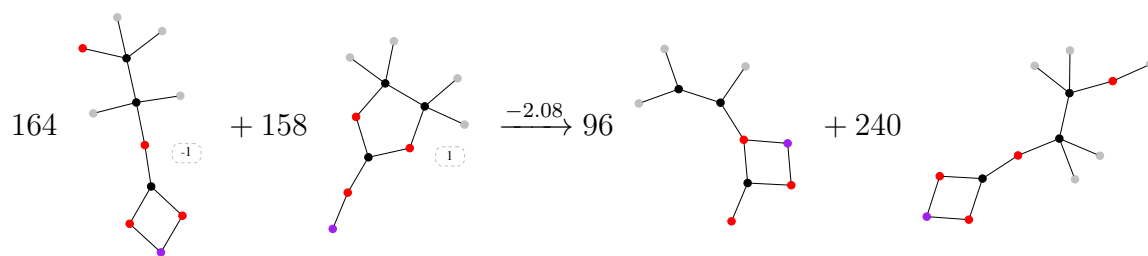
Molecules are drawn as graphs and different colors on the nodes correspond to different atomic species as follows:

- Red: oxygen
- Black: carbon

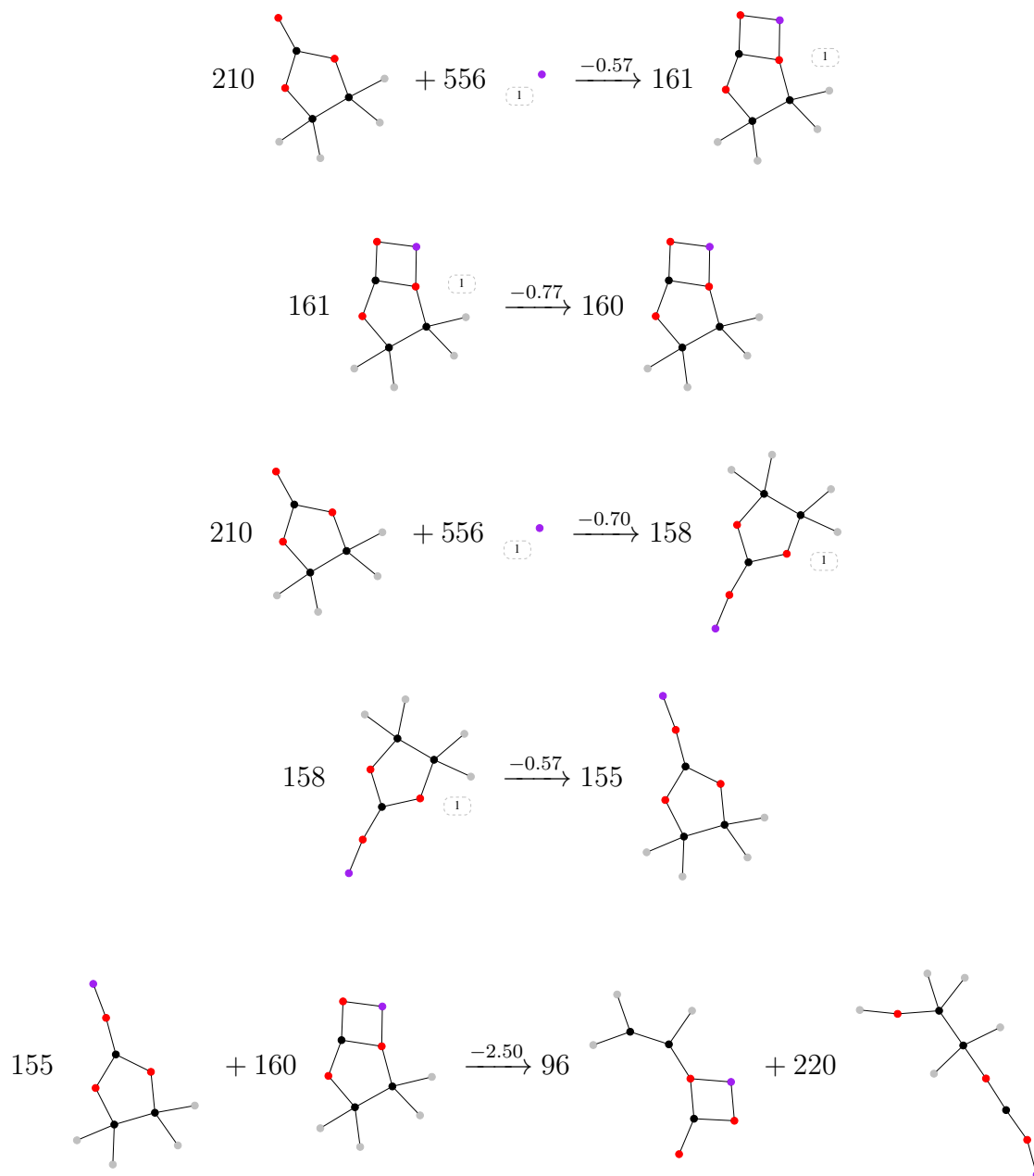
Grey: hydrogen
Pink: lithium.

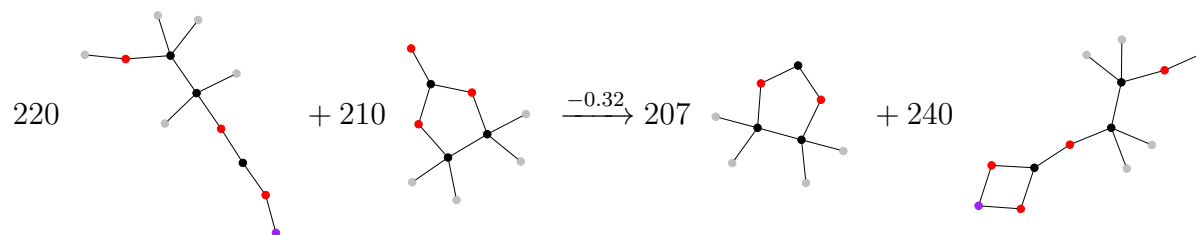
Non-water pathway 1 (ranking 179):



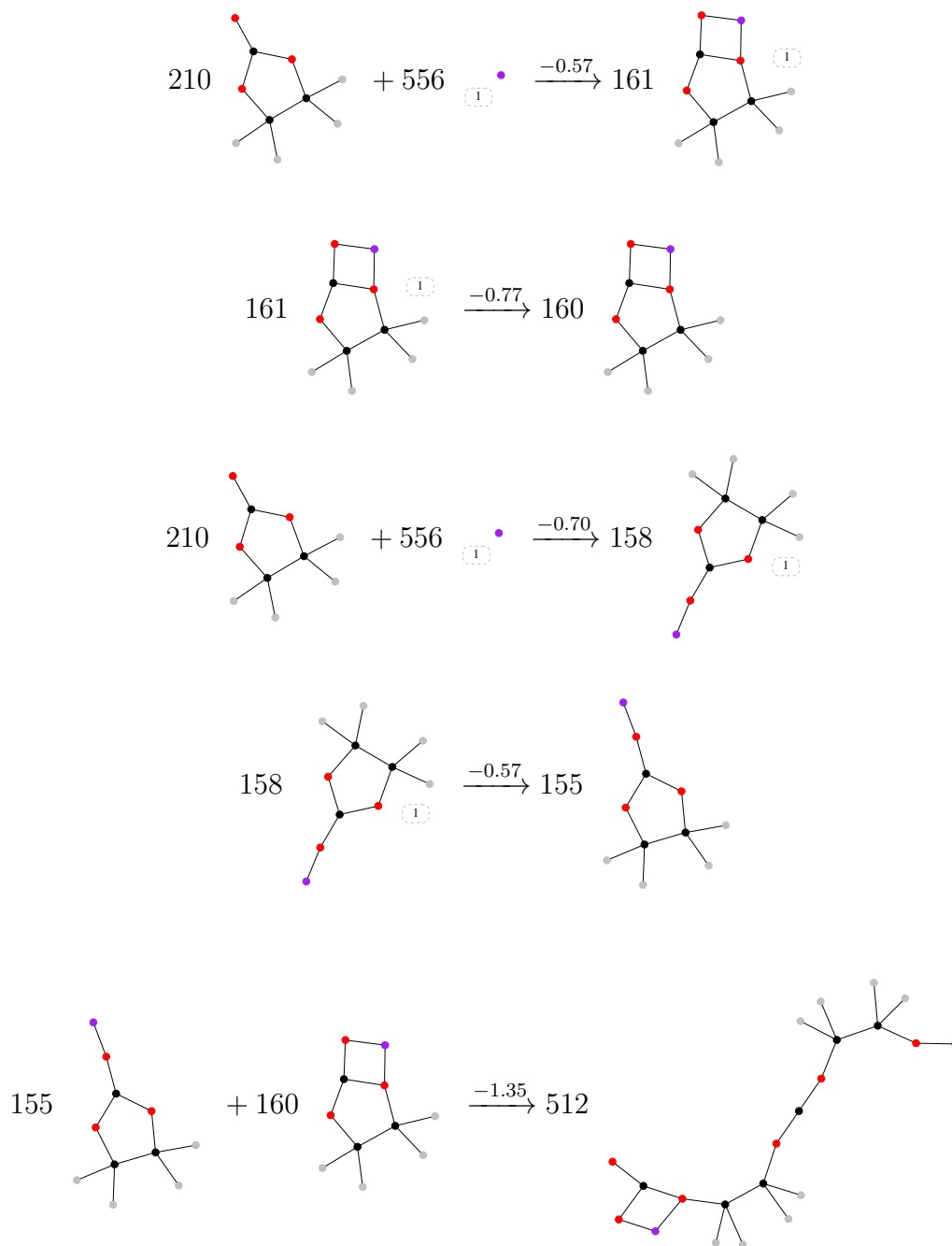


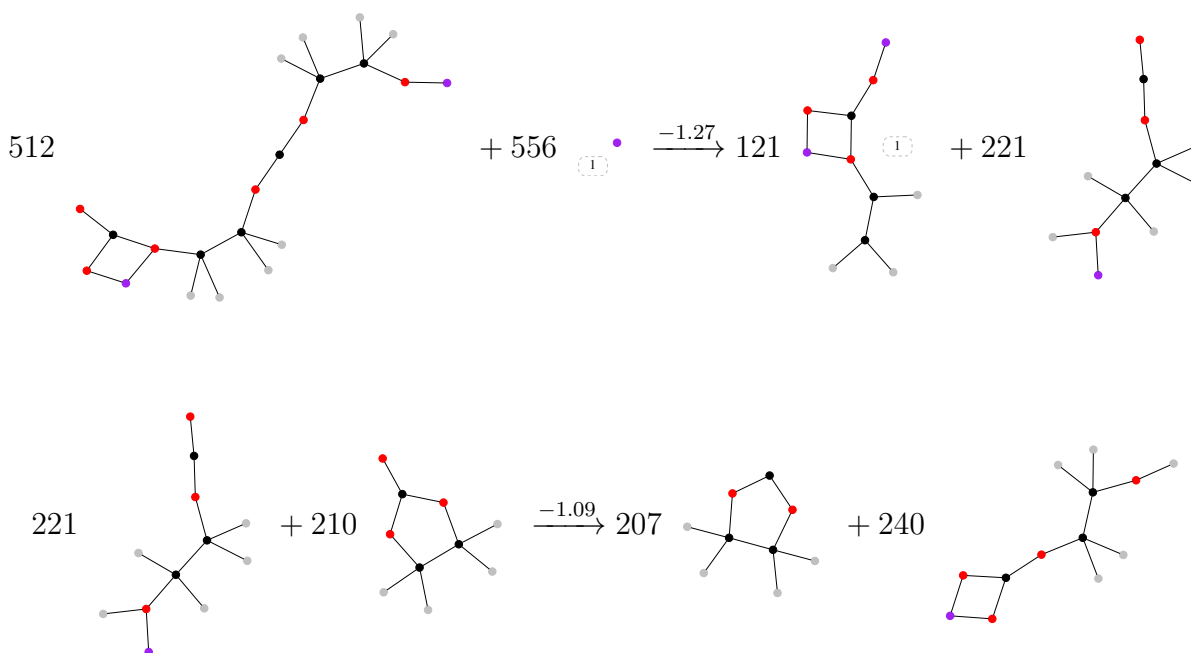
Non-water pathway 2 (ranking 247):



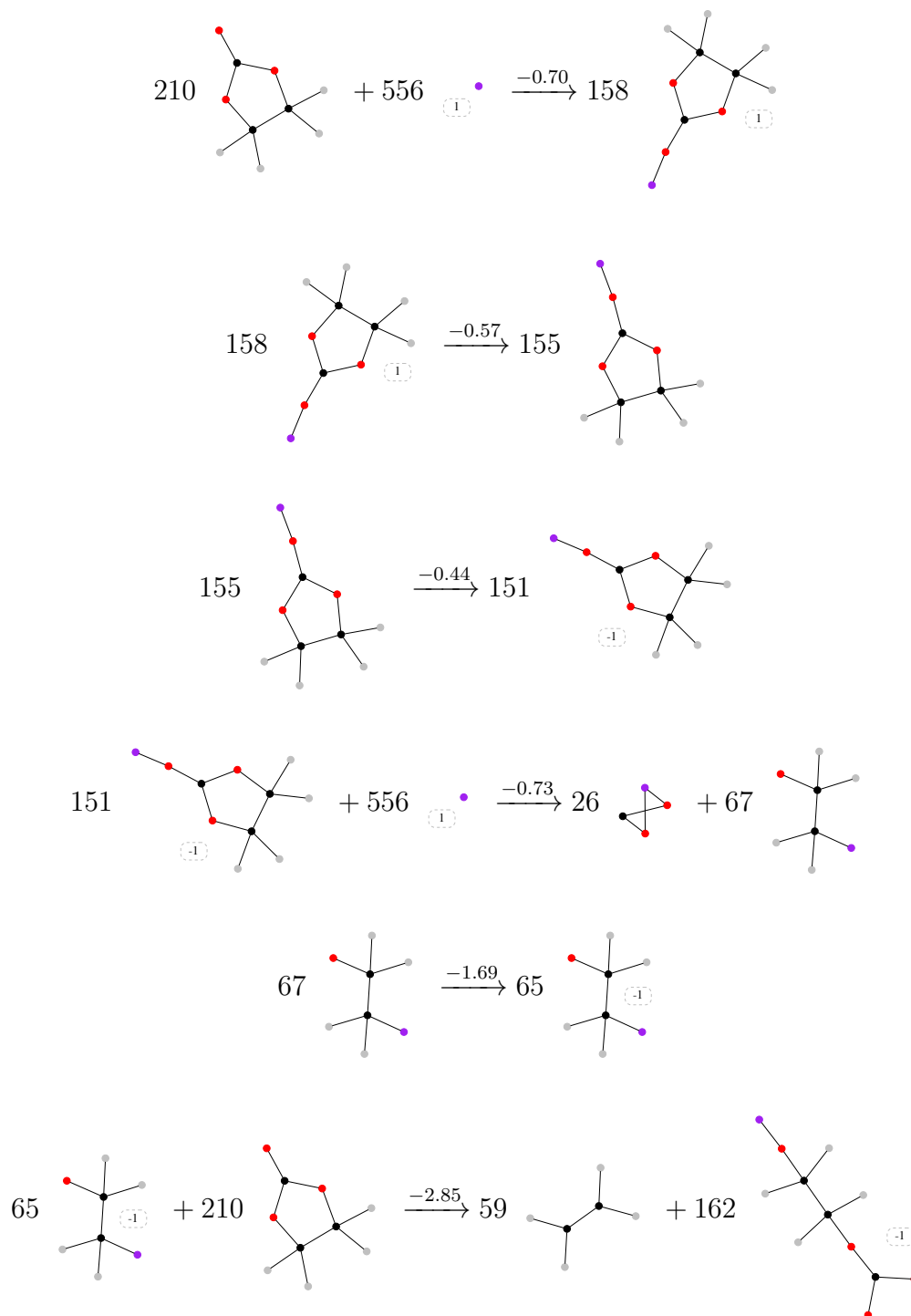


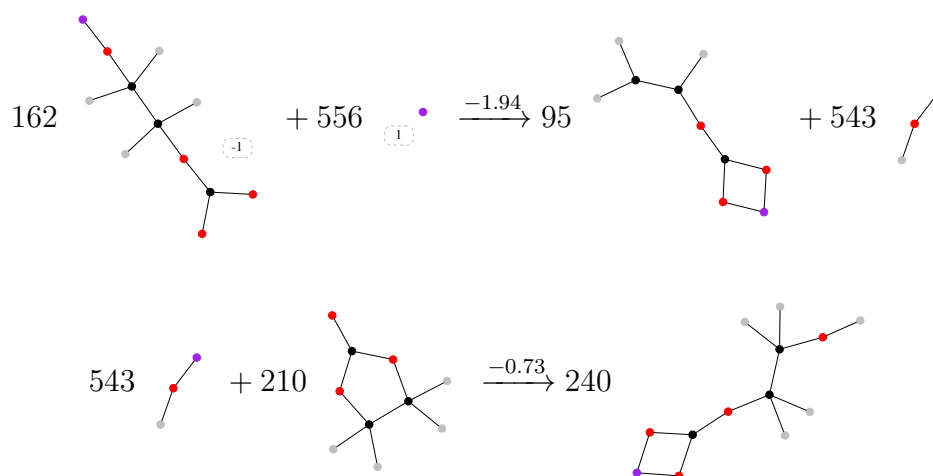
Non-water pathway 3 (ranking 309):



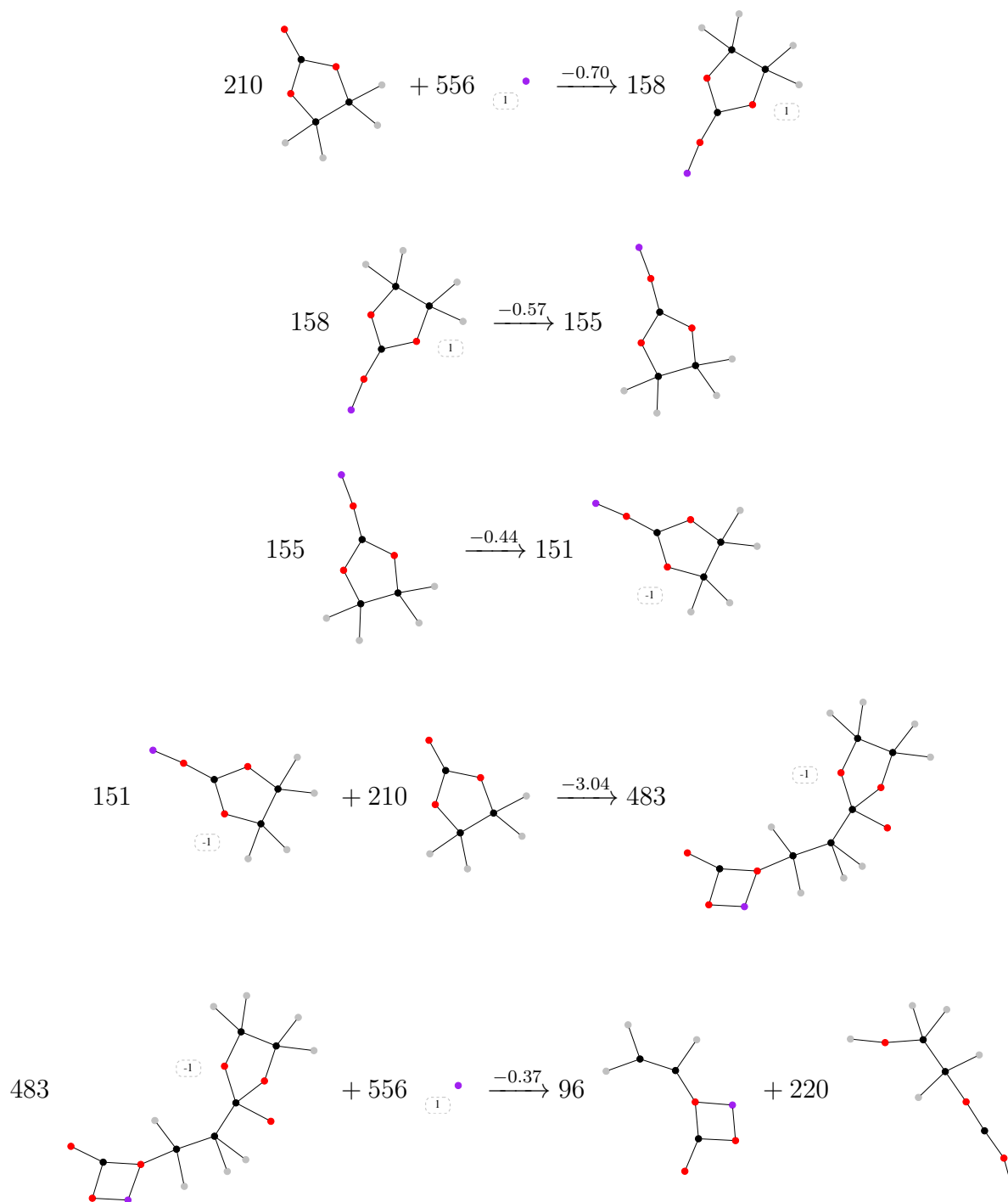


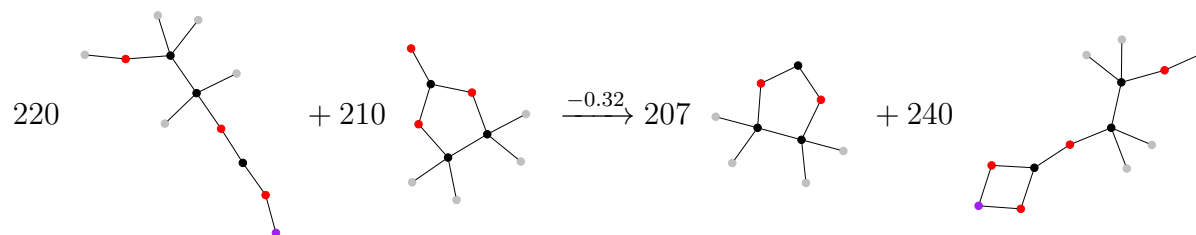
Non-water pathway 4 (ranking 343):



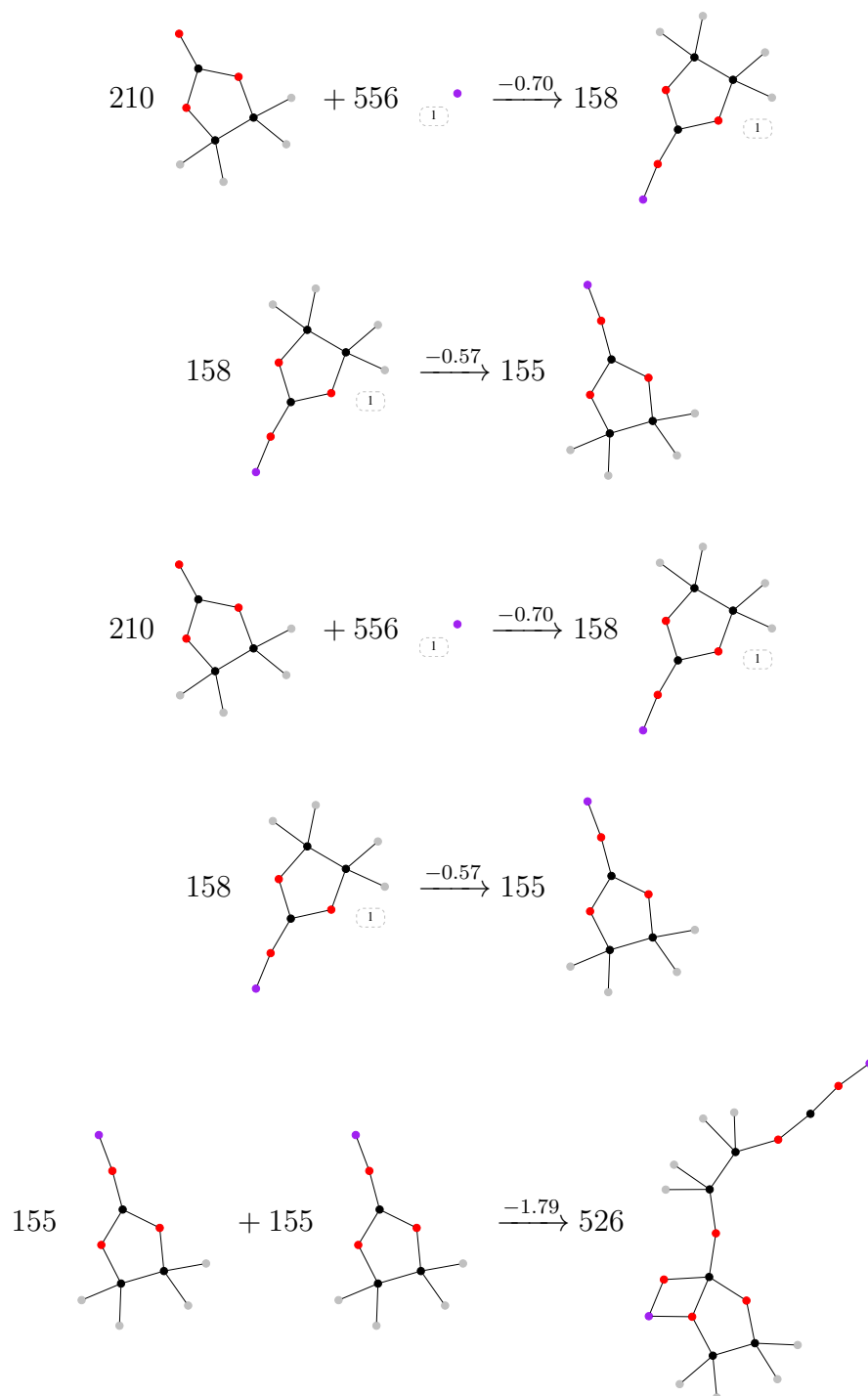


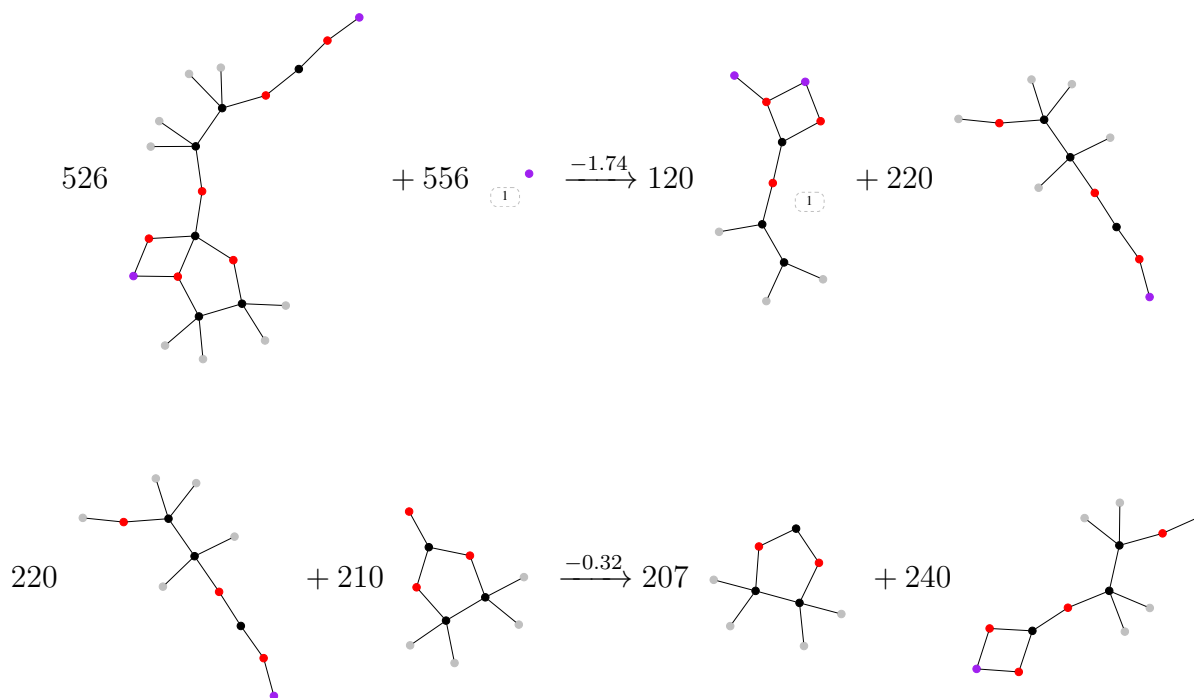
Non-water pathway 5 (ranking 351):



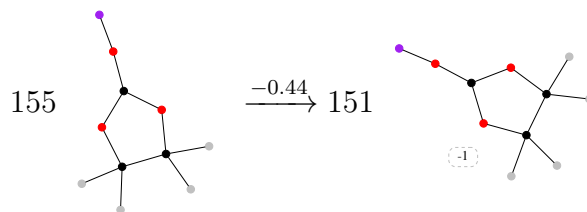
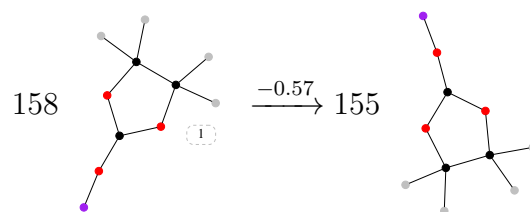
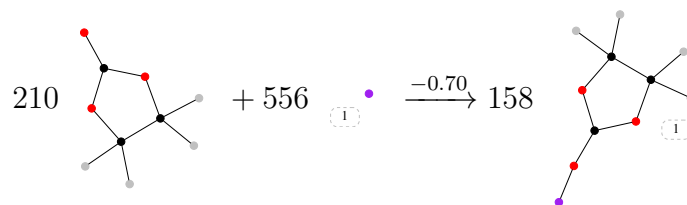
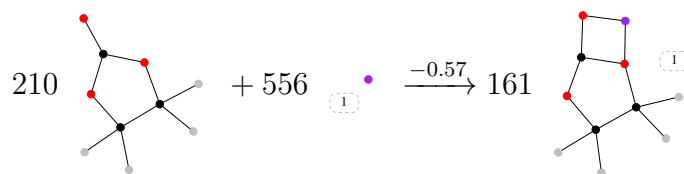
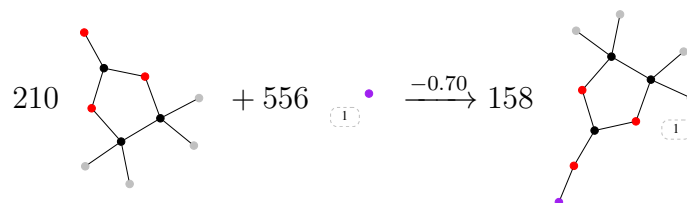


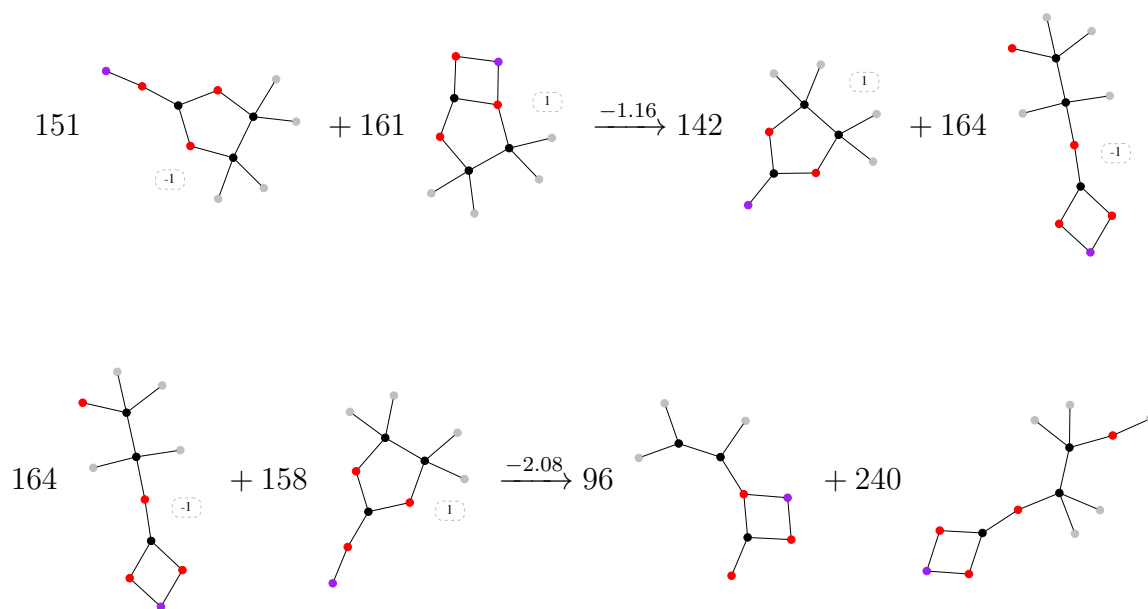
Non-water pathway 6 (ranking 391):



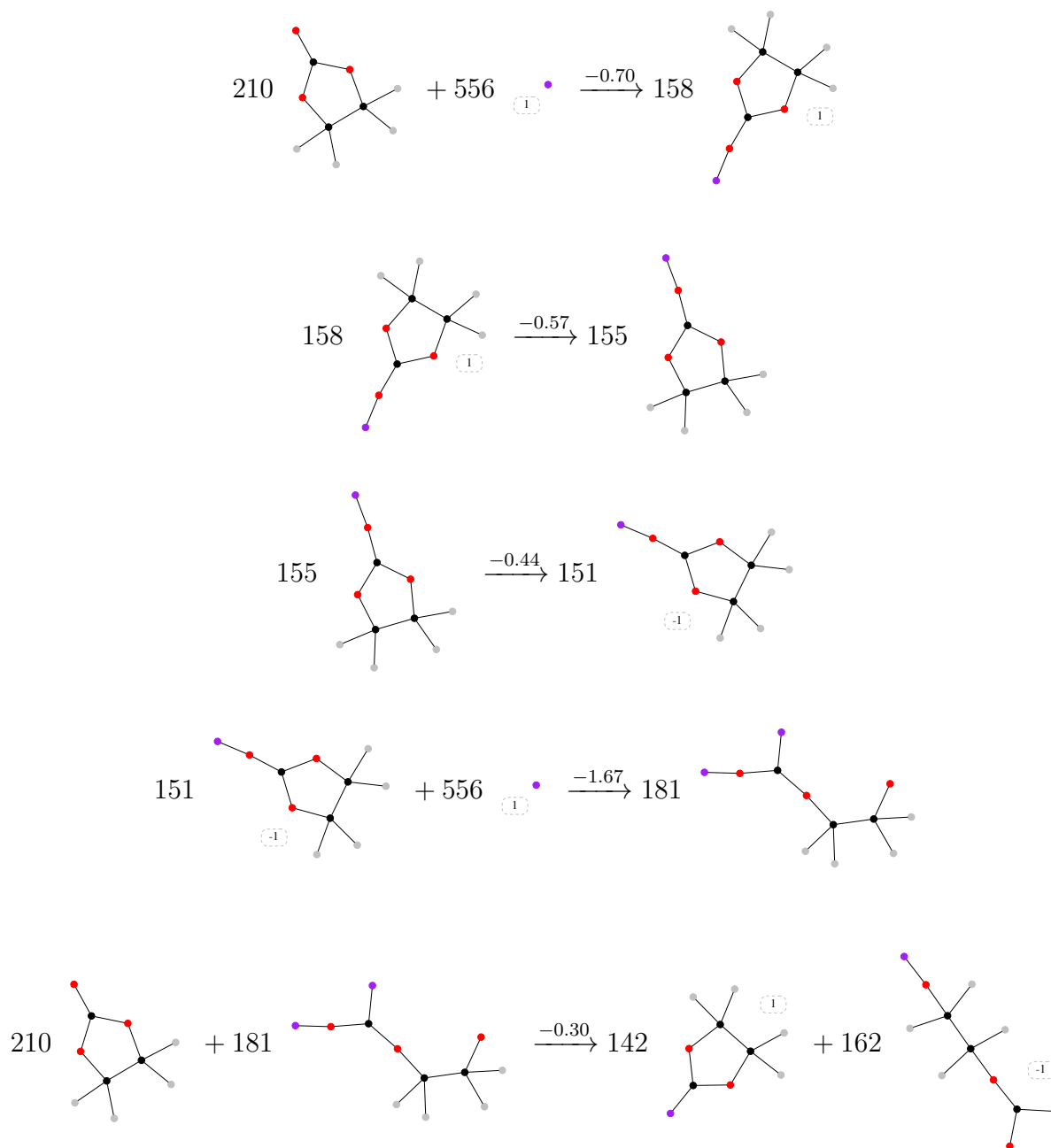


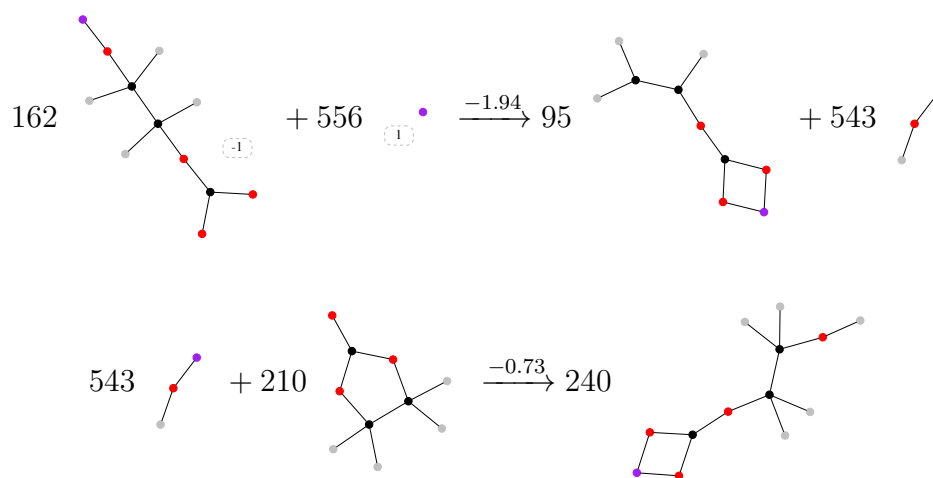
Non-water pathway 7 (ranking 401):

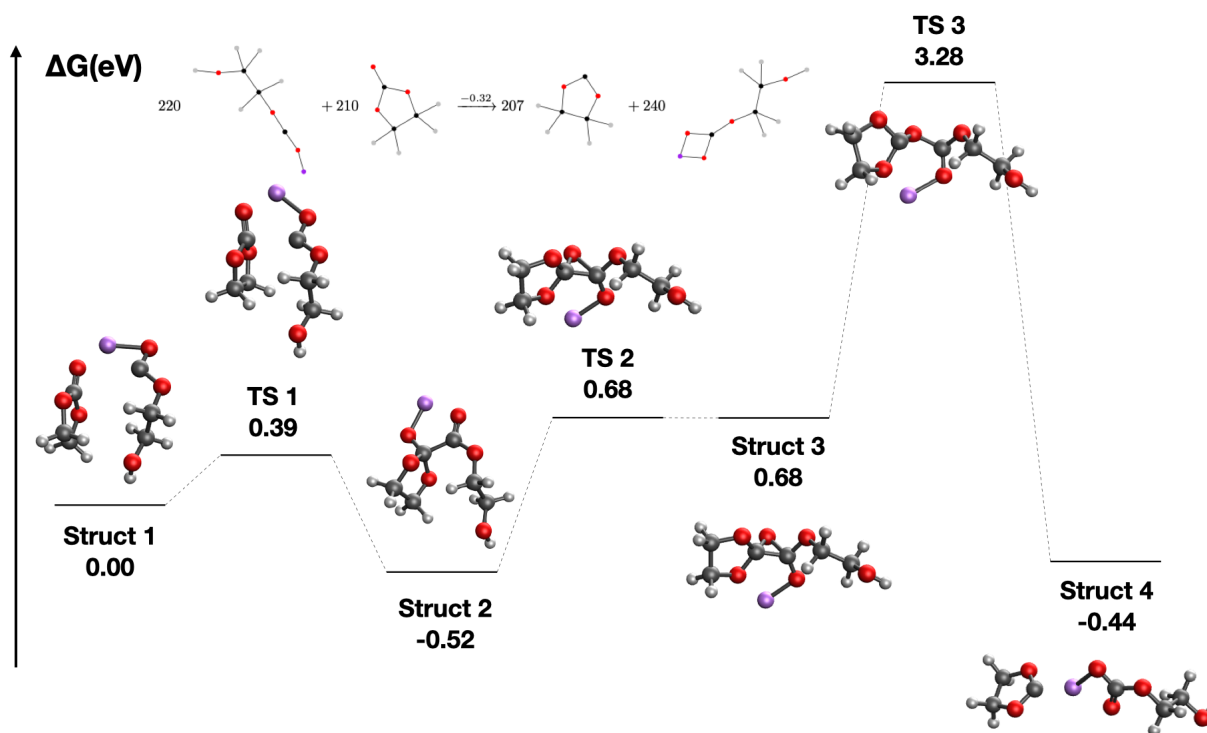




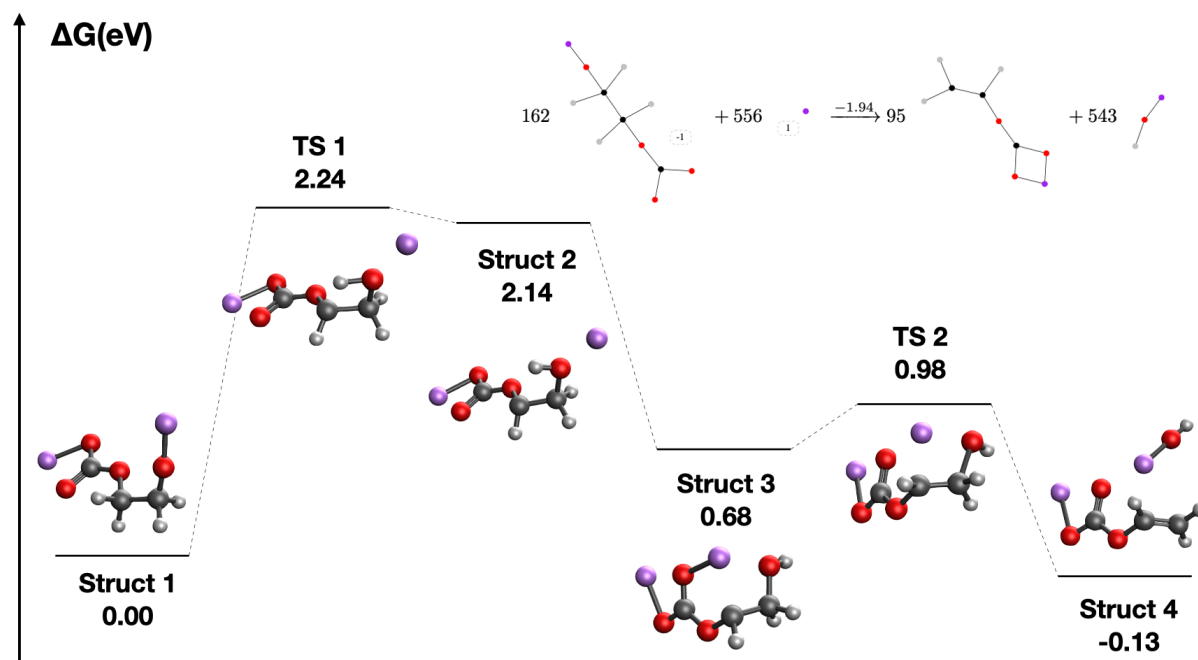
Non-water pathway 8 (ranking 418):







(a) Reaction 1



(b) Reaction 2

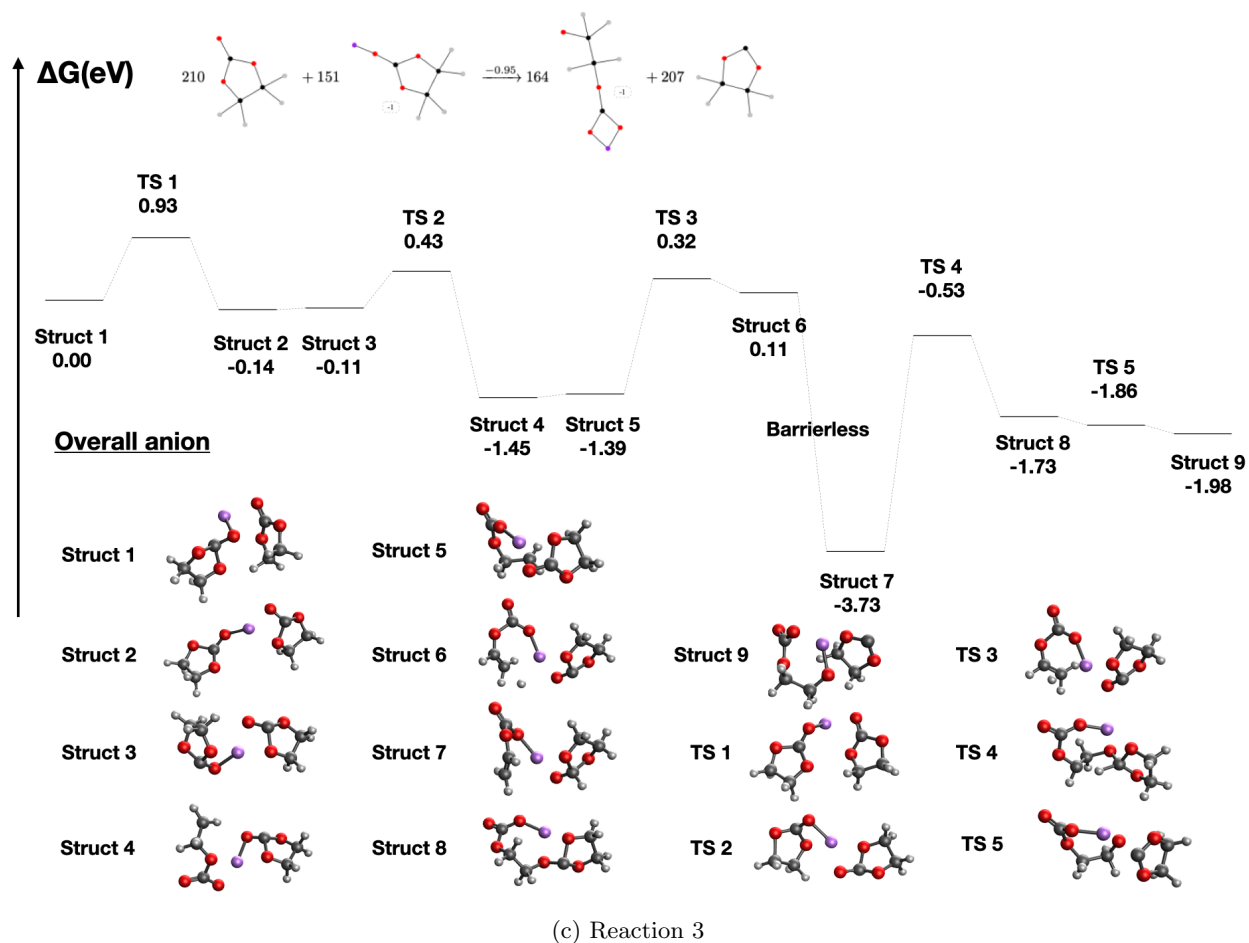


Figure 4.13: Energy diagrams for the 3 major reactions involved in the 8 non-water LEMC paths in the 5-bond-change reaction network

5.5 Effects of explicit solvation on the reaction paths

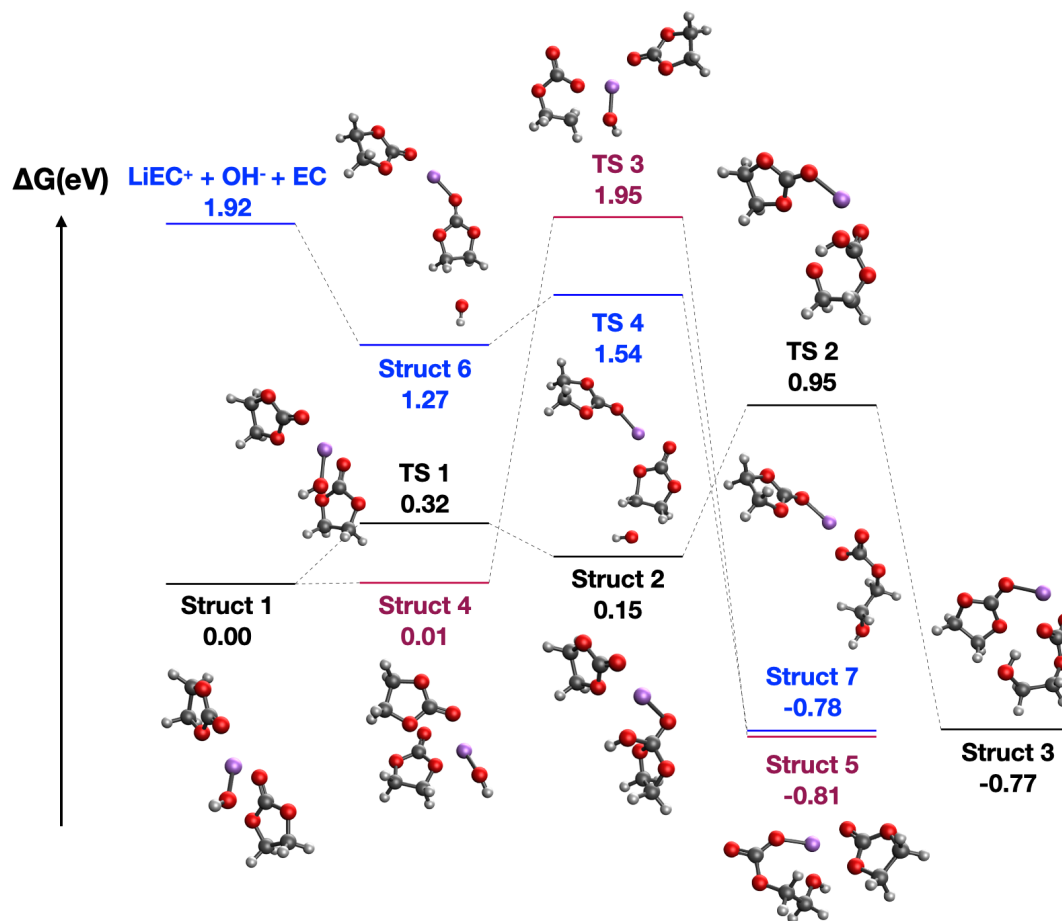


Figure 4.14: The EC hydrolysis paths with one explicit solvent molecule (EC)

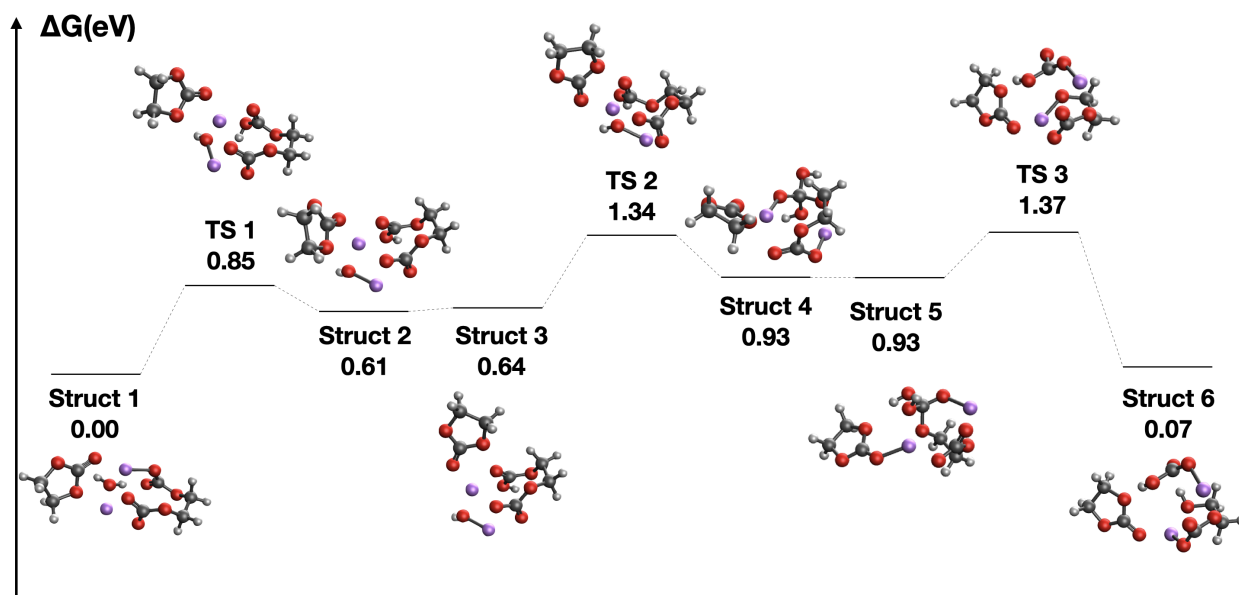


Figure 4.15: The LEDC hydrolysis path with one explicit solvent molecule (EC)

In Figure 4.14 is shown the EC hydrolysis path with an explicit solvent molecule (EC), which can be compared to the right part of Figure 4.4 (a). The rate-limiting barrier for the black path is 0.95 eV with an explicit EC molecule (Struct 1 $\rightarrow$ TS 2 in Figure 4.14) and 1.33 eV without an explicit EC molecule (Struct 3 $\rightarrow$ TS 4 in Figure 4.4 (a)). The rate-limiting barrier for the purple path is 1.95 eV with an explicit EC molecule (Struct 1 $\rightarrow$ TS 3 in Figure 4.14) and 2.24 eV without an explicit EC molecule (Struct 3 $\rightarrow$ TS 5 in Figure 4.4 (a)). The rate-limiting barrier for the blue path is 0.27 eV with an explicit EC molecule (Struct 6 $\rightarrow$ TS 4 in Figure 4.14) and 0.24 eV without an explicit EC molecule (Struct 7 $\rightarrow$ TS 6 in Figure 4.4 (a)).

In Figure 4.15 is shown the LEDC hydrolysis path with an explicit solvent (EC), which can be compared to Figure 4.6 (a). There are subtle differences in the atomic arrangements in the intermediate or TS structures, but the overall rate-limiting barrier (1.48 eV vs 1.37 eV) did not change appreciably.

Overall, one explicit solvent molecule (EC) tend to lower the barrier if it is close to the reaction center (as seen in the black and blue path in the EC hydrolysis paths and the LEDC hydrolysis path) and will not change the barrier appreciably if it is rather far away from the reaction center (as seen in the blue path in the EC hydrolysis path). How much the barrier can be lowered by including the explicit solvent molecule(s) is presumably a case-by-case scenario and need to be investigated in depth. However, at least in the cases we studied, this does not qualitatively change the assessment of the kinetic feasibilities of the reaction paths.

5.6 Effects of using different dielectric constants in the implicit solvation model

We show in Figure 4.16 the LEDC hydrolysis path calculated with implicit solvation models with different solvent parameters. The two solvents considered are 3:7 EC/EMC (dielectric 18.5, optical dielectric 2.00) and DMSO (dielectric 46.8, optical dielectric 2.18). Note that the geometries are optimized with CPCM with different dielectric constants and optical dielectric constants, and single point corrections are performed at a higher level of theory and SMD model (again using the parameters for 3:7 EC/EMC or DMSO). The reaction energies and barriers only differ by sub kcal/mol between the two paths, which indicates changing parameters in the implicit solvation model would not have significant impact on the calculated reaction pathways.

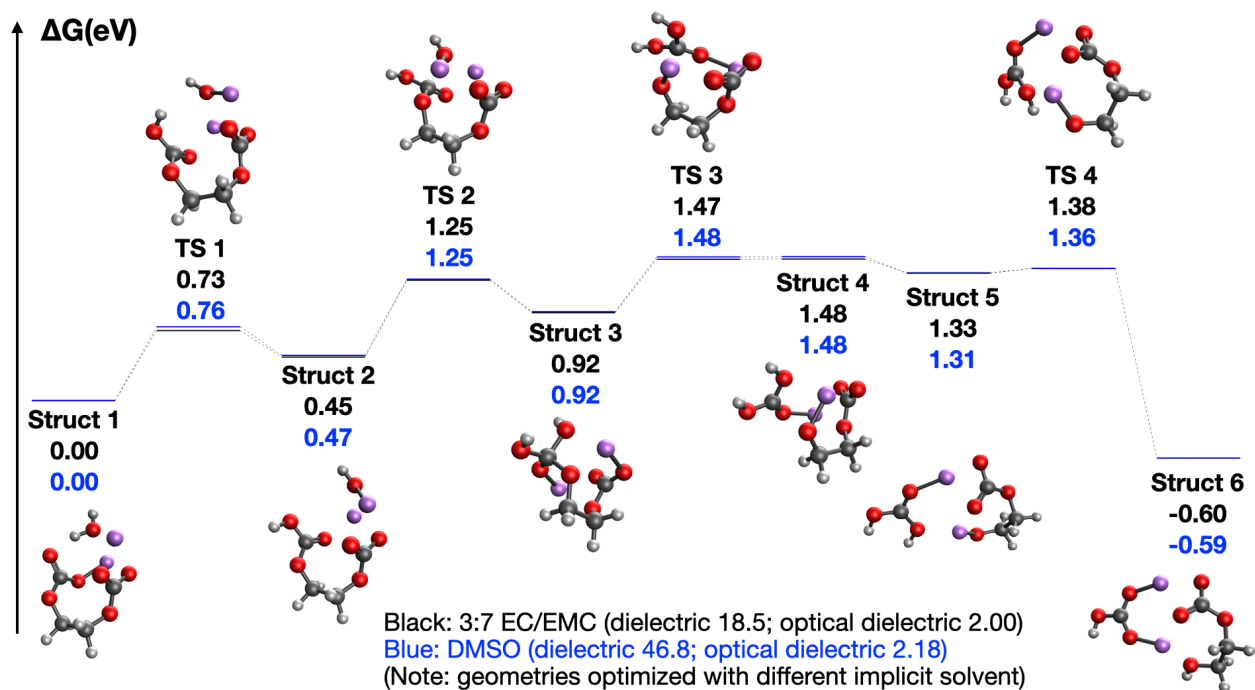


Figure 4.16: The LEDC hydrolysis path computed with implicit solvation model with different dielectric constants

5.7 Analysis of the BonDNet prediction error

The BonDNet prediction error on the recombinant molecules is 0.454 eV, higher than that reported in the BondNet paper. We have conducted error analysis for these predictions, as

shown in Figure 4.17 (distribution of the error by reactant charge), Figure 4.18 (distribution of the error by bond dissociation type) and Figure 4.19 (10 reactions with the largest prediction errors).

In order to analyze how sensitive our procedure for selecting recombinant molecules is to the BonDNet prediction error, we have calculated with DFT a new set of recombinant molecules. There were 1804 recombinants that we filtered out based on the old BonDNet prediction (endergonic upon recombination) and we have calculated here a random subset of those. Out of 384 successful calculations, 56 molecules did not change their connectivities during geometry optimization. These 56 molecules can be formed by 72 unique one-bond reactions. There were only 7 recombinant molecules (out of 384) that we should have included in the reaction network (but have not due to the prediction error with BondNet). In other words, for only 7 molecules did BonDNet predict an endergonic recombination when the recombination reaction was in reality exergonic. We acknowledge there is always a possibility of missing important species/reactions. However, in reality, it is impossible to exhaustively include all possible species and reactions, and the 7 molecules neglected here represent a small percentage loss which is unlikely to significantly change our conclusions. We have included the optimized geometries for these 7 molecules in `geometries.zip` for reference.

In addition, we have trained a new BonDNet model by adding the recombinant molecules from the old test set (the test set in the main text) into the training, and observed an reduction of MAE from 0.496 eV to 0.355 eV when predicting on the new test set described above. This result indicates the performance of BonDNet can be iteratively improved by adding more, representative training data. We plot the DFT vs BonDNet predicted bond dissociation energies (BDEs) in Figure 4.20. The reactions with the 10 largest prediction errors for the new test set from the new BonDNet model are given in Figure 4.21.

The large error from BonDNet for these molecules (either from the old or new test set) could originate from two aspects: (1) they are slightly larger molecules than those in the training set; (2) some bond types are underrepresented in the training. Since the two BondNet models are predicting on different test sets (i.e. the old test set for the current model mainly consists of endergonic bond dissociation reactions and the new test set for the new model mainly consists of exergonic bond dissociation reactions), the bond types that have large prediction errors also differ. Although adding the data from the old test set to the training improves the error for the new test set from 0.496 eV to 0.355 eV, the presence of new bond types limits how far the error can be lowered. For the old test set, the bond types that have high errors (from the current BonDNet model) are bonds between the carbon in $\text{C}(=\text{O})\text{OLi}$ (e.g. 3), 5) in Figure 4.19), $\text{C}(=\text{O})\text{OLiOR}$ (e.g. 1), 2), 4), 6), 9), 10) in Figure 4.19), $\text{C}(=\text{O})(\text{OR})_2$ (e.g. 7), 8) in Figure 4.19) and another carbon/oxygen of various types. For the new test set, those with high errors mostly involve the bonds between the oxygen in $\text{C}(=\text{O})\text{OLiOR}$ (e.g. 1), 2), 3), 4), 5), 6), 7), 9), 10) in Figure 4.21). All of the high-error cases described above involve forming a bond between already sterically hindered carbons and oxygens, which are underrepresented in the training. However, the error (for predicting

on these bond types) can be readily improved by including these cases back into the training set.

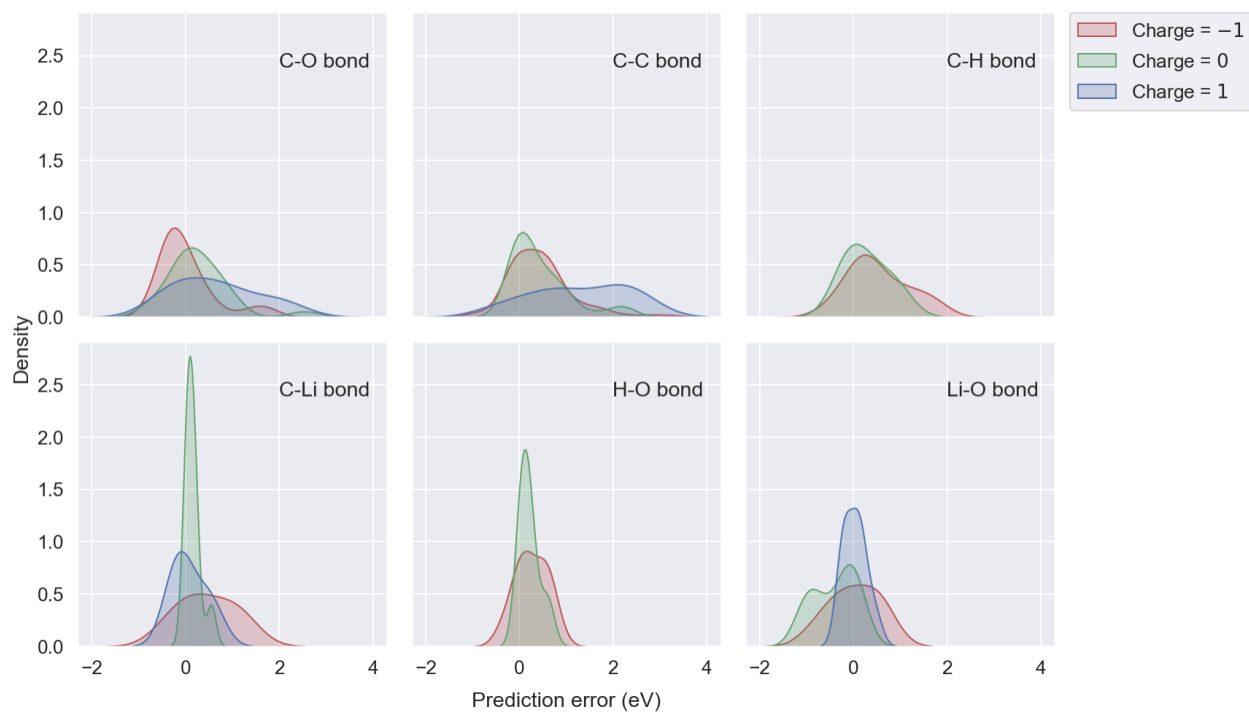


Figure 4.17: Distribution of BonDNet prediction error by reactant charge for the old recombinant molecules

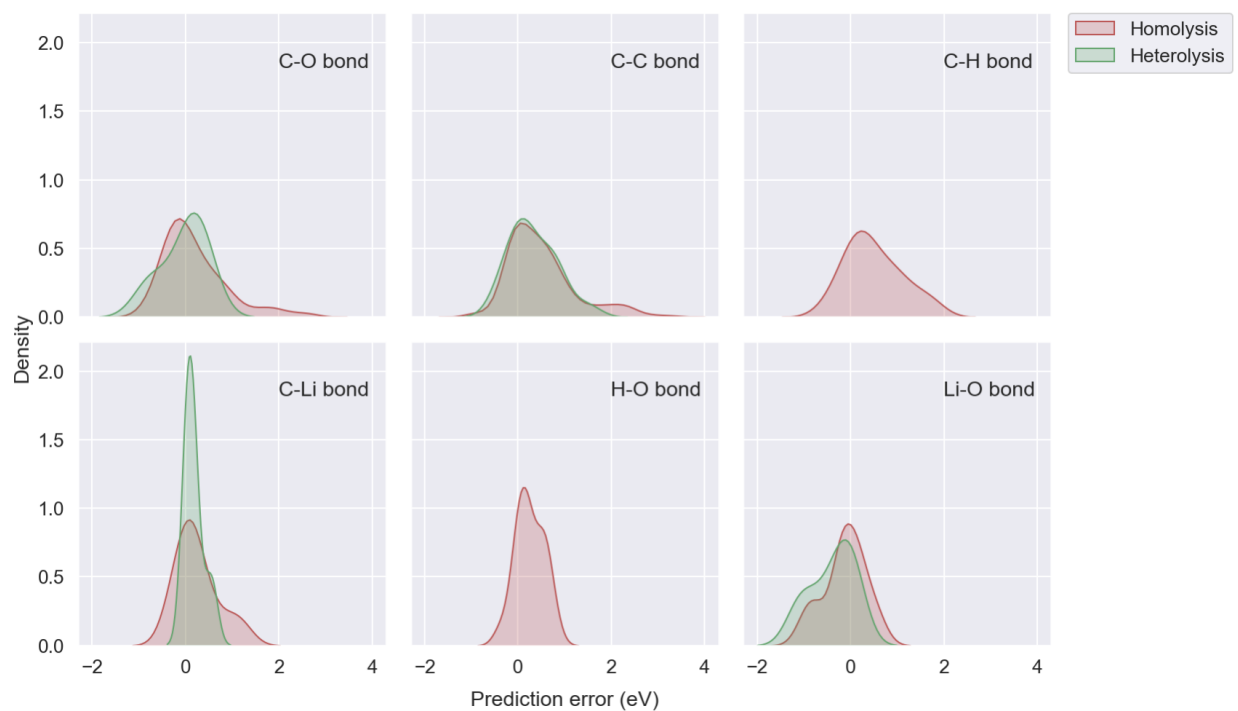


Figure 4.18: Distribution of BonDNet prediction error by bond dissociation type for the old recombinant molecules

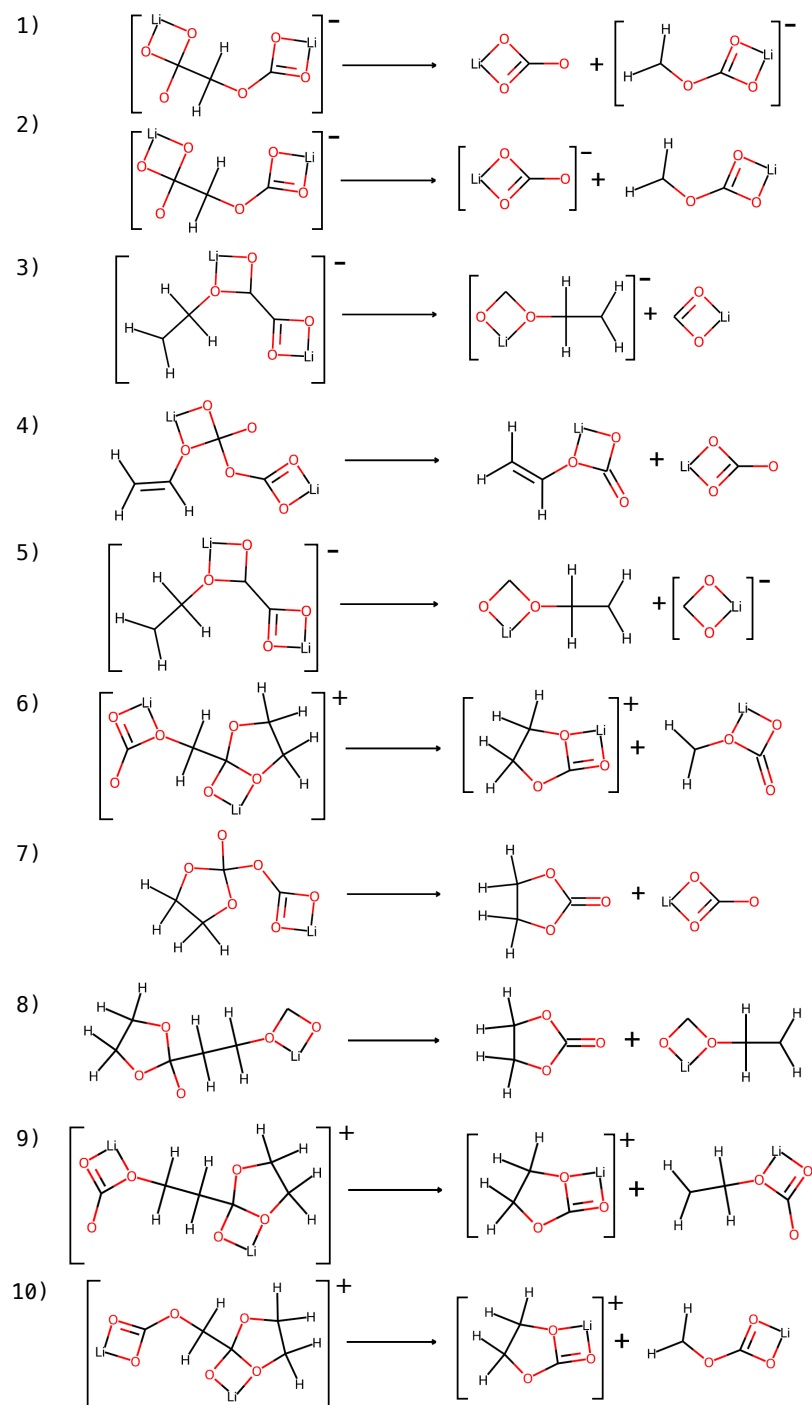


Figure 4.19: Reactions with large prediction error for the old recombinant molecules from the current BonD-Net model

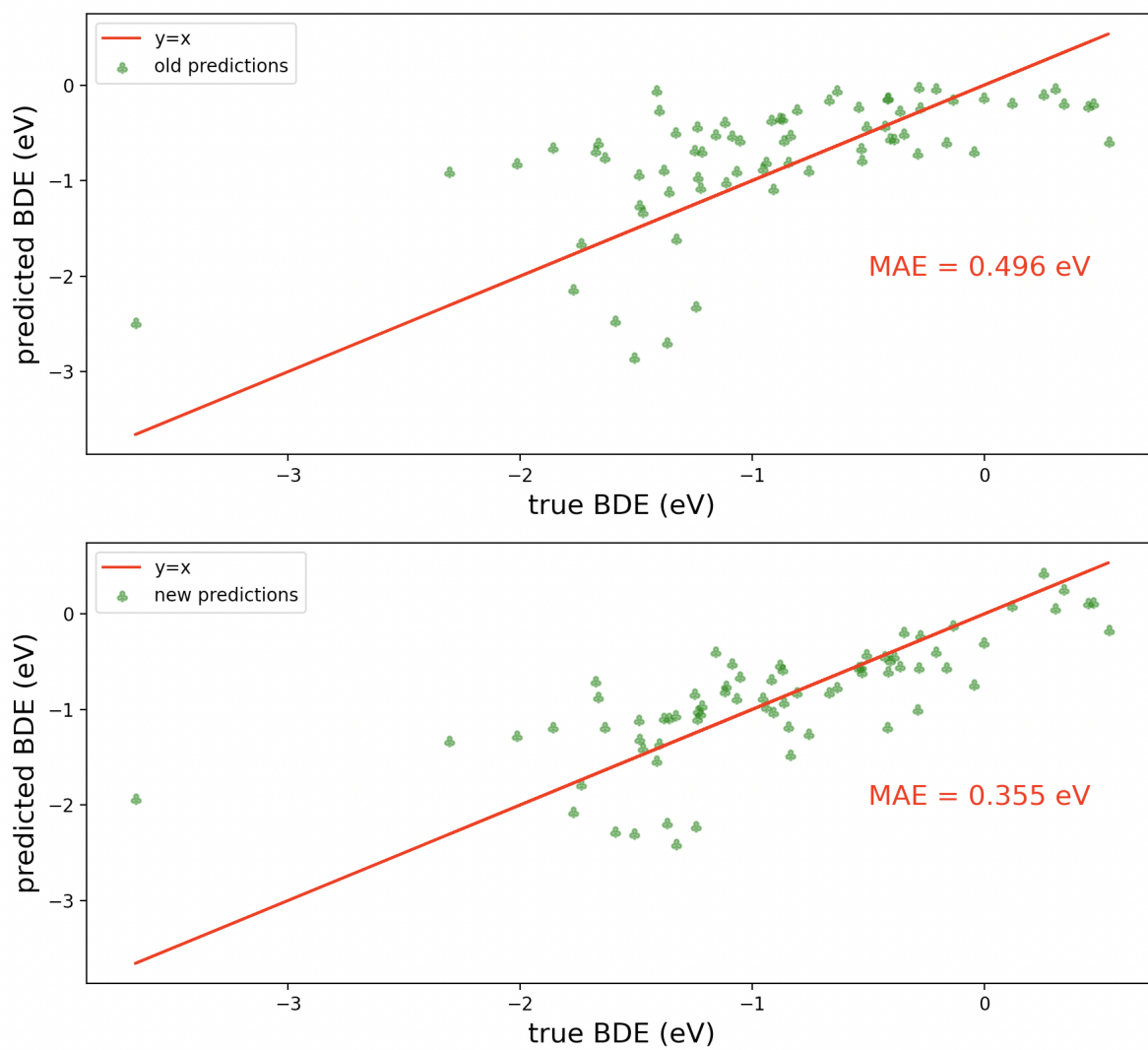


Figure 4.20: Prediction errors for the new test set from the current model and the new model

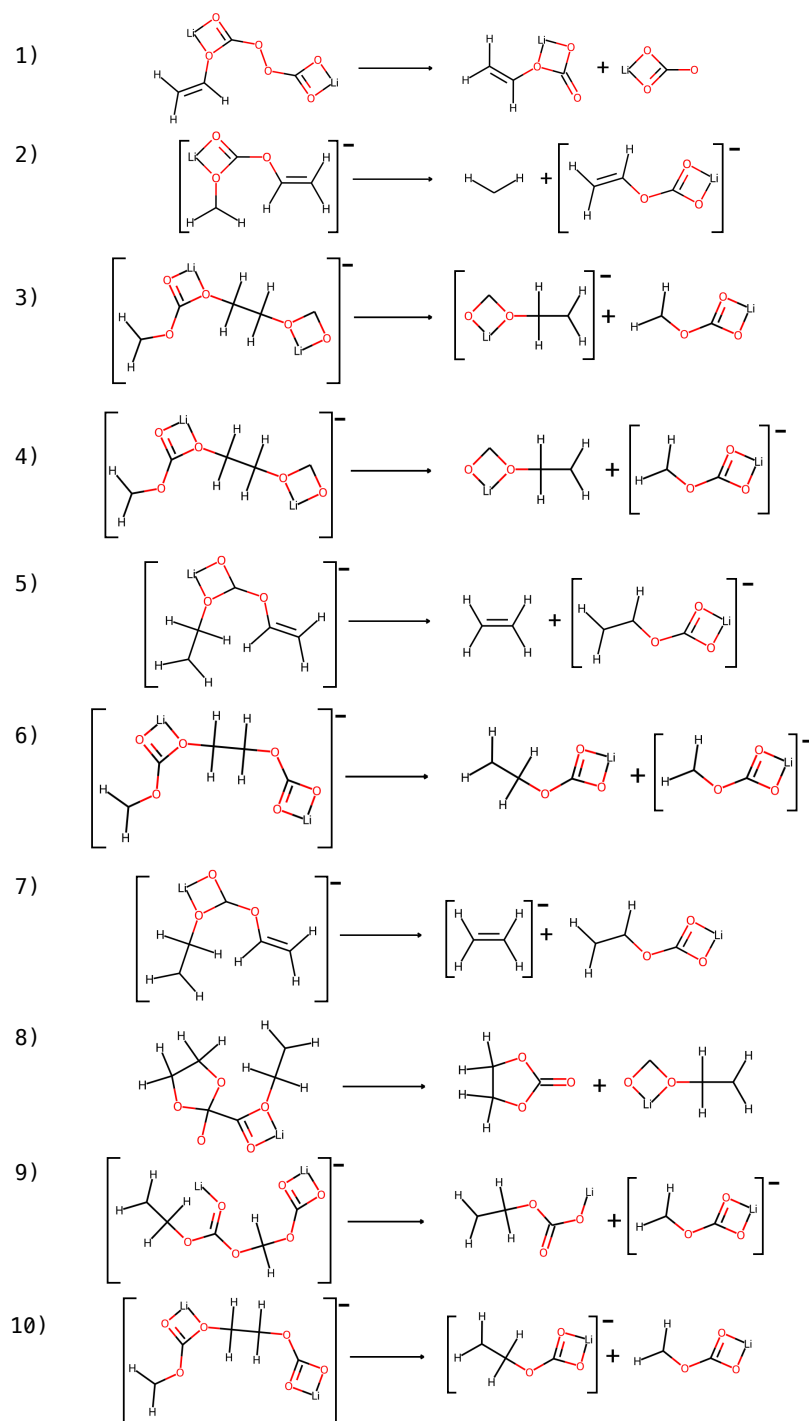


Figure 4.21: Reactions with large prediction error for the new recombinant molecules from the new BonDNet model

Bibliography

- [1] Martin Winter. “The solid electrolyte interphase—the most important and the least understood solid electrolyte in rechargeable Li batteries”. In: *Z. Phys. Chem.* 223.10-11 (2009), pp. 1395–1406.
- [2] Haiping Wu et al. “Recent Progress in Understanding Solid Electrolyte Interphase on Lithium Metal Anodes”. In: *Advanced Energy Materials* (2020), p. 2003092. ISSN: 1614-6832. DOI: 10.1002/aenm.202003092.
- [3] Seong Jin An et al. “The state of understanding of the lithium-ion-battery graphite solid electrolyte interphase (SEI) and its relationship to formation cycling”. In: *Carbon* 105 (2016), pp. 52–76.
- [4] E Peled and S Menkin. “SEI: past, present and future”. In: *J. Electrochem. Soc.* 164.7 (2017), A1703–A1719.
- [5] Kang Xu. “Electrolytes and interphases in Li-ion batteries and beyond”. In: *Chemical reviews* 114.23 (2014), pp. 11503–11618.
- [6] Emanuel Peled. “The electrochemical behavior of alkali and alkaline earth metals in nonaqueous battery systems—the solid electrolyte interphase model”. In: *J. Electrochem. Soc.* 126.12 (1979), p. 2047.
- [7] Fredrik Lindgren et al. “SEI formation and interfacial stability of a Si electrode in a LiTDI-salt based electrolyte with FEC and VC additives for Li-ion batteries”. In: *ACS applied materials & interfaces* 8.24 (2016), pp. 15758–15766.
- [8] Anna M Andersson et al. “Electrochemically lithiated graphite characterised by photoelectron spectroscopy”. In: *Journal of Power Sources* 119 (2003), pp. 522–527.
- [9] D Aurbach et al. “A comparative study of synthetic graphite and Li electrodes in electrolyte solutions based on ethylene carbonate-dimethyl carbonate mixtures”. In: *Journal of the Electrochemical Society* 143.12 (1996), p. 3809.
- [10] Guorong V Zhuang et al. “A study of electrochemical reduction of ethylene and propylene carbonate electrolytes on graphite using ATR-FTIR spectroscopy”. In: *Electrochemical and Solid State Letters* 8.9 (2005), A441.

- [11] Alison L Michan, Michal Leskes, and Clare P Grey. “Voltage dependent solid electrolyte interphase formation in silicon electrodes: monitoring the formation of organic decomposition products”. In: *Chemistry of Materials* 28.1 (2016), pp. 385–398.
- [12] Alison L Michan et al. “Solid electrolyte interphase growth and capacity loss in silicon electrodes”. In: *Journal of the American Chemical Society* 138.25 (2016), pp. 7918–7931.
- [13] AM Andersson and Kristina Edström. “Chemical composition and morphology of the elevated temperature SEI on graphite”. In: *Journal of the Electrochemical Society* 148.10 (2001), A1100.
- [14] Mengyun Nie et al. “Lithium ion battery graphite solid electrolyte interphase revealed by microscopy and spectroscopy”. In: *J. Phys. Chem. C* 117.3 (2013), pp. 1257–1267.
- [15] Michael J Zachman et al. “Cryo-STEM mapping of solid–liquid interfaces and dendrites in lithium-metal batteries”. In: *Nature* 560.7718 (2018), pp. 345–349.
- [16] Doron Aurbach and Yosef Gofer. “The behavior of lithium electrodes in mixtures of alkyl carbonates and ethers”. In: *Journal of the Electrochemical Society* 138.12 (1991), p. 3529.
- [17] Doron Aurbach et al. “The behaviour of lithium electrodes in propylene and ethylene carbonate: The major factors that influence Li cycling efficiency”. In: *Journal of Electroanalytical Chemistry* 339.1-2 (1992), pp. 451–471.
- [18] Doron Aurbach et al. “The study of electrolyte solutions based on ethylene and diethyl carbonates for rechargeable Li batteries: II. Graphite electrodes”. In: *Journal of The Electrochemical Society* 142.9 (1995), p. 2882.
- [19] D Aurbach et al. “Identification of surface films formed on lithium in propylene carbonate solutions”. In: *Journal of The Electrochemical Society* 134.7 (1987), p. 1611.
- [20] CR Yang, YY Wang, and CC Wan. “Composition analysis of the passive film on the carbon electrode of a lithium-ion battery with an EC-based electrolyte”. In: *Journal of power sources* 72.1 (1998), pp. 66–70.
- [21] Guorong V Zhuang et al. “Lithium ethylene dicarbonate identified as the primary product of chemical and electrochemical reduction of EC in 1.2 M LiPF₆/EC: EMC electrolyte”. In: *The Journal of Physical Chemistry B* 109.37 (2005), pp. 17567–17573.
- [22] R Dedryvère et al. “Characterization of lithium alkyl carbonates by X-ray photoelectron spectroscopy: experimental and theoretical study”. In: *The Journal of Physical Chemistry B* 109.33 (2005), pp. 15868–15875.

-
- [23] Kang Xu et al. “Syntheses and characterization of lithium alkyl mono-and dicarbonates as components of surface films in Li-ion batteries”. In: *The Journal of Physical Chemistry B* 110.15 (2006), pp. 7708–7719.
- [24] S-H Kang et al. “Investigating the solid electrolyte interphase using binder-free graphite electrodes”. In: *Journal of Power Sources* 175.1 (2008), pp. 526–532.
- [25] Daniel M Seo et al. “Reduction reactions of carbonate solvents for lithium ion batteries”. In: *ECS Electrochemistry Letters* 3.9 (2014), A91.
- [26] Feifei Shi et al. “A catalytic path for electrolyte reduction in lithium-ion cells revealed by in situ attenuated total reflection-Fourier transform infrared spectroscopy”. In: *Journal of the American Chemical Society* 137.9 (2015), pp. 3181–3184.
- [27] Luning Wang et al. “Identifying the components of the solid–electrolyte interphase in Li-ion batteries”. In: *Nature chemistry* 11.9 (2019), pp. 789–796.
- [28] Aiping Wang et al. “Review on modeling of the anode solid electrolyte interphase (SEI) for lithium-ion batteries”. In: *Npj Comput. Mater.* 4.1 (2018), pp. 1–26.
- [29] Yunsong Li, Kevin Leung, and Yue Qi. “Computational exploration of the Li-electrode—electrolyte interface in the presence of a nanometer thick solid-electrolyte interphase layer”. In: *Accounts of chemical research* 49.10 (2016), pp. 2363–2370.
- [30] Yixuan Wang et al. “Theoretical studies to understand surface chemistry on carbon anodes for lithium-ion batteries: reduction mechanisms of ethylene carbonate”. In: *J. Am. Chem. Soc.* 123.47 (2001), pp. 11708–11718.
- [31] Yixuan Wang et al. “Theoretical studies to understand surface chemistry on carbon anodes for lithium-ion batteries: how does vinylene carbonate play its role as an electrolyte additive?” In: *Journal of the American Chemical Society* 124.16 (2002), pp. 4408–4421.
- [32] Jiamei Yu et al. “Hybrid DFT functional-based static and molecular dynamics studies of excess electron in liquid ethylene carbonate”. In: *Journal of The Electrochemical Society* 158.4 (2011), A400.
- [33] Mahsa Ebadi, Daniel Brandell, and C Moyses Araujo. “Electrolyte decomposition on Li-metal surfaces from first-principles theory”. In: *The Journal of Chemical Physics* 145.20 (2016), p. 204701.
- [34] Ken Tasaki et al. “Theoretical studies on the reductive decompositions of solvents and additives for lithium-ion batteries near lithium anodes”. In: *Journal of the Electrochemical Society* 153.12 (2006), A2192.
- [35] Kevin Leung et al. “Using atomic layer deposition to hinder solvent decomposition in lithium ion batteries: first-principles modeling and experimental studies”. In: *Journal of the American Chemical Society* 133.37 (2011), pp. 14741–14754.

-
- [36] Kevin Leung. “Two-electron reduction of ethylene carbonate: A quantum chemistry re-examination of mechanisms”. In: *Chemical Physics Letters* 568 (2013), pp. 1–8.
- [37] Kevin Leung and Joanne L Budzien. “Ab initio molecular dynamics simulations of the initial stages of solid–electrolyte interphase formation on lithium ion battery graphitic anodes”. In: *Physical Chemistry Chemical Physics* 12.25 (2010), pp. 6583–6586.
- [38] Jan P Unsleber and Markus Reiher. “The exploration of chemical reaction networks”. In: *Annual review of physical chemistry* 71 (2020), pp. 121–142.
- [39] Connor W Coley, Natalie S Eyke, and Klavs F Jensen. “Autonomous discovery in the chemical sciences part I: Progress”. In: *Angewandte Chemie International Edition* 59.51 (2020), pp. 22858–22893.
- [40] Satoshi Maeda et al. “Artificial force induced reaction (AFIR) method for exploring quantum chemical potential energy surfaces”. In: *The Chemical Record* 16.5 (2016), pp. 2232–2248.
- [41] Xiao-Jie Zhang and Zhi-Pan Liu. “Reaction sampling and reactivity prediction using the stochastic surface walking method”. In: *Physical Chemistry Chemical Physics* 17.4 (2015), pp. 2757–2769.
- [42] Lee-Ping Wang et al. “Discovering chemistry with an ab initio nanoreactor”. In: *Nature chemistry* 6.12 (2014), pp. 1044–1048.
- [43] Alessandro Laio and Michele Parrinello. “Escaping free-energy minima”. In: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12562–12566.
- [44] Amanda L Dewyer, Alonso J Argüelles, and Paul M Zimmerman. “Methods for exploring reaction space in molecular systems”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 8.2 (2018), e1354.
- [45] Aurelio Rodriguez et al. “tsscads2018: A code for automated discovery of chemical reaction mechanisms and solving the kinetics”. In: *Journal of computational chemistry* 39.23 (2018), pp. 1922–1930.
- [46] Emilio Martinez-Núñez. “An automated method to find transition states using chemical dynamics simulations”. In: *Journal of computational chemistry* 36.4 (2015), pp. 222–234.
- [47] Paul M Zimmerman. “Automated discovery of chemically reasonable elementary reaction steps”. In: *Journal of computational chemistry* 34.16 (2013), pp. 1385–1392.
- [48] Paul M Zimmerman. “Navigating molecular space for reaction mechanisms: an efficient, automated procedure”. In: *Molecular Simulation* 41.1-3 (2015), pp. 43–54.
- [49] Maike Bergeler et al. “Heuristics-guided exploration of reaction mechanisms”. In: *Journal of chemical theory and computation* 11.12 (2015), pp. 5712–5722.

-
- [50] Gregor N Simm and Markus Reiher. “Context-driven exploration of complex chemical reaction networks”. In: *Journal of chemical theory and computation* 13.12 (2017), pp. 6108–6119.
- [51] Yury V Suleimanov and William H Green. “Automated discovery of elementary chemical reaction steps using freezing string and Berny optimization methods”. In: *Journal of Chemical Theory and Computation* 11.9 (2015), pp. 4248–4259.
- [52] Dmitrij Rappoport et al. “Complex chemical reaction networks from heuristics-aided quantum chemistry”. In: *Journal of chemical theory and computation* 10.3 (2014), pp. 897–907.
- [53] Yeonjoon Kim et al. “Efficient prediction of reaction paths through molecular graph and reaction network analysis”. In: *Chemical science* 9.4 (2018), pp. 825–835.
- [54] Scott Habershon. “Sampling reactive pathways with random walks in chemical space: Applications to molecular dissociation and catalysis”. In: *The Journal of Chemical Physics* 143.9 (2015), p. 094106.
- [55] Christopher Robertson, Idil Ismail, and Scott Habershon. “Traversing Dense Networks of Elementary Chemical Reactions to Predict Minimum-Energy Reaction Mechanisms”. In: *ChemSystemsChem* 2 (2019), e1900047.
- [56] Colin A Grambow et al. “Unimolecular reaction pathways of a γ -ketohydroperoxide from combined application of automated reaction discovery methods”. In: *Journal of the American Chemical Society* 140.3 (2018), pp. 1035–1048.
- [57] Tamal Das, Siddharth Ghule, and Kumar Vanka. “Insights into the origin of life: Did it begin from HCN and H₂O?”. In: *ACS central science* 5.9 (2019), pp. 1532–1540.
- [58] Samuel M Blau et al. “A chemically consistent graph architecture for massive reaction networks applied to solid-electrolyte interphase formation”. In: *Chemical Science* 12.13 (2021), pp. 4931–4939.
- [59] Mathew J Boyer and Gyeong S Hwang. “Kinetic Selectivity of Lithium Alkyl Carbonate Formation from Combination Reactions of Ethylene Carbonate Radical Anions”. In: *The Journal of Physical Chemistry C* 124.47 (2020), pp. 25754–25759.
- [60] Mingjian Wen et al. “BonDNet: a graph neural network for the prediction of bond dissociation energies for charged molecules”. In: *Chemical Science* 12.5 (2021), pp. 1858–1868.
- [61] Jay L Banks et al. “Integrated modeling program, applied chemical theory (IMPACT)”. in: *Journal of computational chemistry* 26.16 (2005), pp. 1752–1780.
- [62] *Schrödinger Python API*. accessed 06/2021. URL: http://content.schrodinger.com/Docs/r2020-4/python_api/overview.html.

-
- [63] Evan Walter Clark Spotte-Smith et al. “Quantum chemical calculations of lithium-ion battery electrolyte and interphase species”. In: *Scientific data* 8.1 (2021), pp. 1–15.
- [64] Eric L First, Chrysanthos E Gounaris, and Christodoulos A Floudas. “Stereochemically consistent reaction mapping and identification of multiple reaction mechanisms through integer linear optimization”. In: *Journal of chemical information and modeling* 52.1 (2012), pp. 84–92.
- [65] Sergio Trasatti. “The absolute electrode potential: an explanatory note”. In: *Pure Appl. Chem* 58.7 (1986), pp. 955–966.
- [66] Sergio Trasatti. “The “absolute” electrode potential—The end of the story”. In: *Electrochimica Acta* 35.1 (1990), pp. 269–271.
- [67] Oleg Borodin, Wishvender Behl, and T Richard Jow. “Oxidative stability and initial decomposition reactions of carbonate, sulfone, and alkyl phosphate-based electrolytes”. In: *The Journal of Physical Chemistry C* 117.17 (2013), pp. 8661–8682.
- [68] Kevin Leung et al. “Stability of solid electrolyte interphase components on lithium metal and reactive anode material surfaces”. In: *The Journal of Physical Chemistry C* 120.12 (2016), pp. 6302–6313.
- [69] Yihan Shao et al. “Advances in molecular quantum chemistry contained in the Q-Chem 4 program package”. In: *Mol. Phys.* 113 (2015), pp. 184–215. DOI: 10.1080/00268976.2014.952696. eprint: <http://dx.doi.org/10.1080/00268976.2014.952696>. URL: <http://dx.doi.org/10.1080/00268976.2014.952696>.
- [70] Narbe Mardirossian and Martin Head-Gordon. “ ω B97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy”. In: *Physical Chemistry Chemical Physics* 16.21 (2014), pp. 9904–9924.
- [71] Florian Weigend and Reinhart Ahlrichs. “Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy”. In: *Physical Chemistry Chemical Physics* 7.18 (2005), pp. 3297–3305.
- [72] Dmitrij Rappoport and Filipp Furche. “Property-optimized Gaussian basis sets for molecular response calculations”. In: *The Journal of chemical physics* 133.13 (2010), p. 134105.
- [73] Aleksandr V Marenich, Christopher J Cramer, and Donald G Truhlar. “Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions”. In: *The Journal of Physical Chemistry B* 113.18 (2009), pp. 6378–6396.
- [74] David S Hall, Julian Self, and JR Dahn. “Dielectric constants for quantum chemistry and Li-ion batteries: solvent blends of ethylene carbonate and ethyl methyl carbonate”. In: *The Journal of Physical Chemistry C* 119.39 (2015), pp. 22322–22330.

-
- [75] Xiaohui Qu et al. “The Electrolyte Genome project: A big data approach in battery materials discovery”. In: *Computational Materials Science* 103 (2015), pp. 56–67.
- [76] Jeng-Da Chai and Martin Head-Gordon. “Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections”. In: *Phys. Chem. Chem. Phys.* 10.44 (2008), pp. 6615–6620.
- [77] Vincenzo Barone and Maurizio Cossi. “Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model”. In: *The Journal of Physical Chemistry A* 102.11 (1998), pp. 1995–2001.
- [78] Sebastian Dohm et al. “Comprehensive thermochemical benchmark set of realistic closed-shell metal organic reactions”. In: *Journal of chemical theory and computation* 14.5 (2018), pp. 2596–2608.
- [79] Rajesh K Raju, Ashfaq A Bengali, and Edward N Brothers. “A unified set of experimental organometallic data used to evaluate modern theoretical methods”. In: *Dalton Transactions* 45.35 (2016), pp. 13766–13778.
- [80] Bruce C Garrett et al. “Algorithms and accuracy requirements for computing reaction paths by the method of steepest descent”. In: *The Journal of Physical Chemistry* 92.6 (1988), pp. 1476–1488.
- [81] Pekka Peljo and Hubert H Girault. “Electrochemical potential window of battery electrolytes: the HOMO–LUMO misconception”. In: *Energy & Environmental Science* 11.9 (2018), pp. 2306–2309.
- [82] Chunzhen Yang et al. “Chemical recognition of active oxygen species on the surface of oxygen evolution reaction electrocatalysts”. In: *Angewandte Chemie* 129.30 (2017), pp. 8778–8782.
- [83] Nicolas Dubouis et al. “The role of the hydrogen evolution reaction in the solid–electrolyte interphase formation mechanism for “Water-in-Salt” electrolytes”. In: *Energy & Environmental Science* 11.12 (2018), pp. 3491–3499.
- [84] Bernardine LD Rinkel et al. “Electrolyte oxidation pathways in lithium-ion batteries”. In: *Journal of the American Chemical Society* 142.35 (2020), pp. 15058–15074.
- [85] Doron Aurbach and Idit Weissman. “On the possibility of LiH formation on Li surfaces in wet electrolyte solutions”. In: *Electrochemistry communications* 1.8 (1999), pp. 324–331.
- [86] Zulipiya Shadike et al. “Identification of LiH and nanocrystalline LiF in the solid–electrolyte interphase of lithium metal anodes”. In: *Nature Nanotechnology* (2021), pp. 1–6.
- [87] David R Lide. *CRC handbook of chemistry and physics*. Vol. 85. CRC press, 2004.

-
- [88] Robert S Mulliken. “Electronic population analysis on LCAO–MO molecular wave functions. I”. in: *The Journal of Chemical Physics* 23.10 (1955), pp. 1833–1840.
- [89] Jin Shang and Saeed Shirazian. “Facilitated Dissociation of Water in the Presence of Lithium Metal at Ambient Temperature as a Requisite for Lithium–Gas Reactions”. In: *The Journal of Physical Chemistry C* 122.28 (2018), pp. 16016–16022.
- [90] José LC Fajin, M Natália DS Cordeiro, and José RB Gomes. “Density functional theory study of the water dissociation on platinum surfaces: General trends”. In: *The Journal of Physical Chemistry A* 118.31 (2014), pp. 5832–5840.
- [91] Oleg Borodin et al. “Molecular dynamics simulations and experimental study of lithium ion transport in dilithium ethylene dicarbonate”. In: *The Journal of Physical Chemistry C* 117.15 (2013), pp. 7433–7444.
- [92] Kenzi Hori et al. “Theoretical study on the reaction mechanism for the hydrolysis of esters and amides under acidic conditions”. In: *Tetrahedron* 63.5 (2007), pp. 1264–1269.
- [93] Shinichi Yamabe, Takeshi Fukuda, and Misao Ishii. “Role of hydrogen bonds in acid-catalyzed hydrolyses of esters”. In: *Theoretical Chemistry Accounts* 130.2 (2011), pp. 429–438.
- [94] Chang-Guo Zhan, Donald W Landry, and Rick L Ornstein. “Energy barriers for alkaline hydrolysis of carboxylic acid esters in aqueous solution by reaction field calculations”. In: *The Journal of Physical Chemistry A* 104.32 (2000), pp. 7672–7678.
- [95] Rabindra Nath Manna, Mandira Dutta, and Biman Jana. “Mechanistic study of the ATP hydrolysis reaction in dynein motor protein”. In: *Physical Chemistry Chemical Physics* 22.3 (2020), pp. 1534–1542.
- [96] Fabio Pietrucci and Antonino Marco Saitta. “Formamide reaction network in gas phase and solution via a unified theoretical approach: Toward a reconciliation of different prebiotic scenarios”. In: *Proceedings of the National Academy of Sciences* 112.49 (2015), pp. 15030–15035.
- [97] Woon Ih Choi et al. “Reductive reactions via excess Li in mixture electrolytes of Li ion batteries: an ab initio molecular dynamics study”. In: *Physical Chemistry Chemical Physics* 21.10 (2019), pp. 5489–5498.
- [98] Ilya A Shkrob et al. “Reduction of carbonate electrolytes and the formation of solid-electrolyte interface (SEI) in lithium-ion batteries. 1. Spectroscopic observations of radical intermediates generated in one-electron reduction of carbonates”. In: *The Journal of Physical Chemistry C* 117.38 (2013), pp. 19255–19269.
- [99] Ilya A Shkrob et al. “Reduction of carbonate electrolytes and the formation of solid-electrolyte interface (SEI) in Lithium-Ion Batteries. 2. Radiolytically induced poly-

- merization of ethylene carbonate”. In: *The Journal of Physical Chemistry C* 117.38 (2013), pp. 19270–19279.
- [100] Fernando A Soto et al. “Formation and growth mechanisms of solid-electrolyte interphase layers in rechargeable batteries”. In: *Chemistry of Materials* 27.23 (2015), pp. 7990–8000.
- [101] Satu Heiskanen, Nina Laszczynski, and Brett L Lucht. “Perspective—Surface reactions of electrolyte with LiNixCoyMnzO2 cathodes for lithium ion batteries”. In: *Journal of the Electrochemical Society* 167.10 (2020), p. 100519.
- [102] Michael Metzger et al. “Origin of H2 evolution in LIBs: H2O reduction vs. electrolyte oxidation”. In: *Journal of The Electrochemical Society* 163.5 (2016), A798.
- [103] Anubhav Jain et al. “The Materials Project: A materials genome approach to accelerating materials innovation”. In: *APL Materials* 1.1 (2013), p. 011002. ISSN: 2166532X. DOI: 10.1063/1.4812323. URL: <http://link.aip.org/link/AMPADS/v1/i1/p011002/s1%5C&Agg=doi>.
- [104] *The Materials Project*. accessed 05/2021. URL: <https://materialsproject.org>.
- [105] Charles Kittel, Paul McEuen, and Paul McEuen. *Introduction to solid state physics*. Vol. 8. Wiley New York, 1996.
- [106] K Doll, NM Harrison, and VR Saunders. “A density functional study of lithium bulk and surfaces”. In: *Journal of Physics: Condensed Matter* 11.26 (1999), p. 5007.
- [107] Luis E Camacho-Forero et al. “Reactivity at the lithium–metal anode surface of lithium–sulfur batteries”. In: *The Journal of Physical Chemistry C* 119.48 (2015), pp. 26828–26839.

Chapter 5

The reductive decomposition kinetics and thermodynamics that govern the design of fluorinated alkoxyaluminate/borate salts for Mg-ion and Ca-ion batteries

Parts of this chapter were reprinted with permission from: “The reductive decomposition kinetics and thermodynamics that govern the design of fluorinated alkoxyaluminate/borate salts for Mg-ion and Ca-ion batteries” Xiaowei Xie, Noel J. Leon, David W. Small, Evan Walter Clark Spotte-Smith, Chen Liao, Kristin A. Persson. *J. Phys. Chem. C*. 2022. <https://doi.org/10.1021/acs.jpcc.2c06653>.

I Introduction

Electrolyte design is a fundamental challenge in battery electrochemistry. Numerous interactions must be considered to optimize both the transport within the bulk electrolyte as well as interfacial reactivity. In the ideal case, an electrolyte should allow for reversible metal plating and stripping (if in contact with a pure metal electrode) or reversible intercalation and deintercalation (if in contact with an intercalation electrode) and should exhibit a wide electrochemical window. This stability can either be intrinsic or can be enabled by the formation of an appropriately passivating solid electrolyte interphase (SEI) layer. Most commonly, electrolytes are designed based on trial and error, guided by previous results for analogous systems, however, rational design based on computed electrolyte properties is a desirable alternative to accelerate the advancement of battery technologies. While Li-ion batteries remain dominant in applications ranging from electric vehicles to portable consumer electronics, multivalent batteries, such as Mg- or Ca-ion batteries, are receiving growing attention as next-generation alternatives.[1–3] However, to realize the promise of earth-abundant Mg or Ca energy storage, improved understanding of the processes that govern electrolyte stability in these systems is needed.

As multivalent-ion systems tend to form more stable ion pairs in solution than their monovalent alkali metal ion counterparts,[4, 5] metal ion de-solvation and undesirable metal ion-assisted anion decomposition reactions, especially from partially reduced metal ions such as Mg^+/Ca^+ , present a formidable challenge. Theoretical investigations on magnesium bis(trifluoromethanesulfonyl)imide ($\text{Mg}(\text{TFSI})_2$) indicated that the ion pairs formed between partially reduced Mg^+ cations and TFSI^- facilitate the decomposition of the anion through an energetically favorable C-S bond breaking.[6] It is also notable that simple salts for Mg-ion batteries, e.g. $\text{Mg}(\text{TFSI})_2$ require a significant concentration of chloride (Cl^-) to enable plating/stripping at low overpotential.[7, 8] Early investigations found that chloride (Cl^-) protects the passivation layer by oxidizing impurities like H_2O via the formation of $\text{Mg}-\text{Cl}$ (ad) and/or MgCl_2 . [8, 9] More recently, cooperative anion effects between TFSI^- and Cl^- have been observed in Mg-, Zn-, and Cu-based systems to enable more reversible metal dissolution and deposition, showcasing the importance of the relative association strength between the constituent anions and the working cations in tailoring the electrochemical performance of electrolytes. [10] In spite of their promise, chloride-based electrolytes are not compati-

ble with high voltage oxide cathode materials and usually suffer from corrosion problems, hindering practical applications of this strategy.

Recently, the fluorinated alkoxyaluminate/borate salts have emerged as a family of promising electrolyte candidates for Mg- and Ca-ion batteries.[11–17] Due to the presence of multiple electron-withdrawing fluorinated sites, the resulting anions tend to be weakly coordinating, leading to higher ionic conductivity and perhaps reducing the likelihood of undesirable Mg^+/Ca^+ -assisted anion decomposition reactions. [6] Following the initial success of one salt in this family, magnesium bis-tetrakis(hexafluoroisopropoxy)borate ($\text{Mg}[\text{B}(\text{hfip})_4]_2$),[13, 18, 19] it was shown by Li[14] and Shyamsunder[15] that reversible Ca^{2+} deposition/stripping was possible at room temperature through the calcium version of this salt, $\text{Ca}[\text{B}(\text{hfip})_4]_2$. In fact, based on both experimental and computational solvation structure analysis, the contact ion pair (CIP) population was found to be extremely low for $\text{Ca}[\text{B}(\text{hfip})_4]_2$ in various ethereal solvents, [20, 21] validating the design principle based on anion association strength.

Though initial electrochemical results show promise of fluorinated alkoxyborate and alkoxyaluminate salts, the structure-property relationships within this family of electrolytes remain unclear. In particular, insufficient attention has been paid to the electrochemical, thermodynamic, and kinetic stability of these salts. Along these lines of research, Jankowski et al. [22] reported free energies for decomposition reactions involving the breaking of B-O, C-O, and C-F bonds in a variety of Mg-borate salts. They concluded that the delicate balance between the electron-withdrawing effects and the degree of ligand substitution in $\text{Mg}[\text{B}(\text{hfip})_4]_2$ renders it the most promising Mg electrolyte. However, the complete decomposition pathways and the associated kinetic barriers were not addressed. Investigations of thermodynamic as well as kinetic stability are critical in order to ensure long cycling lifetimes. Although the CIP population can be constrained to be low by designing weakly-coordinating anions, one can imagine that if there exist (thermodynamically and kinetically) favorable pathways for consuming the CIPs, the bulk equilibrium will restore the CIP population. Recently, we reported a Ca electrolyte, $\text{Ca}[\text{TPFA}]_2$, that is capable of room temperature plating and stripping, and we found that the computed kinetic barriers for the reductive decompositions of $\text{Mg}[\text{TPFA}]_2$ and $\text{Ca}[\text{TPFA}]_2$ correlated well with an overall lower Coulomb efficiency of $\text{Ca}[\text{TPFA}]_2$ compared to $\text{Mg}[\text{TPFA}]_2$. [17] This hints at the possibility of discovering even more stable electrolytes via computational guidance.

In the present combined theoretical-experimental study, we systematically evaluate the redox stabilities for ten fluorinated alkoxyaluminate/borate salts coupled with either Mg^{2+} or Ca^{2+} by examining their redox potentials and possible decomposition pathways through B/Al-O, C-O and C-F bond cleavage. The alkoxy groups (as part of the anions) investigated here include the trifluoro-*tert*-isopropoxy (tfip= $-\text{OCHMe}(\text{CF}_3)$), trifluoro-*tert*-butoxy (tftb= $-\text{OCMe}_2(\text{CF}_3)$), hexafluoro-*tert*-isopropoxy (hfip= $-\text{OCH}(\text{CF}_3)_2$), hexafluoro-*tert*-butoxy (hftb= $-\text{OCMe}(\text{CF}_3)_2$) and perfluoro-*tert*-butoxy (pftb= $-\text{OC}(\text{CF}_3)_3$) groups. These abbreviations for these groups will be used throughout the remaining text.

We use kinetic barriers of reductive decomposition reactions as a metric for assessing and predicting electrolyte performance with respect to stability during cycling. Specifically, we argue that a high energy barrier for salt decomposition is desirable. Though it is possible for selective electrolyte degradation to positively contribute to electrode passivation and enable high Coulombic efficiency and capacity retention through SEI formation, very few SEI layers in Mg-ion or Ca-ion batteries have possessed high ionic conductivity, often requiring high overpotentials to plate or strip.[7, 23, 24] $\text{Mg}[\text{B}(\text{hfp})_4]_2$ is believed to form an SEI layer, however, the functional properties of the film may not be sufficient without the addition of sacrificial additives.[25] Here, we suggest possible promising salts for both Mg-ion and Ca-ion batteries based on their computed properties. We further analyze the correlations between the decomposition barriers and the central element of the anions (Al vs B) and the alkoxy functional groups. These findings offer a path forward for deriving useful design principles for developing novel high-performance electrolytes.

II Results and discussion

2.1 Electrochemical windows

We calculate the theoretical electrochemical windows from the adiabatic ionization potentials (IP) and electron affinities (EA) for the anions, with and without the metal ion present. It has been shown that metal-coordinating anions (CIPs) are more susceptible to reduction than the bare anions, whereas the bare anions are more easily oxidized on the cathode. [16, 26] Illustrative structures of CIPs and bare ions are outlined in Figure 5.1(a). We show the computed electrochemical windows in Figure 5.1(b), using the oxidation potentials for bare anions as the upper limit and the reduction potentials for CIPs as the lower limit. The complete electrochemical windows for CIPs and bare anions are given in Appendix Figure 5.9 and 5.10 respectively; we observe, as expected, that the electrochemical windows are consistently shifted upwards when the metal ion (Mg^{2+} or Ca^{2+}) is present.

The oxidation potentials for the bare anions indicate great oxidative stability for all the anions, although there exist some variations across the salts. The reduction potential of the CIP is a measure of how thermodynamically favorable reduction of the metal-coordinating anion is, at a certain potential. From Figure 5.1(b) we find that all the ion pairs show positive EAs at the Mg/Ca electrode potential, indicating they are all exergonic for single-electron reduction. It is worth noting that different levels of theory (B3LYP-D3(BJ)[27–29] vs $\omega\text{B97X-D3}$)[30] results in fairly different absolute numbers, especially for the reduction potentials, which could differ by up to ~ 0.5 eV, but the trend across the different ligands remains. The differences between the two levels of theory in oxidation potentials for the bare anions mostly originate from discrepancies in the optimal geometries for the two levels of theory (details given in Appendix Sec. 4.1).

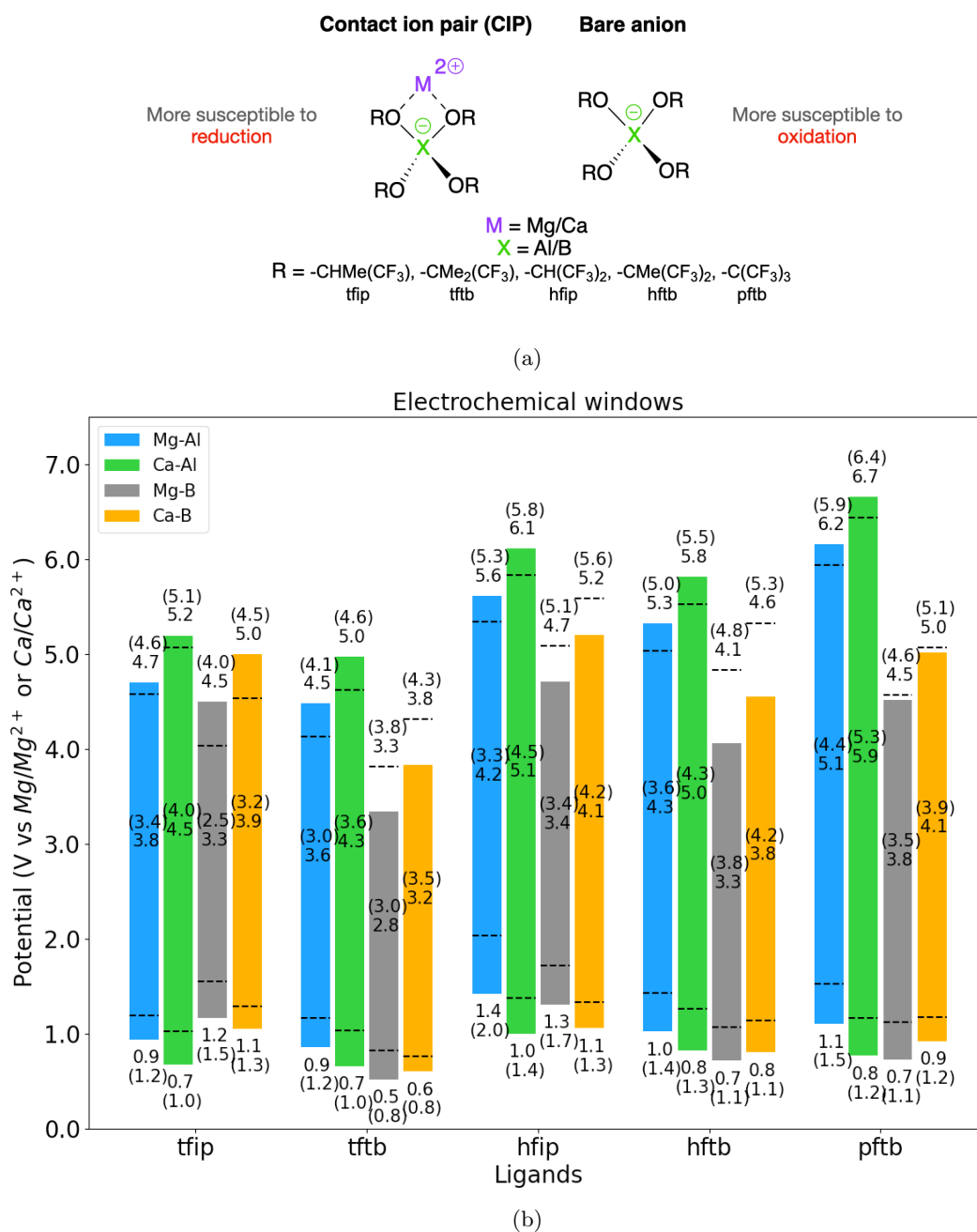


Figure 5.1: (a) Concept of contact ion pairs (CIP) and bare anions. (b) Electrochemical windows for the anions. Bare anions (X^- = monoanionic ligands) were used for oxidation potentials (upper limit) and contact ion pairs ($[M^{2+}X^-]^+$, $M = \text{Mg/Ca}$) were used for reduction potentials (bottom limit). Two levels of theory were used for comparison: the redox potentials shown in the regular bars are based on $\omega\text{B97X-D3/def2-TZVPPD/CPCM/RIJCOSX}/\omega\text{B97X-D3/def2-SVPD/CPCM/RIJCOSX}$ level of theory, whereas the black dashed lines and the numbers in parenthesis are from $\text{B3LYP-D3BJ/def2-TZVPPD/CPCM}/\text{B3LYP-D3BJ/def2-SVP/CPCM/RIJCOSX}$ level of theory. The numbers in the middle of the bars are the energy gaps between the two redox potentials (stability windows).

Across the ligands examined, we find that the tftb/hftb ligands have the potential to show relatively higher reductive stability while maintaining the weakly-coordinating behavior. However, redox potentials alone do not determine performance, e.g. $\text{Mg}^+[\text{B}(\text{HFIP})_4]^-$ shows relatively high reduction potential, yet it exhibits excellent electrochemical behavior experimentally. [13] High kinetic barriers for the subsequent decomposition reactions (after the initial one-electron reduction) can effectively inhibit detrimental decomposition, as has been hypothesized for the case of the acetonitrile (ACN) solvent in Mg systems. [31] In the following sections we present a comprehensive investigation of reduction-mediated decomposition pathways, motivated by NMR spectroscopy and elucidated by structure-chemical trends.

2.2 Decomposition of the salts

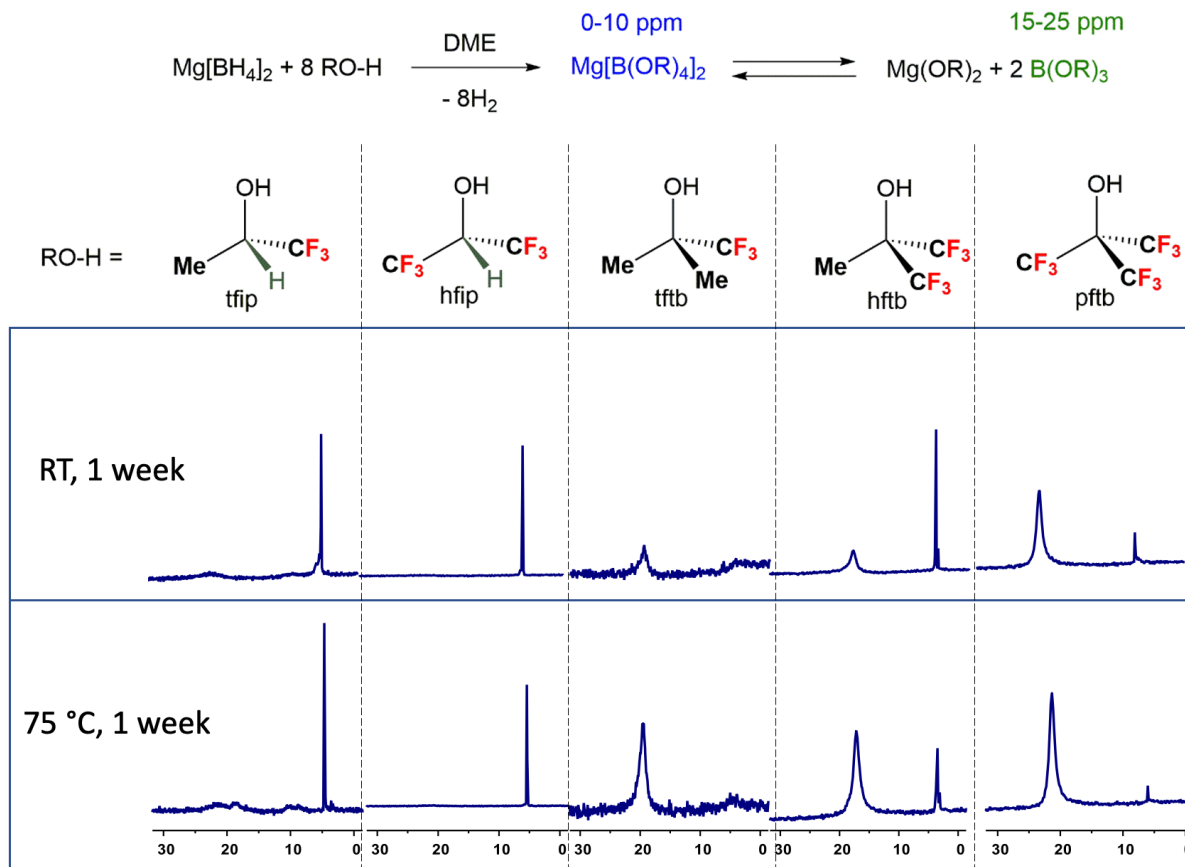
2.2.1 ^{11}B NMR experiments

To gain insight into the possible decomposition mechanisms of the salts, we conducted ^{11}B NMR experiments on the Mg borate salts in this family (Figure 5.2). Briefly, cooled solutions of $\text{Mg}(\text{BH}_4)_2$ in dimethoxyethane (DME) were reacted with its respective alcohol ($\text{ROH} = \text{tfip}, \text{hfip}, \text{tftb}, \text{hftb}$ and pftb), until no effervescence was observed. The reaction mixtures were heated mildly overnight to ensure complete conversion of borohydride starting material, as evidenced by the absence of doublets (0-20 ppm) in ^{11}B NMR spectra. Tfip is an exception as minor doublet signals were seen due to its low relative acidity. The mixtures were then left at either room temperature or 75 °C for one week and ^{11}B NMR spectra were recorded. For some cases (where $\text{ROH} = \text{tftb}, \text{hftb}$, and pftb), two unique peaks were observed, with characteristic signals corresponding to the anionic 4-coordinate B (0-10 ppm) and the neutral 3-coordinate B (15-25 ppm), [32–34] respectively.

The presence of both tri-/tetra-coordinate boron species indicates a rearrangement of alkoxide ligand from the boron center to a magnesium cation, when the alkoxide is sterically bulky enough to overcrowd the tetra-coordinate boron center. At elevated temperature, more $\text{B}(\text{OR})_3$ is produced through B-O bond breaking for the tftb, hftb and pftb ligands, accompanied by precipitation of solid byproducts and lowered signal to noise in the spectra. $\text{Mg}[\text{B}(\text{OR})_4]_2$ is stable at 75 °C when $\text{R} = \text{tfip}/\text{hfip}$. This result provides experimental evidence for B-O bond breaking decomposition even without reducing conditions, which motivated us to investigate the Al/B-O bond decomposition further computationally.

2.2.2 Decomposition barriers for the neutral ion pairs

We examine three types of decomposition patterns for the one-electron reduced CIPs, namely from breaking the bond that is closest to the metal ion in each of the following bond categories: (1) Al/B-O, (2) C-O, and (3) C-F. The Al/B-O bond breaking decomposition is

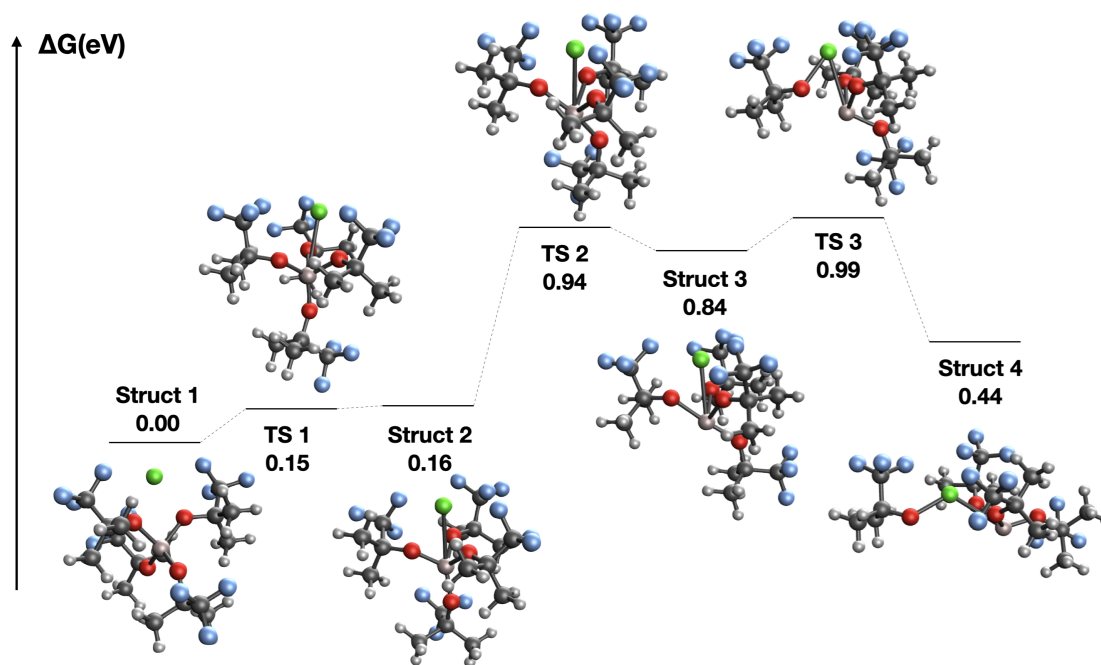
Figure 5.2: ^{11}B NMR experiments on the Mg borate salts.

discussed in Sec 2.2.1, the C-O bond breaking decomposition was proposed in ref.[16] and the C-F bond breaking decomposition is considered because MgF_2 and CaF_2 are observed in the SEI with this family of electrolytes.[15] Figure 5.3 shows exemplary free energy diagrams for the Al/B-O bond breaking decomposition, i.e. for $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$ and $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$. Figure 5.4 shows free energy diagrams for examples of the C-O and C-F bond breaking decompositions, i.e. for $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$. A complete set of energy diagrams for other salt variations are given in the Appendix, for all combinations of $\text{M}^+[\text{X}(\text{Y})_4]^-$ ($\text{M}=\text{Mg}/\text{Ca}$, $\text{X}=\text{Al}/\text{B}$, $\text{Y}=\text{tfip}, \text{tftb}, \text{hfip}, \text{hftb}, \text{pftb}$); a simplified version (showing energy levels only) of those energy diagrams for Al/B-O bond breakage is shown in Figure 5.5(a). We note that although all the ligand variations exhibit qualitatively similar Al/B-O breaking decomposition pathways for aluminates and borates, they can differ in the number of reaction steps and intermediates; in such cases, in Figure 5.5(a), we have aligned the horizontal positions of the energy levels for analogous structures across the series (e.g. where the degree of the Al-O/B-O bond dis-

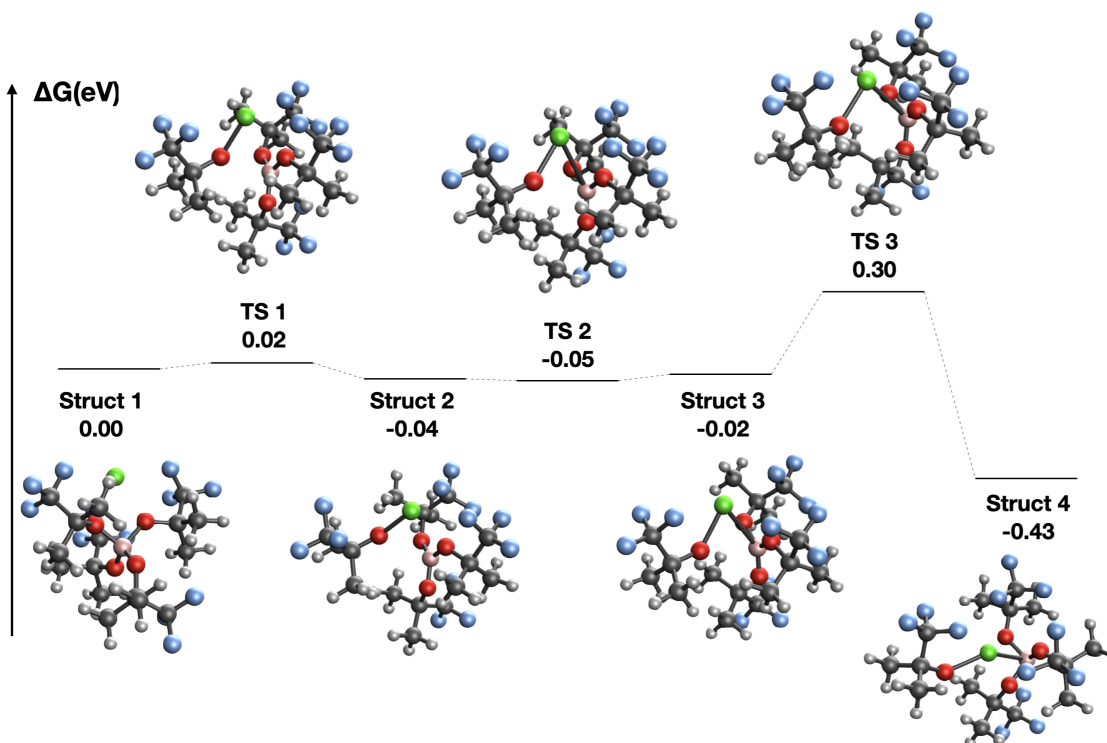
sociation matches, etc.). The rate-limiting step of a decomposition pathway is determined by the energy difference between the starting material (or the lowest-energy intermediate, if lower in energy than the starting material) and the highest-energy TS. The rate-limiting decomposition barriers for all considered bond breakages and ligand variations are given in Figure 5.5(b). For the C-O bond breaking paths, the rate-limiting barrier is shown in Figure 5.5(b) only when the barrier is less than 1.0 eV (see the discussions later in this section).

To identify the origin of the differences in the energy profiles, we analyze the Al/B-O bond breaking pathways from a structural perspective. In the case of the aluminates, taking $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$ as an example (Figure 5.3(a)), the initial bi-dentate (Al coordinating to two oxygens in the anion) rearranges over a low barrier (0.15 eV) to a tri-dentate CIP (Struct 2), followed by a rotation of the axial Al-O bond (Struct 2 $\rightarrow$ TS 2 $\rightarrow$ Struct 3). Subsequently the Al-O bond opposite (180°) to the aforementioned rotated Al-O bond breaks to create the decomposed product, $\text{Ca}^{2+}(\text{tftb})^-[\text{Al}(\text{tftb})_3]^-$ (Struct 3 $\rightarrow$ TS 3 $\rightarrow$ Struct 4). The initial rotation of the Al-O bond exhibits low barriers for all the ligand variations (< 0.3 eV), consistent with the longer Al-O bonds (~ 1.8 Å) as compared with the B-O bonds (~ 1.5 Å) which allows the aluminates to deviate more easily from the tetrahedral geometry, creating space for an extra electron. For borates, on the contrary, we see a more straightforward dissociation of one B-O bond, although in some cases, such as for $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$, minor geometrical-arrangement steps may precede the rate-limiting TS (TS 3) (Figure 5.3(b)). Unique among the borate salts are the complexes with the pftb ligand, where the B-O bond is already significantly dissociated at the intermediate geometries: see Struct 2 in Figure 5.25 for $\text{Mg}^+[\text{B}(\text{pftb})_4]^-$ where the B-O bond is 3.52 Å and Struct 2 in Figure 5.30 for $\text{Ca}^+[\text{B}(\text{pftb})_4]^-$ where the dissociated B-O bond is 4.05 Å. These dissociation lengths are reached before the principle TS, i.e. the one that is qualitatively conserved across all ligands and labeled as “TS” in Figure 5.5(a). Thus we consider the TS before those intermediates (TS 1 in Figures 5.25 and 5.30) the rate-limiting TS with very low barriers of 0.03 eV and 0.07 eV, respectively.

Comparing the full energy diagrams across the ligand variations shown in Figure 5.5(a) and the rate-limiting barriers shown in Figure 5.5(b), we note that the tftb ligand exhibits the highest barrier for decomposition across aluminate salts, whereas the hfip ligand exhibits the highest barrier across borate salts, regardless of whether Mg^+ or Ca^+ is the counter ion. Experimentally, there has been remarkable success with the $[\text{B}(\text{hfip})_4]^-$ anion, in both Mg-[13] and Ca-[14, 15] systems, consistent with the high decomposition barriers observed in our calculations. Specifically, the $\text{Mg}[\text{B}(\text{hfip})_4]_2$ salt has shown exceptional electrochemical performance in terms of oxidative stability (~ 4 V vs Mg), ionic conductivity, and coulombic efficiency. [13] The $\text{Ca}[\text{B}(\text{hfip})_4]_2$ salt also supports reversible Ca^{2+} deposition/stripping, although the accumulation of CaF_2 on the Ca surface indicates reductive salt instability. [15] For aluminates, there are some experimental investigations on the Mg salts while the Ca version of the salts are less reported on. $\text{Mg}[\text{Al}(\text{hfip})_4]_2$ demonstrated excellent coulombic efficiency, [11, 12] and $\text{Mg}[\text{Al}(\text{pftb})_4]_2$ showed oxidative and reductive stability[16]; they are



(a)



(b)

156
 Figure 5.3: Representative decomposition pathways through Al/B-O bond breakage. (a) $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$ decomposition pathway. (b) $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$ decomposition pathway.

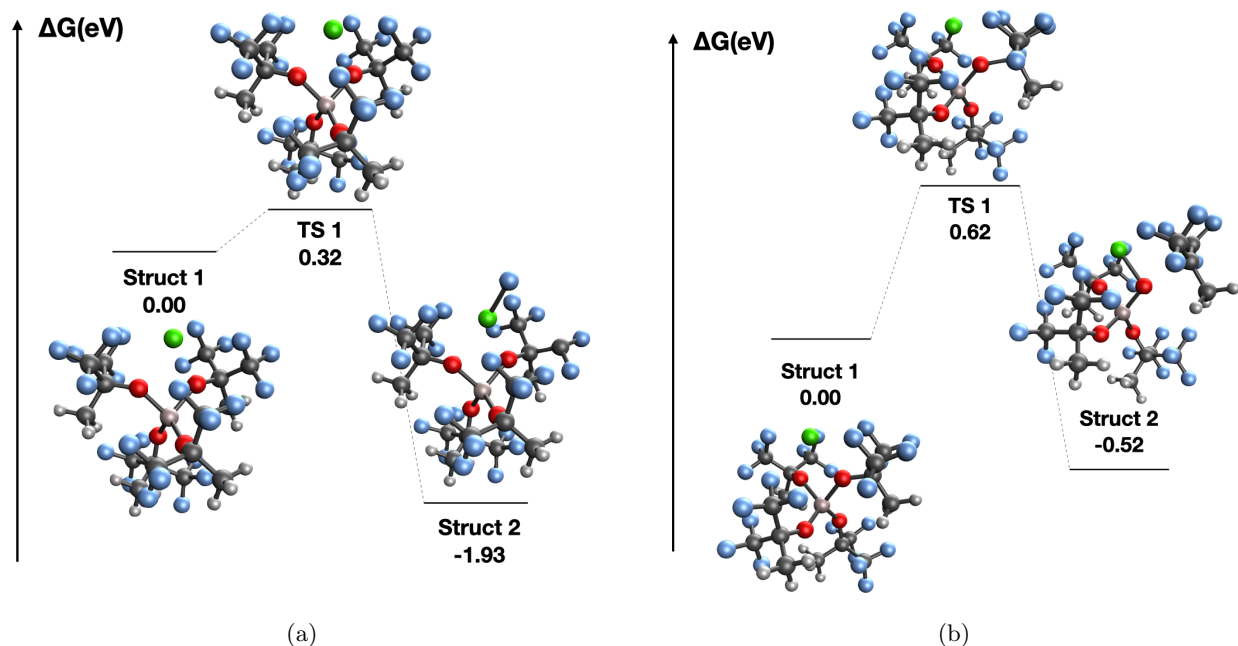
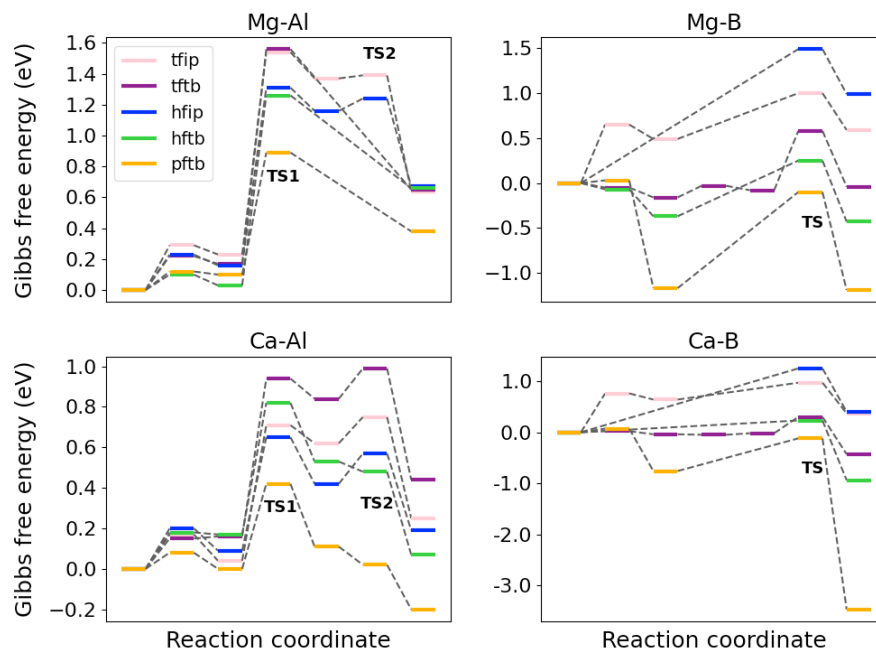


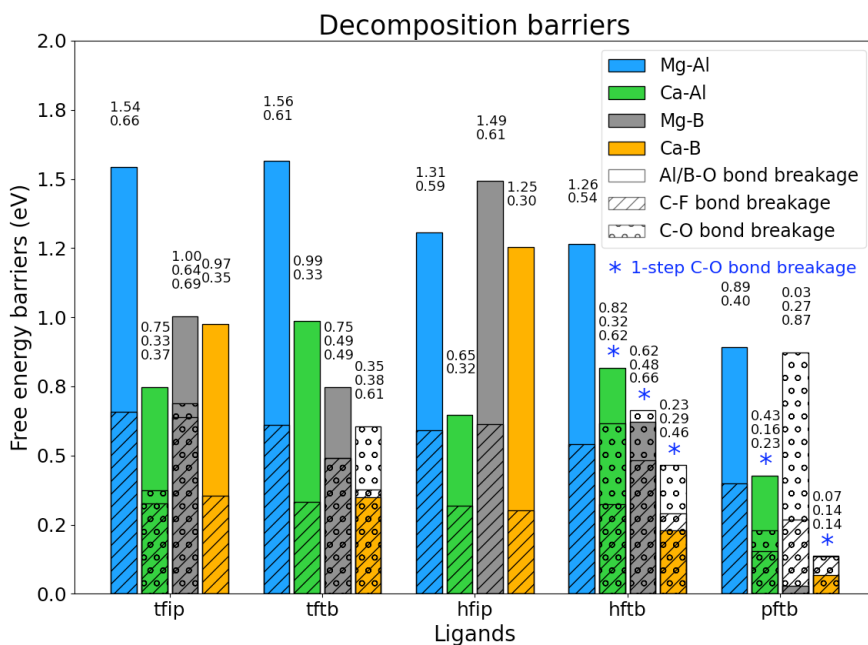
Figure 5.4: Representative decomposition pathways through C-O and C-F bond breakage. (a) $\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$ decomposition through C-F bond breakage. (b) $\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$ decomposition through C-O bond breakage.

predicted from our calculations to possess decomposition barriers of 1.31 eV and 0.89 eV, respectively. As the Ca salts (ion pairs) trend towards lower barriers than their Mg counterparts, we predict their performance will be the most affected by the specific ligand choice. For instance, our calculations indicate that the tftb ligand needs to overcome a 0.99 eV barrier to decompose whereas the pftb ligand only needs to overcome a 0.43 eV barrier; the difference in these numbers implies a significant difference in rate constants for decomposition at room temperature. We note that the barriers may not necessarily correspond quantitatively with the experimental rate constants, as DFT limitations and effects such as explicit solvation and solvent rearrangement are unaccounted for here, however, the predicted qualitative trend can be employed to guide the comparison between the ligand variations.

Examining the C-F bond breaking decomposition pathways, all variations of salts exhibit a straightforward one-step path for Ca^+ to abstract a fluorine atom from the $-\text{CF}_3$ group in the nearby alkoxy group. In Figure 5.4(a), we show one example of such a path, i.e. for $\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$, which showcases a low barrier of 0.32 eV. As shown in Figure 5.5(b) (and the energy diagrams in Appendix Sec. 4.4), such fluorine-abstracting decomposition pathways are highly exergonic ($\Delta G \sim -2.0$ eV) with low barriers for all salt variations (the largest barrier is 0.66 eV for $\text{Mg}^+[\text{Al}(\text{tfip})_4]^-$). Our calculations also indicate that the Ca salt barriers



(a)



(b)

Figure 5.5: Decomposition pathways and rate-limiting barrier heights. (a) Decomposition pathways for Al-O/B-O bond breakage. (b) Rate-limiting decomposition free energy barriers for Al-O/B-O, C-F, and C-O bond breakage. The numbers on each bar from the top to bottom are the rate-limiting barriers for Al/B-O bond breakage, C-F bond breakage, and C-O bond breakage. Note that the C-O bond breaking barrier is only shown when it is less than 1.0 eV.

generally trend lower than their Mg counterparts, consistent with experimental observations that increased CaF_2 formation on the anode surface, as compared to MgF_2 . [15] Although low barriers and large negative reaction ΔG s are observed here, we note that the first coordination sphere of the metal ion is saturated with solvents, e.g. Mg^{2+} is sixfold coordinated by oxygen atoms from both solvent molecules and the salt anions for $\text{Mg}(\text{TFSI})_2$ in diglyme, [35] and this aspect is missed in our calculations. We also speculate that mitigating reactions such as hydrogen abstraction may provide avenues to restore the salt efficacy in this case, in contrast to e.g. the more detrimental Al/B-O bond breaking reactions which will likely cause an irreversible salt breakdown.

Finally, a single-step C-O bond breaking decomposition pathway was identified for some salt variations, mainly Ca salts with the hftb or pftb ligands ($\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$, $\text{Ca}^+[\text{Al}(\text{pftb})_4]^-$, $\text{Ca}^+[\text{B}(\text{hftb})_4]^-$, $\text{Ca}^+[\text{B}(\text{pftb})_4]^-$ and $\text{Mg}^+[\text{B}(\text{hftb})_4]^-$, see bars labeled with blue stars in Figure 5.5(b) and the complete energy diagrams are given in Appendix Sec. 4.3). For borates, a B-O bond would break simultaneously with the C-O bond breaking, consistent with observations by Jankowski et al. [22] For single TS reactions, the barriers for C-O bond breaking decomposition are comparable in magnitude (usually very low) to that for Al/B-O bond decomposition. Hence, C-O bond mediated decomposition is expected to manifest for these salts. For the other salts, a multistep path involving the product from C-F bond breakage was identified, possessing a high rate-limiting barrier due to a stable intermediate state (see Appendix Sec. 4.3). Nevertheless, we have included the rate-limiting barriers in Figure 5.5(b) if they are less than 1.0 eV.

Considering the kinetic barriers for all types of decomposition pathways, we identified the tftb ligand as the most promising ligand for aluminate salts because it exhibits the highest barrier for decomposition under reductive conditions, whereas the hfip ligand shows the most promise in the borate salt series, for both Mg- and Ca- systems.

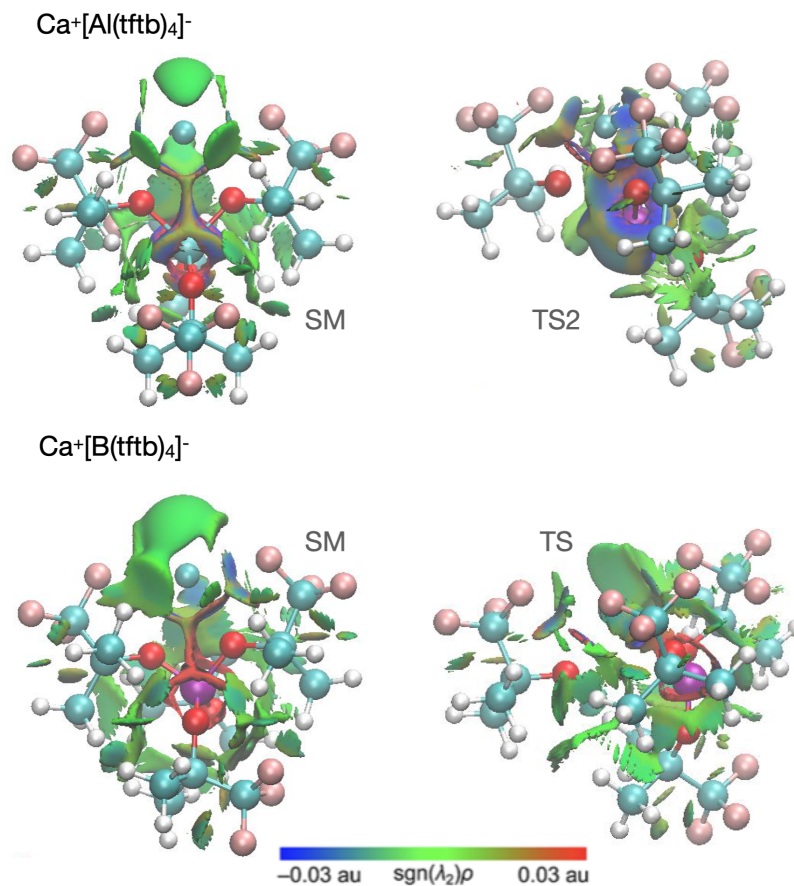
2.3 DFT investigation of the difference in decomposition kinetics between aluminates and borates

DFT-based analysis was undertaken to elucidate the apparent differences in the kinetic barriers across the ligand variations for the borate/aluminate salts, i.e. the high barrier of the tftb (hfip) ligand for aluminates (borates), respectively.

We divided the ion pair complexes into different structural components to qualitatively isolate the substructure that is responsible for the barrier variations. In the current investigation, we performed two kinds of division, as shown in Figure 5.7(a). The aluminum ion pair complexes were divided into the M-OR (M=Ca/Mg) and the $\text{Al}(\text{OR})_3$ substructures (Figure 5.7(a): left); the boron ion pair complexes were divided into the core (B, O, and Ca/Mg) and the ligands (Figure 5.7(a): right). In the following section, for simplicity, we focus on the Ca salts with the following three ligands (tftb, hfip, hftb) for investigation of the barrier differ-

ences. We consider these three examples to be representative, as they occur in the middle of the electronegativity/bulkiness range of the five ligands. We note the conclusions obtained here can be extended to the tfip and pftb ligands as well as the Mg version of these salts.

We analyzed the charge distribution across these complexes for both the starting material (SM) and the TS geometries, based on the natural atomic charges obtained from the Natural Bond Orbital (NBO) analysis.[36] In Figure 5.7(b) we plot the change in natural charge for the Ca atom (on the left) and the change in the sum of the natural charges on the CaOR fragment (on the right) for the SM, TS, and product geometries. Note that the set of aluminate rate-limiting TSs, i.e. one for each ligand variant, comprises two distinct structure types, corresponding to “TS1” and “TS2” on the energy diagram shown in Figure 5.5(a). However, there is only one rate-limiting TS structure type for the borates (plotted as “TS” in Figure 5.5(a) or “TS2” on Figure 5.7(b)). As the reaction proceeds, the Ca oxidation state changes from Ca^+ to Ca^{2+} , which is expected and aligns well with previous studies. [6] The Ca natural charges for the borates at the TSs are significantly smaller than those for the aluminates (both TS1 and TS2), indicating that the aluminate TSs show more charge-separated character. This is also reflected in the sum of natural charges on the CaOR fragments: the aluminate TSs are close to $[\text{CaOR}]^+[\text{Al}(\text{OR})_3]^-$ in character while the CaOR components in borate TSs are approximately neutral. The amount of charge transfer can also be deduced from plotting the iso-surface of the spin density, which is more well-defined than the atomic partial charges as there are numerous possible definitions of the latter. The spin density plots are shown in Appendix Figures 5.60-5.65. To further elucidate why the aluminate TS exhibits an enhanced charge transfer character, we calculated the EAs for $\text{Al}(\text{OR})_3$ and $\text{B}(\text{OR})_3$. In Appendix Sec. 4.7 we show three types of such values, namely the adiabatic electron affinity (EA), vertical attachment energy (VAE), and vertical detachment energy (VDE), the latter two representing the vertical electron affinities at the optimized neutral and anion geometries respectively.[37] An apparent distinction between $\text{B}(\text{OR})_3$ and $\text{Al}(\text{OR})_3$ (see Figure 5.59) is that all the $\text{B}(\text{OR})_3$ compounds show negative adiabatic EAs whereas the $\text{Al}(\text{OR})_3$ EAs are positive, indicating the $\text{B}(\text{OR})_3$ compounds are less amenable to reduction. Even in the case of the optimal anion geometries (bent), the VDEs for $\text{B}(\text{OR})_3$ present more than 1 eV lower than their $\text{Al}(\text{OR})_3$ counterparts. Even though full charge transfer must be reached in the end product, the borate salts evolve through neutral $\text{B}(\text{OR})_3$, which lowers the energy. This phenomenon can be indeed observed from the energy diagrams: e.g. in Figure 5.3(b) the BO_3 moieties are close to planar (as is optimal for isolated neutral $\text{B}(\text{OR})_3$) for all TS geometries whereas for aluminates the AlO_3 moieties are easily bent (i.e. tetrahedral-like, which is optimal for isolated anionic $\text{Al}(\text{OR})_3$) from early on (Figure 5.3(a)).

Figure 5.6: NCI plot of the $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$ and $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$ complexesTable 5.1: Selected parameters (The Laplacian of the density $\nabla^2\rho$, density at the Bond Critical Point ρ @BCP(3,-1), Mayer bond order m , Bader atomic charges q) for the QTAIM analysis of the starting complexes)

structure	bond path	bond length (Å)	$\nabla^2\rho$ (a.u.) @BCP(3,-1)	ρ (a.u.) @BCP(3,-1)	m	$q(\text{Al/B}), q(\text{O})$
$\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$	Al-O	1.802	0.602	0.075	0.584	2.553, -1.373
$\text{Ca}^+[\text{Al}(\text{hftp})_4]^-$	Al-O	1.803	0.599	0.074	0.578	2.566, -1.365
$\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$	Al-O	1.798	0.608	0.075	0.592	2.565, -1.363
$\text{Ca}^+[\text{B}(\text{tftb})_4]^-$	B-O	1.540	0.445	0.126	0.715	2.369, -1.312
$\text{Ca}^+[\text{B}(\text{hftp})_4]^-$	B-O	1.506	0.548	0.138	0.765	2.369, -1.320
$\text{Ca}^+[\text{B}(\text{hftb})_4]^-$	B-O	1.568	0.378	0.116	0.671	2.378, -1.305

The joint analysis of the bonding structure of the SMs (the CIPs) by the quantum theory

of atoms in molecules (QTAIM[38]) and the more qualitative noncovalent interaction (NCI) plot[39, 40] method reveals the covalent and non-covalent nature of the Al-O and B-O bonds. Table 5.1 lists the information extracted from the QTAIM study, i.e., electron-density values ρ and values of the laplacian of the density $\nabla^2\rho$ at bond critical points (BCPs, i.e. (3,-1)-type points)[38], and from the analysis of the corresponding Bader atomic charges q and Mayer bond orders m . The electron density at the BCP indicates the amount of shared density between the bonded atoms and the Laplacian of the electron density depicts the extent of charge concentration or depletion.[38, 41] The fact that the ρ values at the critical points are of the magnitude of 7×10^{-2} (for aluminates) and 1×10^{-1} (for borates) and that the $\nabla^2\rho$ values are larger than zero indicate the Al-O and B-O bonds in the starting tetrahedral ion pairs are both in the ionic regime,[42] with B-O being more covalent in character.

To further visualize the types of interactions involved in the Al-O or B-O bonding, we employed the NCIPLOT method which utilizes the reduced density gradient as an indicator of non-covalent interactions, as characterized by large density gradients in the regions where the density itself is small. The type of NCI can be further classified by the sign of the second eigenvalue λ_2 of the Hessian of ρ , i.e. stabilizing and bonding-like interactions are present when $\lambda_2 < 0$, destabilizing or repulsive interactions when $\lambda_2 > 0$, and finally dispersion interactions when λ_2 is close to 0. These regions can be conveniently visualized from the blue, red, and green colors, respectively, on an NCI plot. Figure 5.6 shows NCI plots for the $\text{Ca}^+[\text{Al}(\text{hfip})_4]^-$ and $\text{Ca}^+[\text{B}(\text{hfip})_4]^-$ complexes, where in the case of $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$ an attractive NCI isosurface is clearly present in between the Al-O bonds, whereas the NCI isosurface of $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$ is repulsive, corroborating the more ionic nature of the Al-O bond vs B-O bond. The blue attractive region is enlarged in the TS2 for $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$ compared with the SM, mostly likely due to the increased favorable electrostatic interactions from the charge separation as discussed earlier.

We reiterate that even at TS1, the amount of charge transfer in the aluminates is significantly larger than for the borates, as reflected in both the Ca natural charges and the spin density plots shown in Appendix Figures 5.60-5.65. The results indicate that almost an entire electron is transferred from Ca^+ to the $\text{Al}(\text{OR})_3$ component at the stage of the TS (regardless of whether it is TS1 or TS2). To estimate the electronic contribution of the CaOR and $\text{Al}(\text{OR})_3$ fragments to the energy barrier of the reaction, we perform single point calculations for each fragment at the geometry it has within the full-complex for SM, TS1 and TS2, at the B3LYP-D3(BJ)[27–29]/def2-TZVPPD[43, 44]/CPCM[45] level of theory. Different total charges were imposed for the fragments in the SM and TS to account for the charge separation, e.g. for the CaOR fragment, a total charge of 0 was used for the SM configuration whereas a total charge of +1 was used for the TS configuration. The fragment-only barrier is then the difference between the fragment energy in its TS configuration and the fragment energy in its SM configuration (Eqn 5.1 and 5.2). We also computed the binding energies of these fragments in the SM and TS geometries. For example, the SM binding energy is the difference between the full-complex energy and the sum of the fragment energies in their SM

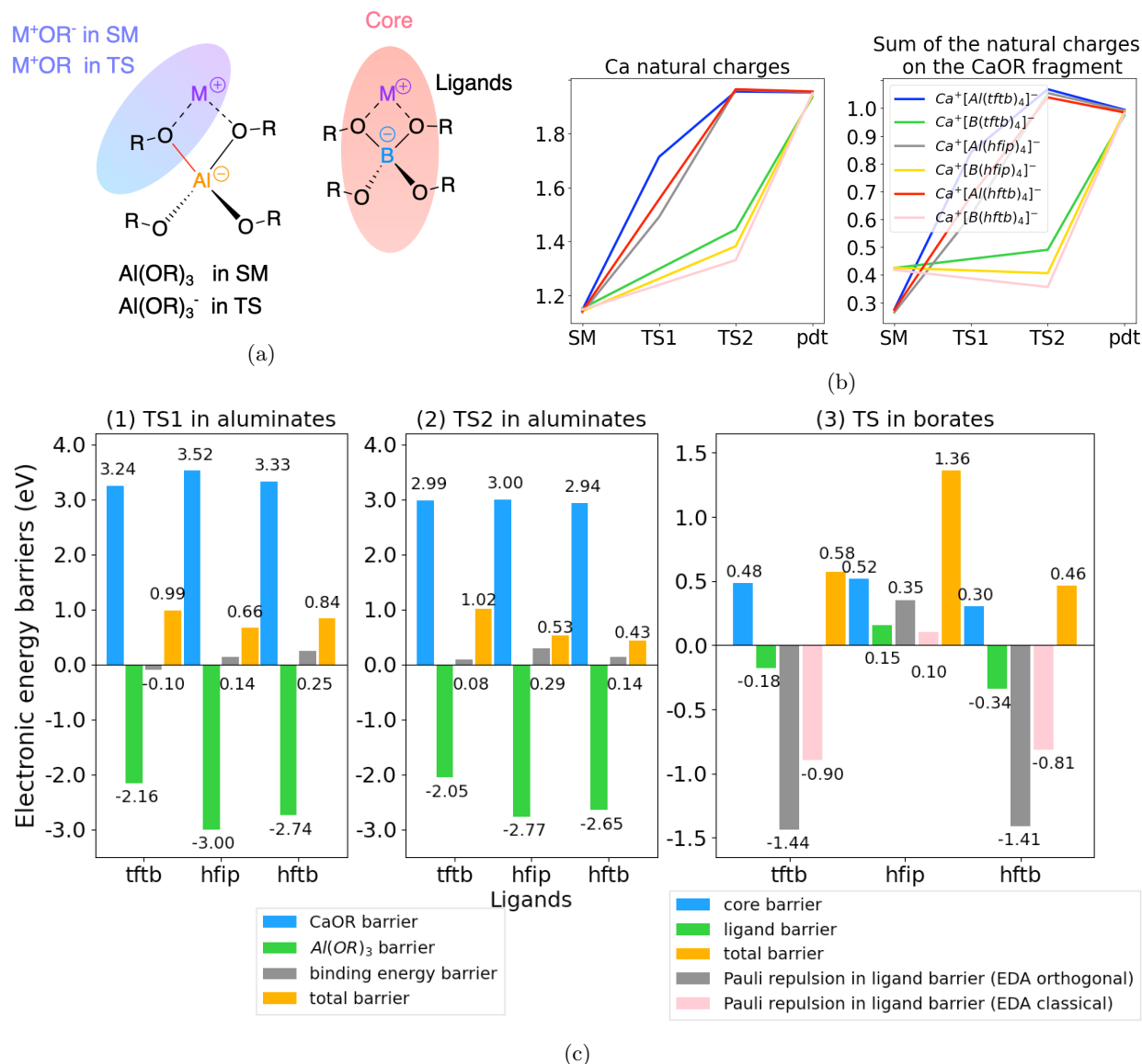


Figure 5.7: Al/B-O bond breaking barrier analysis for aluminates/borates. (a) Illustrative picture of how the complexes are divided into two regions for probing the origin of the barrier differences. (b) Natural charge analysis for the Ca and the CaOR (OR=tftb,hfip,hftb) fragment. (c) Electronic energy barriers for different fragments. (1) The electronic energy barriers for the CaOR and the $Al(OR)_3$ fragments for the TS1 geometry (see Figure 5.5(a)). Note that different total charges are used for the SM and TS in computing the barriers for each fragment. The CaOR barriers are calculated from the $[CaOR]^0 \rightarrow [CaOR]^+$ transition; the $Al(OR)_3$ barriers are calculated from the $[Al(OR)_3]^0 \rightarrow [Al(OR)_3]^-$ transition. Similarly, the binding energy barriers correspond to subtracting the binding energy between $[CaOR]^0 - [Al(OR)_3]^0$ in the SM from the binding energy between $[CaOR]^+ - [Al(OR)_3]^-$ in the TS. The CaOR barrier, the $Al(OR)_3$ barrier, and the binding energy barrier sum to the total energy barrier. (2) is the same as (1) but is for the TS2 geometry in aluminates. (3) Barriers for the borates calculated with the core part of the ligand part taken from the original SM and TS geometries.

configurations. The difference between the binding energies for the TS and SM geometries is a binding energy barrier (Eqn 5.3).

$$\Delta E_{\ddagger}(\text{CaOR}) = E([\text{CaOR}]^+ \text{ at TS geom}) - E([\text{CaOR}]^0 \text{ at SM geom}) \quad (5.1)$$

$$\Delta E_{\ddagger}(\text{Al(OR)}_3) = E([\text{Al(OR)}_3]^- \text{ at TS geom}) - E([\text{Al(OR)}_3]^0 \text{ at SM geom}) \quad (5.2)$$

$$\begin{aligned} \Delta E_{\ddagger}(\text{binding}) = & (E(\text{TS}) - E([\text{CaOR}]^+ \text{ at TS geom}) - E([\text{Al(OR)}_3]^- \text{ at TS geom})) \\ & - (E(\text{SM}) - E([\text{CaOR}]^0 \text{ at SM geom}) - E([\text{Al(OR)}_3]^0 \text{ at SM geom})) \end{aligned} \quad (5.3)$$

We plot these energies, as well as the fragment-only barriers in the first two diagrams in Figure 5.7(c). Since the rate-limiting TS is different for $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$, $\text{Ca}^+[\text{Al}(\text{hfip})_4]^-$ and $\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$, it is important to consider both TS1 and TS2 in the analysis. We observe from Figure 5.7(c) that the total electronic energy barrier calculated from both TS1 and TS2 is distinctly higher for $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$ than that for $\text{Ca}^+[\text{Al}(\text{hfip})_4]^-$ and $\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$. The same trend is found for the fragment-only electronic energy barriers for the $[\text{Al(OR)}_3]^0(\text{SM geom.}) \rightarrow [\text{Al(OR)}_3]^- (\text{TS geom.})$ transition (the green bars in Figure 5.7(c)(1) and (2)), while the corresponding data for the CaOR fragment is roughly constant for the ligand series. Hence, we suggest that the Al(OR)_3 fragment effectively determines the barriers and that there exists a significant ligand effect for stabilizing the $[\text{Al(OR)}_3]^-$ fragment. The more electron-withdrawing ligands (hfip and hftb) exhibit lower barriers, consistent with the expected anion-stabilizing effects of such groups.

We performed analogous fragment-based calculations on the three borates (see Figure 5.56), however in this case the fragments of the SM and TS configurations were assigned neutral charges to reflect the low amount of charge separation in the full-complex TSs. Similar to the aluminate case, the B(OR)_3 fragment-only barriers roughly correlate with the overall barriers, but not with the electron-withdrawing nature of the ligands. This is unsurprising given that the fragments in the calculations are neutral.

Therefore, the charge transfer analysis does not explain the high barrier of the hfip anion within the borate family. To move forward, we note that the hfip ligand possesses a single H atom in the alkoxy group, while the other two borate anions have a methyl group there, which increases the steric hindrance. To quantify the inter-ligand steric repulsions, we divided the SM and TS structures into the core and ligand substructures (Figure 5.7(a)) and analyzed the electronic barrier for each component, similar to the above procedure for the CaOR and Al(OR)_3 fragments. Each broken covalent bond was compensated with H atoms, whose positions were optimized while holding the other atoms fixed.

We find that the resulting ligand barriers correlate well with the actual barriers (green bars in Figure 5.7(c)(3)), and the ligand barrier differences between $\text{Ca}^+[\text{B}(\text{hfip})_4]^-$ and the other two salts explain $\sim 50\%$ of the actual barrier differences. Furthermore, importantly, we observe that the barrier of the ligand-only subsystem of $\text{Ca}^+[\text{B}(\text{hfip})_4]^-$ is significantly

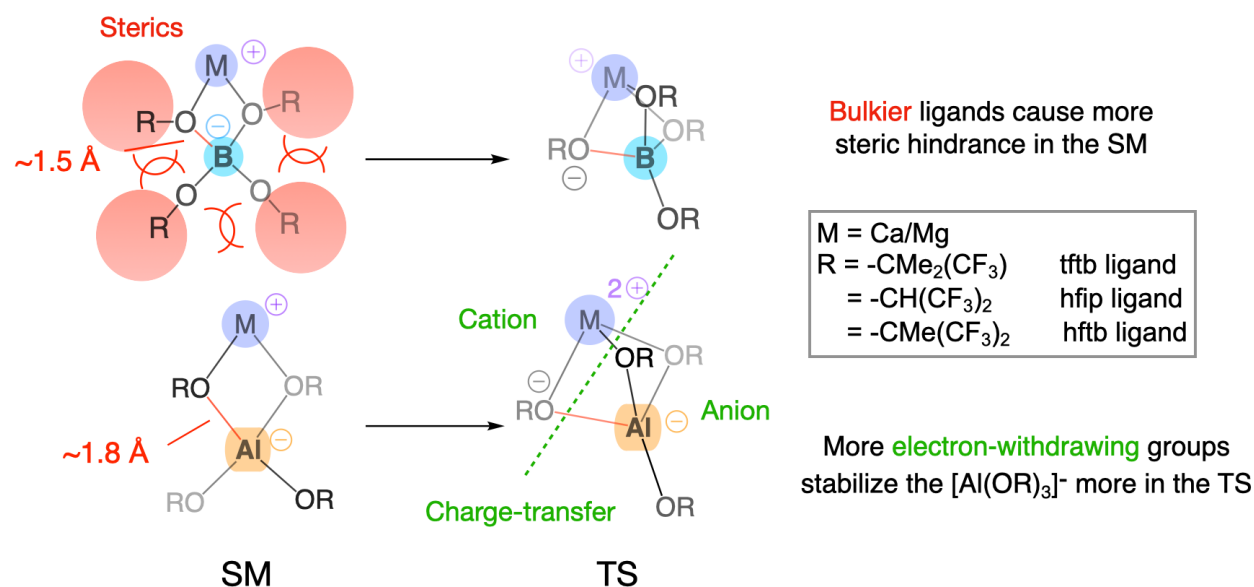


Figure 5.8: Speculative scheme for the kinetic control of the decomposition pathways (Al-O/B-O bond breakage)

different from those for $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$ and $\text{Ca}^+[\text{B}(\text{hftb})_4]^-$, indicating a distinction in the inter-ligand interactions in the hfip complex. We used the second-generation absolutely localized molecular orbital energy decomposition analysis (ALMO-EDA2)[46–48] technique to further analyze the various energy components of the interactions between the four bare ligands. The EDA analysis divides the fragment interaction energy into frozen, polarization, and charge transfer contributions. The frozen part can be further divided into several other physically relevant terms, i.e. electrostatics, Pauli repulsion, and dispersion, either through the “classical” or “orthogonal” approaches,[47] where the Pauli repulsion is the term that corresponds most naturally to the concept of steric repulsion. Indeed, we find among the EDA terms, the Pauli repulsion term correlates well with the decomposition barrier for borates: for $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$ and $\text{Ca}^+[\text{B}(\text{hftb})_4]^-$ there was an apparent decrease in the Pauli repulsion term going from the SM to TS (-1.44 eV with the classical approach or -0.90 eV with the orthogonal approach for tftb, -1.41 or -0.81 eV for hftb) whereas $\text{Ca}^+[\text{B}(\text{hfip})_4]^-$ showed a slight increase (0.35 or 0.1 eV). These results support the hypothesis that more steric repulsion exists for the SMs of $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$ and $\text{Ca}^+[\text{B}(\text{hftb})_4]^-$, which resolves once one of the ligands dissociates from the tetrahedral geometry and releases the steric strain. We note that the analogous EDA analysis performed on the aluminates did not show any meaningful correlation with the barriers, as the magnitude of all the EDA components is small (detailed numbers from the EDA analysis are given in the Appendix Sec. 4.6).

To summarize our findings, Figure 5.8 illustrates the origin of the kinetic behaviors of the borate and aluminate salt decomposition pathways through Al-O/B-O bond cleavage. For borate salts, since the B-O bonds are short ($\sim 1.5 \text{ \AA}$), steric hindrance is significant for bulky ligands confined in the tetrahedral geometry. Comparing the tftb, hfip, and hftb ligands, the hfip ligand is the least bulky, thus the starting material is less sterically hindered and the highest barrier for breaking the B-O bond is observed. In the case of aluminates where Al-O bonds are longer ($\sim 1.8 \text{ \AA}$), steric factors are less important and the electronic effects play an essential role in determining the barrier. This is due to the highly charge-transferred nature of the TS (compared to that of the borates), which is stabilized more by functional groups that are more electron-withdrawing. The tftb ligand is the least electron-withdrawing among the three ligands, and therefore the highest barrier is observed.

III Conclusion

Motivated by the recent interest in fluorinated alkoxyaluminate/borate electrolytes for multivalent energy storage applications, we performed NMR spectroscopy and an extensive computational investigation of electrochemical windows and the reductive decomposition pathways through various bond breaking patterns for a series of fluorinated alkoxyaluminate/borate salts. Kinetic barriers are obtained, for the first time, for the decomposition reactions which rationalize the reductive behavior and hence electrochemical performance. Across the ligand variations studied, we predict the Al-tftb and B-hfip combinations to be promising candidates based on the calculated high barriers for decomposition through Al/B-O bond breaking under reductive conditions. Our theoretical predictions support previous experimental reports of excellent electrochemical behavior in both Mg and Ca systems for borate salts with hfip ligands. To our knowledge, aluminate salts with tftb ligands have not yet been experimentally studied and investigations are currently underway. In addition to the thermodynamic and kinetic results presented here, our mechanistic investigations suggest a significant difference between alkoxy-borate and alkoxy-aluminate salts in factors controlling the kinetics for Al/B-O bond breaking decomposition: variations in decomposition barriers for borate salts are mostly determined by steric factors, whereas electronic effects are more important in aluminate salts, suggesting that different design rules should be considered under different scenarios, which will aid in the future design of functional electrolytes.

IV Appendix

4.1 Supplementary electrochemical windows

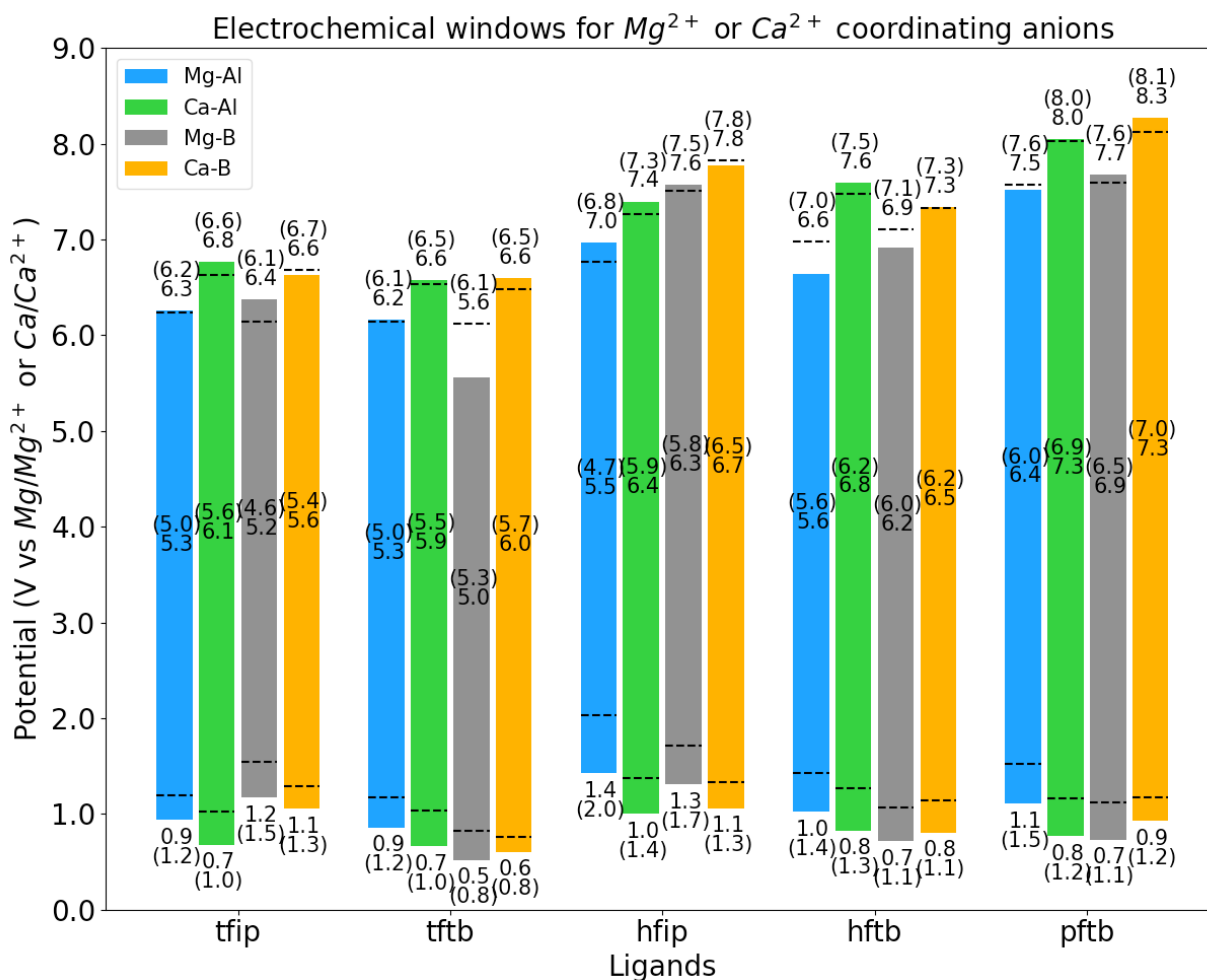


Figure 5.9: Electrochemical windows for CIPs. Two level of theories were used for comparison: the redox potentials shown in the regular bars are based on ω B97X-D3/def2-TZVPPD/CPCM/RIJCOSX// ω B97X-D3/def2-SVPD/CPCM/RIJCOSX level of theory, whereas the black dashed lines and the numbers in parenthesis are from B3LYP-D3BJ/def2-TZVPPD/CPCM//B3LYP-D3BJ/def2-SVP/CPCM level of theory. The numbers in the middle of the bars are the energy gaps between the two redox potentials (stability windows).

During geometry optimization, several CIPs changed their connectivities. A C-C bond breakage was observed for $[MgAl(pftb)_4]^{2+}$, $[MgB(pftb)_4]^{2+}$, $[CaAl(pftb)_4]^{2+}$ $[CaB(pftb)_4]^{2+}$ at

both levels of theory and for $[\text{MgB}(\text{tftb})_4]^{2+}$ at the $\omega\text{B97X-D3/def2-TZVPPD/CPCM/RIJCOSX//}\omega\text{B97X-D3/def2-SVPD/CPCM/RIJCOSX}$ level of theory. Since the oxidation reactions of these CIPs are already very endergonic i.e. oxidation potentials are on the order of ~ 7 eV, we consider these geometrical changes not of much significance.

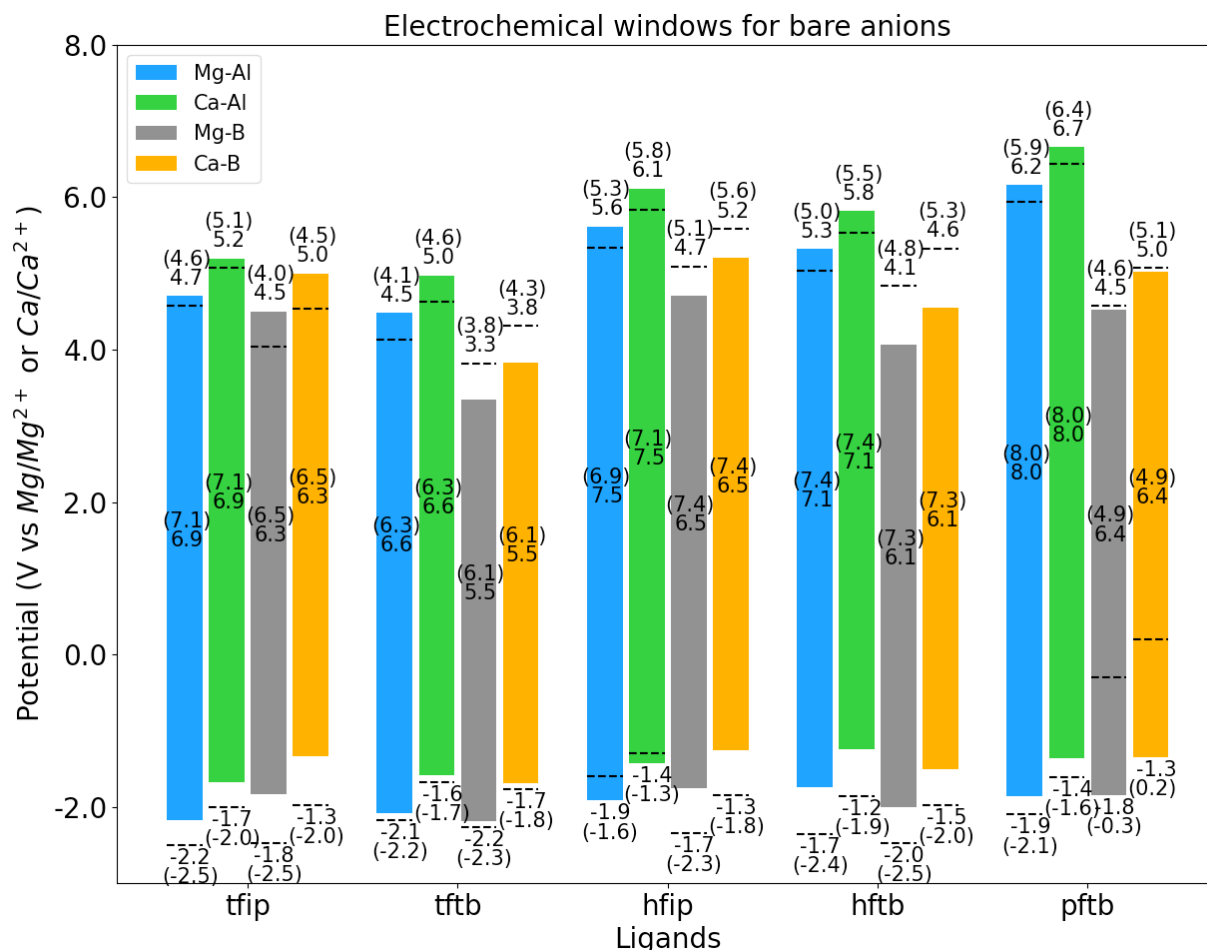


Figure 5.10: Electrochemical windows for bare anions. Two levels of theory were used for comparison: the redox potentials shown in the regular bars are based on $\omega\text{B97X-D3/def2-TZVPPD/CPCM/RIJCOSX//}\omega\text{B97X-D3/def2-SVPD/CPCM/RIJCOSX}$ level of theory, whereas the black dashed lines and the numbers in parenthesis are from $\text{B3LYP-D3BJ/def2-TZVPPD/CPCM//B3LYP-D3BJ/def2-SVP/CPCM}$ level of theory. The numbers in the middle of the bars are the energy gaps between the two redox potentials (stability windows).

Similar to the previous figure, we observed connectivity changes for the following anions.
 $[\text{B}(\text{tftb})_4]^0, [\text{B}(\text{hftb})_4]^0$: B-O bond breakage at the $\omega\text{B97X-D3}$ level of theory;
 $[\text{B}(\text{pftb})_4]^0$: B-O bond breakage at both levels of theory;

$[\text{B}(\text{hfp})_4]^{2-}$: C-H bond breakage at the level of theory;

$[\text{B}(\text{pftb})_4]^{2-}$: B-O bond breakage at the B3LYP-D3BJ level of theory.

These geometrical differences can partially account for the energy discrepancies seen between the two different levels of theory.

4.2 Decomposition pathways through Al-O/B-O bond breakage

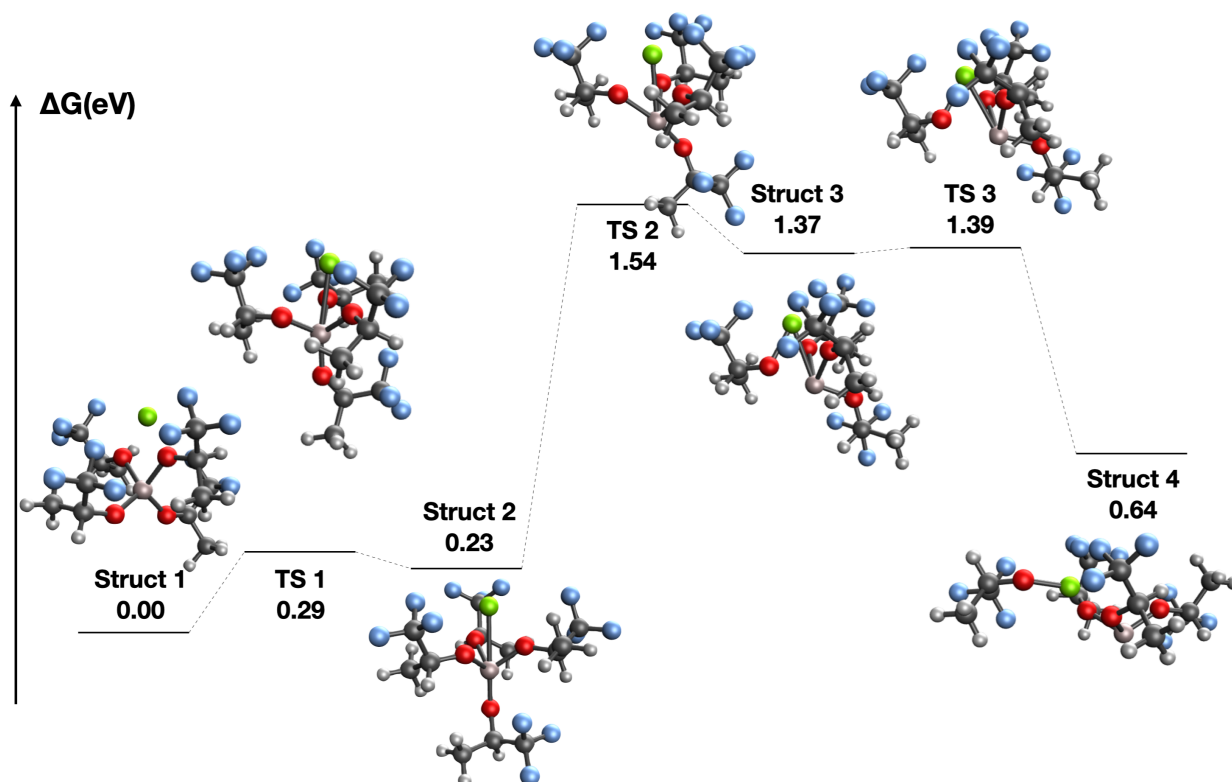


Figure 5.11: $\text{Mg}^+[\text{Al}(\text{tfip})_4]^-$ decomposition pathway through Al-O bond breakage

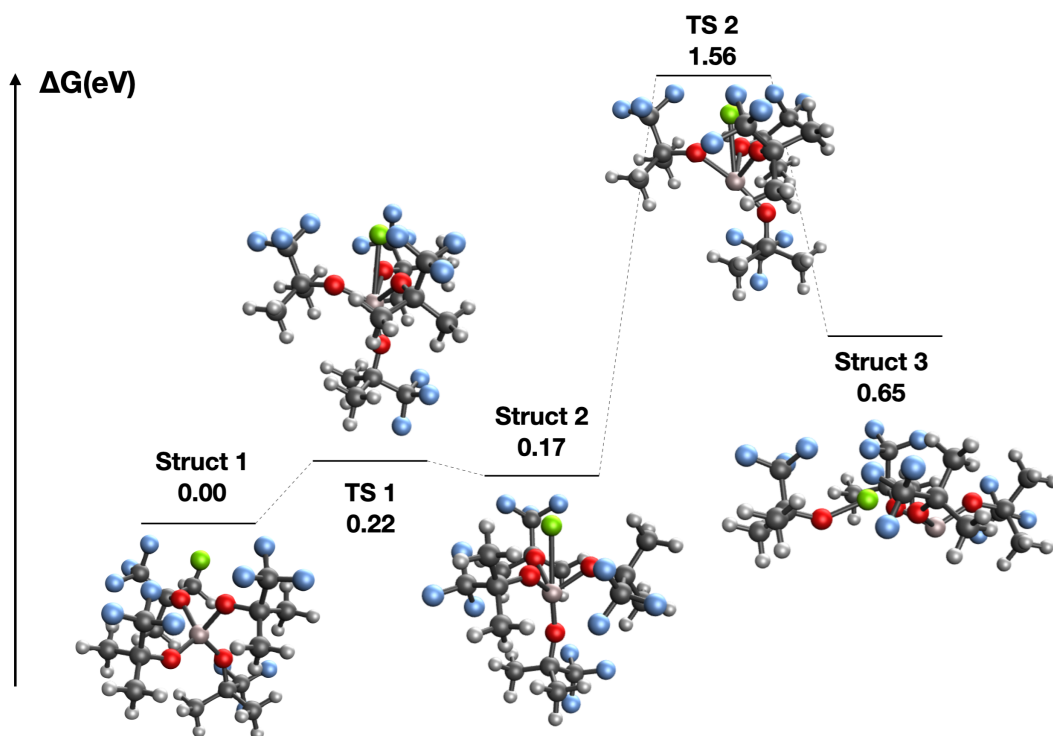


Figure 5.12: $\text{Mg}^+[\text{Al}(\text{tftb})_4]^-$ decomposition pathway through Al-O bond breakage

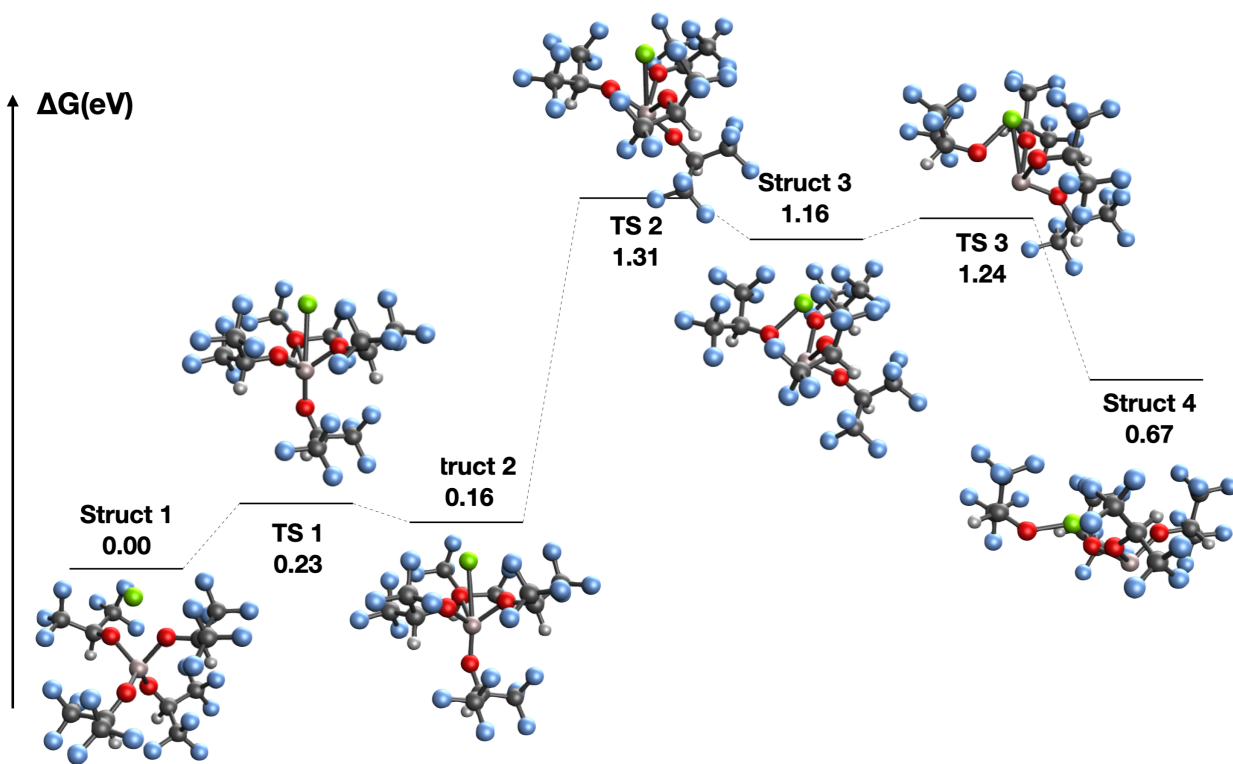


Figure 5.13: $\text{Mg}^+[\text{Al}(\text{hfip})_4]^-$ decomposition pathway through Al-O bond breakage

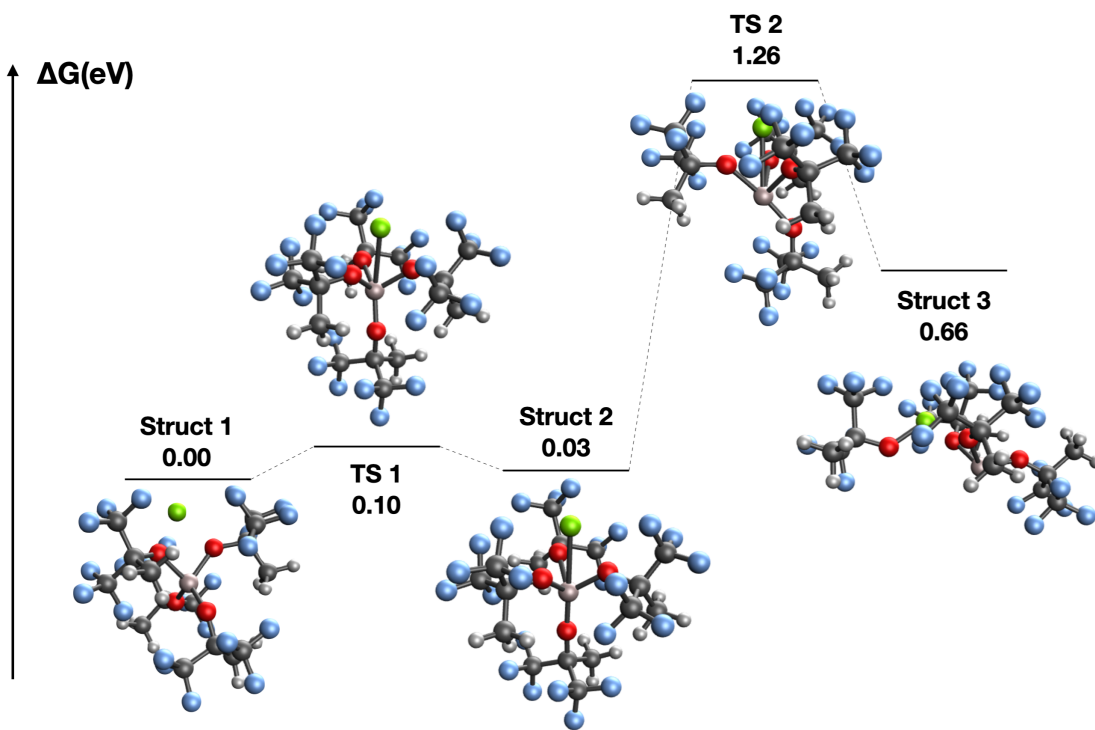


Figure 5.14: $\text{Mg}^+[\text{Al}(\text{hftb})_4]^-$ decomposition pathway through Al-O bond breakage

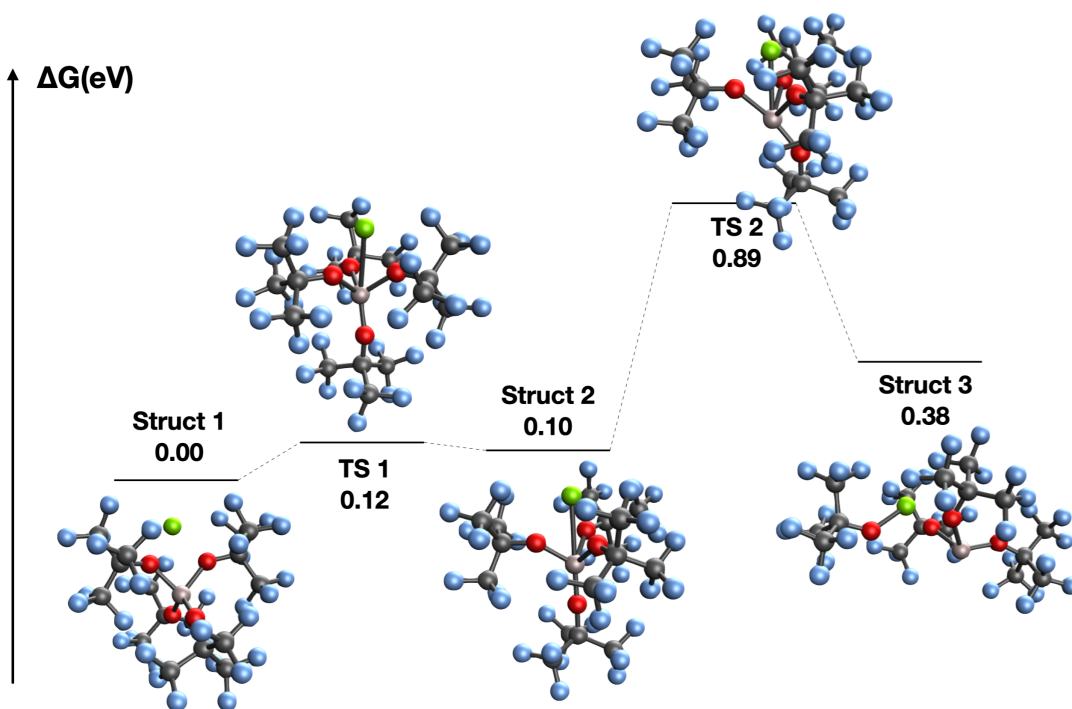


Figure 5.15: $\text{Mg}^+[\text{Al}(\text{pftb})_4]^-$ decomposition pathway through Al-O bond breakage

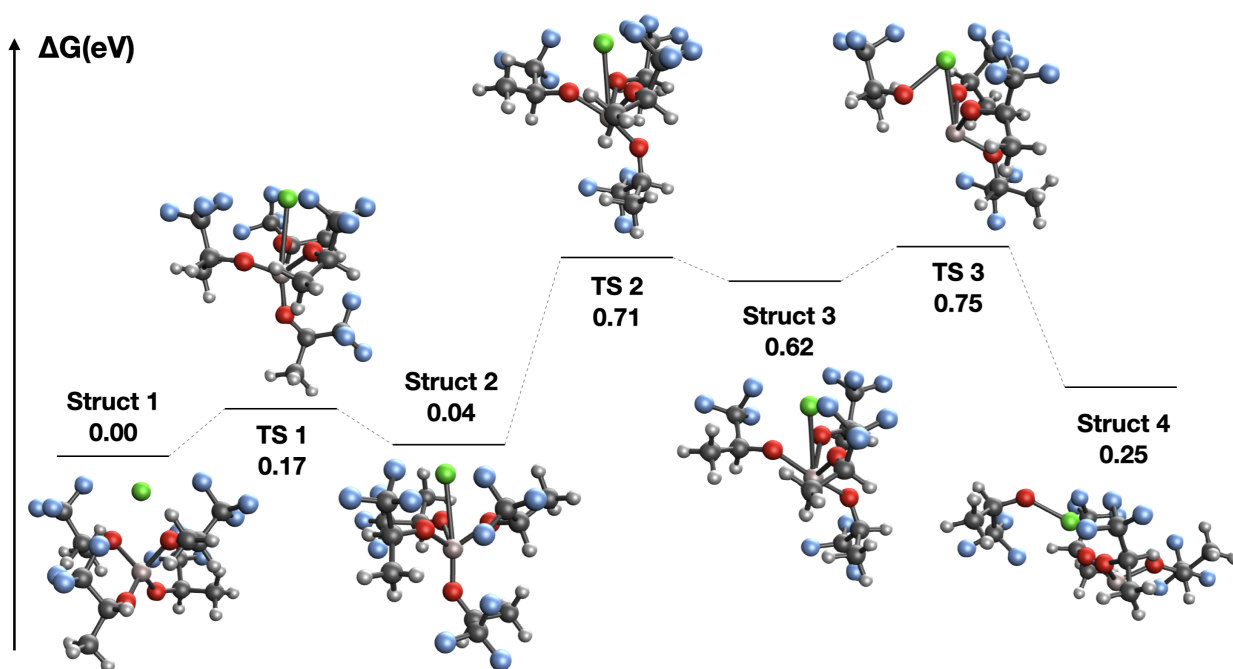


Figure 5.16: $\text{Ca}^+[\text{Al}(\text{tfip})_4]^-$ decomposition pathway through Al-O bond breakage

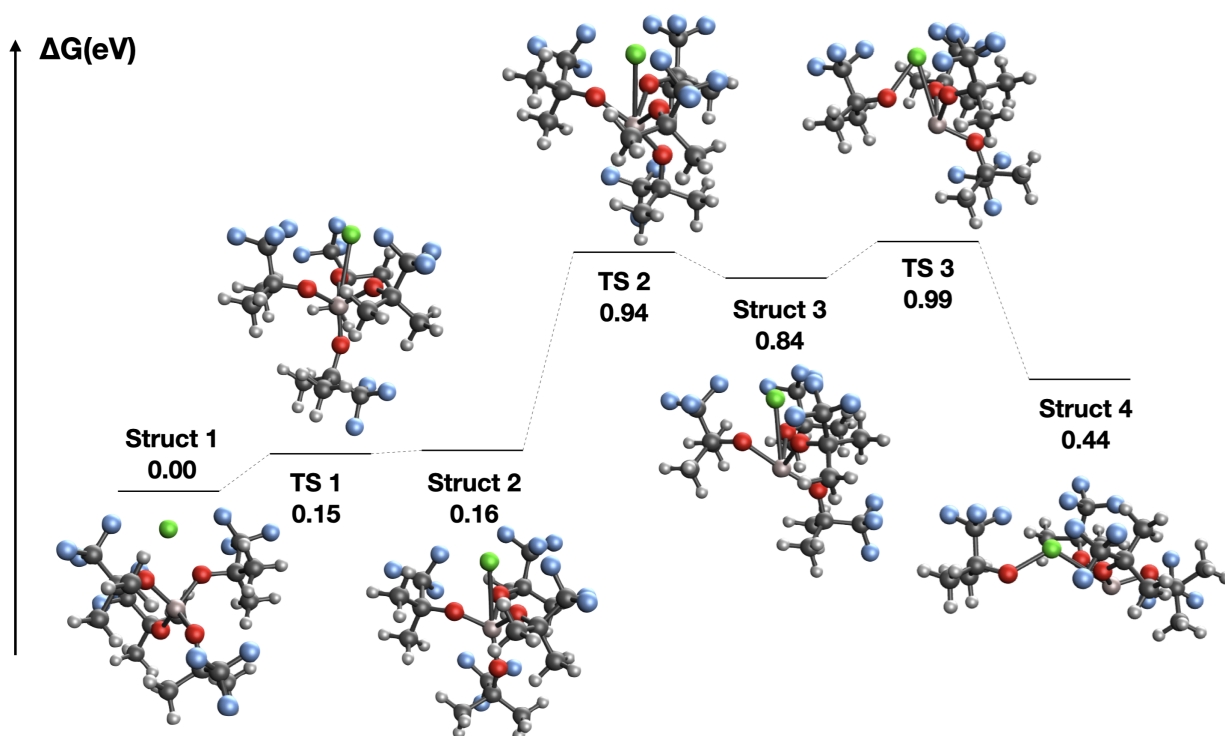


Figure 5.17: $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$ decomposition pathway through Al-O bond breakage

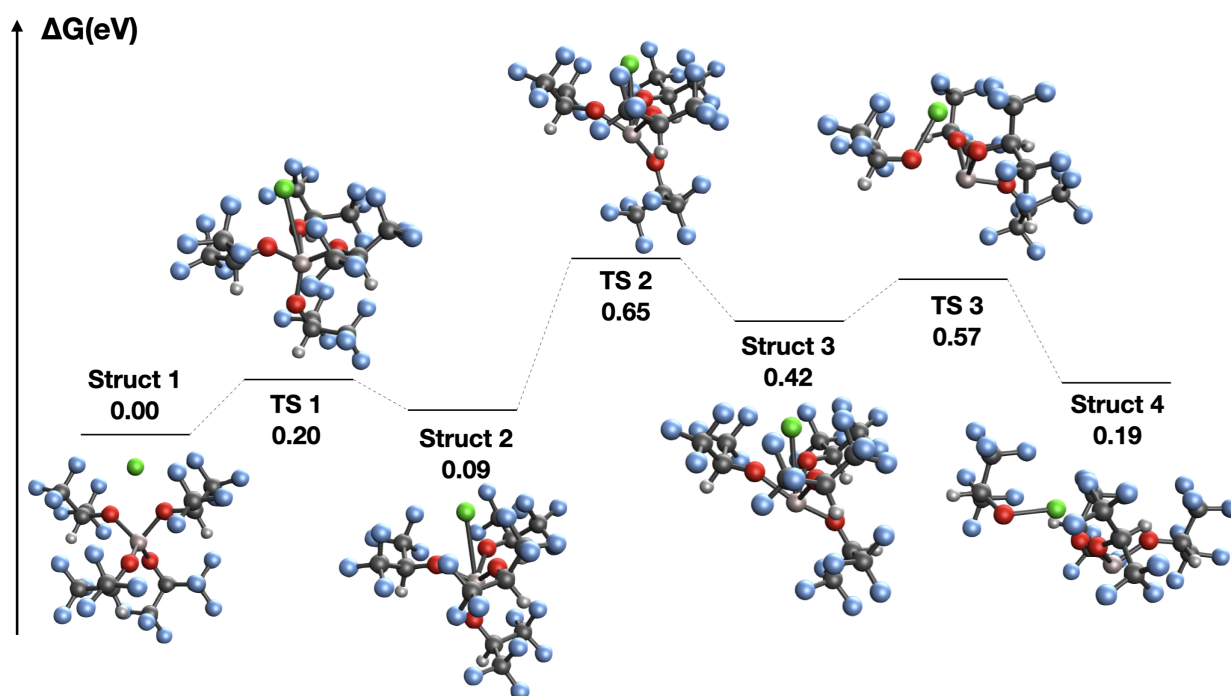


Figure 5.18: $\text{Ca}^+[\text{Al}(\text{hfip})_4]^-$ decomposition pathway through Al-O bond breakage

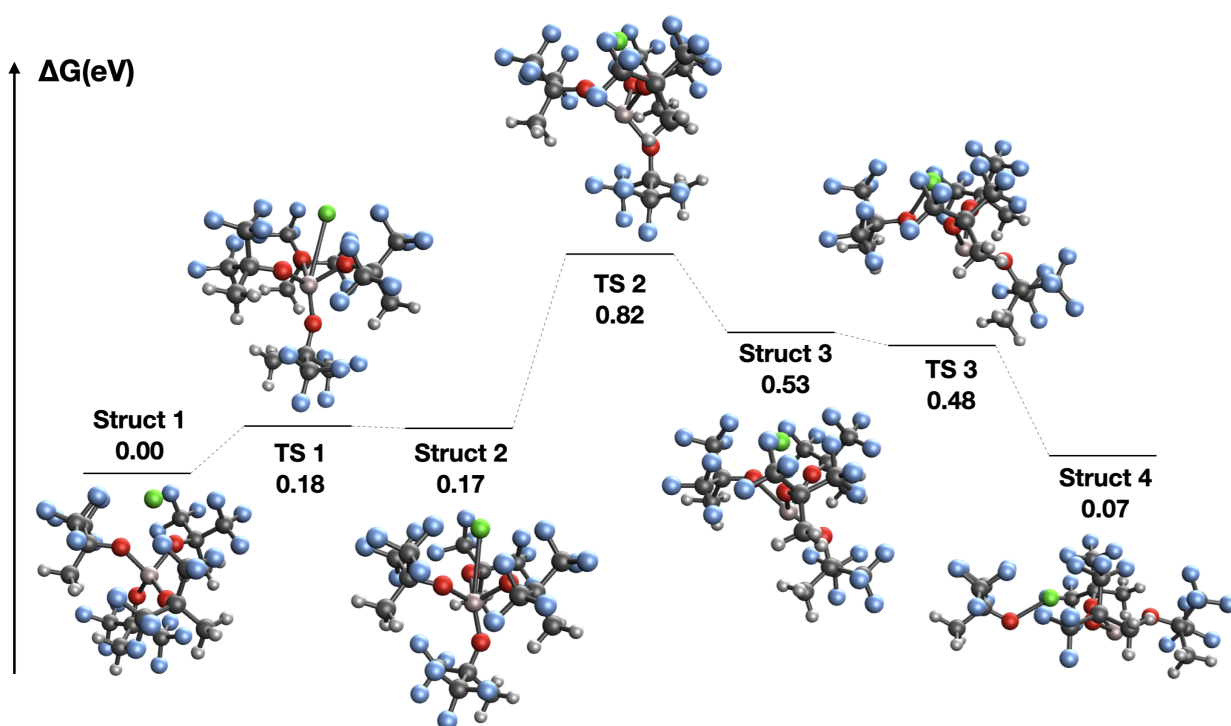


Figure 5.19: $\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$ decomposition pathway through Al-O bond breakage

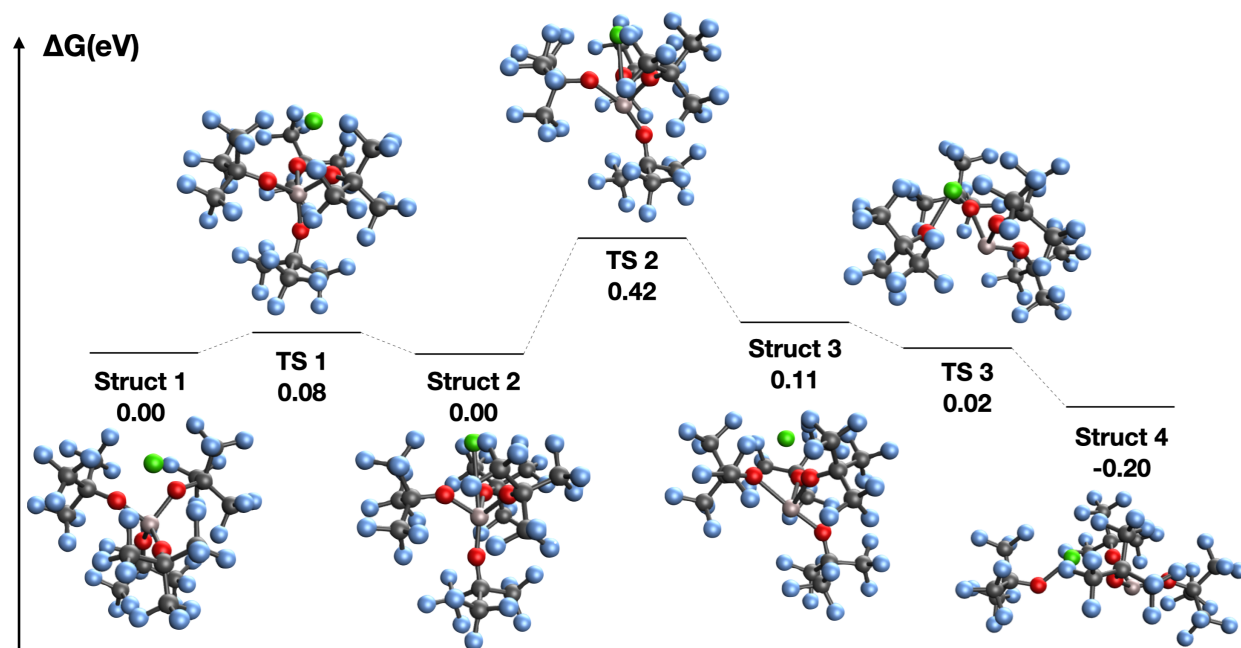


Figure 5.20: $\text{Ca}^+[\text{Al}(\text{pftb})_4]^-$ decomposition pathway through Al-O bond breakage

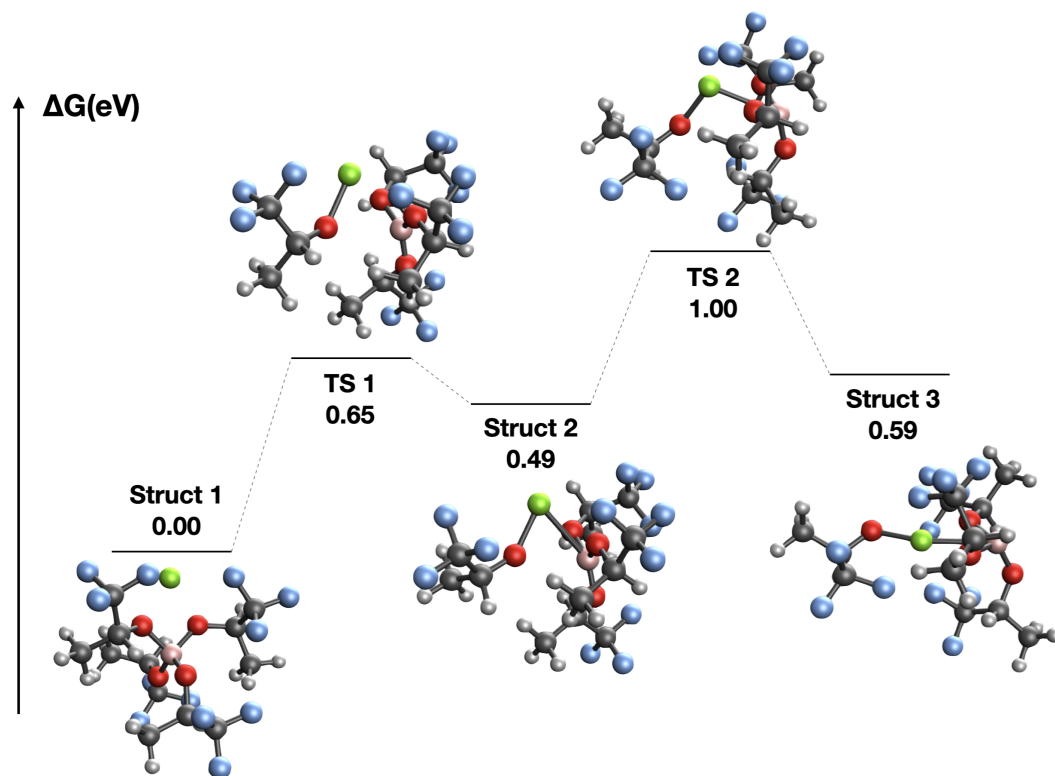


Figure 5.21: $\text{Mg}^+[\text{B}(\text{tfp})_4]^-$ decomposition pathway through B-O bond breakage

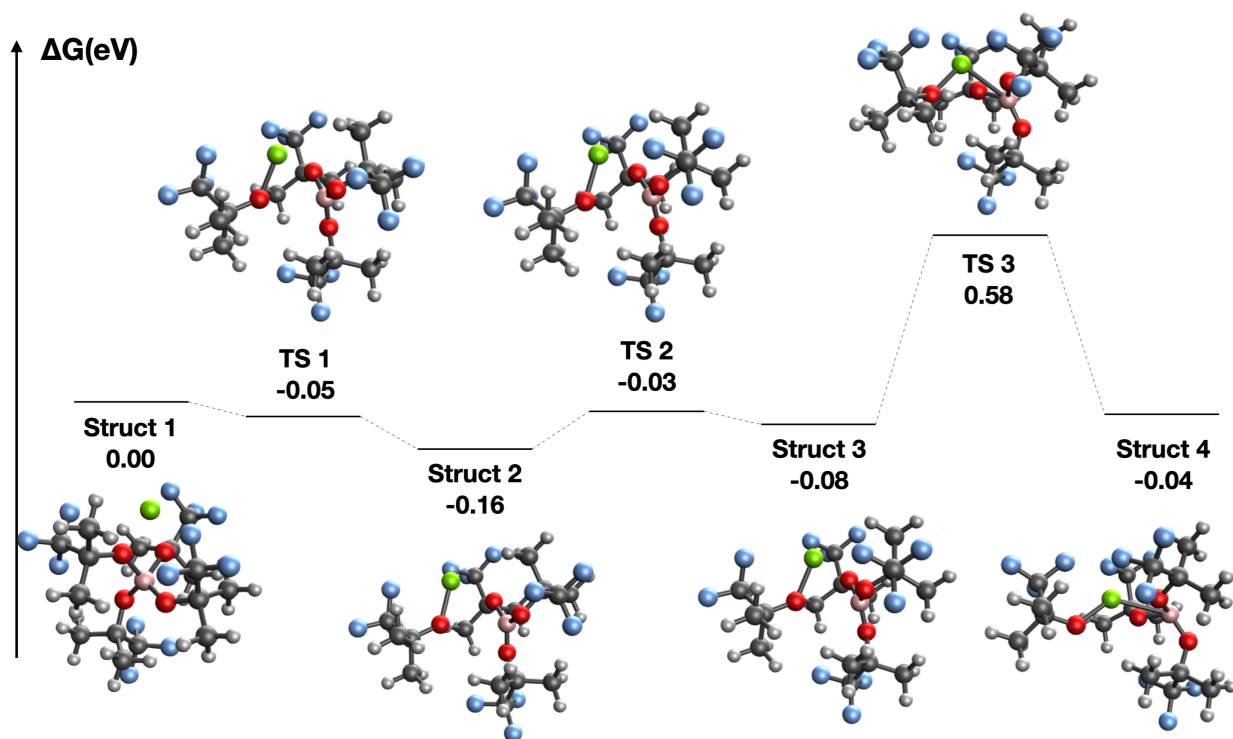


Figure 5.22: $\text{Mg}^+[\text{B}(\text{tftb})_4]^-$ decomposition pathway through B-O bond breakage

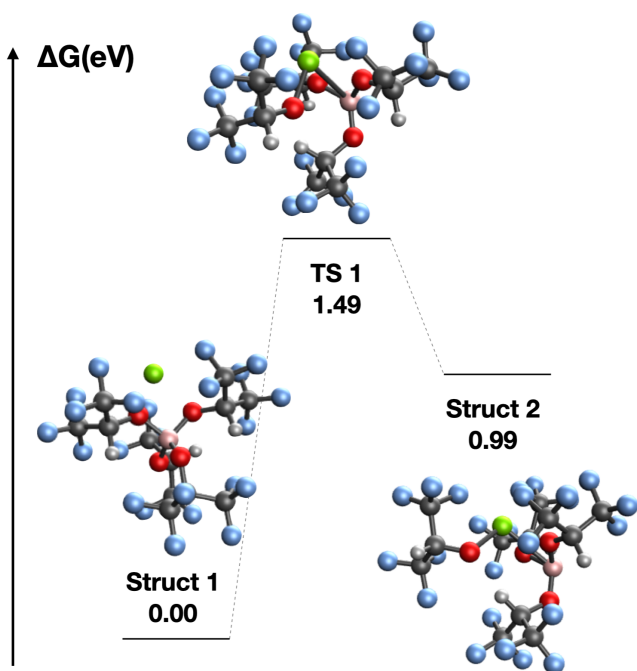


Figure 5.23: $\text{Mg}^+[\text{B}(\text{hfip})_4]^-$ decomposition pathway through B-O bond breakage

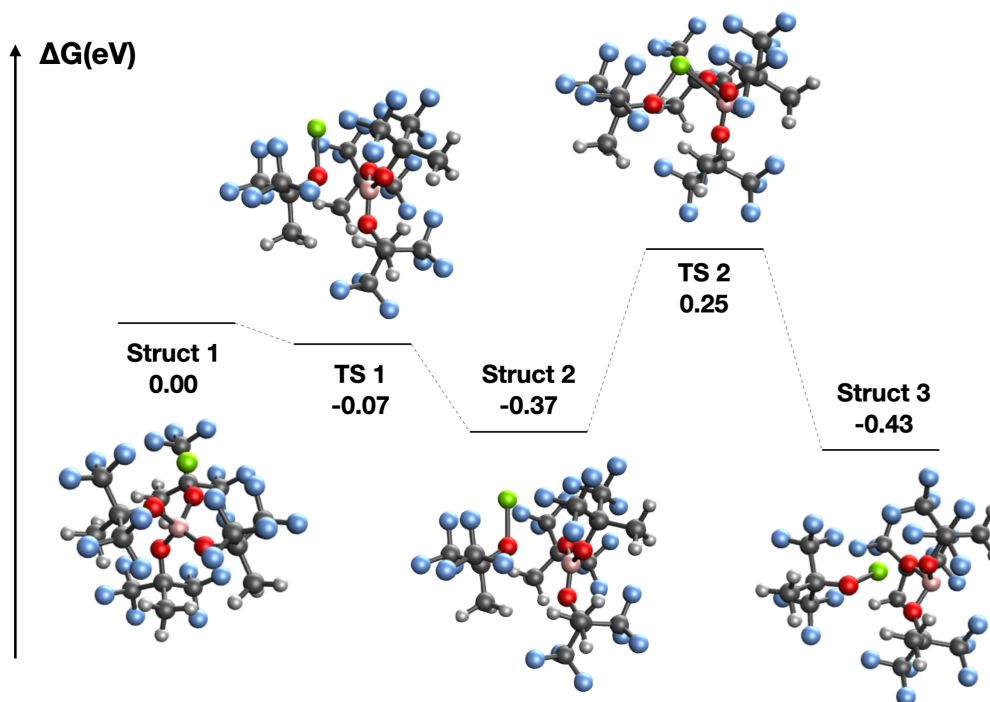


Figure 5.24: $\text{Mg}^+[\text{B}(\text{hftb})_4]^-$ decomposition pathway through B-O bond breakage

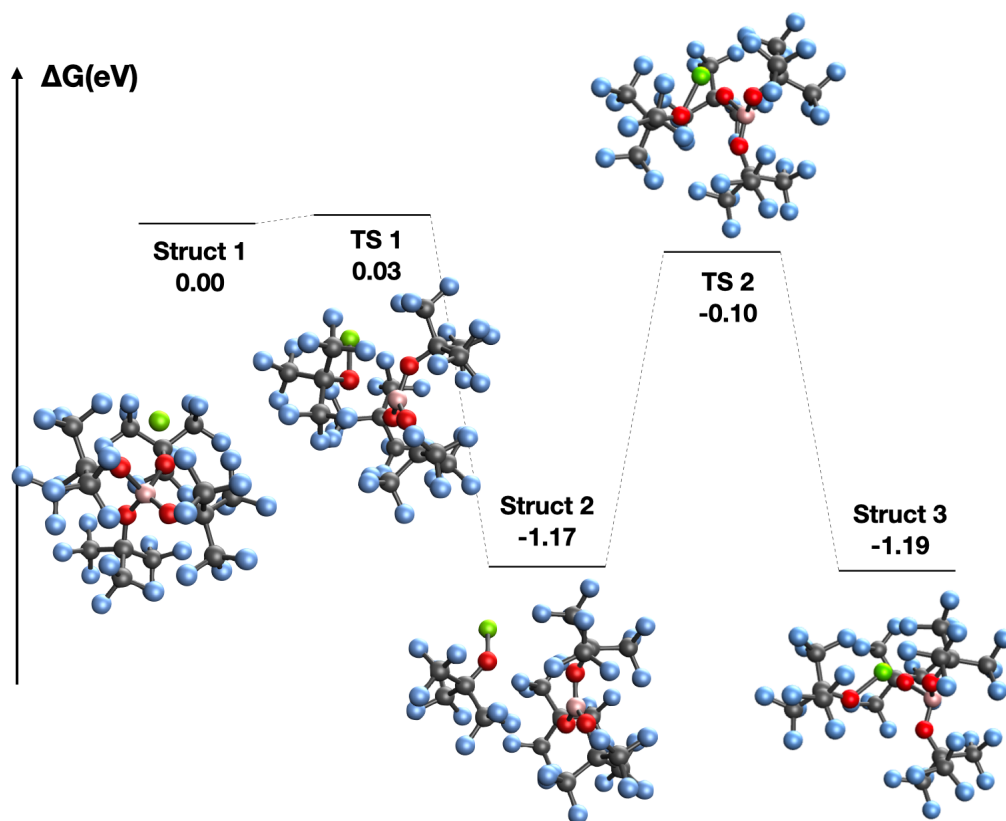


Figure 5.25: $\text{Mg}^+[\text{B}(\text{pftb})_4]^-$ decomposition pathway through B-O bond breakage

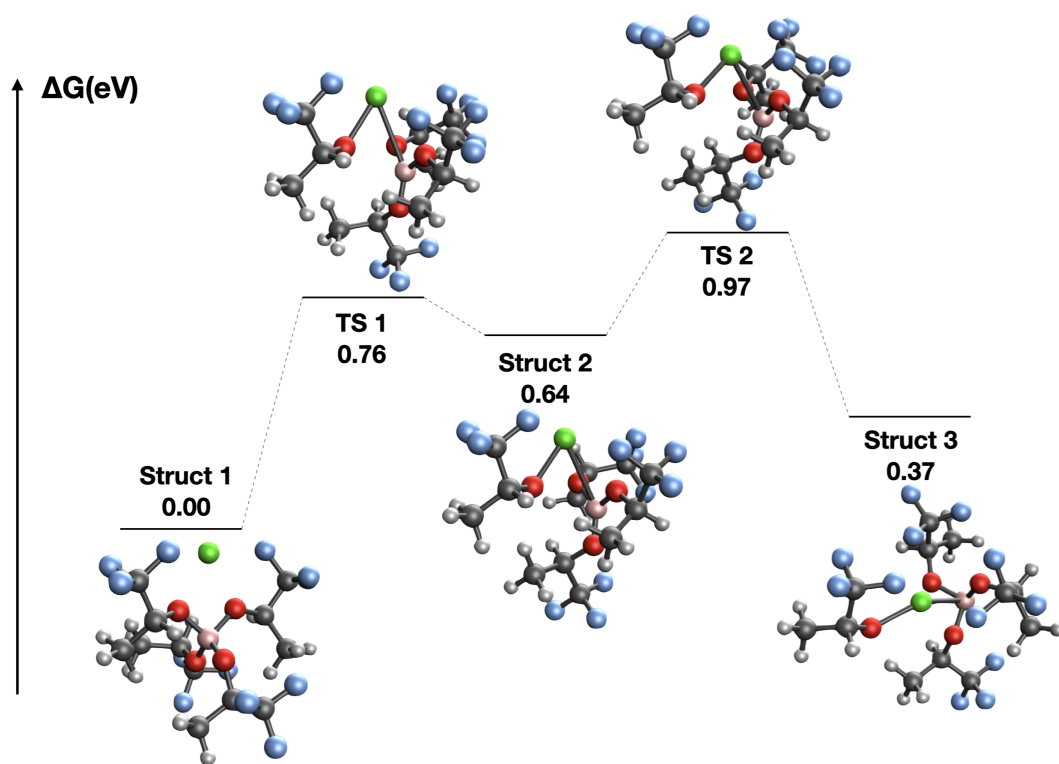


Figure 5.26: $\text{Ca}^+[\text{B}(\text{tfp})_4]^-$ decomposition pathway through B-O bond breakage

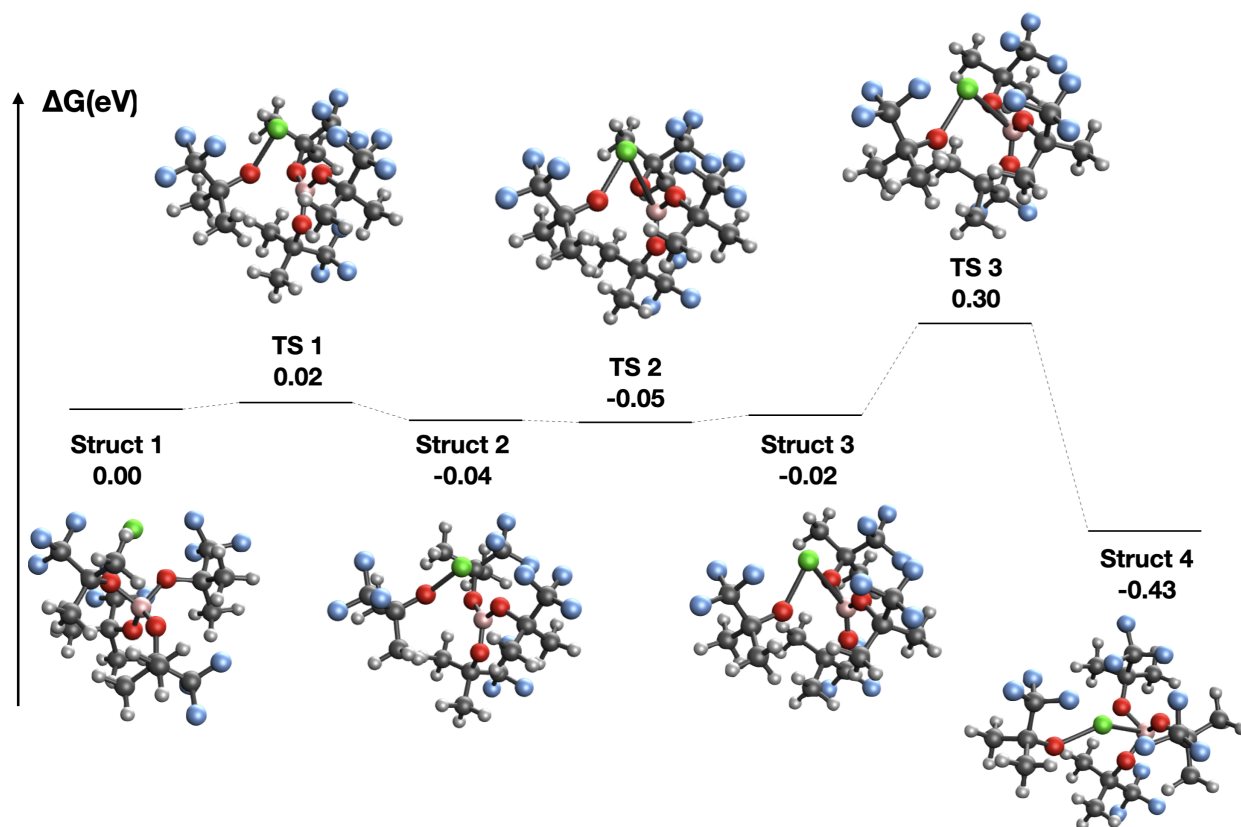


Figure 5.27: $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$ decomposition pathway through B-O bond breakage

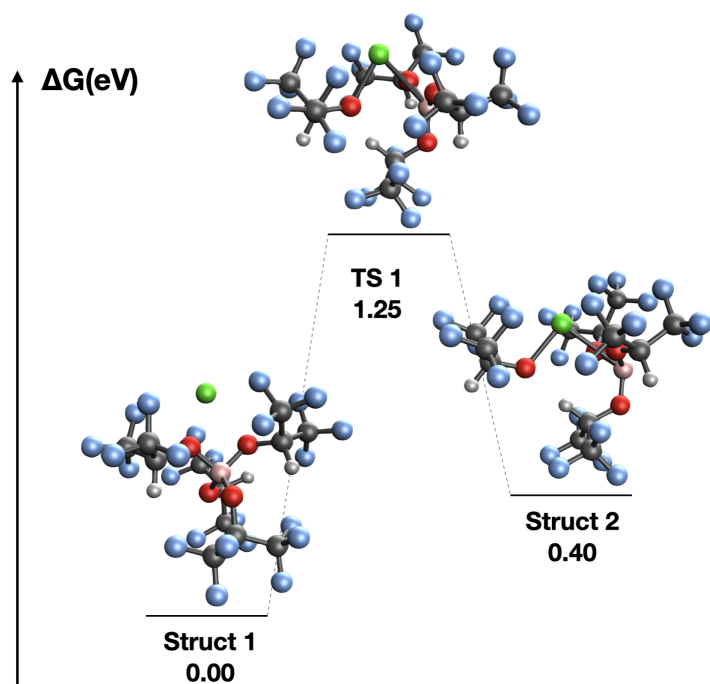


Figure 5.28: $\text{Ca}^+[\text{B}(\text{hfp})_4]^-$ decomposition pathway through B-O bond breakage

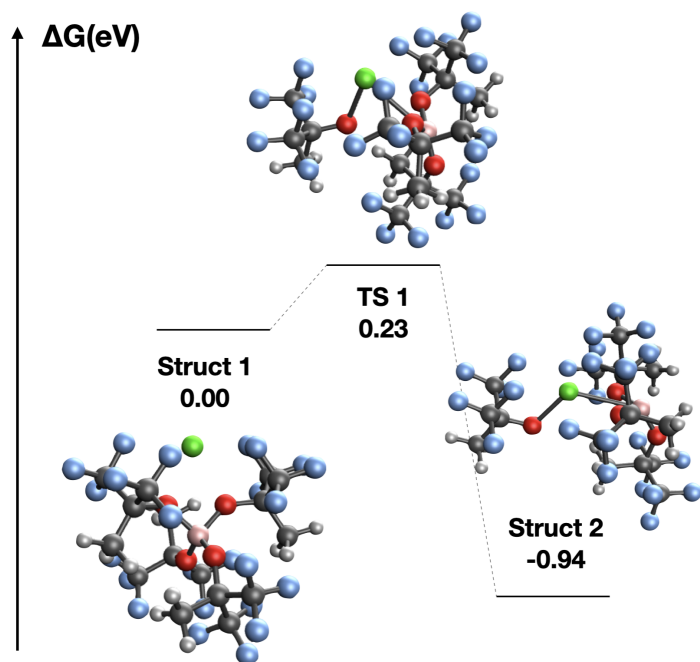


Figure 5.29: $\text{Ca}^+[\text{B}(\text{hftb})_4]^-$ decomposition pathway through B-O bond breakage

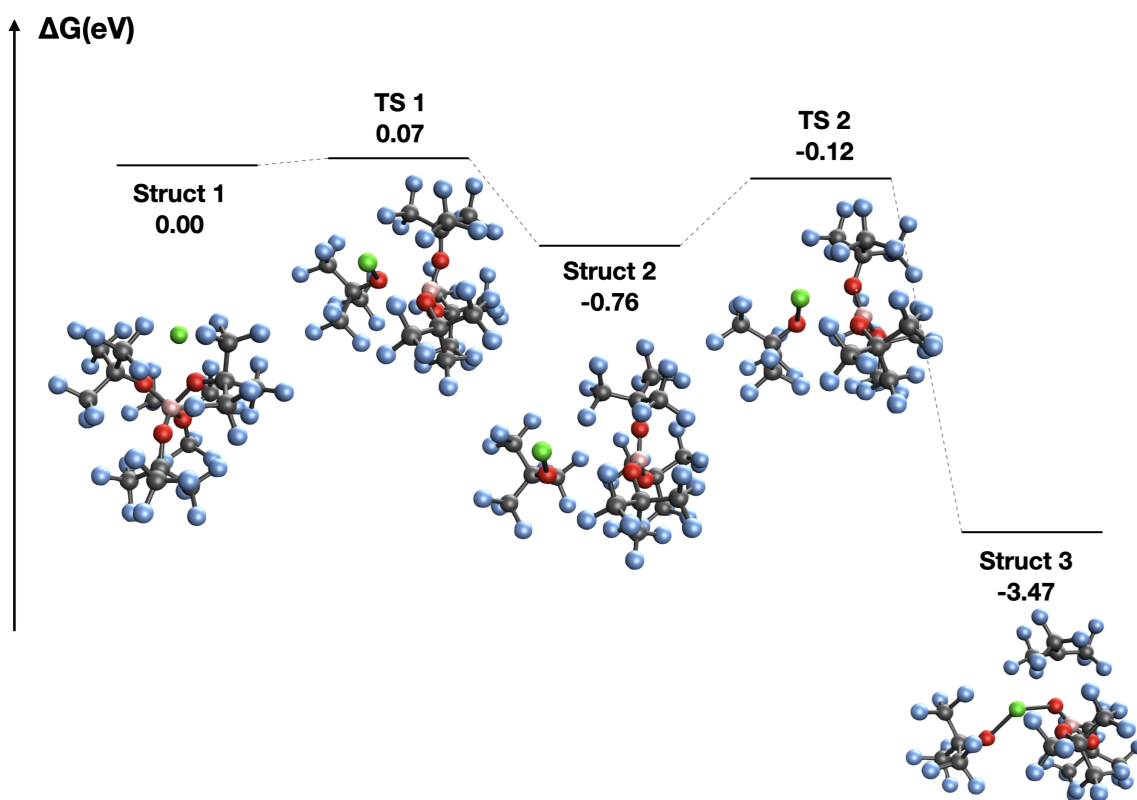


Figure 5.30: $\text{Ca}^+[\text{B}(\text{pftb})_4]^-$ decomposition pathway through B-O bond breakage

4.3 Decomposition pathways through C-O bond breakage

Note that for $\text{Mg}^+[\text{Al}(\text{tftb})_4]^-$, a local minimum that corresponds to the product from C-O bond breakage was not found within the level of theory employed.

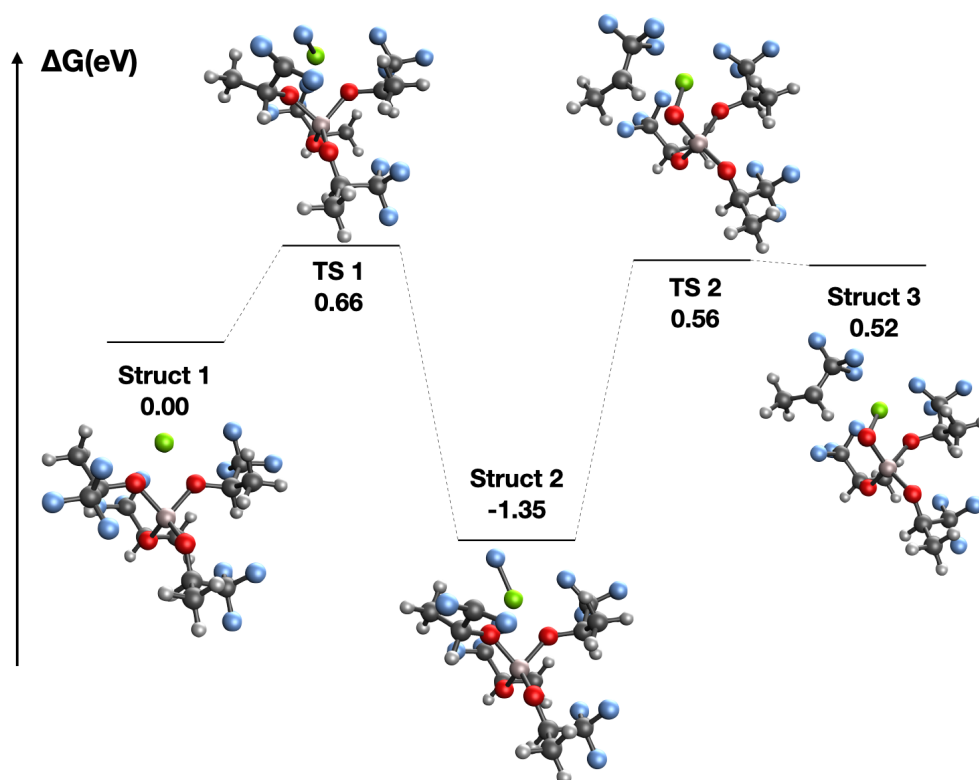


Figure 5.31: $\text{Mg}^+[\text{Al}(\text{tfp})_4]^-$ decomposition pathway through C-O bond breakage

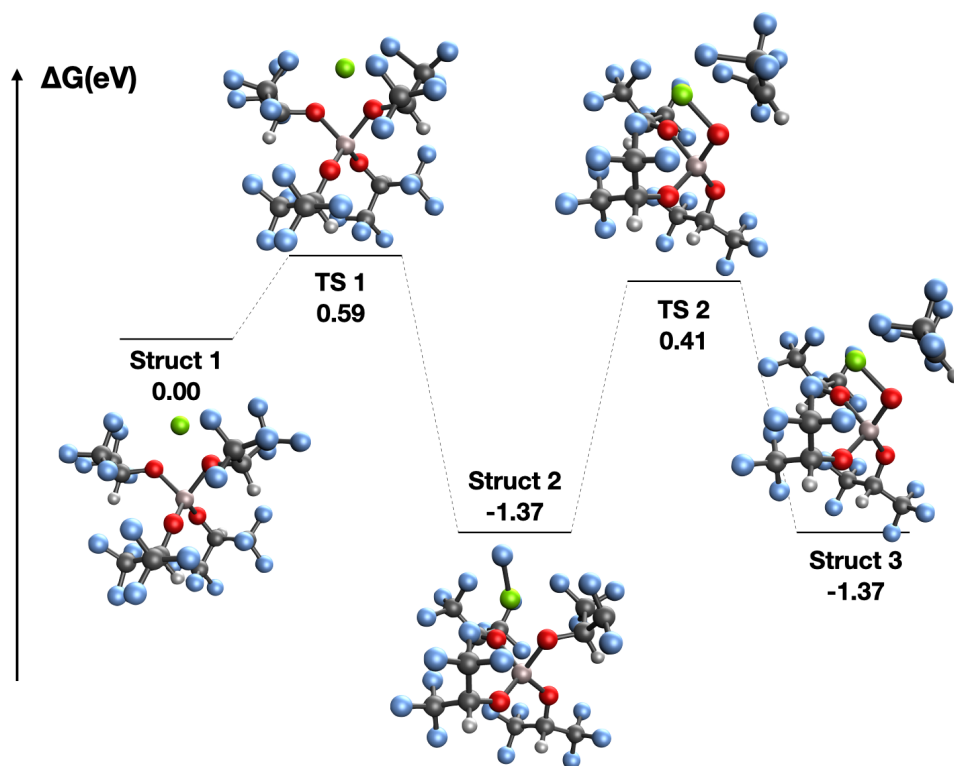


Figure 5.32: $\text{Mg}^+[\text{Al}(\text{hfip})_4]^-$ decomposition pathway through C-O bond breakage

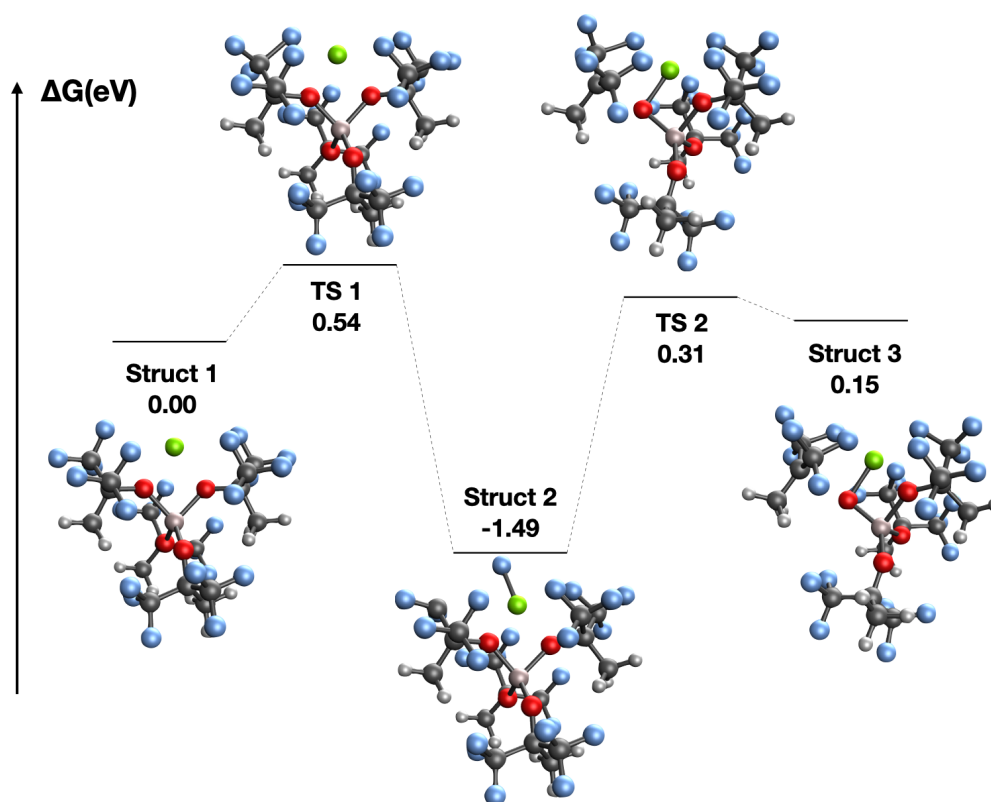


Figure 5.33: $\text{Mg}^+[\text{Al}(\text{hftb})_4]^-$ decomposition pathway through C-O bond breakage

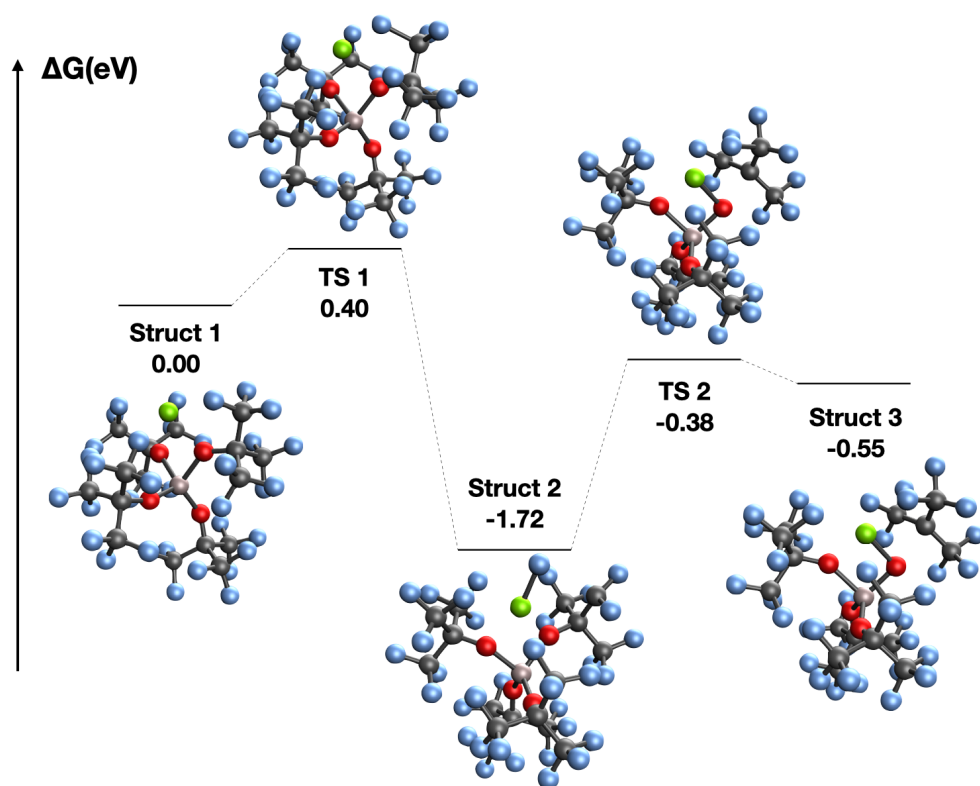


Figure 5.34: $\text{Mg}^+[\text{Al}(\text{pftb})_4]^-$ decomposition pathway through C-O bond breakage

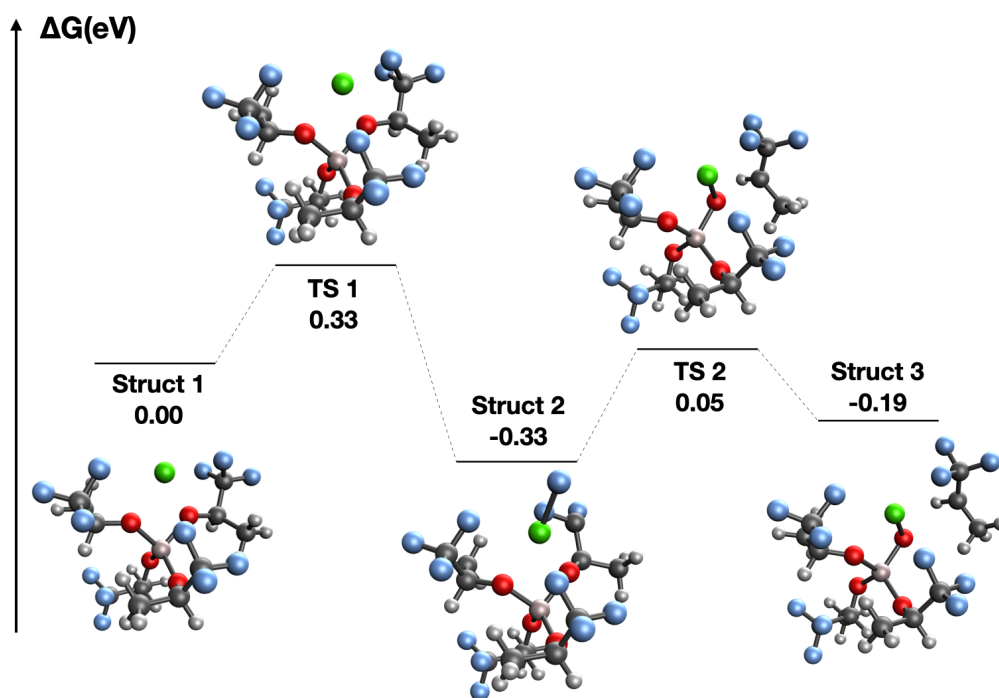


Figure 5.35: $\text{Ca}^+[\text{Al}(\text{tfip})_4]^-$ decomposition pathway through C-O bond breakage

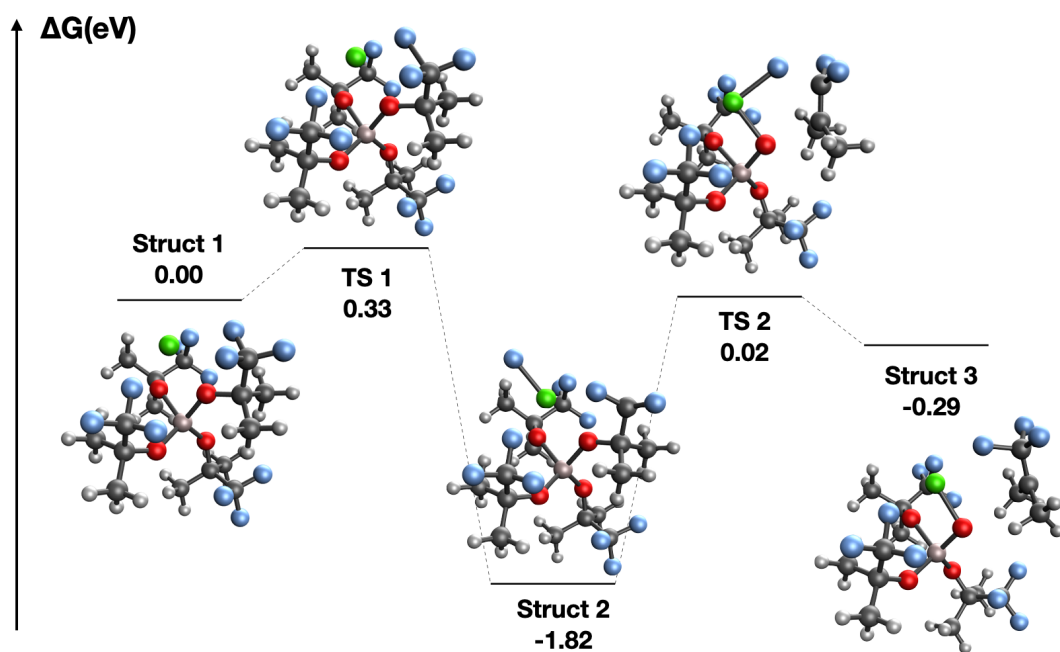


Figure 5.36: $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$ decomposition pathway through C-O bond breakage

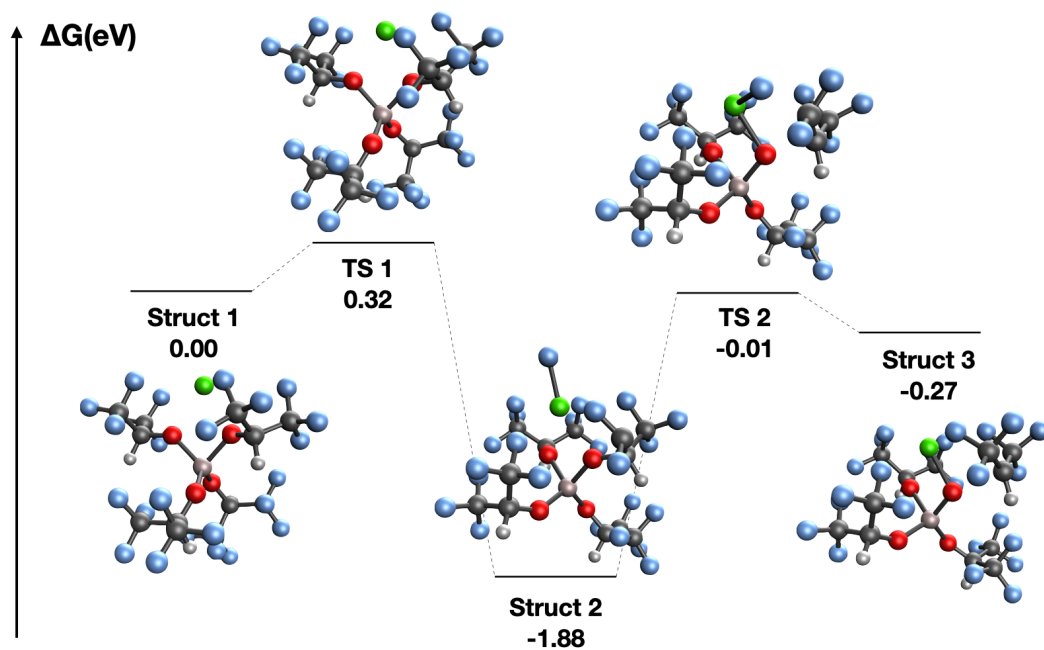


Figure 5.37: $\text{Ca}^+[\text{Al}(\text{hfip})_4]^-$ decomposition pathway through C-O bond breakage

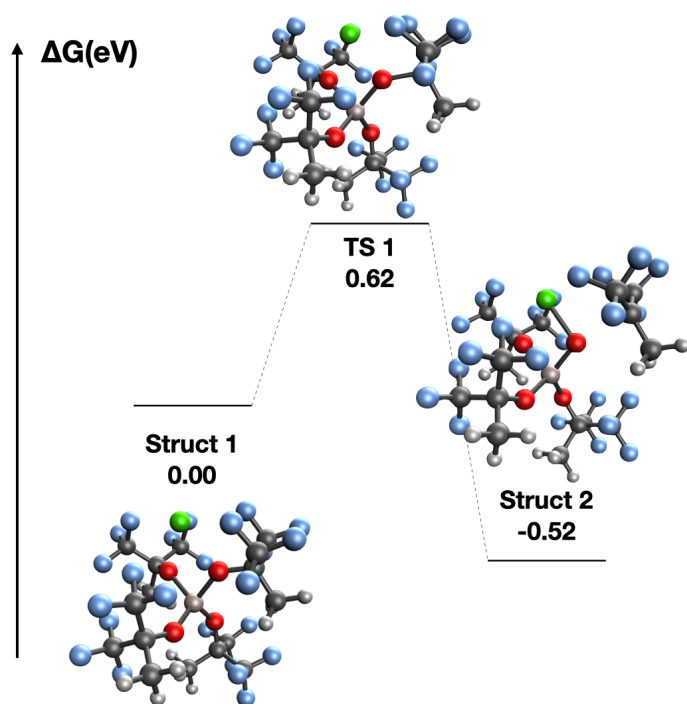


Figure 5.38: $\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$ decomposition pathway through C-O bond breakage

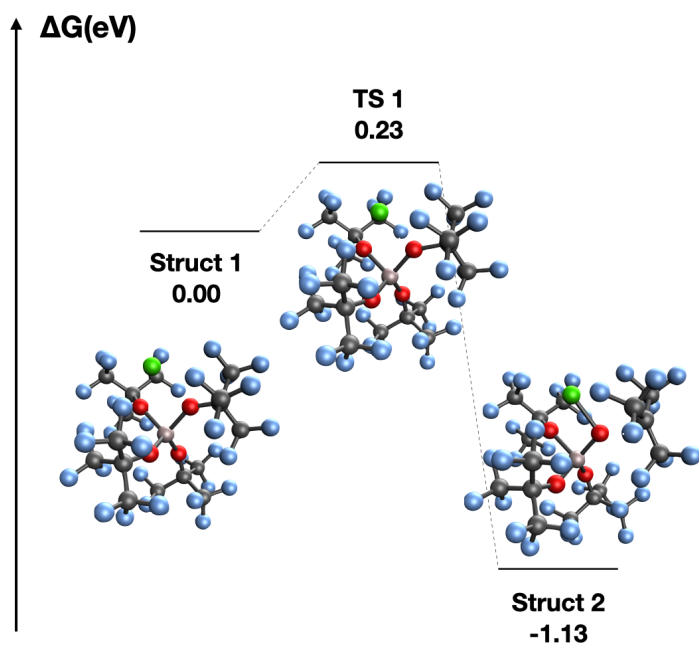


Figure 5.39: $\text{Ca}^+[\text{Al}(\text{pftb})_4]^-$ decomposition pathway through C-O bond breakage

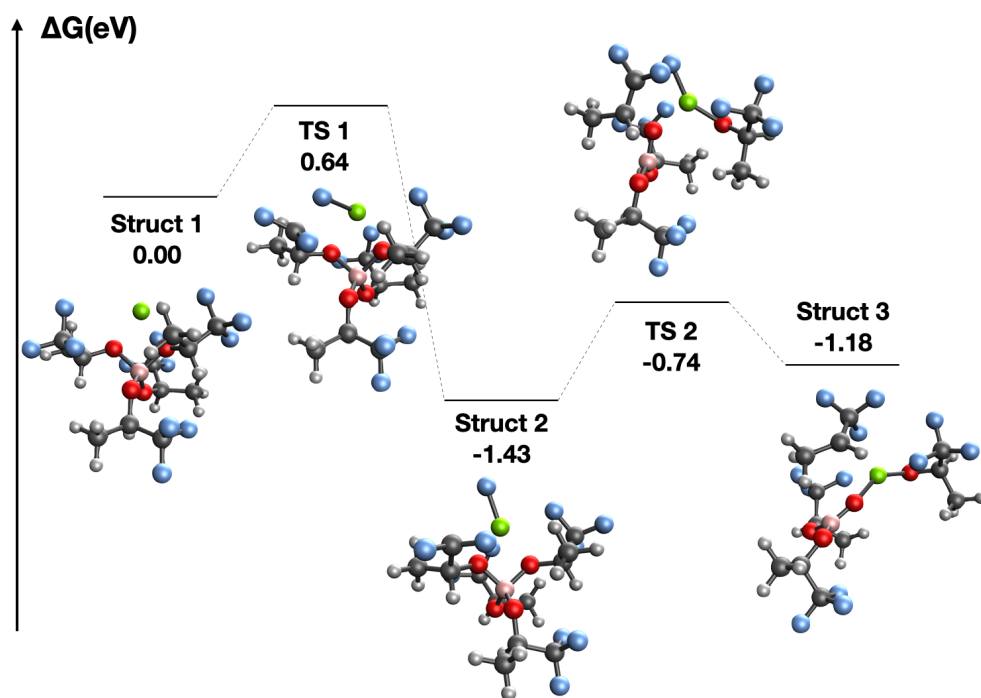


Figure 5.40: $\text{Mg}^+[\text{B}(\text{tfip})_4]^-$ decomposition pathway through C-O bond breakage

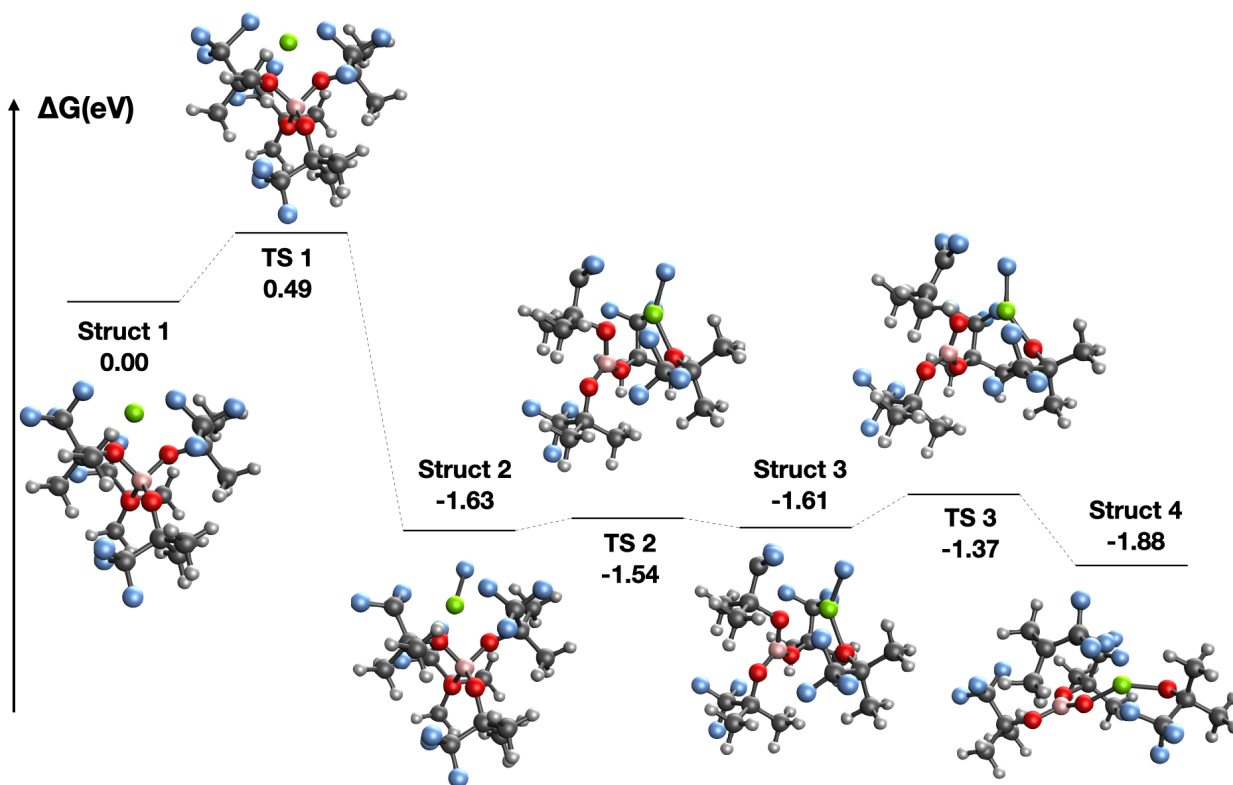


Figure 5.41: $\text{Mg}^+[\text{B}(\text{tftb})_4]^-$ decomposition pathway through C-O bond breakage

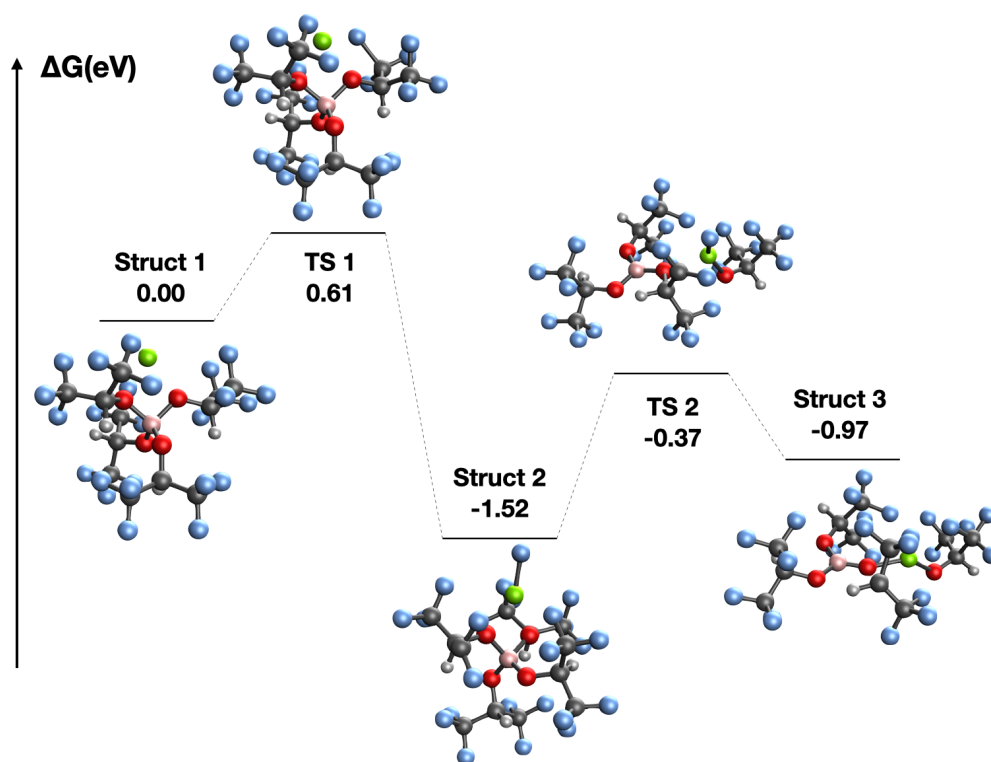


Figure 5.42: $\text{Mg}^+[\text{B}(\text{hfip})_4]^-$ decomposition pathway through C-O bond breakage

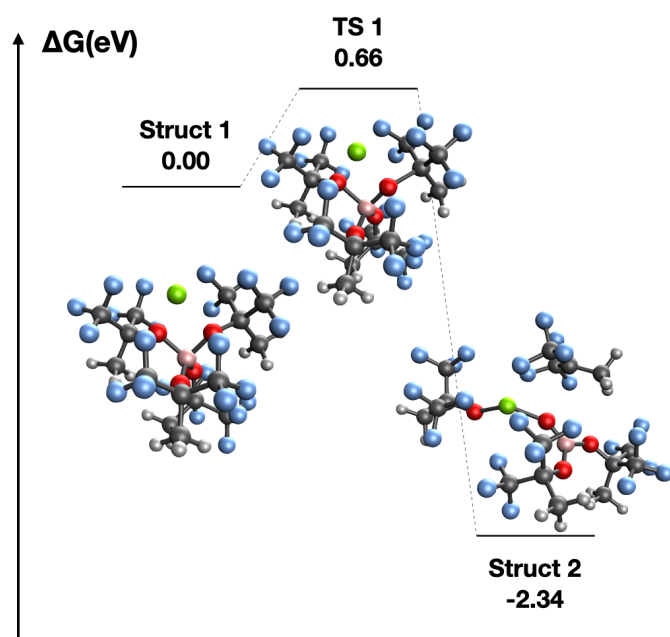


Figure 5.43: $\text{Mg}^+[\text{B}(\text{hftb})_4]^-$ decomposition pathway through C-O bond breakage

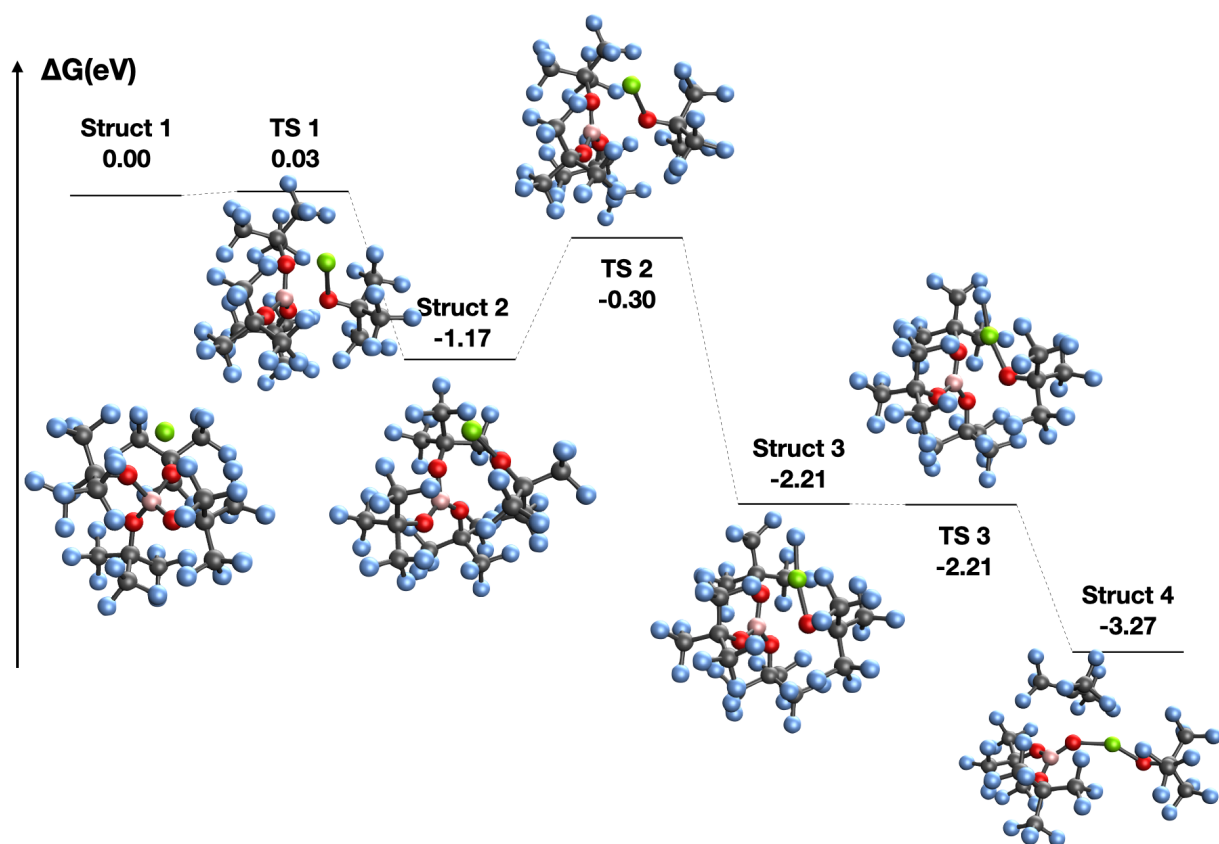


Figure 5.44: $\text{Mg}^+[\text{B}(\text{pftb})_4]^-$ decomposition pathway through C-O bond breakage

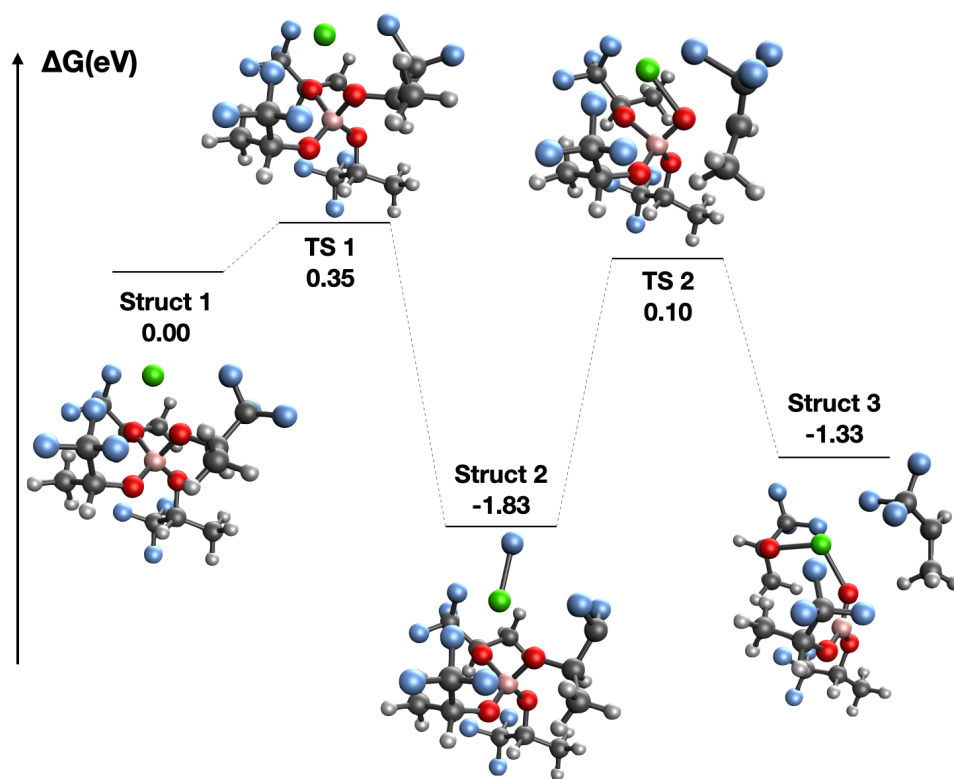


Figure 5.45: $\text{Ca}^+[\text{B}(\text{tfp})_4]^-$ decomposition pathway through C-O bond breakage

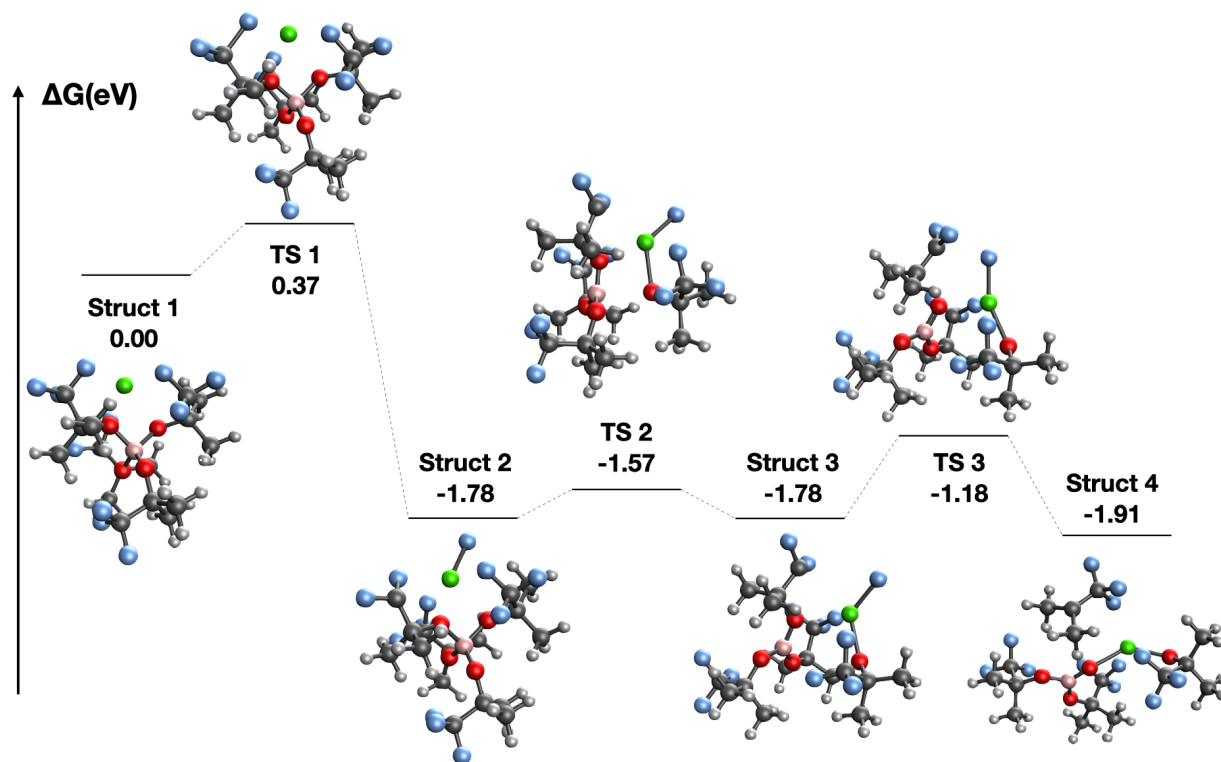


Figure 5.46: $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$ decomposition pathway through C-O bond breakage

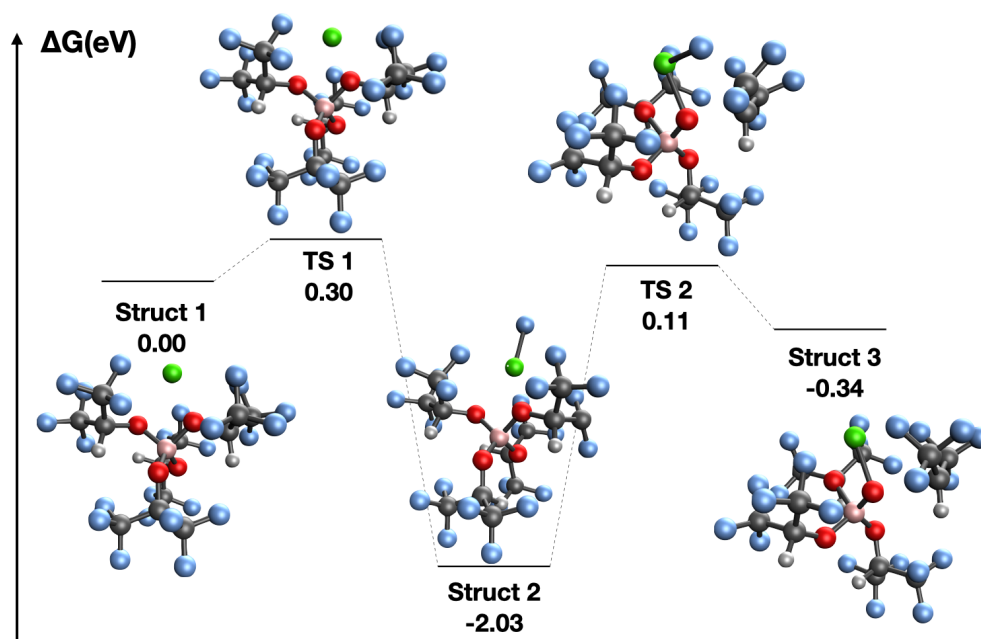


Figure 5.47: $\text{Ca}^+[\text{B}(\text{hfip})_4]^-$ decomposition pathway through C-O bond breakage

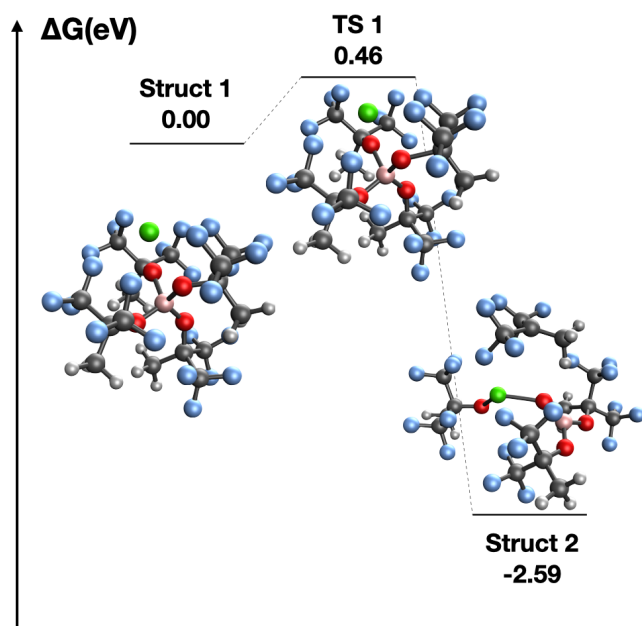


Figure 5.48: $\text{Ca}^+[\text{B}(\text{hftb})_4]^-$ decomposition pathway through C-O bond breakage

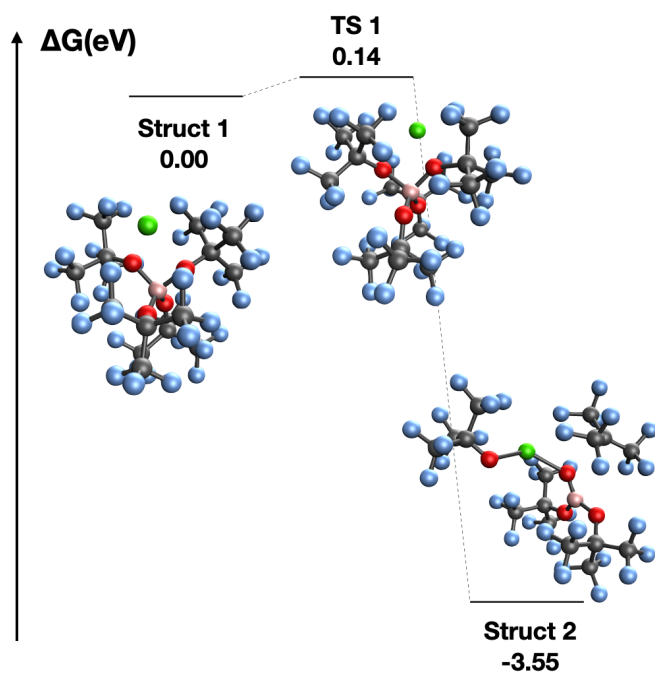


Figure 5.49: $\text{Ca}^+[\text{B}(\text{pftb})_4]^-$ decomposition pathway through C-O bond breakage

4.4 Decomposition pathways through C-F bond breakage

If not shown here, the C-F bond breaking decomposition pathway can be found as the first step in the C-O bond breaking decomposition pathway as shown in Sec 4.3.

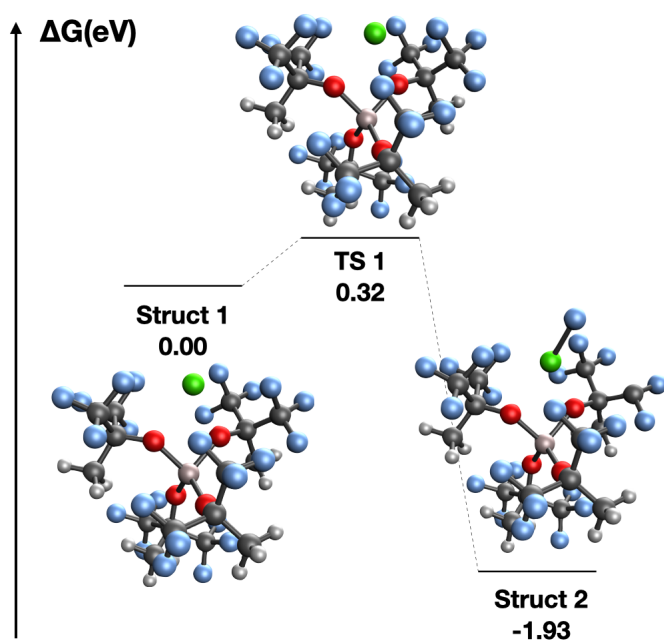


Figure 5.50: $\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$ decomposition pathway through C-F bond breakage

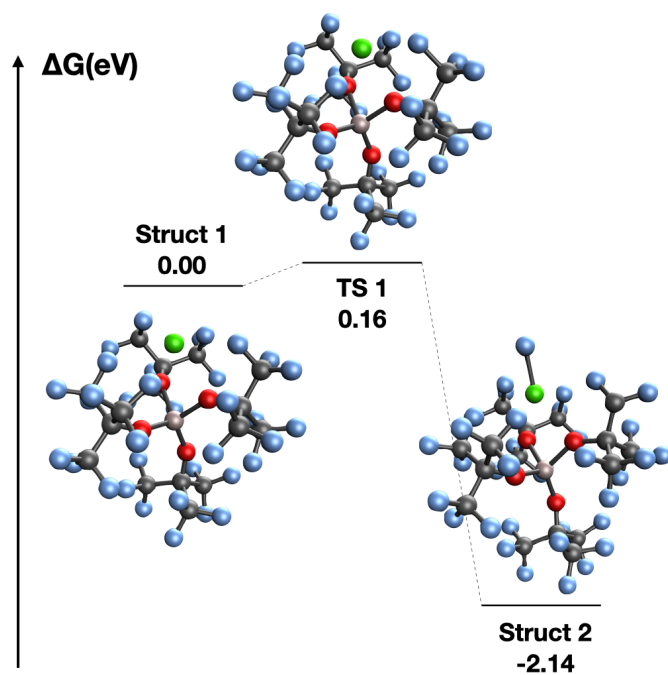


Figure 5.51: $\text{Ca}^+[\text{Al}(\text{pftb})_4]^-$ decomposition pathway through C-F bond breakage

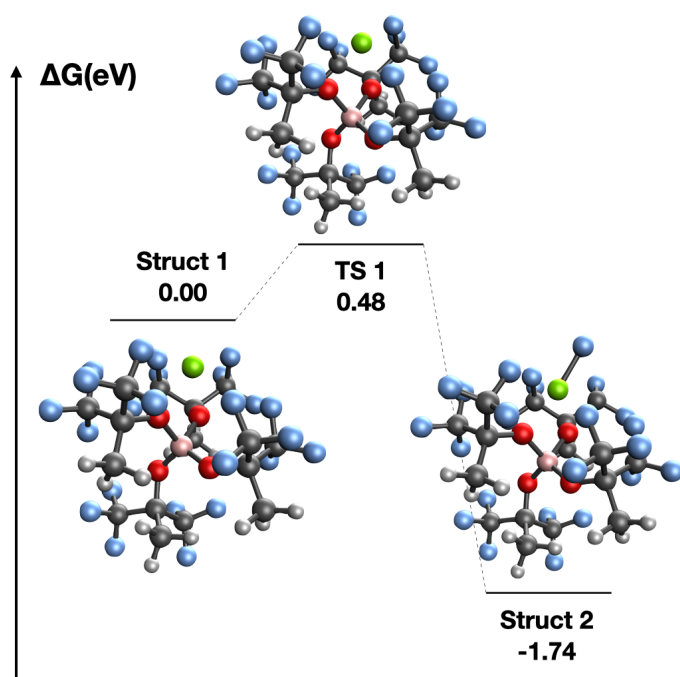


Figure 5.52: $\text{Mg}^+[\text{B}(\text{hftb})_4]^-$ decomposition pathway through C-F bond breakage

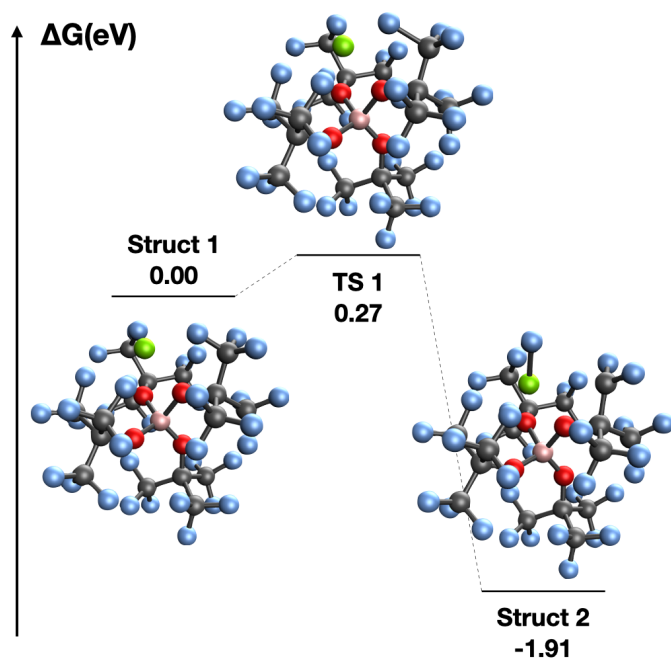


Figure 5.53: $\text{Mg}^+[\text{B}(\text{pftb})_4]^-$ decomposition pathway through C-F bond breakage

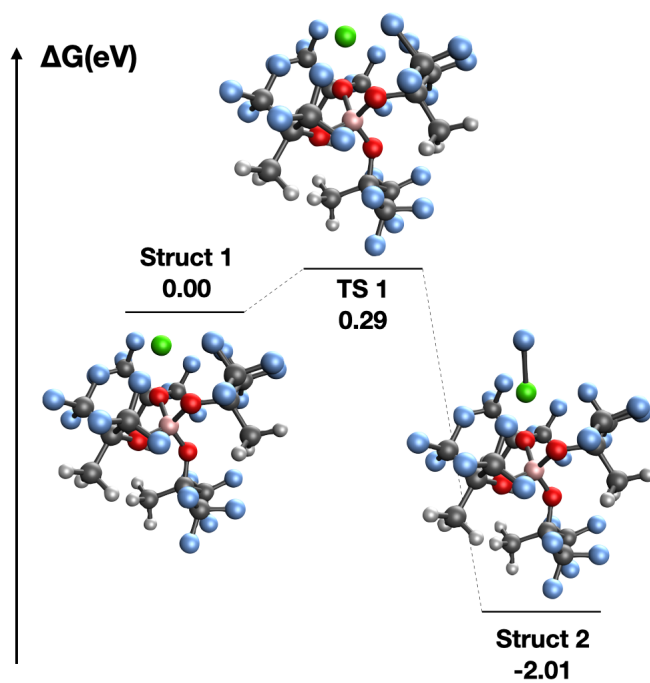


Figure 5.54: $\text{Ca}^+[\text{B}(\text{hftb})_4]^-$ decomposition pathway through C-F bond breakage

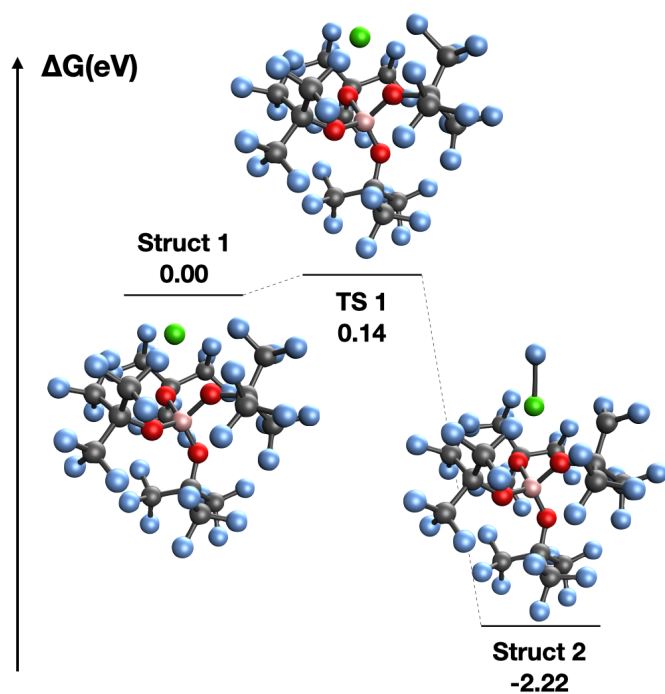


Figure 5.55: $\text{Ca}^+[\text{B}(\text{pftb})_4]^-$ decomposition pathway through C-F bond breakage

4.5 Results from CaOR-B(OR)_3 separation for borates

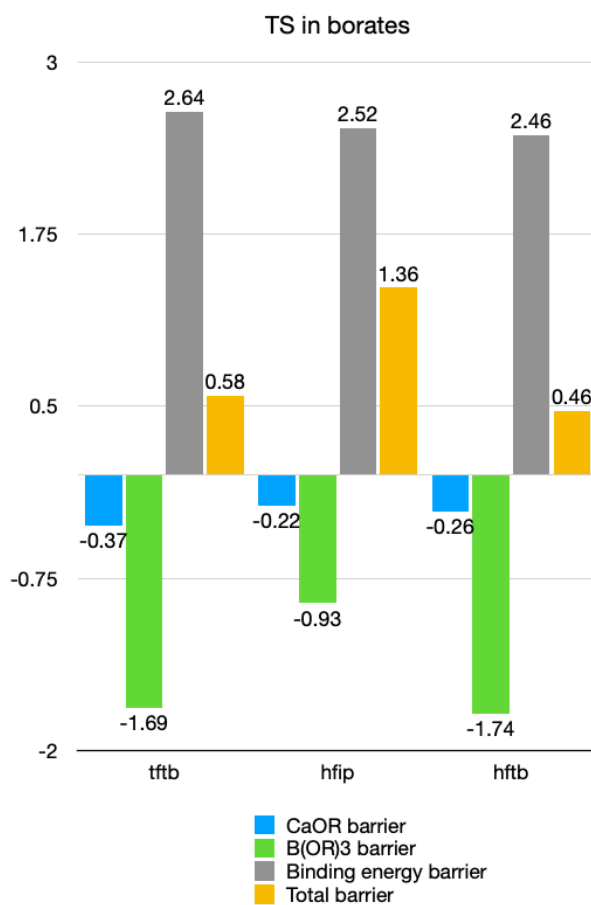


Figure 5.56: Barrier analysis for borates from CaOR-B(OR)_3 separation

4.6 Results from EDA analysis

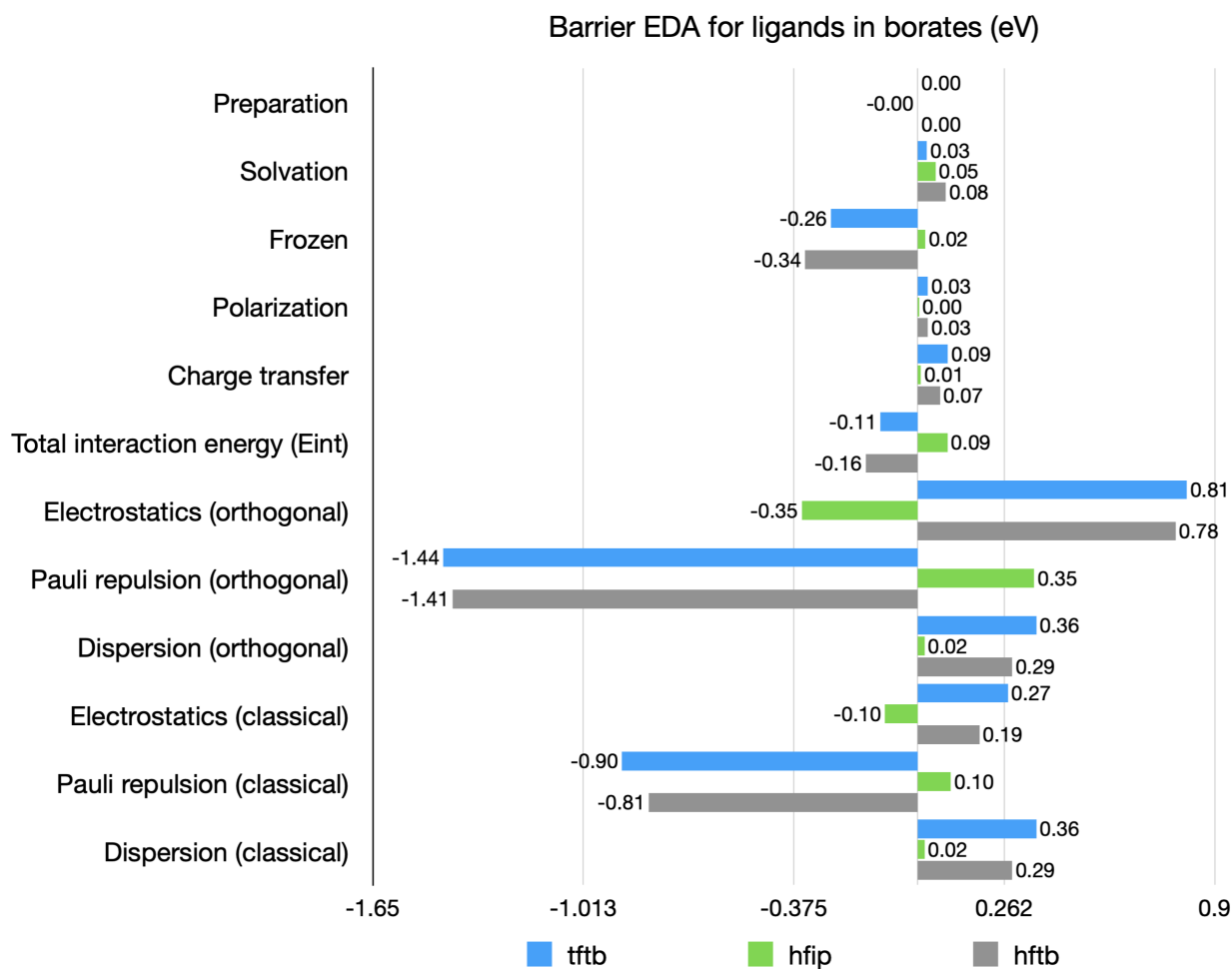


Figure 5.57: EDA analysis of the interaction energy barriers between the four ligands for the B-O bond breaking decomposition for borates

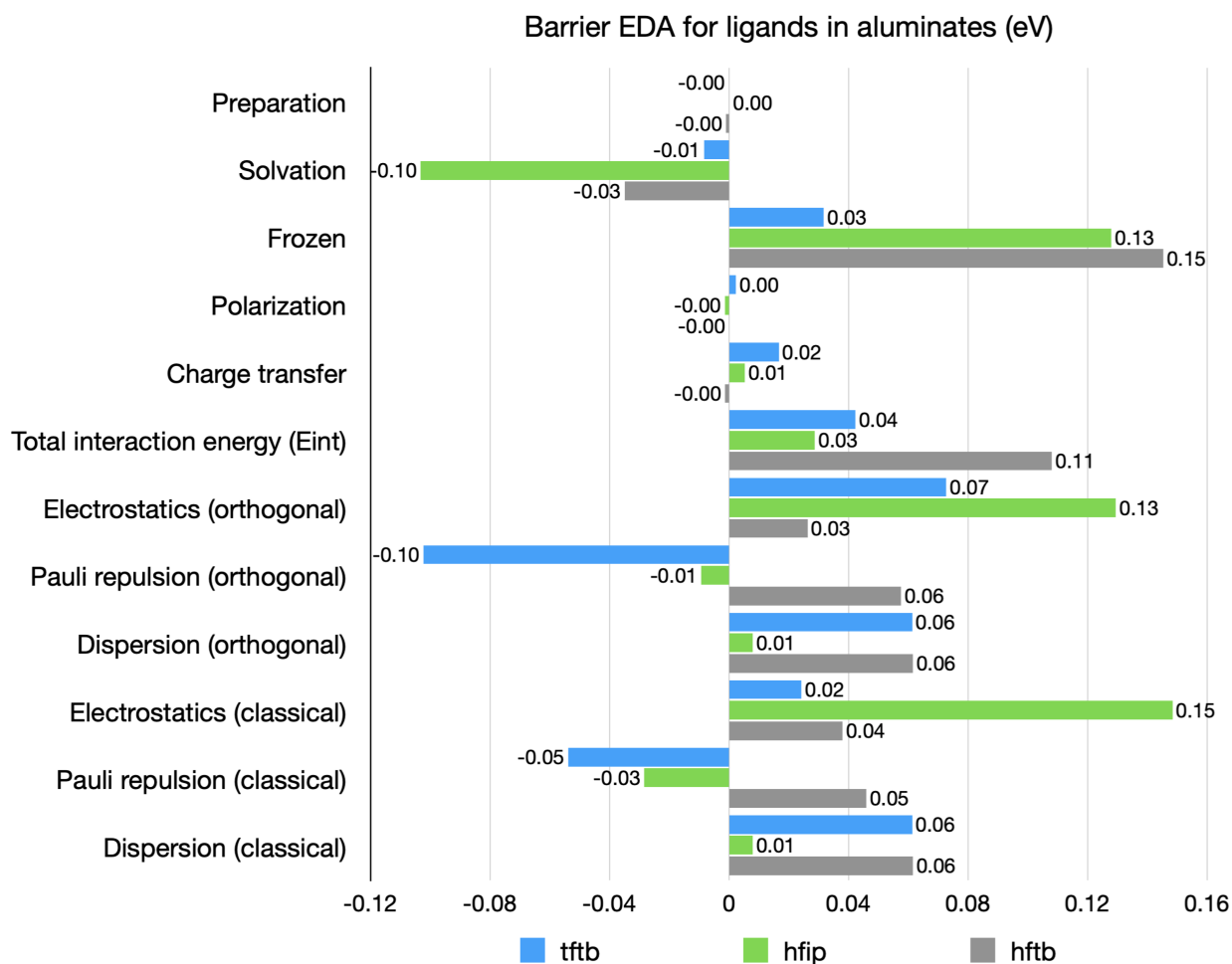


Figure 5.58: EDA analysis of the interaction energy barriers between the four ligands for the Al-O bond breaking decomposition (TS2) for aluminates

4.7 Electron affinities for $\text{Al}(\text{OR})_3$ and $\text{B}(\text{OR})_3$ complexes

The adiabatic electron affinity (EA), vertical attachment energy (VAE), and vertical detachment energy (VDE) were computed for the $\text{Al}(\text{OR})_3$ and $\text{B}(\text{OR})_3$ complexes (OR = tftb, hfip, hftb) at B3LYP-D3(BJ)/def2-SVP level of theory. The EA, VAE, and VDE follow the definitions below.[37]

$$\text{VAE} = E(\text{optimized neutral}) - E(\text{anion at optimized neutral geometry})$$

$$\text{VDE} = E(\text{neutral at optimized anion geometry}) - E(\text{optimized anion})$$

$$\text{EA} = E(\text{optimized neutral}) - E(\text{optimized anion})$$

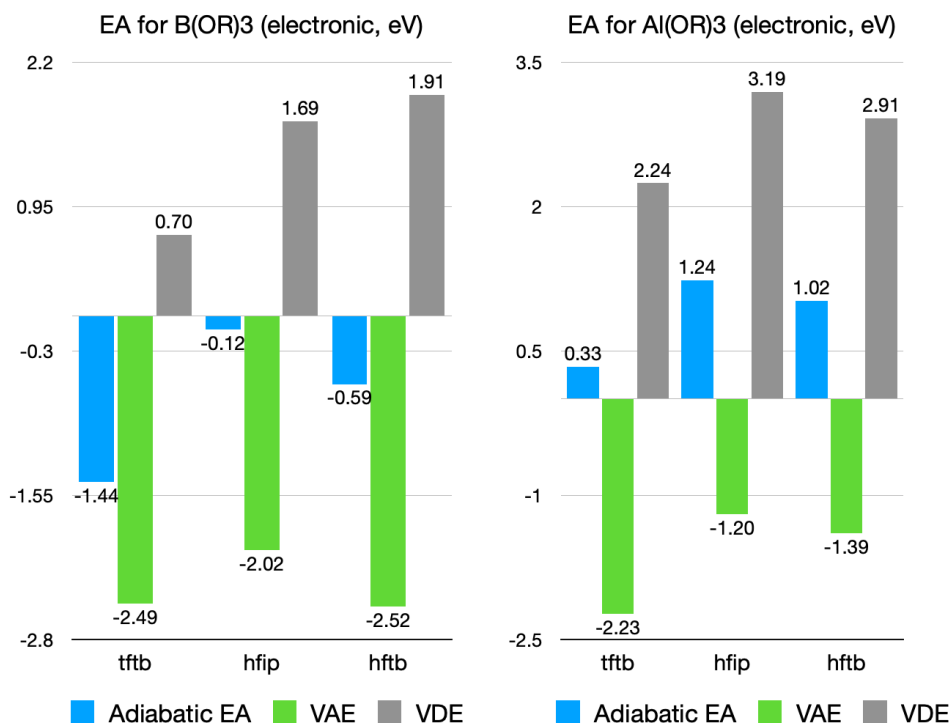


Figure 5.59: Various electron affinity values for Al(OR)₃ and B(OR)₃

4.8 Spin density plots for structures along the Al/B-O bond breaking decomposition pathways

For aluminates, TS1 and TS2 correspond to the two possible rate-limiting TSs as shown in Figure 5.5(a) in the main text. In the detailed energy diagrams, SM, TS1, TS2, and PDT correspond to Struct1, TS2, TS3 and Struct4 in Figure 5.17, Figure 5.18 and Figure 5.19. For borates, TS corresponds to the rate-limiting TS as shown in Figure 5.5(a) in the main text. In the detailed energy diagrams, SM, TS, PDT correspond to Struct1, TS3, Struct4 for $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$ (Figure 5.27), and Struct1, TS1, Struct2 for $\text{Ca}^+[\text{B}(\text{hfip})_4]^-$ (Figure 5.28) and

$\text{Ca}^+[\text{B}(\text{hftb})_4]^-$ (Figure 5.29). The iso-surface for spin density at each geometry was plotted at an isovalue 0.0004 (e/Bohr^3). As can be seen from the spin density plots, for aluminates the spin densities are partially transferred from Ca^+ to $\text{Al}(\text{OR})_3$ (mostly Al atom) at TS1 and fully transferred at TS2; for borates, the spin densities are partially transferred from Ca^+ to $\text{B}(\text{OR})_3$ (mostly B atom) at TS. The degree of charge transfer is larger for TS1 in aluminates than TS in borates.

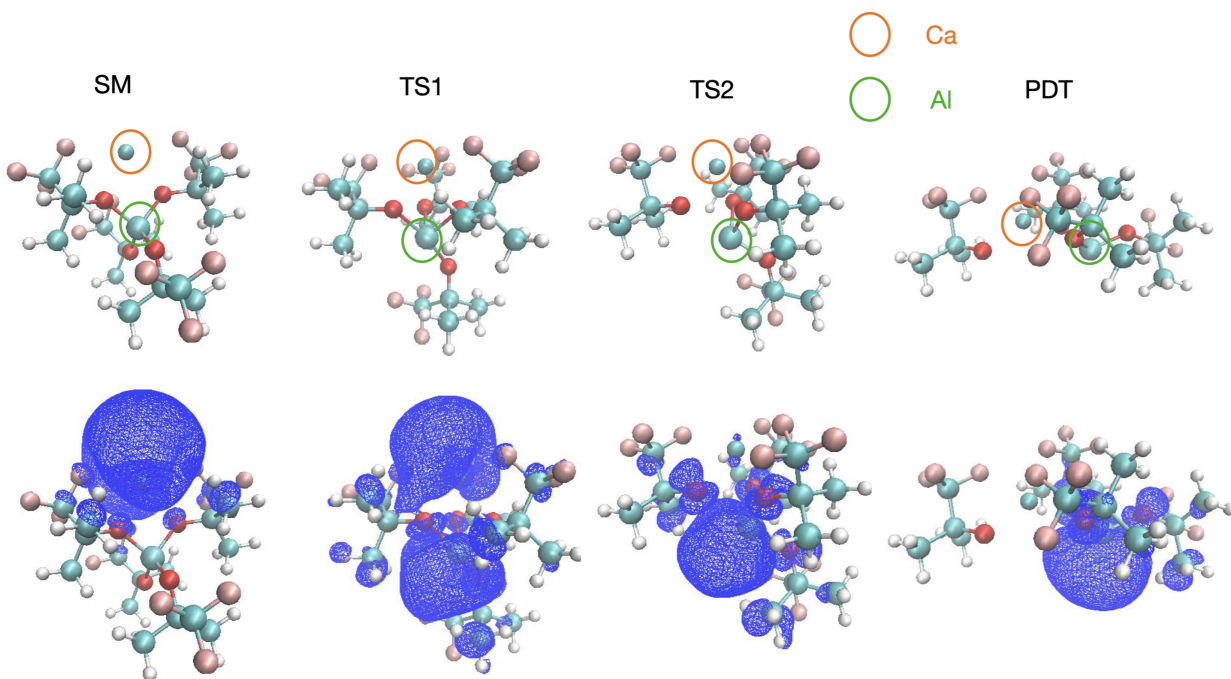


Figure 5.60: Spin density plot for $\text{Ca}^+[\text{Al}(\text{tftb})_4]^-$

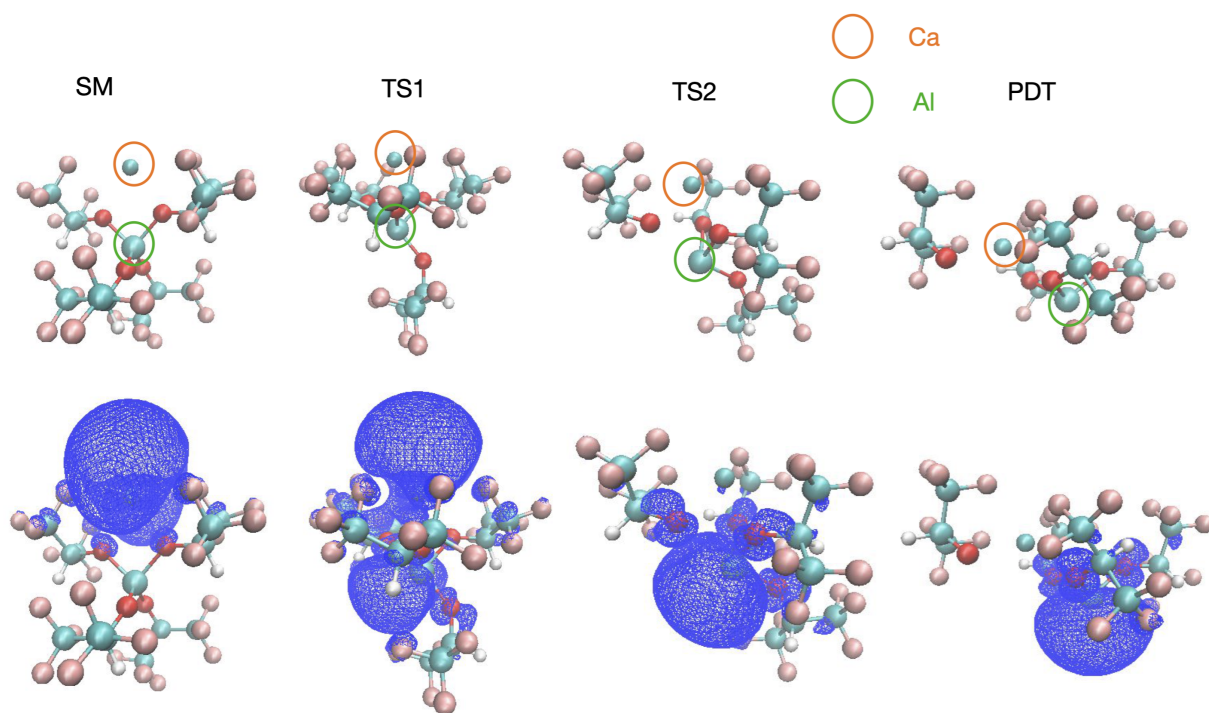


Figure 5.61: Spin density plot for $\text{Ca}^+[\text{Al}(\text{hfip})_4]^-$

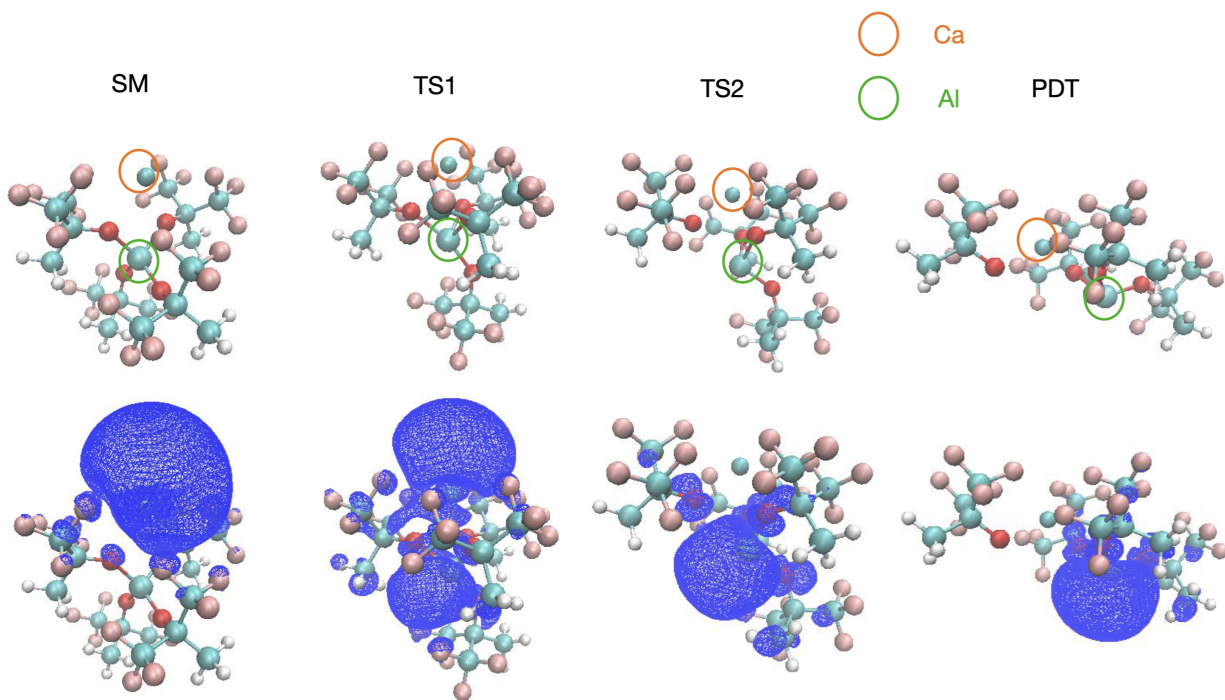


Figure 5.62: Spin density plot for $\text{Ca}^+[\text{Al}(\text{hftb})_4]^-$

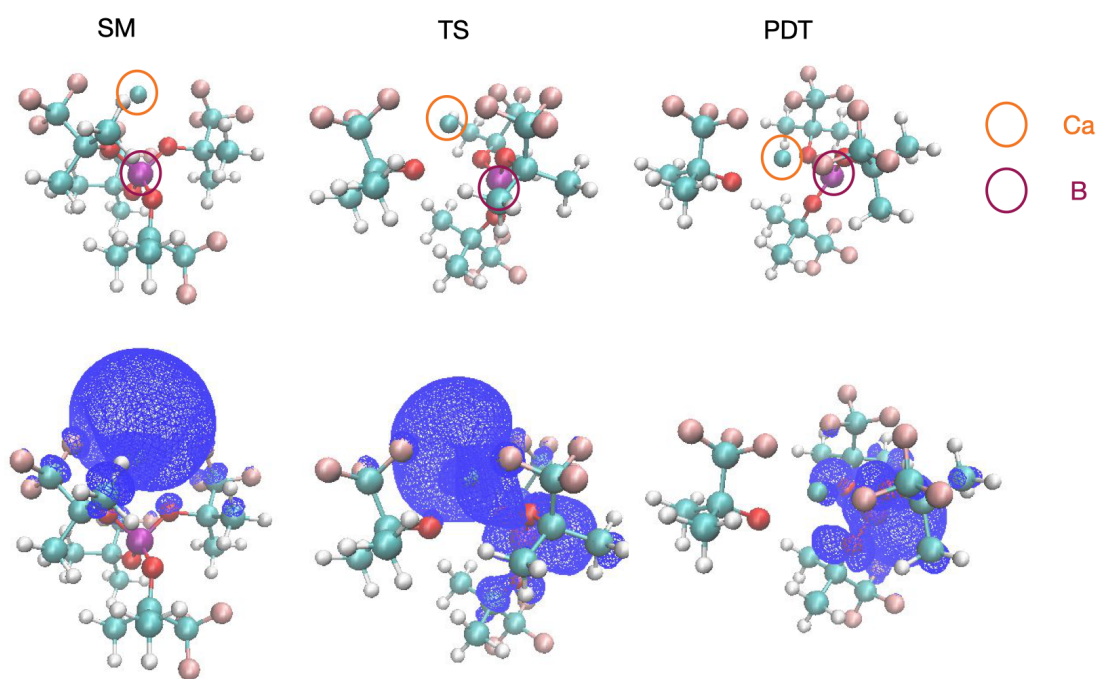


Figure 5.63: Spin density plot for $\text{Ca}^+[\text{B}(\text{tftb})_4]^-$

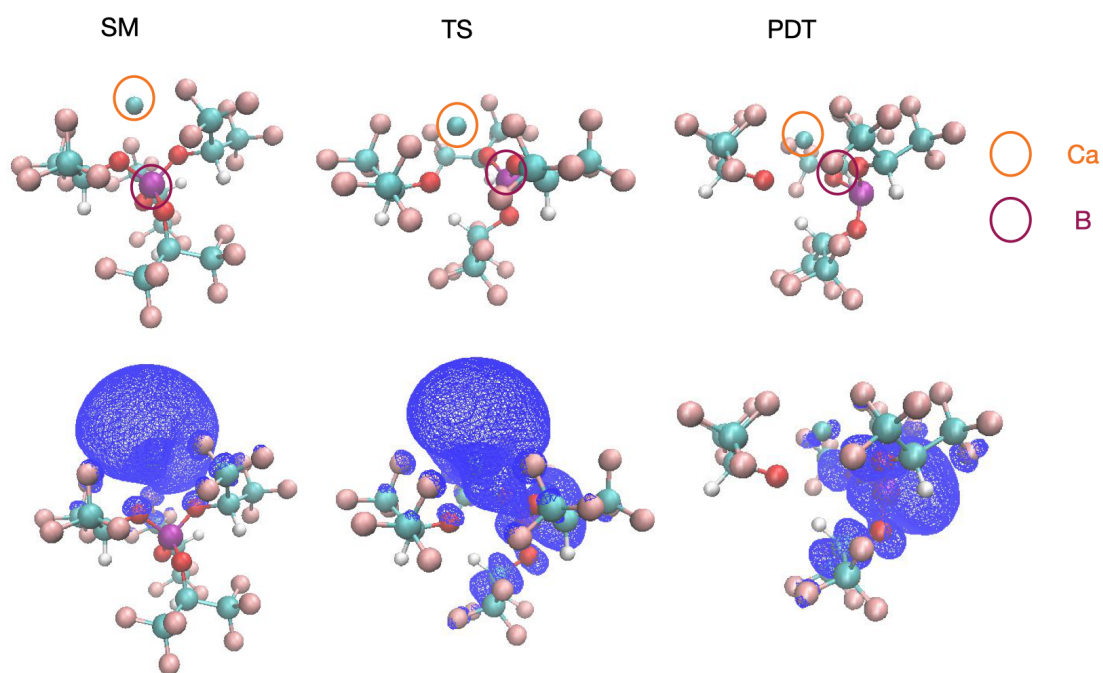


Figure 5.64: Spin density plot for $\text{Ca}^+[\text{B}(\text{hfp})_4]^-$

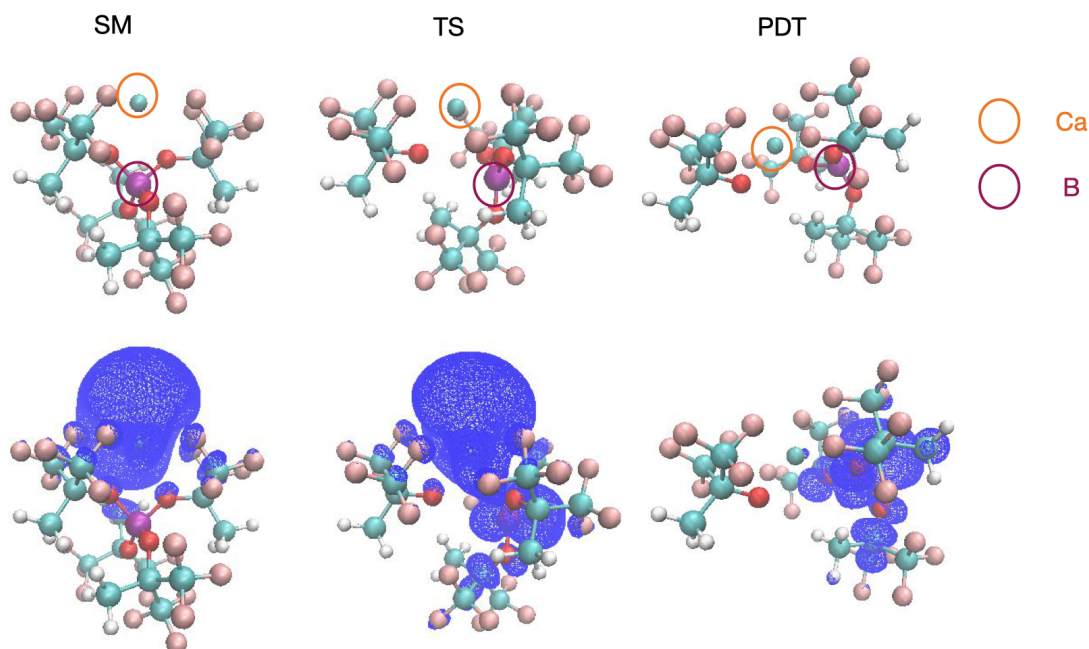


Figure 5.65: Spin density plot for $\text{Ca}^+[\text{B}(\text{hftb})_4]^-$

4.9 Computational and experimental details

4.9.1 Computational details

DFT calculations for decomposition pathways were performed using the QChem[49] 5.3 electronic structure code. Geometry optimizations were performed at the B3LYP-D3(BJ)[27–29]/def2-SVP[44] level of theory, while single point energies at the stationary points were obtained using the B3LYP-D3(BJ)/def2-TZVPPD[43, 44] level of theory with the C-PCM implicit solvation model[45] using the dielectric constant and optical dielectric constant of dimethoxyethane (DME). We used the lower level of theory for computational efficiency in exploring the potential energy surface (PES); TS calculations are very cost prohibitive as they often require numerous iterations, encounter convergence difficulties, and can require multiple hessian calculations to successfully converge. Vibrational frequencies were computed at the stationary points using the theory level that produced them (i.e. the lower level), as required for obtaining the associated zero-point energies and thermal corrections. A procedure based on taking small fixed-length steps in the steepest descent direction (the Euler’s method[50]) was performed to verify the TSs connect to the expected reactant(s) and product(s) through a smooth energy-descent path.[51]

For computing the oxidation/reduction potentials of the CIPs and the bare anions, we employed two levels of theory for comparison. The first one is the same as what we used for the reaction pathway calculations mentioned above, except that the C-PCM model was used during geometry optimization because the involved dication or dianion species would not usually remain stable in the gas phase. For the second level of theory, geometries were optimized at the ω B97X-D3[30]/def2-SVPD[44]/CPCM level of theory using the RIJCOSX[52] approximation as implemented in the ORCA[53] 5.0.2 code, while single point energies were obtained using the ω B97X-D3/def2-TZVPPD/CPCM/RIJCOSX level of theory.

4.9.2 Experimental details

All experiments were performed using a dry box under an argon atmosphere. Perfluorotert-butanol (pftb, Oakwood Chemical), 1,1,1,3,3,3-hexafluoro-2-methyl-2-propanol (hftb, Sigma-Aldrich), 2-trifluoromethyl-2-propanol (tftb, Sigma-Aldrich), 1,1,1,3,3,3-hexafluoro-2-propanol (hfip, Sigma-Aldrich), and 1,1,1-trifluoro-2-propanol (tfip, Sigma-Aldrich) were purchased and dried over 3 Å molecular sieves for two days before use. All solvents were purchased from Sigma Aldrich and dried over 3 Å molecular sieves. $\text{Mg}(\text{BH}_4)_2$ was purchased from Sigma Aldrich and used as received. NMR spectra were recorded on a 300 MHz Bruker Avance spectrometer at ambient temperature. ^{11}B NMR spectra were referenced and locked externally to $\text{BF}_3 \cdot \text{Et}_2\text{O}$ solution in CDCl_3 (δ 0 ppm).

The target $\text{Mg}[\text{B}(\text{OR})_4]_2$ salts were synthesized in one pot using adapted procedures similar to that of $\text{Mg}[\text{B}(\text{hfip})_4]_2$ and $\text{Mg}[\text{B}(\text{pftb})_4]_2$. [13, 54] In 20 mL borosilicate scintillation vials, $\text{Mg}(\text{BH}_4)_2$ powder (27 mg, 0.5 mmol) was dissolved in DME to form mixtures of 0.1M. The solutions were cooled to -20°C in a glovebox freezer loaded into aluminum heating blocks. To the stirring, cooled mixtures were added the respective alcohols (ROH, 8.2 eq, 4.1 mmol), dropwise, at which point, very slow hydrogen evolution occurred. The reaction mixtures were allowed to warm to room temperature until no evolution was observed. Each reaction was stirred and heated to 60°C overnight to ensure full conversion of borohydride starting material. Leftover B-H borohydrides can be observed by non-decoupled ^{11}B NMR with characteristic doublets, often between 0-20 ppm. Aliquots of each reaction mixture were loaded into J-young NMR tubes and held at either room temperature or 75°C for one week before NMR spectra were recorded. In the reaction mixtures containing tftb, hftb, and pftb, small amounts of solid precipitate were observed at the bottoms of the J-young NMR tubes.

Chemical shifts:

$\text{Mg}[\text{B}(\text{tfip})_4]_2$:

(RT, 1 week) ^{11}B NMR (96 MHz, DME-in situ): δ 22.18 (minor), 4.73.

(75°C , 1 week) ^{11}B NMR (96 MHz, DME-in situ): δ 22.22 (minor), 18.86 (minor), 9.51(d, minor), 4.67 (s).

$\text{Mg}[\text{B}(\text{hfip})_4]_2$:

(RT, 1 week) ^{11}B NMR (96 MHz, DME-in situ): δ 5.42.

(75°C, 1 week) ^{11}B NMR (96 MHz, DME-in situ): δ 5.39.

$\text{Mg}[\text{B}(\text{tftb})_4]_2$:

(RT, 1 week) ^{11}B NMR (96 MHz, DME-in situ): δ 19.90.

(75°C, 1 week) ^{11}B NMR (96 MHz, DME-in situ): δ 19.50.

$\text{Mg}[\text{B}(\text{hftb})_4]_2$:

(RT, 1 week) ^{11}B NMR (96 MHz, DME-in situ): δ 17.50, 3.82.

(75°C, 1 week) ^{11}B NMR (96 MHz, DME-in situ): δ 17.15, 3.46.

$\text{Mg}[\text{B}(\text{pftb})_4]_2$:

(RT, 1 week) ^{11}B NMR (96 MHz, DME-in situ): δ 21.33, 6.03.

(75°C, 1 week) ^{11}B NMR (96 MHz, DME-in situ): δ 21.37, 6.04.

Bibliography

- [1] John Muldoon, Claudiu B Bucur, and Thomas Gregory. “Quest for nonaqueous multivalent secondary batteries: magnesium and beyond”. In: *Chemical reviews* 114.23 (2014), pp. 11683–11720.
- [2] Pieremanuele Canepa et al. “Odyssey of multivalent cathode materials: open questions and future challenges”. In: *Chemical reviews* 117.5 (2017), pp. 4287–4341.
- [3] Yaosen Tian et al. “Promises and challenges of next-generation “Beyond Li-ion” batteries for electric vehicles and grid decarbonization”. In: *Chemical reviews* 121.3 (2020), pp. 1623–1669.
- [4] Bengt Elgquist and Margareta Wedborg. “Stability of ion pairs from gypsum solubility degree of ion pair formation between the major constituents of seawater”. In: *Marine Chemistry* 3.3 (1975), pp. 215–225.
- [5] Masaki Okoshi et al. “Theoretical analysis on de-solvation of lithium, sodium, and magnesium cations to organic electrolyte solvents”. In: *Journal of the Electrochemical Society* 160.11 (2013), A2160.
- [6] Nav Nidhi Rajput et al. “The coupling between stability and ion pair formation in magnesium electrolytes from first-principles quantum mechanics and classical molecular dynamics”. In: *Journal of the American Chemical Society* 137.9 (2015), pp. 3411–3420.
- [7] Robert E Doe et al. “Novel, electrolyte solutions comprising fully inorganic salts with high anodic stability for rechargeable magnesium batteries”. In: *Chemical Communications* 50.2 (2014), pp. 243–245.
- [8] Ivgeni Shterenberg et al. “Evaluation of (CF₃SO₂)₂N-(TFSI) based electrolyte solutions for Mg batteries”. In: *Journal of The Electrochemical Society* 162.13 (2015), A7118.
- [9] Justin G Connell et al. “Tuning the reversibility of Mg anodes via controlled surface passivation by H₂O/Cl[−] in organic electrolytes”. In: *Chemistry of Materials* 28.22 (2016), pp. 8268–8277.

- [10] Justin G Connell et al. “Anion Association Strength as a Unifying Descriptor for the Reversibility of Divalent Metal Deposition in Nonaqueous Electrolytes”. In: *ACS Applied Materials & Interfaces* 12.32 (2020), pp. 36137–36147.
- [11] Jake T Herb, Carl A Nist-Lund, and Craig B Arnold. “A fluorinated alkoxyaluminate electrolyte for magnesium-ion batteries”. In: *ACS Energy Letters* 1.6 (2016), pp. 1227–1232.
- [12] Toshihiko Mandai, Yong Youn, and Yoshitaka Tateyama. “Remarkable electrochemical and ion-transport characteristics of magnesium-fluorinated alkoxyaluminate–diglyme electrolytes for magnesium batteries”. In: *Materials Advances* 2.19 (2021), pp. 6283–6296.
- [13] Zhirong Zhao-Karger et al. “A new class of non-corrosive, highly efficient electrolytes for rechargeable magnesium batteries”. In: *Journal of Materials Chemistry A* 5.22 (2017), pp. 10815–10820.
- [14] Zhenyou Li et al. “Towards stable and efficient electrolytes for room-temperature rechargeable calcium batteries”. In: *Energy & Environmental Science* 12.12 (2019), pp. 3496–3501.
- [15] Abhinandan Shyamsunder et al. “Reversible calcium plating and stripping at room temperature using a borate salt”. In: *ACS Energy Letters* 4.9 (2019), pp. 2271–2276.
- [16] Ka-Cheong Lau et al. “Widening electrochemical window of Mg salt by weakly coordinating perfluoroalkoxyaluminate anion for Mg battery electrolyte”. In: *Journal of The Electrochemical Society* 166.8 (2019), A1510.
- [17] Noel J Leon et al. “Room-Temperature Calcium Plating and Stripping Using a Perfluoroalkoxyaluminate Anion Electrolyte”. In: *The Journal of Physical Chemistry C* (2022).
- [18] Zhonghua Zhang et al. “Novel Design Concepts of Efficient Mg-Ion Electrolytes toward High-Performance Magnesium–Selenium and Magnesium–Sulfur Batteries”. In: *Advanced Energy Materials* 7.11 (2017), p. 1602055.
- [19] Aobing Du et al. “An efficient organic magnesium borate-based electrolyte with non-nucleophilic characteristics for magnesium–sulfur battery”. In: *Energy & Environmental Science* 10.12 (2017), pp. 2616–2625.
- [20] Darren M Driscoll et al. “Rationalizing Calcium Electrodeposition Behavior by Quantifying Ethereal Solvation Effects on Ca^{2+} Coordination in Well-Dissociated Electrolytes”. In: *Journal of the Electrochemical Society* 167.16 (2020), p. 160512.
- [21] Nathan T Hahn et al. “Concentration-dependent ion correlations impact the electrochemical behavior of calcium battery electrolytes”. In: *Physical Chemistry Chemical Physics* (2022).

- [22] Piotr Jankowski et al. “Development of Magnesium Borate Electrolytes: Explaining the Success of Mg [B (hfp) 4] 2 Salt”. In: *Energy Storage Materials* (2021).
- [23] Z Lu et al. “On the electrochemical behavior of magnesium electrodes in polar aprotic electrolyte solutions”. In: *Journal of Electroanalytical Chemistry* 466.2 (1999), pp. 203–217.
- [24] Seoung-Bum Son et al. “An artificial interphase enables reversible magnesium chemistry in carbonate electrolytes”. In: *Nature Chemistry* 10.5 (2018), pp. 532–539.
- [25] Kun Tang et al. “A stable solid electrolyte interphase for magnesium metal anode evolved from a bulky anion lithium salt”. In: *Advanced Materials* 32.6 (2020), p. 1904987.
- [26] Feilure Tuerxun et al. “Effect of interaction among magnesium ions, anion, and solvent on kinetics of the magnesium deposition process”. In: *The Journal of Physical Chemistry C* 124.52 (2020), pp. 28510–28519.
- [27] Axel D Beck. “Density-functional thermochemistry. III. The role of exact exchange”. In: *J. Chem. Phys* 98.7 (1993), pp. 5648–6.
- [28] Chengteh Lee, Weitao Yang, and Robert G Parr. “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density”. In: *Physical review B* 37.2 (1988), p. 785.
- [29] Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. “Effect of the damping function in dispersion corrected density functional theory”. In: *J. Comput. Chem.* 32.7 (2011), pp. 1456–1465.
- [30] You-Sheng Lin et al. “Long-range corrected hybrid density functionals with improved dispersion corrections”. In: *Journal of Chemical Theory and Computation* 9.1 (2013), pp. 263–272.
- [31] Trevor J Seguin et al. “Elucidating non-aqueous solvent stability and associated decomposition mechanisms for Mg energy storage applications from first-principles”. In: *Frontiers in chemistry* 7 (2019), p. 175.
- [32] HS Lee et al. “The synthesis of a new family of boron-based anion receptors and the study of their effect on ion pair dissociation and conductivity of lithium salts in nonaqueous solutions”. In: *Journal of the Electrochemical Society* 145.8 (1998), p. 2813.
- [33] Hannes Böhrer et al. “From unsuccessful H 2-activation with FLPs containing B (Ohfp) 3 to a systematic evaluation of the Lewis acidity of 33 Lewis acids based on fluoride, chloride, hydride and methyl ion affinities”. In: *Dalton Transactions* 44.16 (2015), pp. 7489–7499.

- [34] Francis A LeBlanc et al. “Synthesis, Characterization, and Properties of Weakly Coordinating Anions Based on tris-Perfluoro-tert-Butoxyborane”. In: *Inorganic Chemistry* 56.2 (2017), pp. 974–983.
- [35] Saul H Lapidus et al. “Solvation structure and energetics of electrolytes for multivalent energy storage”. In: *Physical Chemistry Chemical Physics* 16.40 (2014), pp. 21941–21945.
- [36] Eric D Glendening, Clark R Landis, and Frank Weinhold. “NBO 6.0: Natural bond orbital analysis program”. In: *Journal of computational chemistry* 34.16 (2013), pp. 1429–1437.
- [37] Jonathan C Rienstra-Kiracofe et al. “Atomic and molecular electron affinities: photoelectron experiments and theoretical computations”. In: *Chemical reviews* 102.1 (2002), pp. 231–282.
- [38] Richard FW Bader and TT Nguyen-Dang. “Quantum theory of atoms in molecules–Dalton revisited”. In: *Advances in Quantum Chemistry*. Vol. 14. Elsevier, 1981, pp. 63–124.
- [39] Julia Contreras-Garcia et al. “NCIPLOT: a program for plotting noncovalent interaction regions”. In: *Journal of chemical theory and computation* 7.3 (2011), pp. 625–632.
- [40] Erin R Johnson et al. “Revealing noncovalent interactions”. In: *Journal of the American Chemical Society* 132.18 (2010), pp. 6498–6506.
- [41] Tsong-Song Hwang and Yu Wang. “Topological Analysis of the Electron Density Distribution of Bis (diiminosuccinonitrilo) nickel, Ni (C₄N₄H₂)₂: Comparison between Experiment and Theory”. In: *The Journal of Physical Chemistry A* 102.21 (1998), pp. 3726–3731.
- [42] Paul Lode Albert Popelier, FM Aicken, and SE O’Brien. *Atoms in molecules*. Vol. 188. Prentice Hall Manchester, 2000.
- [43] Florian Weigend and Reinhart Ahlrichs. “Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy”. In: *Physical Chemistry Chemical Physics* 7.18 (2005), pp. 3297–3305.
- [44] Dmitrij Rappoport and Filipp Furche. “Property-optimized Gaussian basis sets for molecular response calculations”. In: *The Journal of chemical physics* 133.13 (2010), p. 134105.
- [45] Vincenzo Barone and Maurizio Cossi. “Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model”. In: *The Journal of Physical Chemistry A* 102.11 (1998), pp. 1995–2001.

- [46] Paul R Horn and Martin Head-Gordon. “Polarization contributions to intermolecular interactions revisited with fragment electric-field response functions”. In: *The Journal of chemical physics* 143.11 (2015), p. 114111.
- [47] Paul R Horn, Yuezhi Mao, and Martin Head-Gordon. “Defining the contributions of permanent electrostatics, Pauli repulsion, and dispersion in density functional theory calculations of intermolecular interaction energies”. In: *The Journal of chemical physics* 144.11 (2016), p. 114107.
- [48] Paul R Horn, Yuezhi Mao, and Martin Head-Gordon. “Probing non-covalent interactions with a second generation energy decomposition analysis using absolutely localized molecular orbitals”. In: *Physical Chemistry Chemical Physics* 18.33 (2016), pp. 23067–23079.
- [49] Evgeny Epifanovsky et al. “Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package”. In: *The Journal of chemical physics* 155.8 (2021), p. 084801.
- [50] Bruce C Garrett et al. “Algorithms and accuracy requirements for computing reaction paths by the method of steepest descent”. In: *The Journal of Physical Chemistry* 92.6 (1988), pp. 1476–1488.
- [51] Xiaowei Xie et al. “Data-Driven Prediction of Formation Mechanisms of Lithium Ethylene Monocarbonate with an Automated Reaction Network”. In: *Journal of the American Chemical Society* 143.33 (2021), pp. 13245–13258.
- [52] Frank Neese et al. “Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A ‘chain-of-spheres’ algorithm for the Hartree–Fock exchange”. In: *Chemical Physics* 356.1-3 (2009), pp. 98–109.
- [53] Frank Neese et al. “The ORCA quantum chemistry program package”. In: *The Journal of chemical physics* 152.22 (2020), p. 224108.
- [54] Jian Luo et al. “A stable, non-corrosive perfluorinated pinacolatoborate Mg electrolyte for rechargeable Mg batteries”. In: *Angewandte Chemie* 131.21 (2019), pp. 7041–7045.

Chapter 6

Conclusions and Future Work

Computational modeling of battery chemistry is always filled with great promises and challenges. There are so many parameters to be tuned only from the modeling perspective: effects of electrolyte/salt/additive composition including both their identities and concentrations, electrode surfaces, cathode materials, etc. A slight change in a single parameter could result in a change in microscopic properties such as the cation/anion solvation structures, the interfacial reactions, and the chemical composition of the surface film, which can further affect the macroscopic properties such as ionic conductivity, electrical conductivity, dielectric properties, viscosity, and those properties will eventually be linked to the overall battery performance. During my doctoral studies, I have demonstrated how to apply data-driven approaches to investigate battery chemistry, including the fundamental ML force field development for handling species seen in such systems, the usage of a massive reaction network with DFT data to probe interfacial reactions, and computational screening of novel electrolytes.

Within chapters 2 and 3, I discussed the development of BpopNN, a physics-based ML potential applicable to molecules with varying electronic states. ML approximations to DFT PESs are showing great promise for reducing the computational cost of accurate molecular simulations, but at present, they are not applicable to large-scale simulations of battery electrolytes or interfacial reactivities. An accurate description of long-range electrostatics and charge transfer is needed to achieve this goal. BpopNN addresses a common problem for most ML potentials based on local descriptors: they are not suitable for describing molecular systems in which the local electronic structure is sensitive to the medium to long-range electronic environment. Conceptually, the method is based on approaching the true DFT energy as a function of electron populations on atoms; in practice, this is realized with available density functionals and CDFT. The new approach creates approximations to this function with neural networks. These approximations thereby incorporate electronic information naturally into an ML approach, and optimizing the model energy with respect to populations allows the electronic terms to self-consistently adapt to the environment, as in DFT. The effectiveness of this approach was confirmed with a variety of calculations on LinHn clusters (chapter 2) and Li-organic systems (chapter 3). Interestingly, multiple SCF solutions can also be obtained with BpopNN, making it capable of modeling CT excited states.

In chapter 4, a data-driven approach, i.e. an automated reaction network was used to investigate whether LEMC or LEDC is the major organic component of the LIB SEI. The data-driven, automated method is based on a systematic generation of relevant species using a general fragmentation/recombination procedure which provides the basis for a vast thermodynamic reaction landscape, calculated with DFT. The shortest pathfinding algorithms are employed to explore the reaction landscape and obtain previously proposed formation mechanisms of LEMC as well as several new reaction pathways and intermediates. For example, we identify two novel LEMC formation mechanisms: one which involves LiH generation and another that involves breaking the $(\text{CH}_2)\text{O}-\text{C}(\text{O})\text{OLi}$ bond in LEDC. Most importantly, we find that all identified paths, which are also kinetically favorable under the explored condi-

tions, require water as a reactant. This condition severely limits the amount of LEMC that can form, as compared with LEDC, a conclusion that has direct impacts on the SEI formation in Li-ion energy storage systems. The data-driven framework presented here is also generally applicable to any electrochemical system and is expected to improve our understanding of surface passivation.

Chapter 5 described a computational electrolyte design rationale for Mg/Ca-ion battery applications. The stabilities of the weakly-coordinating fluorinated alkoxyborate and alkoxyaluminate salts are examined, based on ^{11}B NMR experiments (in collaboration with Dr. Noel J. Leon at Argonne national laboratory) and DFT calculations of their redox potentials and detailed reductive decomposition pathways involving the breaking of Al/B-O, C-O or C-F bonds. The kinetic barriers for such decomposition pathways are obtained and they are demonstrated to correlate well with their electrochemical performance for existing electrolytes, e.g. $\text{Mg}[\text{B}(\text{hfip})_4]_2$, $\text{Ca}[\text{B}(\text{hfip})_4]_2$, $\text{Mg}[\text{Al}(\text{hfip})_4]_2$, $\text{Mg}[\text{Al}(\text{pftb})_4]_2$, $\text{Ca}[\text{Al}(\text{pftb})_4]_2$. Based on the decomposition kinetics, the hfip ligand for borates and the tftb ligand for aluminates are identified as promising ligands for constructing the salt anions; the borate prediction corroborates the success of the B-hfip salts. An important distinction between borate and aluminate salts in the factors governing the Al/B-O bond breaking barriers is also discussed, suggesting useful design rules for functional electrolytes.

Rechargeable batteries have already changed the world and are supposed to become an indispensable part of a future zero-carbon society. With more computing power and high-quality data becoming available, beyond conventional trial-and-error approaches, the combination of experimental, theoretical, and data tools is strongly expected to establish a data-driven research paradigm for exploring the nature of chemistry and designing functional materials for next-generation energy storage applications.

There are many open directions for future research:

MD simulations with BpopNN

In this thesis, I presented the application of BpopNN for static calculations, i.e. structure and energy prediction, reaction pathway analysis, etc. In the future, I intend to utilize the NN potential to perform large-scale MD simulations. One ultimate goal is to simulate the formation of the SEI with an ML force field that is capable of capturing possible reactivities but at a much reduced computational cost than DFT. This could bridge the gap between AIMD and classical MD simulations and afford fundamental insights into the atomistic-level SEI formation processes.

Another interesting direction is to use BpopNN to conduct grand-canonical MD simulations. BpopNN, unlike AIMNet or SpookyNet, can naturally deal with fractional number of electrons. In AIMNet[1] and SpookyNet[2], separate NNs (or separate embeddings) are used to predict the energies and forces for molecules at different integer total charge states. The

target energy function for BpopNN, on the contrary, is defined for any possible (integer or fractional) total charge/spin states, and a universal ML model is used for all the total population choices. In the field of battery simulations, we are naturally interested in the impact of electrode potentials on ionic/electron transport and interfacial reaction mechanisms, because batteries are typically operated at a fixed potential. Indeed, a significant change in the reaction profile was observed for the Li intercalation processes with the constant-potential vs constant-charge approaches. [3] This could be a promising application of BpopNN although some challenges need to be overcome, e.g., how to solve for the fermi energy of the electrode within the ML force field.

It is also of interest to use BpopNN to conduct excited state dynamics in general. Literature is scarce on this topic, but there could exist electron transport processes in the SEI formation that involve CT excited states. Very limited excited state interatomic potentials are transferable to different compounds and excited states of different electronic characters. [4, 5] ML functionals based on atomic electron populations like BpopNN have the potential to achieve cheaper but widely applicable excited state ML interatomic potentials.

Improving the scheme for sampling training data

One reason that BpopNN is not yet applicable to large-scale MD simulations is the lack of efficient sampling of structures. As a result, if the geometry encountered in MD simulations falls out of the training regime, the energy and forces deviate a lot from the true values (sometimes even lead to NaN), causing unstable MD trajectories or even unphysical structures. In order to tackle this problem, I need to go beyond the often-used method of using geometry optimization and (or) MD trajectories to sample structures, but to incorporate intelligent methods to actively search and selectively include useful reactive structures, such as meta-dynamics[6] and active learning.[7] This is especially true for complex systems like the SEI where higher energy parts of the potential energy surface are frequently seen in the possible reactions.

Improving the physics in the ML model

The energy expression we used in BpopNN for incorporating long-range electrostatics (a damped tanh function) is presumably too simple, especially considering the Becke populations are not configured to optimally match the electrostatic potentials. Other charge schemes (Hirshfeld,[8] Mulliken,[9] Restrained Electrostatic Potential (RESP) charges[10]) should be tested and compared with Becke charges. More complicated functional forms can also be developed to represent the correct long-range electrostatic behaviors.

The atomistic descriptor employed also needs improvement. Although SOAP[11] is a successful descriptor in this field, it suffers from the loss of angular information when contracting over m , inhibiting its overall expressive power of complicated geometries. Learnable

descriptors[12] or equivariant features[13] are popular options to tackle this problem.

Investigating the impact of different electrolytes/additives/electrodes for the SEI formation with the reaction network

In this thesis, I investigated the formation mechanisms of LEMC when using EC-based electrolytes in LIB, with or without water participation. It is worthwhile to investigate how the reaction mechanisms change with different electrolyte formulations, different additives involved, etc. The effect of the electrode is also important, but it is currently only accounted for through using different electron chemical potentials (to represent the different electrode potentials) and no chemical reactivities involving electrodes are taken into consideration. For a better description of the interfacial reactivities, we hope to incorporate the chemical nature of the electrodes into the reaction network in the future.

Automating TS search for computational electrolyte design

TS searches are expensive and error-prone, especially for the kind of ion pairs for battery research seen in chapter 5. This is presumably why kinetics are less used than thermodynamics as metrics for screening promising electrolytes. Luckily, automated TS search tools exist, e.g., AutoTS, [14] and it would be helpful if the TS search processes can be automated with a high success rate for this kind of molecular clusters and thus rational electrolyte design can be further accelerated. AutoTS is successfully used in our research group for kinetic Monte Carlo simulations for the Li-anode SEI formation. [15]

Bibliography

- [1] Roman Zubatyuk et al. “Teaching a neural network to attach and detach electrons from molecules”. In: *Nature communications* 12.1 (2021), pp. 1–11.
- [2] Oliver T Unke et al. “SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects”. In: *Nature communications* 12.1 (2021), pp. 1–14.
- [3] Tuan Anh Pham et al. “Intercalation of Lithium into Graphite: Insights from First-Principles Simulations”. In: *The Journal of Physical Chemistry C* 124.40 (2020), pp. 21985–21992.
- [4] Simon Axelrod, Eugene Shakhnovich, and Rafael Gómez-Bombarelli. “Excited state, non-adiabatic dynamics of large photoswitchable molecules using a chemically transferable machine learning potential”. In: *arXiv preprint arXiv:2108.04879* (2021).
- [5] Pavlo O Dral and Mario Barbatti. “Molecular excited states through a machine learning lens”. In: *Nature Reviews Chemistry* 5.6 (2021), pp. 388–405.
- [6] John E Herr et al. “Metadynamics for training neural network model chemistries: A competitive assessment”. In: *J. Chem. Phys.* 148.24 (2018), p. 241710.
- [7] Leif D Jacobson et al. “Transferable Neural Network Potential Energy Surfaces for Closed-Shell Organic Molecules: Extension to Ions”. In: *Journal of Chemical Theory and Computation* 18.4 (2022), pp. 2354–2366.
- [8] Fred L Hirshfeld. “Bonded-atom fragments for describing molecular charge densities”. In: *Theoretica chimica acta* 44.2 (1977), pp. 129–138.
- [9] Robert S Mulliken. “Electronic population analysis on LCAO–MO molecular wave functions. I”. in: *The Journal of Chemical Physics* 23.10 (1955), pp. 1833–1840.
- [10] Christopher I Bayly et al. “A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model”. In: *The Journal of Physical Chemistry* 97.40 (1993), pp. 10269–10280.
- [11] Albert P Bartók, Risi Kondor, and Gábor Csányi. “On representing chemical environments”. In: *Phys. Rev. B* 87.18 (2013), p. 184115.

- [12] Kristof T Schütt et al. “SchNet—A deep learning architecture for molecules and materials”. In: *J. Chem. Phys.* 148.24 (2018), p. 241722.
- [13] Nathaniel Thomas et al. “Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds”. In: *arXiv preprint arXiv:1802.08219* (2018).
- [14] Leif D Jacobson et al. “Automated transition state search and its application to diverse types of organic reactions”. In: *Journal of chemical theory and computation* 13.11 (2017), pp. 5780–5797.
- [15] Evan Walter Clark Spotte-Smith et al. “Toward a Mechanistic Model of Solid–Electrolyte Interphase Formation and Evolution in Lithium-Ion Batteries”. In: *ACS Energy Letters* 7.4 (2022), pp. 1446–1453.