Experimental and Computational Electrochemistry to Move Toward Plastics Circularity

by

Joseph Maalouf

M.S., Chemical Engineering Practice, Massachusetts Institute of Technology, 2019

B.S. Chemical Engineering with Honors and Distinction, Stanford University, 2017

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemical Engineering
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2023

© 2023 Joseph Maalouf. All rights reserved.

The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Author . . . .

Joseph Maalouf
Department of Chemical Engineering
February 17, 2023

Certified by . . . .

Karthish Manthiram
Professor in Chemical Engineering
Thesis Supervisor

Certified by . . . .

Yuriy Román
Robert T. Haslam (1911) Professor of Chemical Engineering
Thesis Supervisor

Accepted by . . . .

Patrick S. Doyle
Robert T. Haslam (1911) Professor of Chemical Engineering
Department Committee on Graduate Theses

Experimental and Computational Electrochemistry to Move Toward Plastics Circularity

by

Joseph Maalouf

Submitted to the Department of Chemical Engineering
on February 17, 2023, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Chemical Engineering

Abstract

This thesis work is grounded primarily in the goal of leveraging the potent yet fine tunable nature of an electrochemical driving force to tackle key issues in augmenting the chemical recyclability of plastics, namely the synthesis of plastic monomers and the deconstruction of existing plastics. Increasing plastic circularity will be crucial in decarbonizing the 400 Mt of plastic generated annually and the associated climate and environmental effects that result from producing plastics on this scale. While the key chemical reaction involved in the synthesis of plastics is the polymerization of monomers, the goal of this thesis is to demonstrate that electrochemistry – both experimental and computational – has a role to play in the synthesis of novel plastic monomers in addition to allowing for new potential decomposition pathways for the plastics in use today. This thesis can be broken down into three parts: (1) the experimental demonstration of sustainable synthesis of circular monomers using electrochemistry (2) the computational study of organic redox mediators with the potential for polystyrene deconstruction(3) and the implementation of data driven models to improve the throughput of computational screening.

Thesis Supervisor: Karthish Manthiram
Title: Professor in Chemical Engineering

Thesis Supervisor: Yuriy Román
Title: Robert T. Haslam (1911) Professor of Chemical Engineering

Acknowledgments

This long journey is not one that was completed alone. I have had the pleasure of meeting so many new amazing people during my time at MIT, while also being supported by those that I met beforehand. I'd like to start off by thanking my scientific mentor for the last 5 years - Karthish. Your positivity and scientific support made this journey possible for me. I recall many times leaving your office feeling as though I could take on the world, and that ability to motivate people is such a rare skill. Your willingness to trust your students and give them intellectual freedom is not something I took for granted, as I was able to learn and study topics that I did not think would be possible when starting in your lab. You have also built a great community of lab members at Cal Tech (shout out to them!), I know you are in good hands – thank you. I also need to thank the Manthiram lab members throughout my time here at MIT. Kindle, Zack, Nik, Nathan, Joy, Minju, Katie, Fang-Yu, Trent, Simar, Sayandeep, Kyoungsuk, Hee Jo, Dengtao, Ruquan, Minghui, you all provided a wonderful environment to conduct science in, and I'm looking forward to the work you guys put out in the future. A special thanks to Joy and Minju for being some of the best office mates during the last year. Additionally, I would like to thank Yuriy and the Roman lab for providing a social and scientific community to work in during my final year. I learned a lot from everyone in the lab and had a blast.

Tennis was a large part of my life during my PhD, and so was the club tennis team. I'd like to thank everyone on the club for providing such a great community, and for all the fun nights. The PhD is not an easy thing, so I also want to thank the team for letting me come out to practice and therapeutically beat down on them on the court.

A lot of people made my time outside the lab filled with some amazing memories. I want to thank my roommate for the last 5 years, Alex. I certainly did not expect us to have spent so much time together after meeting in undergrad, but I'm so glad we did. We have had so many fun nights out, and I always enjoy hitting with you, I look forward to seeing where our friendship goes. To my other friends, Becki, Daniel, Andrew, Irene, Sarah, Chloe, Mary, Ameya, Kevin, you all made the PhD fly by, so thank you for the support and fun times we have all shared.

I also want to thank my partner Maddie for the support throughout these last couple years. We have shared so many wonderful times together, and I look forward to what the future holds for us. Seeing how hard you work as a labor organizer is always inspiring and makes me keep my own work in perspective. I can't wait to see what amazing things you do.

Lastly, to my family. Truly without their support I would not be here. To Renee and Gaby, somehow we are all still in school, but that is all changing this year. I can't wait for us to all enter the next phase of our lives together, and I can't wait to see your success. To my mom and dad, thank you for always being there and keeping everything in perspective. You always made sure to remind me that you were proud of me and that I would always have your support, no matter where my journey took me.

Contents

1 Introduction 19
1.1 Climate Change and Decarbonization 19
1.2 The Plastics Problem and Circularity 20
1.3 Lactones 24
1.4 Electrochemistry for Chemical Synthesis 25
1.4.1 Water as an Oxygen Atom Source 27
1.5 Polystyrene Upcycling 28
1.6 Rapid Discovery of Electro-active Molecules 29
1.6.1 Machine Learning for the acceleration of high throughput computational screening 33
1.7 Brief Overview of Thesis Work 36

2 Electrochemical Synthesis of Lactones using Water as the Oxygen Atom Source 39
2.1 Introduction 40
2.2 Data and Discussion 42
2.3 Experimental Methods for Lactonization Chemistry 50
2.3.1 Materials 50
2.3.2 Methods 50
2.3.3 Supplementary Figures 54

3 Computational Screening of Organic Hydrogen Atom Transfer Redox Mediators 59

3.1 Introduction 60
3.2 Results and Discussion 62
    3.2.1 Virtual Library Generation and Calculating Properties 62
    3.2.2 Experimental Characterization 68
3.3 Conclusion 68
3.4 Supporting Information 70
    3.4.1 Methods 70
    3.4.2 DFT Calculations 73
    3.4.3 Calculated Thermodynamic Quantities 74
    3.4.4 Machine Learning Models 75
    3.4.5 N-Ammonium Ylide Synthesis 76
    3.4.6 NMR 78
3.5 Supplementary Figures 79
    3.5.1 NMR Spectra 83

4 Incorporating Machine Learning into the prediction of Organic PCET Redox Potentials in Multiple Solvents 87

4.1 Introduction 88
4.2 Datasets 94
    4.2.1 E^o 94
    4.2.2 ΔH_f, ΔS^o, and ΔG_{solv} 94
4.3 Methods 95
    4.3.1 DFT Calculations 95
4.4 Results and Discussion 95
    4.4.1 Machine Learning Models 95
    4.4.2 Prediction of Redox Potential 100
4.5 Conclusion 105
4.6 Supporting Information 107
4.7 Figures 107
4.8 Tables 111

5 Conclusions and Future Directions **115**

5.1 Conclusions 115

5.2 Future Directions 115

5.2.1 Lactonization 116

5.2.2 vHTP screening of redox mediators for plastic deconstruction 117

List of Figures

1-1 Scheme of closed loop recycling. 22

1-2 Scheme of open loop recycling and chemical upcycling. 23

1-3 a. The overall Baeyer-Villiger reaction using peroxyacids. Cyclohexanone is transformed into ϵ-caprolactone. b. The mechanism of the Baeyer-Villiger reaction, showing 3 crucial steps that facilitate the reaction.$\mathrm{R_M}$ = migrating group. 25

1-4 The BV reaction mechanism using Pt homogeneous complexes. 26

1-5 a. Direct electrochemical organic oxidation. b. Mediated electrochemical organic oxidation. [65] 27

1-7 Virtual high throughput screening workflow. 31

1-6 Jacobs ladder of DFT functionals. By including more and more information about the electron density into the expression for $E_{xc}[\rho]$, higher accuracy can typically be obtained. 32

1-8 General scheme of how GNNs work. 35

2-1 Scheme of a typical Baeyer-Villiger reaction using cyclohexanone and a peroxy acid as the oxygen atom source in the top reaction. The corresponding carboxylic acid is produced as a stoichiometric byproduct. The middle and bottom arrow show electrochemical lactone formation from cyclohexanone using water as the oxygen atom source with an applied potential (This work). Hydrogen gas is produced at the cathode. 42

2-2 a,b) Mass spectra of γ-caprolactone and δ-hexanolactone using either 1 M $^{16}OH_2$ and $^{18}OH_2$. In the case of $^{18}OH_2$, $^{16}O_2$ was also flowed in at 10 sccm to test to see if O_2 from air was acting as an oxygen atom source. 200 mM cyclohexanone and 1 M $^{18}OH_2$ were used. (c) Quantification of H_2 via GC was performed, showing near 100% FE toward H_2 production. 400 mM cyclohexanone and 10 M H_2O were used for H_2 quantification. (d) The proposed pathways for each of the observed M/Z shifts from experiments using labeled water is shown. M/Z shift of +2 can be explained by two different pathways. For simplicity, only the formation of δ-hexanolactone is shown. 44

2-3 (a) Cyclic voltammograms collected with all combinations of the presence of cyclohexanone and water at a scan rate of 50 mV/s. When present, 400 mM cyclohexanone and 10 M H_2O are used. (b) The Tafel slope is shown with 100% manual IR correction, described in the SI. A value of 139 mV is obtained at 400 mM cyclohexanone, 10 M H_2O and 350 mM $TBABF_4$. (c) Cyclohexanone order dependence study, collected with 10 M water and 350 mM $TBABF_4$ present, at a potential of 2.15 V vs Fc/Fc+. (d) Water order dependence study, collected at 400 mM cyclohexanone and 350 mM TBABF4, at a potential of 2.15 V vs Fc/Fc+. 46

2-4 Products observed from reactions employing different cyclic lactone substrates. All runs are conducted at 400 mM Substrate, 10 M H_2O, 350 mM TBA BF_4, applying 2.15 V vs Fc/Fc+ and passing 50 C using Pt foil electrodes. 48

2-5 (a) Mechanism consistent with previous literature of similar transformations. The literature suggests an outersphere pathway that begins with an initial electron transfer from cyclohexanone. (b) Proposed mechanism based off of the collected electrochemicalkinetic data. Shows an inner sphere mechanism on Pt that undergoes an initial electron transfer involving water as the rate determining step, followed by reaction with cyclohexanone and another electron transfer. 49

2-6 Electrochemical sandwich cell used for all experiments. (A) Aluminum foil current collector. (B) Platinum foil counter electrode. (C) Ag/AgCl reference electrode. (D) Aluminum foil current collector. (E) Polished platinum foil working electrode. 52

2-7 Redox Peak of 1 mM ferrocene in ACN with 350 mM TBABF4, 400 mM cyclohexanone with varying amounts of water (1, 5, 10, 12 M). 54

2-8 Calibration curves for γ-caprolactone (a) and δ-hexanolactone (b) using GCMS. The fitted equation for (a) was calculated to be y=1.066E-4x + 0.511 and for (b) was calculated to be y=2.566E-4x + 0.410. 54

2-9 Calibration curves for γ-caprolactone (a) and δ-hexanolactone (b) using GC-FID. The fitted equation for (a) was calculated to be y=4.744E-4x + 0.191 and for (b) was calculated to be y=5.923E-4x + 0.244. 55

2-10 (a) Faradaic efficiencies measured with GCFID as functions of the applied potential. (b) as functions of cyclohexanone concentration, and (c) as functions of water concentration. Standard conditions are 400 mM cyclohexanone, 10 M H2O, and 2.15 V vs Fc/Fc+. (d) Table with the plotted data from a-c. The reported potential has been IR corrected. 56

2-11 Cyclohexanone order dependence collected at 2.15 V Vs Fc/Fc+ and 1 M H2O. 57

2-12 Cyclic voltammetry with 400 mM cyclohexanone, 10 M H2O, 350 mM TBABF4, comparing a platinum foil working electrode with a glassy carbon electrode. . 57

2-13 XPS spectra for Pt 4f conducted on a Pt foil before and after reaction. The pre-reaction spectrum is shown in blue, and is consistent with reported spectra of Pt metal. The spectra after passing 15 C of charge with 10 M H2O, 400 mM Cyclohexanone, 350 mM TBABF4, and 2.15 V vs Fc/Fc+ is shown in red. The development of the shoulder at approximately 76 eV, and has been previously assigned to oxide forms of Pt. 58

3-1 a) Catalytic cycle of ylide redox mediators for HAT reactions. b) Virtual high throughput computational workflow developed for redox mediator design in this work. 63

3-2 a)The two HAT reactions for 1,3-diphenylpropane, identified as (BDE 1) and (BDE 2) with their associated BDE energies. b) Free energy landscape of an ylide redox mediators showing the electron transfer, hydrogen abstraction and deprotonation. R-H = substrate, B = Base. c) Scatter plot of ΔG_{CH} and ΔG_{DP} for all DFT analyzed ylides. 65

3-3 a)The two HAT reactions for 1,3-diphenylpropane, identified as (1) and (2) with their associated BDEs. b) Free energy landscape of an ylide redox mediators showing the electron transfer, hydrogen abstraction and deprotonation. c) Scatter plot of ΔG_{CH} and ΔG_{DP} for all DFT analyzed ylides. 66

3-4 a) Selected Ylide candidates that were chosen for further examination. b) Performance of the selected candidates for HAT chemistry. 68

3-5 Top: structures of ylides that were successfully synthesized in this work. Bottom: Cyclic voltammetry of selected ylides. 100 mM of ylide was added to MeCN with 100 TBA BF_4 as the electrolyte. 69

3-6 Original Reaxys carboxylic acid query. 70

3-7 Molecular connections made using RDkit to form ylide library. 71

3-8 Model substrates considered in this work. Energies correspond to the BDE of removing a hydrogen atom from the indicated C-H bond. 72

3-9 General synthesis steps to make N-ammonium ylides when starting from a carboxylic acid, as shown in 3.4.5. 76

3-10 a) Ylides that contain at least one halogen in the set F, Cl, Br,I. It is possible for molecules in this plot to contain multiple halogens, in which case the ylide was plotted for each case. b) Ylides containing at least one of the amine types specified in the legend. c) Ylides containing alcohols and ethers. It is possible for other functional groups to exist on the molecule, such as amines or halogens. d) Ylides containing P atoms. e) ylides containing sulfur atoms. f) Ylides that contain only carbon atoms (excluding the ylide motif). No other functional groups are present on the molecule in this case. 79

3-11 Histograms of a) E^o, b) ΔG_{DPFE}, and c) ΔG_{H_bind} from calculations performed at the M062X/def2-SVPD level of theory. 80

3-12 Scatter plots from calculations performed at the M062X/def2-SVPD level of theory of a) E^o vs ΔG_{DPFE}, b) E^o vs $\Delta G_{H_{bind}}$, and c) $\Delta G_{H_{bind}}$ vs ΔG_{DPFE}. 80

3-13 Histogram of the wall time of each ylide, ylide radical, and protonated ylide calculation. 81

3-14 Selected candidates that were synthesized in this work. 82

3-15 NMR Spectra of Y1 taken in deuterated MeCN. 83

3-16 NMR Spectra of Y2 taken in deuterated MeCN. 84

3-17 NMR Spectra of Y3 taken in deuterated MeCN. 85

3-18 NMR Spectra of Y6 taken in deuterated MeCN. 86

4-1 Born-Haber cycle, demonstrating how the free energy of a redox reaction can be broken up into its constituent gas phase and solvation free energies. Note that charges on A and B are implicit, such that conservation of charge is maintained. 90

4-2 Summary of all approaches used to calculate redox potentials. Approach 1 uses DFT is used to calculate gas phase thermodynamics along with SMD to account for solvation. Approach 2 uses DFT to calculate gas phase thermodynamics and uses a machine learning model to obtain the solvation free energy (ΔG_{solv}). Approach 3 uses a machine learning model to predict gas phase free energies ($\Delta H_f, \Delta S^o$) and a separate machine model (same as Approach 2) to predict ΔG_{solv} 93

4-3 Architecture of our model to predict solvation free energies. SMILES strings of the solute and solvent are transformed into graph representation where atom and bond features are assigned. Convolutions are performed with a GNN to form embedding that are concatenated and then passed through an MLP for property prediction. 96

4-4 Parity plot of GNN used to fit ΔG_{solv} .RMSE in kcal/mol. 98

4-5 Architecture of our model to predict solvation free energies. SMILES strings of the solute and solvent are transformed into graph representation where atom and bond features are assigned. Convolutions are performed with a GNN to form embeddings that are concatenated and then passed through an MLP for property prediction. 99

4-6 Parity plot of GNN used to fit ΔH_f (a) and S^o (b). 100

4-7 a,c: Parity plots of approach 2 where gas phase free energies are predicted using DFT and solvation free energies are predicted using a machine learning model. DFT calculations are at the **B3LYP/6-31(2df,p)** level of theory. b,d: Parity plots of approach 2 where gas phase free energies are predicted using an ML model and solvation free energies are predicted using a machine learning model. 104

4-8 Comparison of performance of the DFT and DFT + ML approaches. Results for all possible combinations of the functionals B3LYP and M062X and the basis sets 6-31++G** and 6-31G(2df,p) are shown. 107

4-9 Comparison of performance of the DFT + ML approach (approach 1) based on solvent. Results for all possible combinations of the functionals B3LYP and M062X and the basis sets 6-31++G** and 6-31G(2df,p) are shown. . . 108

4-10 Comparison of performance of the DFT + ML approach (approach 1) based on the number of electrons transferred during the reaction (n). Results for all possible combinations of the functionals B3LYP and M062X and the basis sets 6-31++G** and 6-31G(2df,p) are shown. 109

4-11 All molecules used in the E^o dataset with the solvent and redox potential vs NHE indicated below. 110

List of Tables

3.1 Substructures filtered from original Reaxys query. 71

3.2 Summary of the number of failed calculations for the 10,000 catalytic cycles, corresponding to 30,000 DFT calculations 74

3.3 Definitions of ylide thermodynamic quantities used for screening. 75

3.4 Table of the average wall time per molecule of each of the three species in the ylide catalytic cycle from calculations at the M062X/def2-SVPD level of theory with SMD solvation in acetonitrile. 81

4.1 Classes of electron transfer reactions considered in this work. We considered both 1 and 2 electron proton coupled electron transfers. 94

4.2 Summary of datasets for various thermochemical quantities used to build data driven models in this work. 95

4.3 Performance metrics of solvation free energy GNN models. 98

4.4 Results of GNN models to predict gas phase thermochemistry. Error bars are the standard deviation of 10 random splits and trained models. 99

4.5 Performance of different approaches depending on the number of electrons passed during the reaction (n). DFT calculations are at the **B3LYP/6-31(2df,p) level of theory.** 101

4.6 Performance of different approaches on specific solvents. DFT calculations are at the **B3LYP/6-31(2dfp) level of theory.** 105

4.7 Performance of different approaches depending on the number of electrons passed during the reaction (n). MAE= mean average error. MAE_n = mean average error for a specific n value. RMSE = root mean squared error.DFT calculations are at the **B3LYP/6-31++G**** level of theory. 111

4.8 Performance of different approaches depending on the number of electrons passed during the reaction (n). MAE= mean average error. MAE_n = mean average error for a specific n value. RMSE = root mean squared error.DFT calculations are at the **M062X/6-31++G**** level of theory. 111

4.9 Performance of different approaches depending on the number of electrons passed during the reaction (n). MAE= mean average error. MAE_n = mean average error for a specific n value. RMSE = root mean squared error. DFT calculations are at the **M062X/6-31(2dfp)** level of theory. 112

4.10 Performance of different approaches on specific solvents. DFT calculations are at the **B3LYP/6-31++G****level of theory. 112

4.11 Performance of different approaches on specific solvents. DFT calculations are at the **M062X/6-31++G****level of theory. 112

4.12 Performance of different approaches on specific solvents. DFT calculations are at the **M062X/6-31G(2df,p)**level of theory. 113

Chapter 1

Introduction

1.1 Climate Change and Decarbonization

The threat of our society's dependence on fossil fuels has never been so imminent, with rising temperatures and extreme weather events becoming more commonplace due to ever-increasing amounts of anthropogenic carbon dioxide (CO_2). As a result, society is in a race to decarbonize all sectors of our economy, with a common goal being net-zero emissions by 2050. This will be difficult, as every aspect of life results in the emissions of green house gases. Most emitting processes can be placed into one of five categories: how we make things (materials), electricity generation and consumption, growing things (agriculture), transportation, and keeping and cool and staying warm [15]. Each category roughly accounts for 10-30 % of the 51 billion tons of CO_2 equivalents that we emit as a planet each year. However, the role of each sector in total emissions has historically received varying attention. For example, there have been great strides forward in the development of technologies that will allow for the decarbonization of electricity generation, namely wind and solar power. While electricity generated from these renewable sources still has issues pertaining to their intermittency and thus inability to always meet baseload electricity demand, research in the area of energy storage is alleviating these concerns. Additionally, because the grid is becoming increasingly carbon free, electrification has developed as a robust approach to decarbonizing many processes and products. A clear example is the electrification of passenger vehicles where Tesla has dominated the private market and is only aided by the recent Inflation Reduction

Act which offers thousands of dollars in subsidies toward electric vehicles that source their battery materials and produce their batteries in the US.

Electricity generation and transportation are important, but one area that has historically received less attention in the context of decarbonization is the materials sector. The materials sector can be broken up into the production of cement, steel, plastics, and other petrochemicals chemicals, which together account for 31% of our yearly emissions. Unlike, electricity and transportation, technological advances are still required to allow for the carbon free production of these materials. Cement and steel production release CO_2 both stoichiometrically due to certain chemical reactions and due to elevated temperatures that are achieved from burning fossil fuels. Promising start-ups (Sublime Systems, Boston Metal) have developed novel approaches for the production of cement and steel, often utilizing electrification for either the production of process heat through Joule heating or the facilitation of chemical reactions at an electrode. Plastics suffer from similar issues as cement and steel, with emissions projected to increase drastically in the coming decades. Like cement and steel, fossil fuels are burned to produce the necessary heat and energy for their synthesis – this accounts for a majority of their associated emissions. However, there are some key differences pertaining to plastics – as they are made now, they require fossil fuel carbon feedstocks and emissions after post consumer use are also of great concern. This thesis focuses on developing experimental and computational tools that can help address some of the issues associated with plastics, including both their production and their deconstruction after consumer use. The tools used in this thesis are those of electrochemistry and computational chemistry. Electrochemistry serves a powerful yet fine-tunable driving force for that can be used to synthesize novel plastics, and computational chemistry can aid in developing new materials that can leverage clean electricity to deconstruct some of the common plastics used today.

1.2 The Plastics Problem and Circularity

As a society, humans have produced nearly 9 gigatonnes (Gt) of plastic since plastic production began to accelerate in the 1950s [16]. Currently, we produce approximately 400 Mt of plastic resin annually with the most common plastics being polyethylene (PE), polypropylene (PP), polyethylene terephthalate (PET), polyvinyl chloride (PVC), and polystyrene (PS). From packaging, to containers, to textiles – these materials are everywhere and have become pervasive in our society because of their useful properties and low cost. The issues come when we consider how they are produced and what happens to these materials after consumer use. Traditional plastics are synthesized from petroleum and its derivatives. For example, polyolefins are made from ethylene and propylene, with PET and polystyrene coming from benzene, toluene, and xylene (BTX) derivatives. Upon first considerations, plastics may seem like a carbon sink because the carbon is stored in the material, but their synthesis requires non-negligible amounts of heat and energy that is obtained by burning fossil fuels. Life-cycle assessments often quote that the majority of emissions from plastics come from their production.

While plastic production emits large amounts of CO_2, one cannot ignore detrimental effects that plastics bring about after they serve their purpose. Plastic is most often used in a linear fashion meaning that plastic products are discarded after serving their primary purpose, any recycling done is only for a small amount of the total plastic produced (<10%). Nearly half of all plastic produced has ended up in landfills, an additional 10% has been incinerated for cheap process heat, resulting in additional emissions.[16] Even when in landfills bacteria can degrade these materials into potent greenhouse gases such as methane, meaning emissions continue after consumer use. It is clear that society needs to move away from the traditional linear way in which it uses plastics and move toward plastic circularity. Circularity means using plastics (or any resource) more efficiently by keeping the material in use for as long as possible, getting the most we can from the material during its use, and then recovering it to make new products. This could imply scaling up the traditional mechanical recycling that we use now, but this has key limitations. Mechanical recycling involves mechanically degrading products into fine particles that can be processes into pellets of plastic resin. These pellets are often of much lower quality than the original plastic fed into the recycling process, meaning that the products that they can go on to make are quite limited. As a result, materials are typically only mechanically recycled once before ultimately ending up in a landfill. Furthermore, there is a severe limitation on what types of materials can be currently mechanically recycled. Most municipalities are only capable of mechanically recycling PET, PE, and some PP meaning a large fraction of materials only have the option of being incinerated or going to a landfill. Chemical recycling is a new paradigm for recycling that is under extensive research and offers many benefits over more traditional mechanical recycling.

Chemical recycling comes in different forms, but at its core chemically recycling a polymer implies that the bonds in the polymer backbone are being broken, resulting in smaller molecules.

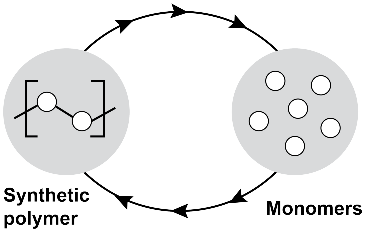


Figure 1-1: Scheme of closed loop recycling.

One form of chemical recycling is referred to as closed loop chemical recycling – in this case plastics would be depolymerized back into their constituent monomers via some process and then these monomers could then be polymerized to make a product of a similar quality to the original plastic, something not possible with mechanical recycling [37]. In this paradigm, the quantity of and emissions associated with virgin plastic resin synthesized would be greatly diminished. As one might expect, this property does not belong to every plastic, especially those most commonly used today (PE, PP, PET, PS). Thus, we will need to design new classes of plastics if we are to adopt this paradigm. A key metric of a polymer that can give insight into whether not it will undergo closed loop chemical recycling is its ceiling temperature (T_c), the ceiling temperature is defined as the temperature at which the monomer and polymer are in equilibrium and thus have the same Gibbs free energy(ΔG).[6, 50] In most cases, below T_c the monomer state is favored and above T_c polymerization is favorable, but this depends on the individual enthalpy and entropy of polymerization as we recall that

$$\Delta G = \Delta H - T\Delta S \tag{1.1}$$

Polymers with a medium T_c of 200 °C – 400 °C , typically for C–X(heteroatom) backboned polymers, can be redesigned to be intrinsically circular polymers (iCPs) or chemically re/up-cycled through engineering reaction/processing conditions and/or catalytic processes. ICPs are polymers that favor the monomer state at room temperature, with a kinetic barrier keeping them in the polymer state. This with the appropriate catalyst they can be depolymerized with much greater ease than traditional plastics [50]. One important class of medium T_cC-X backboned monomers are lactones, which can be polymerized into polyesters via ring opening polymerization (ROP). Great work has been done in designing lactones monomers – by varying their structure polymers with different properties can be synthesized [59, 18, 49, 48]. This is important because the property requirements of plastics are completely application specific, and thus a wide range of polymer properties will be required if iCPs are to replace traditionally used plastics. A detailed discussion of lactones for circularity can be found in 1.3.

In contrast to iCPs, traditional plastics have ceiling temperature such as 610 °C for PE and 395 °C for PS, making it difficult to deconstruct these materials into their constituent polymers in a controlled fashion. As a result, these plastics are more amenable to a second kind of plastic circularity – open loop circularity. Open loop recycling is defined as the deconstruction of waste plastics into molecular intermediates that are manufactured into products that differ from the starting polymer feedstock [37]. In the case that the manufactured product is of higher value than the starting plastic, we refer to this as chemical upcycling. Open loop thermal processes like gasification and pyrolysis are typically used at scale today – these produce simpler products that will likely end up as fuels or chemicals, with some fraction going to making new polymers. Catalytic technologies could allow for plastic upcycling of materials beyond polyolefins and polyesters, which are the polymer classes of primary focus to date. Developing new methods for recycling polymers that are not recy-

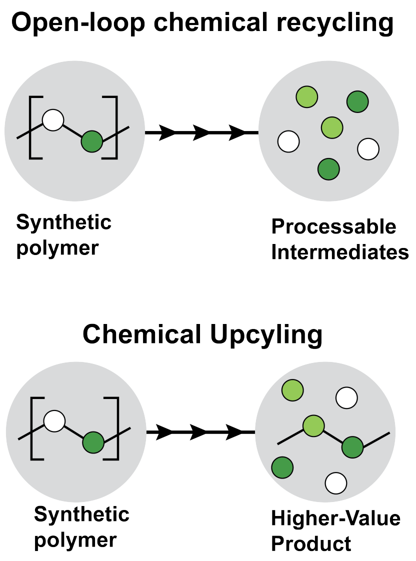


Figure 1-2: Scheme of open loop recycling and chemical upcycling.

clable today is a key step toward circularity. Furthermore, due to the scale at which plastics are produced and used, it is likely that multiple upcycling processes will end up being used, even for a given type a plastic. For example, for PS there are thermal, photocatalytic, and electrochemical approaches that are currently under investigation. The details of PS deconstruction are discussed in more detail in 1.5

1.3 Lactones

Lactones are a key class of molecules that have the potential to serve as scalable iCPs for closed loop circularity.[59] These cyclic esters, can be polymerized via ring opening polymerization into polyesters whose properties are dictated by the structure of the starting monomer. Various lactone derivatives have been shown to serve as good monomers for synthesizing iCPs. Often the ring size and substitution are altered to augment the polymerization thermodynamics and kinetics, with the inclusion of fused rings also being a common tactic. [47, 6, 49] If this class of molecule is to be used at scale to replace traditional plastics, it is important to understand how they are typically synthesized today. In an industrial setting, lactones can be synthesized from cycloalkanones via the Bayer-Villiger(BV) reaction, with one of the most common lactones being ϵ-caprolactone synthesized from cyclohexanone and peroxy-acetic acid.

The reaction proceeds by first activating the ketone oxygen in an acidic manner, followed by a nucleophilic attack from the oxygen on the peroxy group, forming the Criegee intermediate which facilitates the migration of an R group and thus forming the lactone product (Figure 1-3). The BV reaction can also be facilitated catalytically, for example in the homogeneous case Pt complexes have found great success using hydrogen peroxide. They operate via mechanisms as show in Figure 1-4. Again, one can see that the ketone is activated in a Lewis acidic manor, after which a nucleophilic Pt-OOH oxygen species attacks the carbonyl carbon forming a Criegee like intermediate. Lastly, the BV reaction has also been shown to run heterogeneously on materials such as β-Sn zeolites. In this context, the Sn active sites are able to form nucleophilic oxygen and activate the ketone oxygen to allow the formation of a Criegee like intermediate. While these catalytic materials have shown great success, at

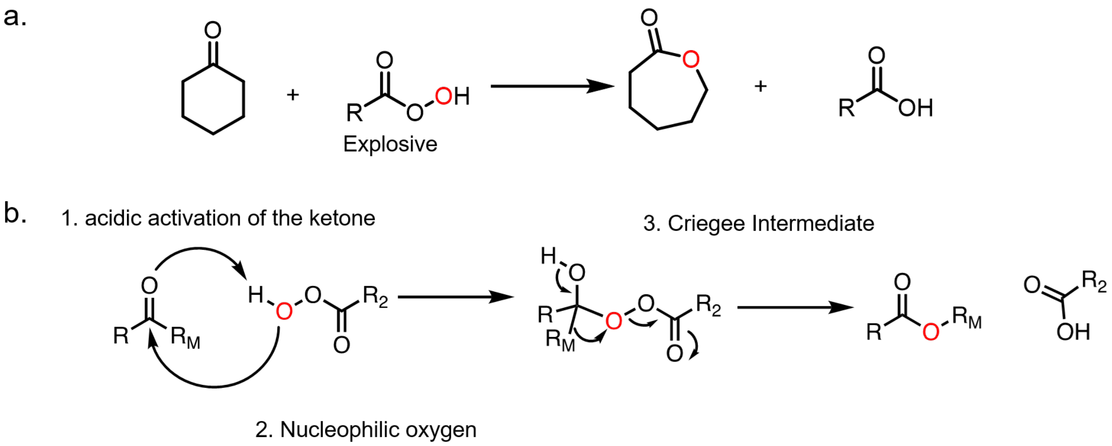


Figure 1-3: a. The overall Baeyer-Villiger reaction using peroxyacids. Cyclohexanone is transformed into ϵ-caprolactone. b. The mechanism of the Baeyer-Villiger reaction, showing 3 crucial steps that facilitate the reaction.R_M = migrating group.

an industrial scale peroxyacids are still the oxidants of choice, but there are crucial issues with the current process that will only be exacerbated if the scale of lactone production is to increase to meet the future projected demand of circular plastics. The drawbacks mainly pertain to requirement of a strong oxidant such as peroxy-acids, these chemicals pose acute safety hazards as they are extremely reactive, to the point of being an explosion hazard. Additionally, the process generates carboxylic acids as low value stoichiometric byproducts that need to be separated and regenerated into their peroxy acid counterparts using H_2O_2, which is generated via the anthroquinone process in an industrial setting. Lactone synthesis would benefit from the ability to use a safer oxidative driving force that maintains the same potency and a more benign oxygen atom source, both of these issues can be addressed by implementing an electrochemical approach to lactone synthesis that uses water as an oxygen atom source.

1.4 Electrochemistry for Chemical Synthesis

Electrochemical organic synthesis is the process of facilitating a chemical reaction using an applied external potential to transfer electrons to or from an organic molecule. As previously discussed, one of the key benefits of this approach is its potential use of an increasingly

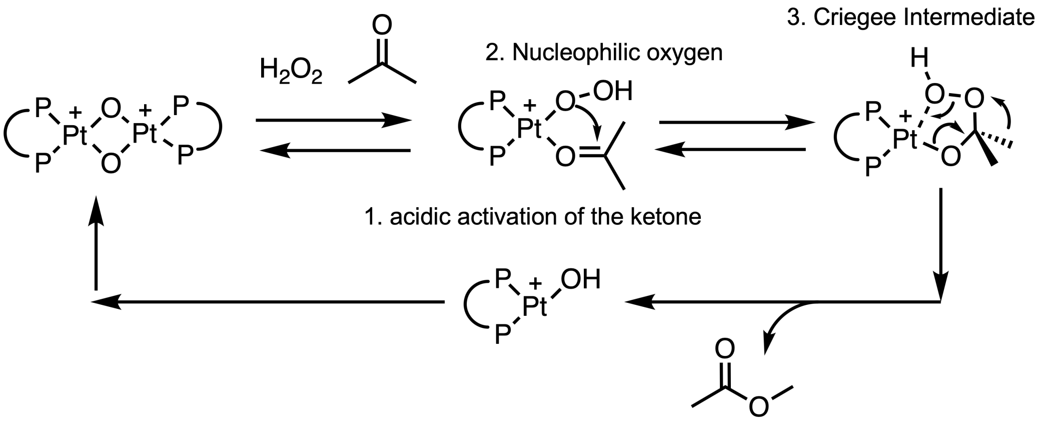


Figure 1-4: The BV reaction mechanism using Pt homogeneous complexes.

carbon free electricity grid, resulting in reduced process emissions. In general, an electron transfer can either be done directly at an electrode or indirectly by using another type of molecule called a redox mediator (Figure 1-5). Both approaches are explored in this thesis. In direct electrochemical oxidation, the molecule undergoes transport to an electrode (anode for oxidation and cathode for reduction) where the molecule experiences a potential drop. The applied voltage alters the energy states of the electrode material until it is thermodynamically favorable for an electron in the highest occupied molecular orbital (HOMO) of the organic molecule to transfer to the lowest occupied molecular orbital (LUMO) of the electrode material. There are instances where a direct electron transfer is difficult, resulting in a high over potential (η). The over potential is defined as the difference between the applied potential and the equilibrium potential E^o, where the equilibrium potential is defined as

$$E^o = -\frac{\Delta G^o}{nF} \tag{1.2}$$

here n is the number of electrons transferred during the reaction, F is Faraday's constant and ΔG^o is the Gibbs free energy of the reaction. With this definition, we can say $\eta = E_{app} - E^o$. When η is large, one can try to use a redox mediator as an alternative, which can be one of many classes or molecules, ranging from organics to metal centered complexes. During mediated oxidation, the redox mediator is first oxidized directly at the electrode, producing an activate form of the mediator - a radical or radical cation, for example. The oxidized mediator then goes on to chemically react with the substrate of interest, producing the desired oxidized substrate and regenerating the mediator. In addition to reduced overpotentials, redox mediators are also used because they can facilitate a wide variety of chemistries. However, mediators make the system more complex and can suffer from deactivation mechanisms such as dimerization. For this reason, the choice to use a redox mediator should be dependent on the chemistry at hand.

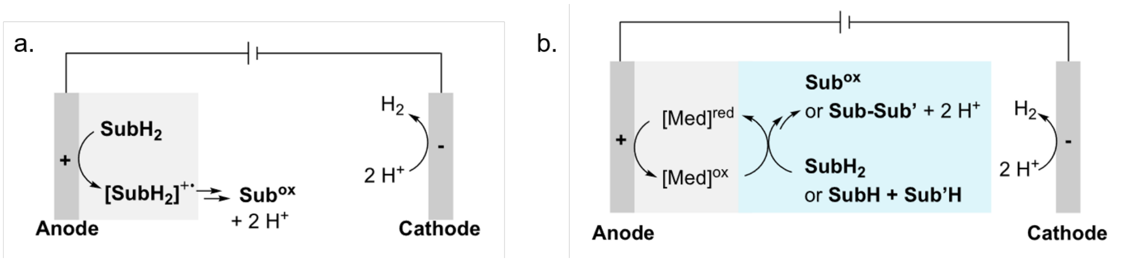


Figure 1-5: a. Direct electrochemical organic oxidation. b. Mediated electrochemical organic oxidation. [65]

1.4.1 Water as an Oxygen Atom Source

One of the most widely studied electrochemical reactants is water, primarily for its use as a source of hydrogen and oxygen through water electrolysis or in the context of fuel cells, where hydrogen and oxygen can react to form water and produce energy. For this reason, there has been a large body of work on the development of water oxidation electrocatalysts. For water oxidation, the typical reaction is

$$2\,H_2O \longrightarrow O_2 + 4\,H^+ + 4\,e^- \tag{1.3}$$

where water is oxidized to molecular oxygen via a 4 electron process. Leveraging this chemistry, an approach to electrifying chemical reactions that transfer oxygen to an organic molecule (epoxidation, lactonization) uses water as the oxygen atom source. This is in contrast to the thermochemical context, where O_2 is more likely to be used. This is because the oxidation of water is often a thermodynamically unfavorable reaction at room temperature, where it would take immensely high temperatures to shift the equilibrium toward the oxidized product. In contrast, potentials on the order of 1–2 volts are typically all that are required to facilitate water oxidation, making electrochemistry the clear tool to use when using water as the oxygen atom source.

$$
\begin{aligned}
&\mathrm{R} + \mathrm{H_2O} \longrightarrow \mathrm{RO} + \mathrm{H_2} \\
&\mathrm{R} + \mathrm{O_2} \longrightarrow \mathrm{RO}
\end{aligned}
\tag{1.4}
$$

Water has various benefits over oxidation with molecular oxygen or other chemical oxidants, like hydrogen peroxide or peroxy acids. First, water tends to be safer, peroxides and peroxy acids are known to be explosive in many contexts and O_2 can also pose flammability issues when it is in the presence of large amounts of organic material. Additionally, if the process is run in the liquid phase, water allows for a higher limiting current density because one is able to achieve higher concentrations. The limiting current is given by,

$$
i_{lim} = \frac{DC}{\delta} \tag{1.5}
$$

where D is the diffusion coefficient, C is the concentration, and δ is the boundary layer in the system.

1.5 Polystyrene Upcycling

As previously mentioned, traditional commodity plastics used today are not as amenable to closed loop chemical circularity as other classes of molecules, such as lactones. For this reason, commodity chemicals are instead thought of as feedstocks for chemical upcycling processes, where the materials can be converted into more valuable products. Because of the differing chemical backbones of PE, PP, PET, and PS, it is likely that various polymer-specific upcycling technologies will need to be developed [23]. A plastic of particular interest is polystyrene (PS), PS is produced on the order of 15 mega tonnes (MT) per year still making it a large scale product, although PE, PP, and PET are all made on larger scales. One can find it in materials such as styrofoam, plastic utensils, and more. Unlike the more inert PE and PP, due to the presence of periodic aromatic groups on the polymer backbone, PS is amenable to functionalization and degradation via hydrogen atom transfer (HAT) reactions. HAT reactions can generally be classified as the following reaction class,

$$\mathrm{R} \longrightarrow \mathrm{R}^{\bullet} + \mathrm{H}^{\bullet} \tag{1.6}$$

where a proton and electron are removed from a molecule in the form of a hydrogen atom. Notably, redox mediators have been previously shown to be able to facilitate these HAT reactions, typically in the presence of molecular oxygen, where the organic radical reacts with O_2 to eventually form ketones and alcohols [25, 70]. Yan and coworkers used a common HAT redox mediator – N-hydroxyphthalimide (NHPI) – to deconstruct styrene into processable intermediates [69]. They found that after H atom abstraction and reaction with O_2 a C-C bond scission mechanism occurred that resulted in the production of smaller oxygenated molecules. Their results demonstrated that mediated electrochemistry is a viable tool for PS deconstruction, but there were some still some key issues that needed to be addressed. A significant portion of the NHPI mediator was lost during the reaction, resulting in NHPI oligomerization, and thus inhibiting long term electrolysis. This can be mitigated through a better matching of the redox mediator energy to the C-H at hand. Typically, the energy of a molecule can be tuned via functionalization, but in the case of NHPI, only a handful of accessible derivatives exists. Recently, a new class of mediators, called N-ammonium ylides were demonstrated to facilitate HAT reactions with the key property of being easily functionalizable [46]. This synthetic accessibility opens up the chance to use valuable tools in the realm of high throughput computational screening that were previously restricted. In short, there is now the opportunity to design ylide mediators based on their calculated energetics – this concept is discussed in more detail in chapter 3.

1.6 Rapid Discovery of Electro-active Molecules

The computational screening tools used to design HAT redox mediators in this thesis have become common place in other fields, such as computational drug discovery. In those fields, the typical work flow involves screening a library of preselected molecular structures by calculating key properties specific to their application. These properties include solubility, free energy, docking energy, etc. Notably, the structures in drug discovery tend to be small neutral molecules (< 10 atoms), which makes them very amenable to the rapid calculation of their properties. The tools used to obtain these molecular properties typically include molecular dynamics, wave function techniques, and density functional theory (DFT). DFT tends to be one of the more widely used tools for calculating properties because of its relatively good accuracy and speed, and for that reason it is the primary computational tool used in this thesis.

DFT attempts to solve the many body Schrödinger equation by attempting to solve for the electron density ρ_e. The time dependent Schrödinger equation is given by

$$i\hbar\frac{\partial\Psi(r,t)}{\partial t} = \hat{H}\Psi(r,t) \tag{1.7}$$

where Ψ is the wave function, and $\hat{H}$ is the Hamiltonian, which is system dependent and accounts for all the forms of energy in the system. For molecules, the Hamiltonian is given by

$$\hat{H} = \hat{T}_N(R) + \hat{T}_e(r) + V_{e,N}(r,R) + V_{NN}(R) + V_{ee}(r) \tag{1.8}$$

$\hat{T}_N$ and $\hat{T}_e$ are the nuclear and electronic kinetic energy, while $V_{e,N}$, V_{NN}, and V_{ee} are the electron-nuclear, nuclear, and electronic coulombic attraction terms. Wave function methods, such as Hartree-Fock, Configuration Interaction (CI), and coupled-cluster singles and doubles (CCSD) attempt to solve for the wave function directly by expressing the wave function as a series of atom centered basis functions, which are typically Gaussian in nature. DFT instead relies on the Hohenberg-Kohn theorems, which state that the electron density ρ_e contains all the information necessary to obtain the ground state properties of a molecule. Thus, in DFT the energy given by 1.8 is written as

$$E[\rho] = T[\rho] + E_{eN}[\rho] + E_{ee}[\rho] \tag{1.9}$$

here $T[\rho]$ and $E_{ee}[\rho]$ are the multielectron interacting kinetic energy and coulombic interaction functional, meaning they map electron density – a function of position – to energy. For most systems, analytical expressions for these terms must be approximated and errors can become quite large. For this reason, the Kohn-Sham formalism is typically applied.

This formalism considers a non-interacting system of particles that is identical to the original. In this case we are able to model the non-interacting kinetic and coulombic terms exactly and the energy associated with particle interaction is lumped into a term called the exchange-correlation functional. Written out, the energy becomes:

$$E[\rho] = T_{ni}[\rho] + E_{ni,eN}[\rho] + E_{ni,ee}[\rho] + E_{xc}[\rho] \tag{1.10}$$

Here, "ni" refers to non-interacting and $E_{xc}[\rho]$ is the exchange-correlation energy functional. In principle, equation 1.10 is exact, but in practice the expression for $E_{xc}[\rho]$ is not known (the other terms are) and must be approximated. Countless approximation for this functional have been developed with ranging accuracy depending on the property. Often, they are organized on "Jacob's Ladder" as is shown in Figure 1-6 where the more accurate functionals tend to include more information on the density such as information about its derivative and second derivative. Some famous functionals include the Minnesota functionals, B3LYP, PBE, etc. Depending on the chosen functional, DFT calculations can scale anywhere from $\mathcal{O}(N^3)$ - $\mathcal{O}(N^4)$ where N is the number of basis set functions used. In contrast, the most accurate wave function techniques scale as $\mathcal{O}(N^7)$, restricting them to very small molecules ($<$10 atoms). It is often the obligation of the researcher to verify which functional is suitable for the problem at hand. For more information on DFT and computational chemistry, the reader can refer to the textbook by Cramer [8].

High throughput computational screening aims to accelerating the discovery of materials for a given purpose when compared to the manual way in which materials have typically been discovered by scientist. The typical workflow is found in Figure 1-7. Structures are first examined with the most rapid techniques, either machine learning models (1.6.1) or techniques like DFT. The number of candidates is then down selected based on these results, and the smaller set is examined with more accurate but time

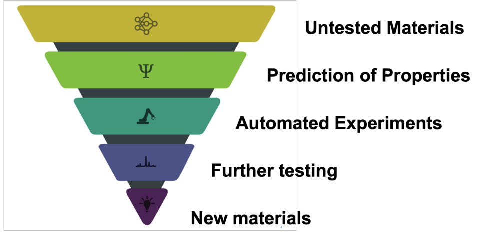


Figure 1-7: Virtual high throughput screening workflow.

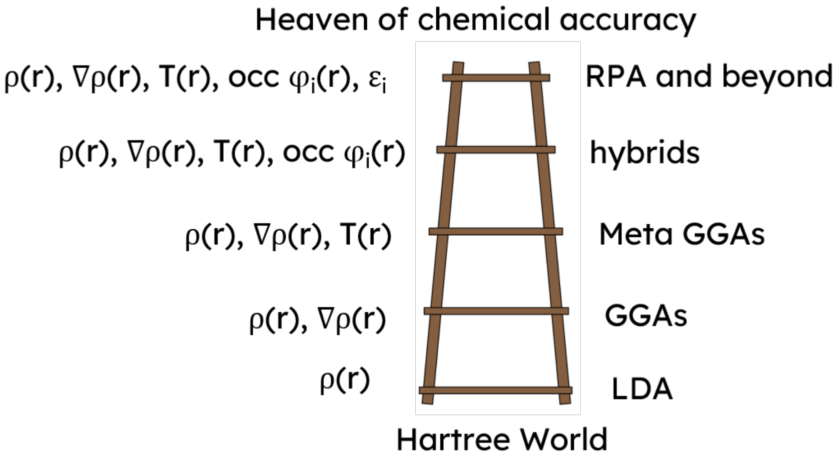

Heaven of chemical accuracy
ρ(r), ∇ρ(r), T(r), occ φi(r), εi
RPA and beyond
ρ(r), ∇ρ(r), T(r), occ φi(r)
hybrids
ρ(r), ∇ρ(r), T(r)
Meta GGAs
ρ(r), ∇ρ(r)
GGAs
ρ(r)
LDA
Hartree World

Figure 1-6: Jacobs ladder of DFT functionals. By including more and more information about the electron density into the expression for $E_{xc}[\rho]$, higher accuracy can typically be obtained.

intensive experimental techniques. The workflow is considered successful if it can produce real working candidates with performance that is better than what a human could produce. This approach has been generally applied extensively in the drug discovery field, as previously discussed, but only a handful of times in electrochemistry contexts. When it is used to discover and design electroactive material, it has typically been in the context of designing redox couples for redox flow batteries – an energy storage technology. [72, 58, 38, 5] Here, properties that of interest include the redox potential and solubility, among others. In electrochemical contexts, species can be charged or exist in their radical state, which can make the calculations slightly less accurate because consideration of the solvent becomes more important. When DFT is used, typically what is known as an implicit solvation technique is used to model the solvation environment. In this case, the structure of the solvent is not considered, and it is modeled as a continuous dielectric medium. For many properties, this approximation is able to recover sufficient accuracy, but if explicit bonding with solvent molecules is important for the property, often large errors are observed. Computational screening of potential RFB redox couple has shown success because the properties of each redox couple can essentially be examined in isolation. If one were to use the same techniques to examine HAT redox mediators, the property of each candidate would need to be taken into account with respect to the substrate of interests. Each time the substrate changes, the "optimal" mediator should also change, effectively resulting in a moving target. This concept is discussed in more detail in chapter 3.

1.6.1 Machine Learning for the acceleration of high throughput computational screening

While DFT is already a rapid technique, when molecular libraries become large ($>10^7$ molecules), the time and hardware required to screen the entire library can become limiting. In the last 5 years, machine learning (ML) has proved to be a valuable technique for accelerating the calculation of many material properties. In DFT, we have a physics based model that is able to map from chemical structure to the properties we desire (section 1.6), ML instead attempts to use large quantities of data to learn a data driven model that is able to rapidly map chemical structures to a given set of properties such as formation energies or solvation free energies [63, 64]. In the context of vHTP, ML models can be used to downselect a large library before DFT calculations are performed.

Data

ML techniques are usually broken up into three classes: supervised learning, unsupervised learning, and reinforcement learning. The most common technique used for predicting molecular properties is supervised learning, in this case the goal is to build a model by using a set of data that has labels associated with the output. Typically, a ML model has many parameters that need to be assigned a value and this is typically done through a process called gradient descent where the labeled data are used to iteratively update the parameters until some local (ideally also global) optimum is reached. This means the quality and quantity of data is immensely important. The amount of data dictates how complex the model architecture can be. With more parameters, more data is needed to avoid over fitting. For some of the most complex model architectures, typically dataset sizes on the order of 10^5 points are sufficient to justify their use. The quality of the data is also very crucial. Typically, experimental data is considered to be among the most reliable, however for many molecular properties this is not available is large quantities as it requires extensive manual labor to collect. For this reason, computational data sets are often used to train ML models. Large datasets can be generated with relative ease, the main issue in this case is that QM techniques often have larger errors associated with them that can sometimes be systematic. Having a better data set is usually more important than the functional form of the model – it is common that very similar results can be achieved with different models [57].

Model Architectures

A typical machine learning model used for supervised learning will map a vector to a scalar or vector output. There are many forms of models that can do this, in the simplest case linear regression can achieve this. The issue is that due to its simplicity, linear regression is often not able to map complex data sets and is thus limited in its use for molecular property prediction. In order to increase the complexity of the model architecture, the models are often made non-linear, with the most famous class of model being neural networks. A class of models that has shown a lot of success recently have been graph convolutional neural networks (GNNs). These models represent a given molecule as a graph, where the atoms are nodes and bonds are edges. Each atom in the molecule is assigned a feature vector of properties. This can be things like the atomic number, the number of attached hydrogens, the hybridization, etc. Likewise, each bond is also assigned a bond feature vector with information such as the degree of the bond or whether it is in a ring. Each of these atom vectors is then updated with information from its neighbors in a predetermined number of convolution steps, (usually 2-6). Finally, the atom vectors are aggregated into a single molecular feature vector and then passed into a normal feed forward neural network to map the vector to the target property (Figure 1-8).

While not physics-based, GNNs are inspired by the natural connectivity of molecules and their highly non-linear nature has allowed them to successfully map large datasets to properties like free energies, solvation energies, and even reaction properties like rate constants. As they have been described thus far, one draw back is that they only consider two-dimensional connectivity with no incorporation of atomic positions, but there has been development in including this information into newer types of GNNs.

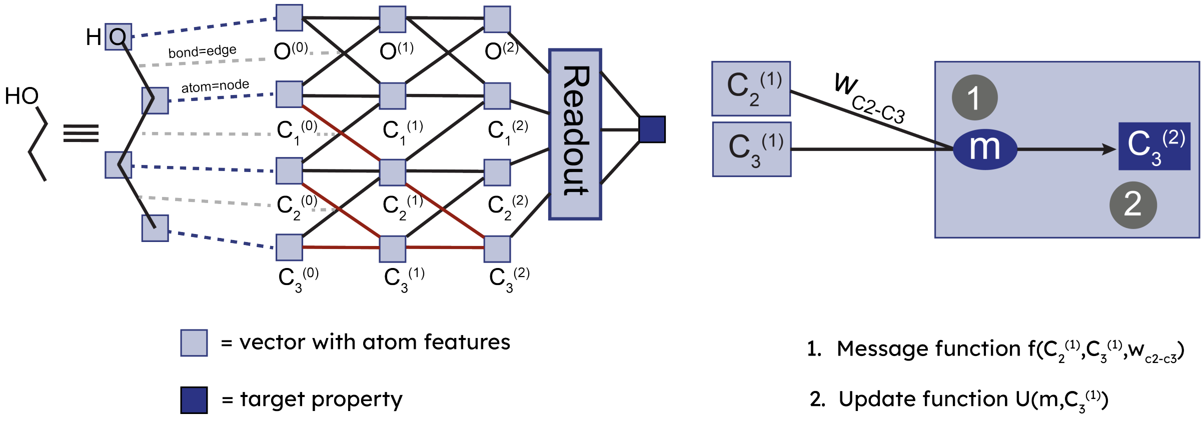

HO
H O
bond=edge
atom=node
O(0)
O(1)
O(2)
C1(0)
C1(1)
C1(2)
C2(0)
C2(1)
C2(2)
C3(0)
C3(1)
C3(2)
Readout
C2(1)
C3(1)
wC2-C3
1
m
C3(2)
2
= vector with atom features
= target property
1. Message function f(C2(1),C3(1),wc2-c3)
2. Update function U(m,C3(1))

Figure 1-8: General scheme of how GNNs work.

Training

The general workflow of validating a machine learning model starts with splitting the available data set into a test and training set. This choice of train and test split is of utmost importance, since it must allow for meaningful conclusions regarding the tasks the ML model will be applied to. Often, this is done randomly. For each train-test split, the model needs to be built using solely the training data. If a series of specific settings of the ML model (so-called hyperparameters) are available, their influence on the performance should be evaluated by so-called nested cross validation. The best-performing model is selected based on its average performance throughout nested cross validation, and with these settings, the ML algorithm is subsequently trained.

In order to train, a loss function must be defined. This gives a sense of how accurate the model predictions currently are and whether any parameter updates have resulted in improvement. For a regression problem, a very common loss function is the least squares error defined as

$$L = (Y_{true} - Y_{predicted})^2 \tag{1.11}$$

Where Y_{pred} is the output of our machine learning model $Y_{pred} = M(X;\theta)$. In order to update the model parameters, θ we need to use gradient descent, moving in the direction opposite of the gradient of the loss function with respect to the model parameters. The way we update our parameters then becomes

$$\theta_{i+i} = \theta_i - \gamma \frac{\partial L}{\partial \theta_i} \tag{1.12}$$

where γ is some step size parameter and i refers to the iteration step. After the parameters are fit based on the training data, the performance of the best parameters θ_{best} are examined with the test set. A good model will have similar accuracy on the test set as the training set. A common error is to build a model that fits the training set very well but performs more poorly on the test set, referred to as overfitting. In this case, the model will not generalize well and should be used with caution.

1.7 Brief Overview of Thesis Work

This work is grounded primarily in the goal of leveraging the potent yet fine-tunable nature of an electrochemical driving force to tackle key issues in augmenting the chemical recyclability of plastics, namely the synthesis of plastic monomers and the deconstruction of existing plastics. We demonstrate that electrochemistry – both experimental and computational – have a role to play in the synthesis of novel plastic monomers, in addition to allowing for new potential decomposition pathways for the plastics in use today.

This thesis can be broken down into three parts: (1) the experimental demonstration of sustainable synthesis of circular monomers using electrochemistry (2) the computational study of organic redox mediators with the potential for polystyrene deconstruction(3) and the implementation of data driven models to improve the throughput of computational screening.

To summarize these parts:

1. Electrochemical synthesis of lactone monomers: Lactones can serve as excellent monomers for circular plastics. Currently, they can typically be synthesized via the Bayer-Villiger reaction, which transforms cyclic alkanones into their corresponding lactones. This process suffers from drawbacks including the use of dangerous peroxyacids and in general should be made more sustainable to operate on the scale needed to produce monomers

that replace current plastics. We demonstrate that by using an applied electrochemical potential, platinum catalysts, and water as the oxygen atom source, we can synthesize lactones with selectivity toward branched products with a Faradaic efficiency of 25%, complementing Baeyer-Villiger products. By collecting electrochemical kinetic data, we propose a mechanism that can explain the observed selectivity of our system. Additionally, we implement a computational chemistry method to better understand the free energy landscape of our system and to investigate the polymerizability of our lactone products, a key metric for their use as circular plastics.

2. Computational Design of organic H- atom transfer redox mediators: Complementary to the synthesis of circular monomers, the deconstruction of current plastics is also a key problem to tackle in the decarbonization of the plastics sector. Leveraging the computational frameworks developed for in the study of the lactonization free energy landscape, we try to address current issues with the electrochemical deconstruction of polystyrene via H atom transfer pathways. Previous attempts at utilizing an organic redox mediator to deconstruct plastics suffered from a mismatch of the bond dissociation energy of the redox mediator and the C-H bond that was to be activated in the plastic. For this reason, we computationally screened tens of thousands of ylide redox mediators by using DFT to calculate all the energies relevant to their catalytic cycle. We were able to then identify and synthesize candidate structures with varying bond dissociation energies to investigate how electrochemical performance changed as the BDE of the redox mediator approached the BDE of the bond in the plastic.

3. Construction of data driven models to aid in the rapid prediction of electrochemical properties to increase the throughput of screening electrochemical materials: The virtual high throughput screening performed for project 2 still suffered from temporal constraints and should be further accelerated for future screening. We tackled this issue by constructing graph-based machine learning models using the DFT calculations from (2) that were able to predict key thermodynamic properties of redox mediators. Specifically, we constructed models that could simultaneously predict the redox potential, bond dissociation energy, and deprotonation energy of a ylide redox mediator

with reasonable accuracy for each property. A key consideration was that all DFT calculations were performed in acetonitrile using a continuum solvation model. Varying solvent is an important parameter in electrochemistry, but performing additional calculations for each structure in a new solvent quickly becomes intractable. For this reason, we developed an additional graph-based machine learning model that was able to map a solute and solvent to its solvation free energy. With this, we were able to take a gas phase DFT calculation and rapidly get its energy in different solvents with reasonable accuracy.

Chapter 2

Electrochemical Synthesis of Lactones using Water as the Oxygen Atom Source

The material in this chapter reproduced with little change with permission from Maalouf, J. H., Jin, K., Yang, D., Limaye, A. M., & Manthiram, K. (2020). Kinetic Analysis of Electrochemical Lactonization of Ketones Using Water as the Oxygen Atom Source. ACS Catalysis, 10(10), 5750–5756. https://doi.org/10.1021/acscatal.0c00931

Abstract

Lactones serve as key synthetic intermediates for the large-scale production of several important chemicals, such as polymers, pharmaceuticals, and scents. Current thermochemical methods for the formation of some lactones rely on molecular oxidants, which yield stoichiometric side products that result in poor atom economy and impose safety hazards when in contact with organic substrates. Electrochemical synthesis can alleviate these concerns by exploiting applied potential to enable a sustainable and safe route for lactonization. In this study, we investigated the mechanism of electrochemical lactone formation from cyclic ketones. When using a platinum anode and cathode in acetonitrile with 10 M H_2O and 400 mM cyclohexanone, we found that non-Baeyer-Villiger products, δ-hexanolactone and γ-caprolactone, are formed with a total Faradaic efficiency of 20%. Isotope labeling experiments support that water is the oxygen atom source for this reaction. In addition, electrochemical kinetic data suggest a 1st order dependence on water at low water concentrations ($<$2 M H_2O) and a 0th order dependence on the substrate, cyclohexanone. A Tafel slope of 139 mV/decade was measured at 400 mM cyclohexanone and 10 M H_2O, implying a first electron transfer as the rate determining step. Literature proposed mechanisms for similar transformations suggest an outer sphere pathway. However, based on the collected electrochemical kinetic data, we propose the possibility that Pt reacts with water in an initial

electron transfer that forms Pt-OH, which can subsequently react with the ketone substrate. A subsequent electron transfer forms a ring opened carboxylic acid cation that can reclose to form either of the observed five- or six-member ring lactone products.

2.1 Introduction

As mentioned previously, lactones are important functional motifs that are found in several organic molecules with myriad applications across the chemical industry. For example, polycaprolactone, which finds use in human tissue scaffolds, long-term drug delivery systems, microelectronics, and adhesives, can be made via ring-opening polymerization (ROP) of ϵ-caprolactone.[28] Importantly, γ-lactones have also been shown to undergo ROP, with the formation of γ-valerolactone and ϵ-caprolactone copolymers being previously demonstrated.[14] Additionally, substituted 5 and 6 membered lactone moieties are often present in pharmaceutical molecules that have fungicidal, antibiotic, and anti-cancer properties[52, 26].

Lactones can be synthesized from several starting materials such as peroxyacids (via intramolecular esterification), alkenes, and cyclic ketones. Cyclic ketones in particular are an attractive option, as they are available on commodity chemical scales and only require the net insertion of a single oxygen atom to produce a lactone product[36]. The Baeyer-Villiger reaction, which inserts an oxygen atom between the more substituted alpha carbon of a carbonyl, typically using a peroxy acid oxidant, is a common chemical method for converting ketones into lactones (Figure 2-1) [43]. (However, this has many drawbacks; peroxy acids are highly reactive and require employing chlorinated solvents, posing significant safety and environmental hazards. Additionally, the stoichiometric by-products of peroxy acids are lower value chemicals and require expensive downstream separation processes [29, 3]. Catalytic approaches to lactone production have been studied extensively, and enable the use of milder oxygen atom sources, such as hydrogen peroxide, in the presence of a homogeneous metal catalyst [11]. Despite these advances, hydrogen peroxide still poses safety concerns, and current catalytic approaches for substituted lactone production still rely on undesirable toxic homogeneous metal catalysts in large quantities [61, 34].

Using an applied potential as an alternative oxidative driving force can avoid the previously mentioned harsh chemical oxidants. In this vein, previous studies have demonstrated electrochemical lactone formation from diverse sets of precursors and oxygen atom sources. For example, the Baeyer-Villiger reaction has been carried out electrochemically using molecular O_2 in ionic liquids [27]. Another study has demonstrated a scheme for C-O bond formation to form aromatic lactones from their carboxylic acid counterparts [60]. Additionally, the formation of substituted lactones from hemioxalate salts and aliphatic carboxylic co-acids has been explored[39]. It should be noted that the latter two methods do not involve the transfer of an oxygen atom to the starting material. While these approaches demonstrate promise, they all suffer key drawbacks, such as the use of exotic solvents, the direct generation of CO_2, and the use of O_2, which can also be unsafe when mixed with organic solvents at scale. Moreover, electrochemical kinetic analyses that shed light on the lactone formation mechanism are generally lacking in the literature.

Issues with electrochemical lactone synthesis could be alleviated by using water as a sustainable and safe oxygen atom source. Thermochemically, the use of water as an oxygen atom source is difficult, as it requires unreasonably high temperatures and pressures. However, an alternative paradigm that involves the use of electrical potential as a driving force in chemical synthesis makes possible the use of water as an oxygen atom source without needing strong chemical oxidants. The successful implementation of a such a scheme would eliminate stoichiometric side products and hazardous reagents, overcoming the aforementioned issues with current routes for lactone synthesis. This concept has been demonstrated in the context of other electrochemical oxygen functionalization reactions, such as olefin epoxidation[24].

In this work we demonstrate that using water as an oxygen atom source, we were able to form lactones from cyclic ketones such as cyclohexanone using Pt foil electrodes, with Faradaic efficiencies (FEs) of approximately 20% toward lactone products. This scheme is shown in Figure 1. These lactone products differ from those produced by the Baeyer-Villiger reaction of cyclohexanone, diversifying the set of lactone products that can be synthesized from a given starting material. Isotope labelling confirms that the oxygen in the lactone products is inserted into cyclohexanone from the water in the electrolyte, with hydrogen gas co-produced at the cathode. Additionally, we performed a detailed electrochemical kinetic study on this reaction to elucidate a candidate mechanism. Similar chemical transformations have been reported in schemes that employ other oxygen atom sources, such as trifluoroacetic

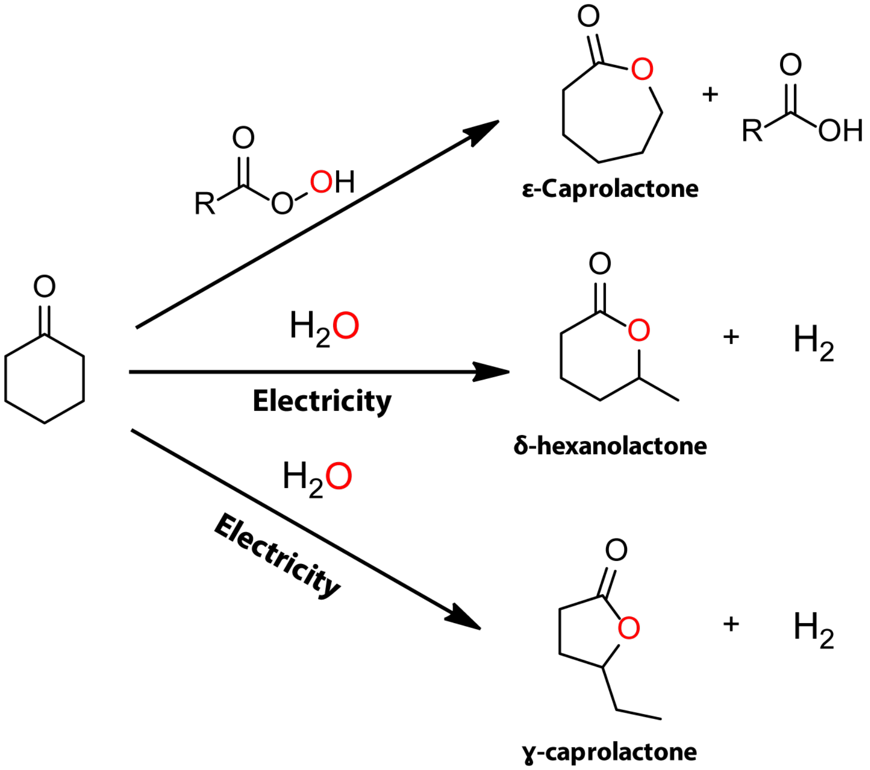


Figure 2-1: Scheme of a typical Baeyer-Villiger reaction using cyclohexanone and a peroxy acid as the oxygen atom source in the top reaction. The corresponding carboxylic acid is produced as a stoichiometric byproduct. The middle and bottom arrow show electrochemical lactone formation from cyclohexanone using water as the oxygen atom source with an applied potential (This work). Hydrogen gas is produced at the cathode.

acid and ethanol, but without any accompanying mechanistic data[13, 32]. Hence, we are able to demonstrate an electrochemical scheme for the proposed overall reaction, where cycloalkanone substrates are oxidized at ambient conditions into lactones using water as the sole source of oxygen in the system. This is accompanied by a mechanistic analysis of the chemical transformation that is supported by electrochemical kinetic data.

2.2 Data and Discussion

First, we sought to ascertain the overall reaction in the electrochemical cell. Experiments using 18OH2 supported that the oxygen atoms inserted into the products were derived from water, rather than from other unintentional oxygen atom sources in the system. During these labelled oxygen experiments, ${}^{16}O_2$ was also flowed into the system at 10 standard cubic centimeters per second (sccm), and while the concentration of $^{16}O2$ was expected to be much lower than that of water, we wanted to explicitly account for any potential dissolved $^{16}O_2$ from the atmosphere during normal conditions. Mass spectrometry analysis of products after electrolysis with 200 mM cyclohexanone and 1 M $18OH_2$ show products with shifts in M/Z of +2 and +4 (Figure 2-2a and 2-2b), relative to the value of M/Z=114 for the unlabeled products. In this case, a low cyclohexanone concentration was used to suppress the overall ratio of ^{16}O:^{18}O in the system, minimizing the amount of introduced unlabeled oxygen. The +4 M/Z shift may seem anomalous, since only shifts of +2 in M/Z would be expected via the insertion of a single ^{18}O atom. However, this peak can be explained by a chemical equilibrium between water and ketones that involves the exchange of O atoms through a ketal intermediate (Figure 2-2d) [30]. The small amount of product with M/Z value of 114 is still consistent with water acting as an oxygen-atom source; the ratio of ^{16}O:^{18}O in the system is approximately 1:5, so an M/Z=114 product may be formed from $^{16}OH_2$ produced in situ from a prior ketone-water oxygen exchange. Additionally, the $^{18}OH_2$ water used was 97% pure, providing another minor source of $^{16}OH_2$. Lactone products formed from ^{18}O insertion into an unexchanged cyclohexanone ring with an M=16 ketone oxygen can account for the M/Z=116 peak, as well as oxygen from $^{16}OH_2$ inserted into an ^{18}O exchanged cyclohexanone. Additionally, when no water is intentionally added to the system, the lactone formation partial current diminished by nearly an order of magnitude (compared to 1M H_2O), and lactone products are not quantifiable. When N_2 gas is flowed through the cell (to remove any O_2 in solution) with water in the system, the FE and partial current toward lactone formation remain unchanged from the case where no N_2 is flowed. From these observations, we propose that water is acting as the sole oxygen atom source.

Additionally, in a separate experiment, the effluent gas from the cell was fed into a GC and quantified. Hydrogen gas was found to be produced in quantities corresponding to approximately 100% FE at the cathode (Figure 2-2). For this reason, we believe that the desired anodic lactonization transformation paired with the cathodic production of hydrogen results in an overall reaction in the electrochemical cell that is depicted in Figure 2-1. Our reaction scheme could thus be utilized to generate valuable chemicals at both electrodes.

A comparison of CV curves with only water, only cyclohexanone, and both cyclohexanone

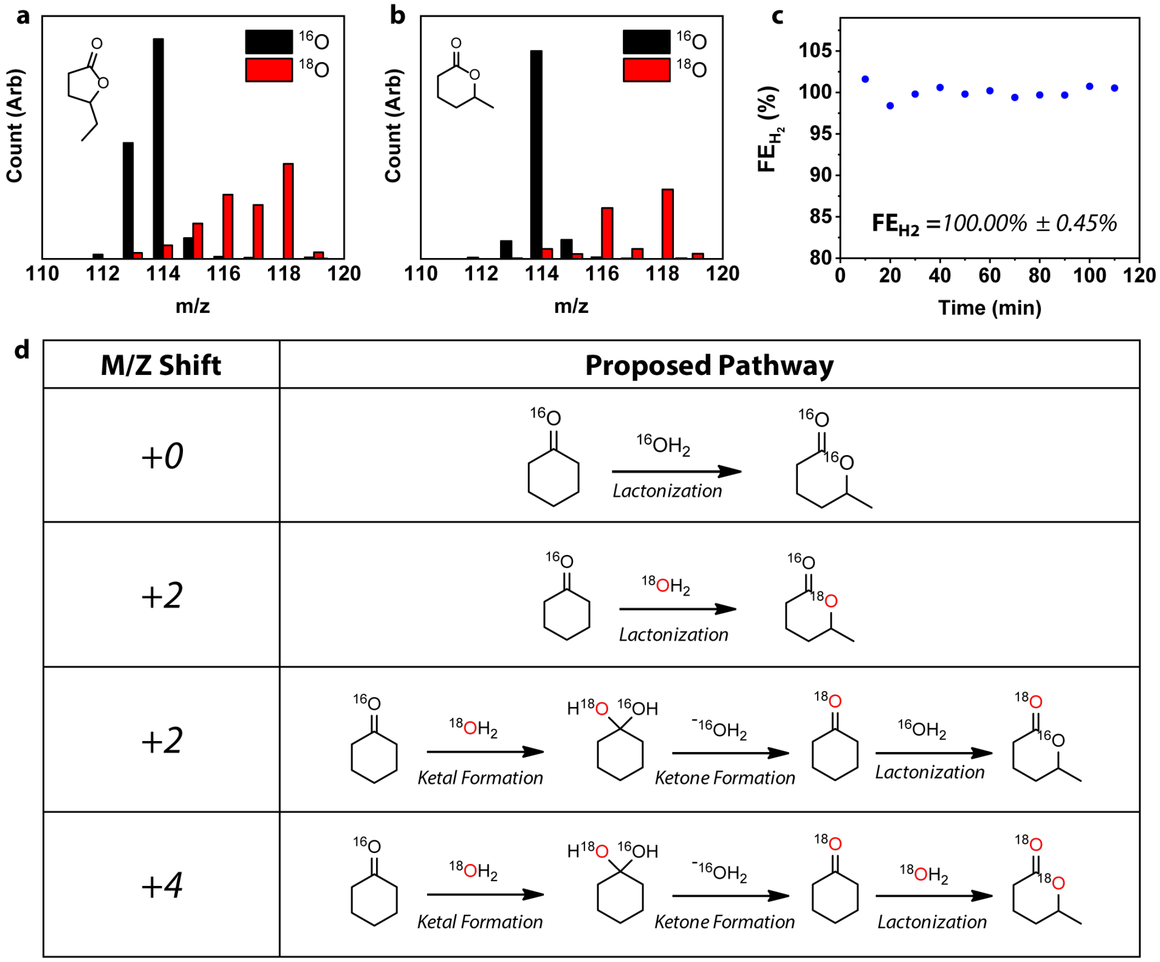


Figure 2-2: a,b) Mass spectra of γ-caprolactone and δ-hexanolactone using either 1 M $^{16}OH_2$ and $^{18}OH_2$. In the case of $^{18}OH_2$, $^{16}O_2$ was also flowed in at 10 sccm to test to see if O_2 from air was acting as an oxygen atom source. 200 mM cyclohexanone and 1 M $^{18}OH_2$ were used. (c) Quantification of H_2 via GC was performed, showing near 100% FE toward H_2 production. 400 mM cyclohexanone and 10 M H_2O were used for H_2 quantification. (d) The proposed pathways for each of the observed M/Z shifts from experiments using labeled water is shown. M/Z shift of +2 can be explained by two different pathways. For simplicity, only the formation of δ-hexanolactone is shown.

and water (400 mM, and 10M respectively) demonstrate that in the presence of water, the observed redox peak with only cyclohexanone present disappears, and the current continues to grow in the explored potential range. The observed redox peak when no water is present can potentially be ascribed to an irreversible oxidative electron transfer involving cyclohexanone; since no explicit O atom source is present in this case, potential products include cyclohexanone coupling derivatives, among others. However, the onset of anodic current reaches 5 mA/cm$_2$ at approximately 2.0 V vs Fc/Fc+ when water is present, suggesting that water is oxidized at lower potentials than cyclohexanone in the system. This supports the notion that in the presence of water, it is the water that will be preferentially oxidized first and this is what is responsible for the onset current. A CV comparison to a glassy carbon electrode indicated that the onset potential of Pt is approximately 0.5 V lower with 10 M H_2O and 400 mM cyclohexanone (Figure 2-12). To further investigate this reaction, we performed chronoamperometry (CA) at various potentials, with subsequent product analysis and quantification using GC-MS and GC-FID. The reported lactone partial current (i_{lac}) comprises the sum of the γ-caprolactone and δ-hexanolactone partial currents.

Electrochemical kinetic data aided in elucidating a possible lactone formation mechanism. While holding the water concentration constant at 10 M H_2O at a potential of 2.15 V vs Fc/Fc+, the concentration of cyclohexanone was varied from 200 mM to 1.5 M, with the change in partial current over this range indicating a 0th order dependence on cyclohexanone as shown in Figure 2-3c. Notably, the 0th order in cyclohexanone concentration is maintained at lower water concentrations as well (1M H_2O) as is shown in Figure 2-11. Varying the water concentration from 0.5 to 10 M H_2O while holding the cyclohexanone concentration constant at 400 mM resulted in changes in the partial current stemming from both changes in the FE (Figure 2-10c) and changes in the total current. At low water concentrations ($<$2 M H_2O) the reaction followed a 1.38 order with respect to water, which we approximate as a unity order dependence. Stabilizing effects of hydrogen bonding between surface adsorbates and water molecules near the Pt electrode surface provide a plausible explanation for the observed elevated water order dependence. Specifically, hydrogen bonding between adsorbed OH species (which we propose exist on the surface at oxidative potentials, and play a role as a reaction intermediate, see below) and water molecules near the Pt electrode has been previously characterized in aqueous systems[35]. Increasing the water concentration should increase the density of hydrogen bond acceptors near the surface, further stabilizing adsorbed OH species and leading to an elevated order dependence. This is further supported by an observed depression in the FE with decreasing water concentration (Figure 2-10c). On the other hand, the observed saturation of partial current at higher water concentrations ($>$2M H_2O) could possibly stem from either saturation of surface sites or from the breakdown of the

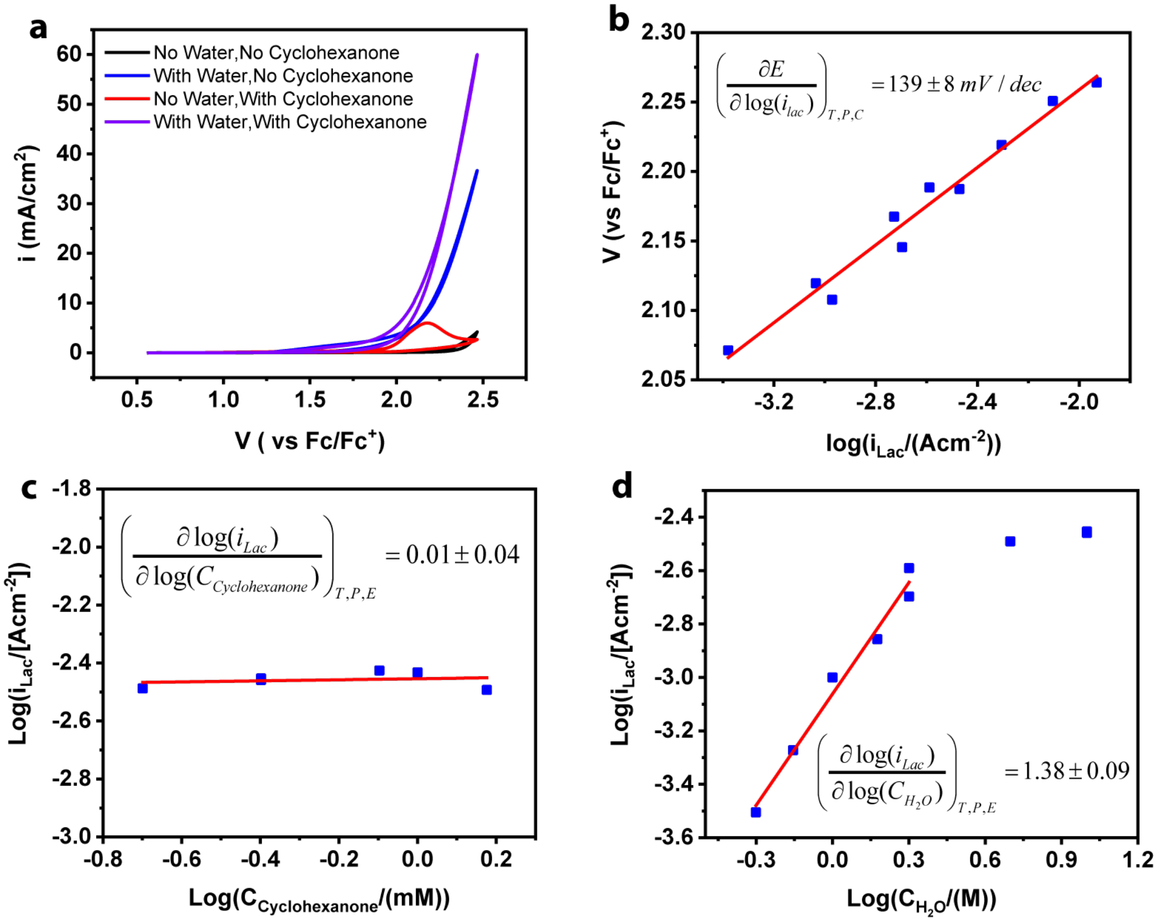


Figure 2-3: (a) Cyclic voltammograms collected with all combinations of the presence of cyclohexanone and water at a scan rate of 50 mV/s. When present, 400 mM cyclohexanone and 10 M H_2O are used. (b) The Tafel slope is shown with 100% manual IR correction, described in the SI. A value of 139 mV is obtained at 400 mM cyclohexanone, 10 M H_2O and 350 mM $TBABF_4$. (c) Cyclohexanone order dependence study, collected with 10 M water and 350 mM $TBABF_4$ present, at a potential of 2.15 V vs Fc/Fc+. (d) Water order dependence study, collected at 400 mM cyclohexanone and 350 mM TBABF4, at a potential of 2.15 V vs Fc/Fc+.

ideal solution approximation, which stipulates that the concentration of water is equivalent to its activity. Thermodynamic data for water and AcN suggests that above mole fractions of 0.1 (roughly 2 M H_2O), increasing the concentration of water only marginally increases its solution activity [53].

Additionally, the current and voltage dependence was collected in the form of a Tafel plot (Figure 2-3b), indicating a Tafel slope of 139 mV/decade. Data was collected both at constant current conditions and constant potential conditions with 100% manual IR compensation

(Supporting Information). While this is formally higher than theoretically predicted values with an assumed transfer coefficient of 0.5, it is in line with Tafel slopes observed for water oxidation on bulk Pt. In the water oxidation literature, Tafel slopes slightly greater than 120 mV/decade have been explained by the formation of a surface PtO2 layer [42]. For this reason, we believe that the Tafel slope suggests that the rate determining step is an initial electron transfer. Taken together, the electrochemical kinetic data suggests the following rate law:

$$i_{lac} \propto C_{H_2O} \exp\left(\frac{F\eta}{2RT}\right) \tag{2.1}$$

We hypothesize that the observed product selectivity is controlled by the stabilization of a carbocation intermediate on a secondary carbon that is formed via ring opening, as well as the driving force for the intermediate to close into predominantly 5- and 6- member ring lactones. To probe this hypothesis, we attempted lactonization of cyclopentanone and cyclohexanone derivatives such as 2 methyl-cyclohexanone, and characterized the observed product distributions. The results are summarized in Table 1. When reacting cyclopentanone in the same system, γ-valerolactone is the sole observed lactone product. This is consistent with our hypothesis; a 6-membered product is not observed because it requires the formation of a highly unfavorable primary carbocation, whereas the 5-membered product proceeds through a secondary carbocation. Reaction of 2-methyl cyclohexanone under the same conditions produced two products, 6-ethyloxan-2-one and 7-methyl-2-oxepanone both of which are in line with a secondary carbocation intermediate. Interestingly, the observed products are consistent with ring opening only on the alpha carbon of 2-methyl cyclohexanone that contains the methyl substituent. This is rationalizable, as the methyl substituted alpha carbon would directly form a secondary carbocation, while the unsubstituted alpha carbon would form a primary carbocation.

Based on the collected kinetic data and the observed product distribution of different substrates, we propose a mechanism that differs from mechanisms proposed in the literature for similar chemical transformations [13, 32]. Prior literature suggests outer sphere pathways that involving an initial electron transfer from the substrate (Figure 2-5a). No explicit

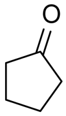
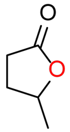
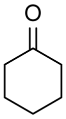
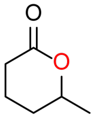
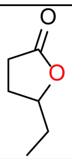
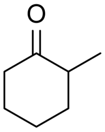
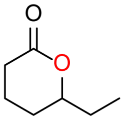
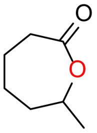
Starting Material	Observed Lactone Products
	
	 
	 

Figure 2-4: Products observed from reactions employing different cyclic lactone substrates. All runs are conducted at 400 mM Substrate, 10 M H_2O, 350 mM TBA BF_4, applying 2.15 V vs Fc/Fc+ and passing 50 C using Pt foil electrodes.

electrochemical kinetic data is reported for the transformation in the literature, including substrate dependence, oxygen source dependence, and Tafel slope; thus, when determining the feasibility of the literature mechanisms, we have compared it to our collected electrochemical kinetic data. The mechanism in Figure 2-5a would exhibit 1st order kinetics with respect to the ketone substrate since this is an outer-sphere mechanism, regardless of the RDS, which is inconsistent with our kinetic data. If we disregard this inconsistency regarding order in ketone substrate and assume a similar Tafel slope as we find in our work, that would point toward the initial electron transfer being the RDS; this, however, would be inconsistent with the observed 1st order dependence on water, which is not possible if water is only involved in the mechanism after the proposed RDS. These observations taken together suggest that the mechanism in Figure 2-5a is inconsistent with the electrochemical kinetic data measured in this study. A possible mechanism more in line with the data is proposed in Figure 2-5b, where water is initially activated by Pt in an electron transfer (ET) to form Pt−OH. After this ET, which we propose is the RDS, the adsorbed OH reacts with adsorbed cyclohexanone to form a tetrahedral intermediate that can undergo an additional ring opening electron transfer to form a carboxylic acid cation. Finally, the carboxylic acid cation can undergo ring closure to form the two observed lactone products. The proposed mechanism is consistent with the kinetic data, and the initial activation of water is consistent with observations from prior studies of water oxidation on Pt surfaces in neutral and alkaline media [12]. Additionally, the proposed adsorption of cyclohexanone to Pt through the carbonyl oxygen is similar to the proposed binding mechanism of cyclohexanone in homogeneous Pt catalysts for the Baeyer-Villiger reaction[11]. Density functional theory studies have also shown that it is favorable for ketones to adsorb to Pt surfaces through the carbonyl moiety in heterogeneous reactions as well[22, 62]. The hypothesized Pt-OH surface species can also be further oxidized into PtO_2, which, as previously stated, can rationalize the observed elevated Tafel slope. While the mechanism presented in Figure 2-5b is consistent with the observed data and is put forth as a plausible one, further in situ and ex situ spectroscopic studies are needed to refine the proposed mechanism.

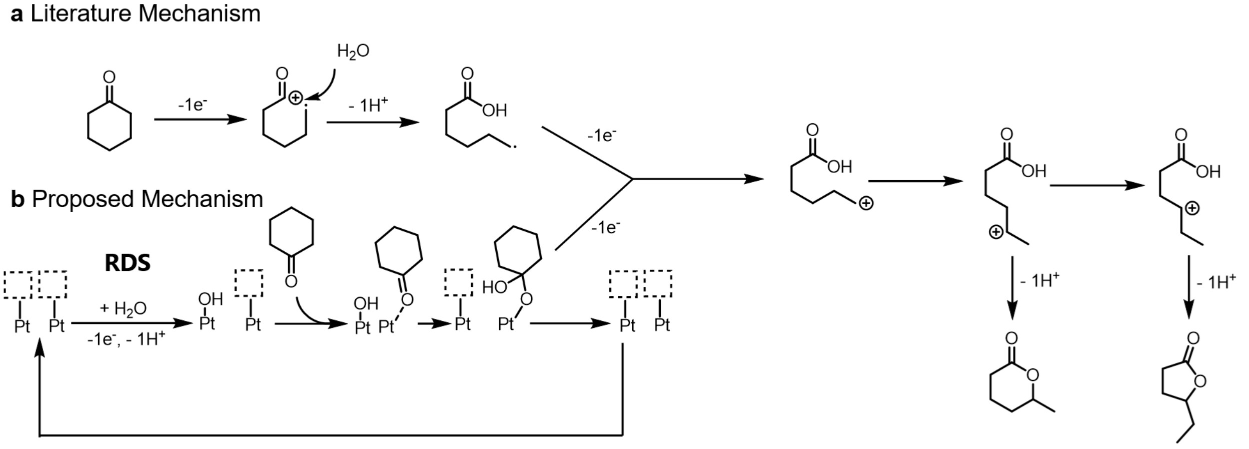


Figure 2-5: (a) Mechanism consistent with previous literature of similar transformations. The literature suggests an outersphere pathway that begins with an initial electron transfer from cyclohexanone. (b) Proposed mechanism based off of the collected electrochemicalkinetic data. Shows an inner sphere mechanism on Pt that undergoes an initial electron transfer involving water as the rate determining step, followed by reaction with cyclohexanone and another electron transfer.

In summary, we have investigated a method for using electricity derived from renewable sources to convert cyclohexanone into 5 and 6 membered substituted lactones with an FE of approximately 20% using Pt electrodes and water as the sole oxygen atom source. These products expand and complement the set of possible lactone products synthesized from cyclohexanone via the Baeyer-Villiger reaction. Additionally, H_2 gas was co-produced at the cathode with a near 100 % FE. Electrochemical kinetic studies allowed us to propose a mechanism that involves the initial inner sphere activation of water. We believe that our concept will provide a new pathway for electrochemically producing important intermediate ring sized lactones that are not accessible via the Baeyer-Villiger reaction from the same starting material.

2.3 Experimental Methods for Lactonization Chemistry

2.3.1 Materials

Platinum foil (Pt Foil) (99.9%, Beantown), cyclohexanone (>99%, Sigma Aldrich), cyclopentanone (>95%, Sigma Aldrich), γ-valerolactone (>99%, Sigma Aldrich), δ-hexanolactone (98%, Alfa Aesar), γ-caprolactone (98%, Sigma Aldrich), acetonitrile (ACN) (anhydrous, 99.8+%, Alfa Aesar), tetrabutylammonium tetrafluoroborate (TBABF4) (98%, TCI), dichloromethane (DCM) (99.5%, Fisher Scientific), nitric acid (70%, Fisher Scientific), alumina polish (0.05 micron, CH Instruments), $^{18}OH_2$ (97%, Sigma Aldrich), were purchased and used without additional treatment.

2.3.2 Methods

Platinum Foil Preparation

Platinum foils were cut into 2.5 cm x 2.5 cm squares. Prior to each experiment, the working electrode and counter electrode foils were rinsed with acetone and water. The foils to act as working electrodes were polished with 0.05 micron size alumina powder for at least 1 minute on each side. They were then rinsed with Milli-Q water and then sonicated in Milli-Q water for 5 minutes to remove remaining alumina powder. The working electrodes were then rinsed in 20% nitric acid and finally one last time with water, then dried with a paper towel. Foils that acted as counter electrodes were rinsed in 20% nitric acid and then with Milli-Q water and dried with a paper towel.

Electrochemical Methods

Electrochemical experiments were conducted with a one compartment electrochemical cell. Platinum foil was used both as the counter and working electrode (Figure 2.3.2). The platinum foil working electrode was prepared as described in section 2.3.2. A Ag/AgCl electrode (3.4 M KCl leak-free 2.0 mm diameter Innovative Instruments) was used as the reference electrode and aluminum foil was used as the current collector. To determine the catalytic activity of Pt foil toward cyclohexanone lactonization, cyclic voltammetry (CV) curves were recorded. Acetonitrile with 0.35M tetrabutylammonium tetrafluoroborate ($TBABF_4$) was used as the solvent, with varying concentrations of cyclohexanone and water. The total volume of electrolyte was set to be 4mL. All the potentials were 85% IR compensated, and the resistance value at open circuit potential (OCP) was measured by EIS techniques. For Tafel analysis, the points were 100% manually IR compensated. The reported potentials were calibrated by measuring a 1 mM ferrocene / ferrocenium redox couple at various water concentrations in acetonitrile (1M, 5M, and 10 M H_2O in 0.35 M TBABF4 with 400 mM cyclohexanone) and calculating the half potential of the redox couple. Examples of ferrocene redox couples at different water concentrations are shown in Figure 2-7. The electrochemical measurements were conducted with a VMP3 multichannel potentiostat from BioLogic.

Prior to each experiment, an ACN solution of 0.5M TBABF4 was treated with molecular sieves (4 x 8 Mesh Type 3Å, Acros Organics) for at least 12 hours to remove residual water in MeCN. For each measurement, 4 mL of electrolyte containing a specified composition of cyclohexanone, water, and TBABF4 in MeCN was added into the 1-compartment cell. A micro-magnetic stir bar was placed into the cell for stirring. For cyclic voltammetry analysis, CV scans (no stirring) were initiated from the open-circuit potential, and two cycles (From OCP to 2.6 V vs. Fc/Fc+) were recorded successively at a scan rate of 50 mV/sec (no stirring in this case). The chronoamperometry analysis was conducted at 2.15 V vs. vs. Fc/Fc+ to obtain electrochemicalkinetic data (with stirring). In all cases where partial current is reported (i_{Lac}), this was calculated by quantifying the Faradaic efficiency toward δ-hexanolactone and *gamma*-caprolactone and summing them up to a single FE toward lactone products. Specifics on the product analysis is given in section C.

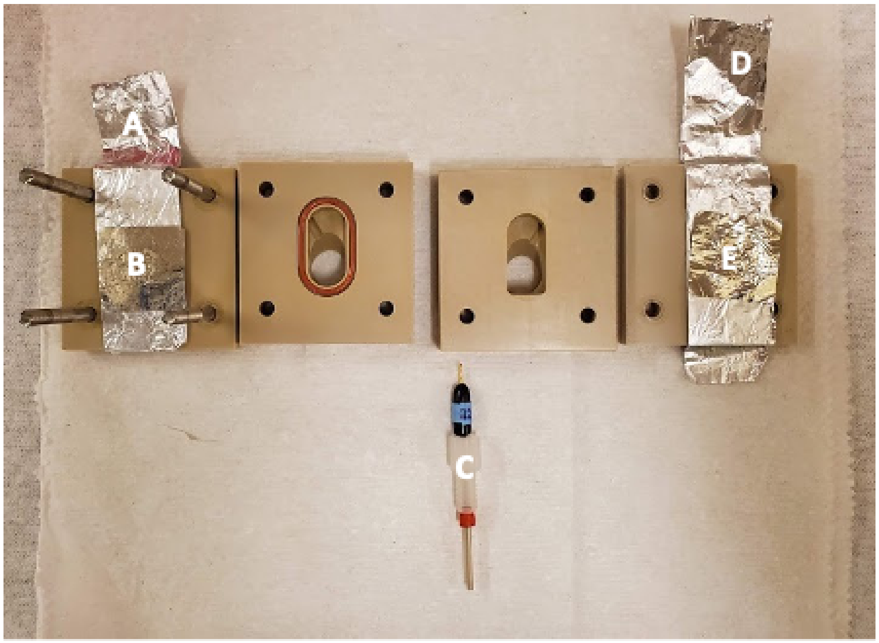


Figure 2-6: Electrochemical sandwich cell used for all experiments. (A) Aluminum foil current collector. (B) Platinum foil counter electrode. (C) Ag/AgCl reference electrode. (D) Aluminum foil current collector. (E) Polished platinum foil working electrode.

Electrochemical Kinetic Measurements

For Tafel analysis, points were collected at both constant potential conditions and constant current conditions. For constant potential conditions, a constant potential was applied between 2.1 V vs Fc/Fc+ and 2.3 V vs Fc/Fc+ with 85% manual IR compensation. The average recorded total current, after passing 50 C total, was used to compensate the remaining 15% of the resistance. For points collected at constant current, a total current density between 5 and 25 mA/cm$_2$ was applied with no IR compensation, but with the resistance recorded. The total current density was then used to 100% manually IR compensate after the experiment (subtracting IR from the recorded potential). Data from both conditions are overlaid in Figure 3b. For chronoamperometry experiments conducted to collect the water order dependence, 50C of charge was passed with 400 mM cyclohexanone and 0.35 M TBABF4. For chronoamperometry experiments conducted to collect the cyclohexanone order dependence, 50 Coulombs (C) of charge was passed, except in the case of 200 mM cyclohexanone where 40 C of charge was passed. In this case less charge was passed in order to reduce the theoretical maximum conversion of cyclohexanone which is necessary for making the approximation that the cyclohexanone concentration is effectively constant throughout the experiment. The water concentration was held constant at 10 M H_2O.

Isotope Labeling Experiments

Isotope labelling of the oxygen atom in water was used in order to determine if the inserted oxygen atom in the lactone products were derived from water. For these experiments, 1M 97% $^{18}OH_2$ was used instead of $16OH_2$. Additionally, 200 mM was used as the cyclohexanone concentration, and 10 sccm of O_2 (Industrial Grade) was also flown through the cell in order to attempt to rule out oxygen gas as the oxygen atom source. Gas chromatography mass spectroscopy (GCMS) was used to quantify the products, and determine any shifts in the mass to charge ratio of these products. 50 C of charge was passed for these experiments.

Product Analysis

All the products generated by electrochemical cyclohexanone oxidation were analyzed by gas chromatography–mass (GC-MS) spectrometry (5977B MSD and 7890B GC, Agilent) and gas chromatography-FID (7890B GC, Agilent). After each electrolysis, we took out 50 μL of electrolyte from the electrochemical cell and diluted it with 950 μl of dichloromethane (DCM). The 20 times diluted solution was analyzed by GCMS and GCFID. External standards of δ-hexanolactone and γ-caprolactone were obtained to quantify the product. All the presented quantification data was obtained based on the GCMS and GCFID analysis. Examples of external standard curves for γ-caprolactone and δ-hexanolactone are show in Figure 2-8 and Figure 2-9.

Gas Chromatography For H_2 Quantification

Gas chromatography was conducted to quantify the cathodic reaction product H_2. An in-line gas chromatograph (SRI Instruments, Inc., MG #5, Model 8610C) with auto-sampling capabilities was used. N_2 gas available in-house was used as the GC carrier gas, and was adjusted to roughly 15 psig on the instrument. The data sample rate was 5 Hz. Gas flow into the GC was normally vented out through a 1-mL sample loop. A sample was collected and was sent onto a 6' HayesepD pre-column. At the same time, another sample was collected and held without flowing. The gas-phase ppm determined from GC was converted into the partial current for a given product, and Faradaic efficiencies were calculated by dividing the partial current by the total current.

2.3.3 Supplementary Figures

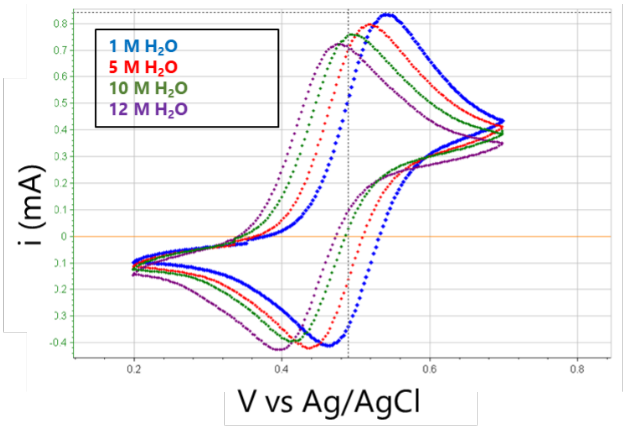


Figure 2-7: Redox Peak of 1 mM ferrocene in ACN with 350 mM TBABF4, 400 mM cyclohexanone with varying amounts of water (1, 5, 10, 12 M).

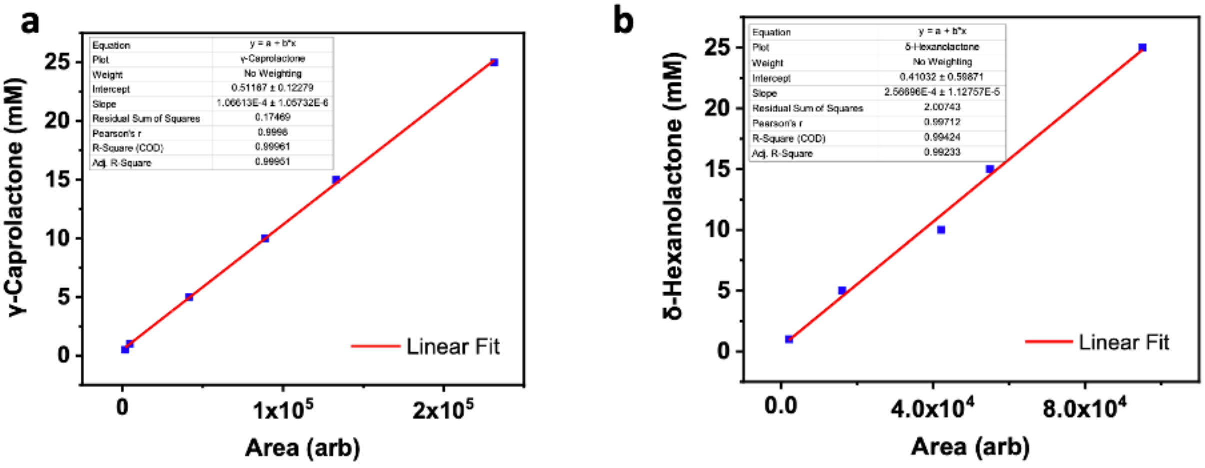


Figure 2-8: Calibration curves for γ-caprolactone (a) and δ-hexanolactone (b) using GCMS. The fitted equation for (a) was calculated to be y=1.066E-4x + 0.511 and for (b) was calculated to be y=2.566E-4x + 0.410.

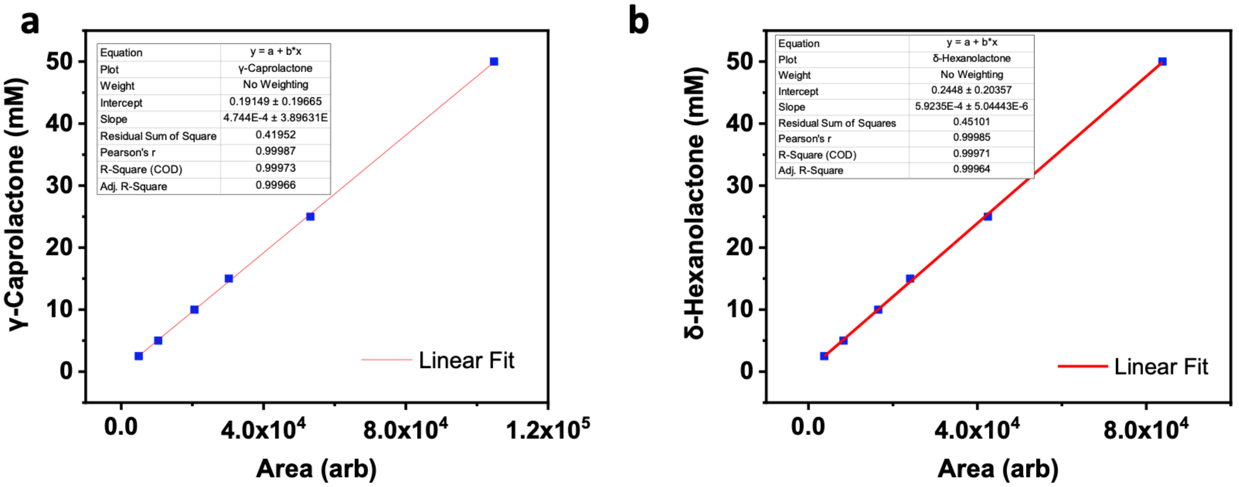
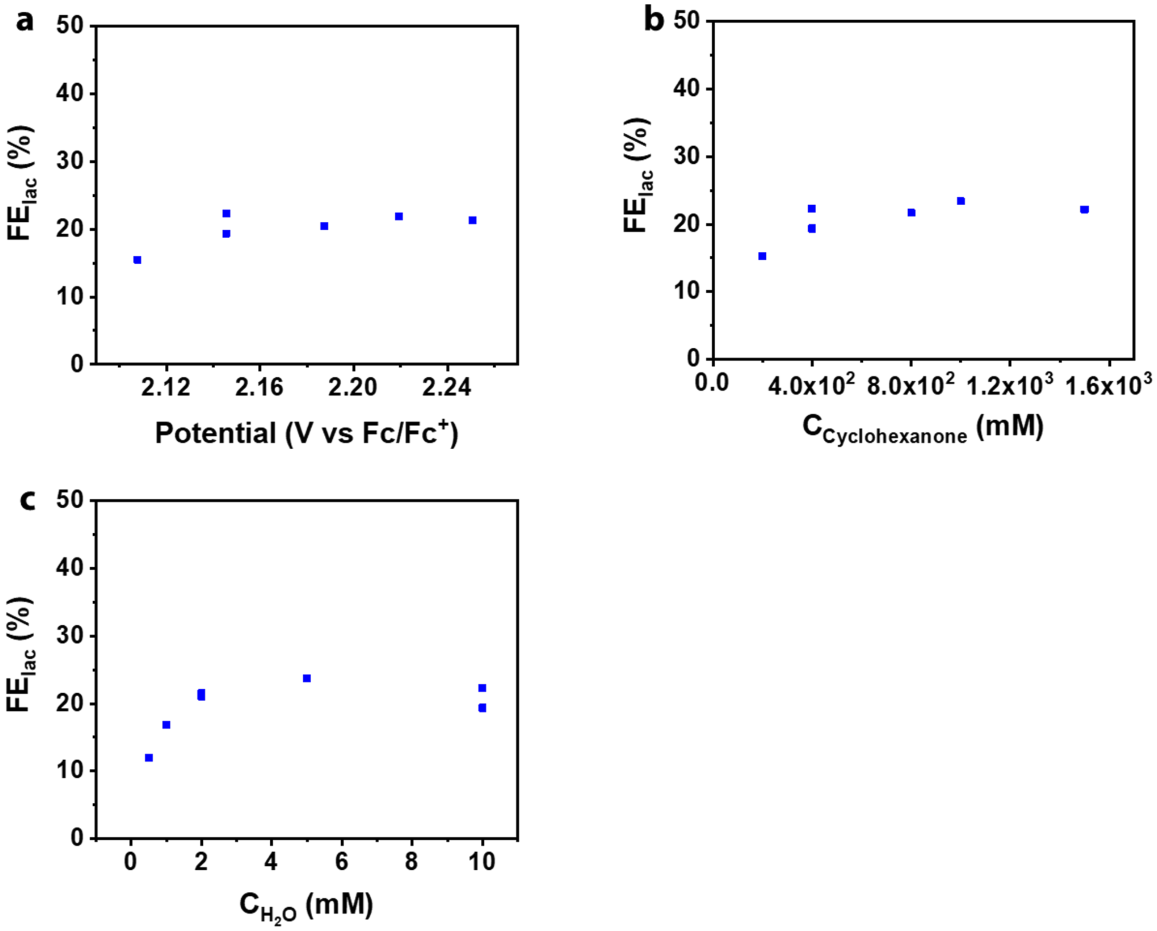


Figure 2-9: Calibration curves for γ-caprolactone (a) and δ-hexanolactone (b) using GCFID. The fitted equation for (a) was calculated to be y=4.744E-4x + 0.191 and for (b) was calculated to be y=5.923E-4x + 0.244.



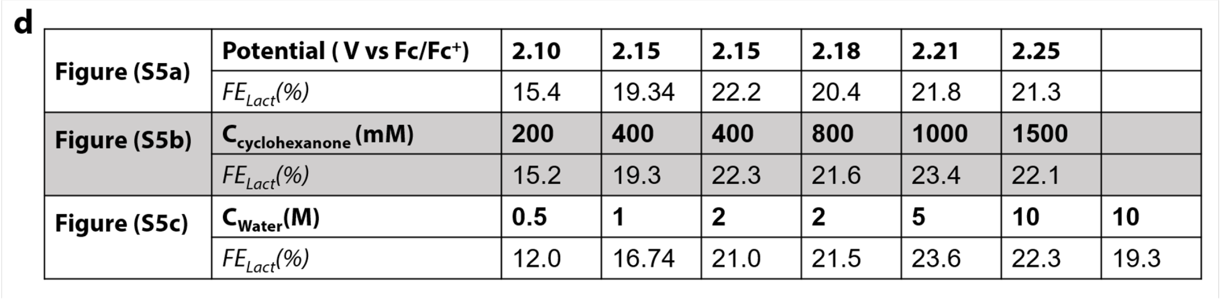

d

Figure (S5a)	**Potential (V vs Fc/Fc$^+$)**	**2.10**	**2.15**	**2.15**	**2.18**	**2.21**	**2.25**	
	FE_{Lact}(%)	15.4	19.34	22.2	20.4	21.8	21.3	
Figure (S5b)	**$C_{cyclohexanone}$ (mM)**	**200**	**400**	**400**	**800**	**1000**	**1500**	
	FE_{Lact}(%)	15.2	19.3	22.3	21.6	23.4	22.1	
Figure (S5c)	**C_{Water}(M)**	**0.5**	**1**	**2**	**2**	**5**	**10**	**10**
	FE_{Lact}(%)	12.0	16.74	21.0	21.5	23.6	22.3	19.3

Figure 2-10: (a) Faradaic efficiencies measured with GCFID as functions of the applied potential. (b) as functions of cyclohexanone concentration, and (c) as functions of water concentration. Standard conditions are 400 mM cyclohexanone, 10 M H2O, and 2.15 V vs Fc/Fc+. (d) Table with the plotted data from a-c. The reported potential has been IR corrected.

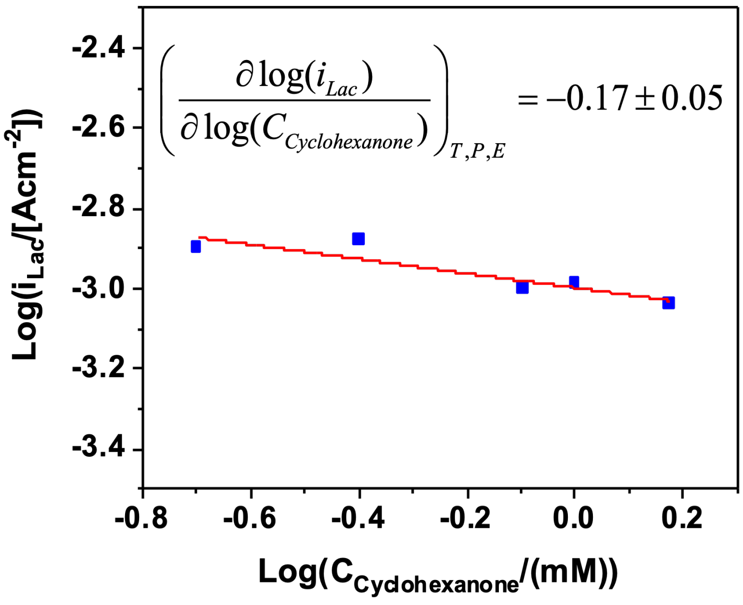


Figure 2-11: Cyclohexanone order dependence collected at 2.15 V Vs Fc/Fc+ and 1 M H2O.

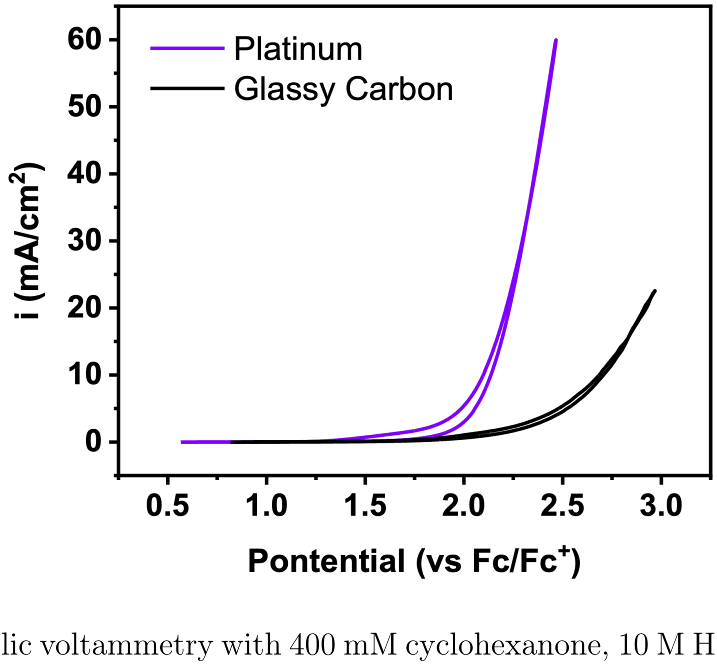


Figure 2-12: Cyclic voltammetry with 400 mM cyclohexanone, 10 M H2O, 350 mM TBABF4, comparing a platinum foil working electrode with a glassy carbon electrode.

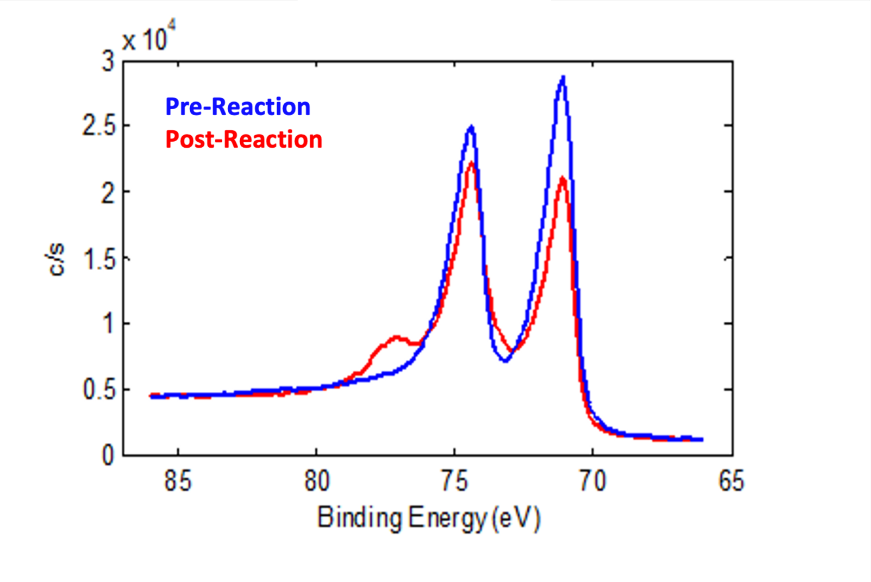


Figure 2-13: XPS spectra for Pt 4f conducted on a Pt foil before and after reaction. The pre-reaction spectrum is shown in blue, and is consistent with reported spectra of Pt metal. The spectra after passing 15 C of charge with 10 M H2O, 400 mM Cyclohexanone, 350 mM TBABF4, and 2.15 V vs Fc/Fc+ is shown in red. The development of the shoulder at approximately 76 eV, and has been previously assigned to oxide forms of Pt.

Chapter 3

Computational Screening of Organic Hydrogen Atom Transfer Redox Mediators

Abstract

Hydrogen atom transfer reactions are ubiquitous and important class of reactions. They can be facilitated by organic p-block radicals, as is seen in mediated electrochemistry, where for example they have been used to deconstruct plastics such as polystyrene. However, when organic HAT redox mediators are used, the energetics of the mediator are typically not considered in the context of the reactive C-H bond. Herein, using N-ammonium ylides as a class of redox mediators, we demonstrate that superior HAT redox mediators can be computationally designed to match the bond dissociation energy of the C-H bond of interest. We generate a library of 20,000 candidate N-ammonium ylide redox mediators and calculate the relevant thermodynamic values present in the ylide HAT catalytic cycle, namely the redox potential (E^o), the hydrogen binding energy ($\Delta G_{H_{Bind}}$) and the deprotonation energy (ΔG_{DPFE}). We use DFT to calculate these properties for the first 10,000 catalytic cycles (30,000 molecules) of our molecular library and use these calculations to build machine learning models that are able to calculate the same thermodynamic values for the remaining half of the library – greatly reducing the required computational cost. We demonstrate, for PS model compounds, we can computationally screen N-ammonium ylide thermodynamics to generate candidate structures that promote the activation of the lowest energy C-H bond in the system. We then validate these predicted ylide structures through electrochemical experiments.

3.1 Introduction

Approximately 400 million metric tonnes (Mt) of plastic are produced annually worldwide, with production set to exceed 500 Mt by 2040 [16]. This will result in unprecedented amounts CO_2 emissions and solid waste in the environment. For this reason, there is a necessary push to increase plastic circularity, either through designing monomers that can be easily chemically recycled or through developing improved methods for chemically upcycling the common plastics that are already in use today. If not recycled into the constituent monomer, plastics can often be deconstructed into small and useful molecules for other applications [37, 23]. However, plastic deconstruction is typically difficult because of the highly inert C-C bonds present in common plastics (PE, PP, PS). One such plastic is polystyrene whose use is found in styrofoam, plastic utensils and more and whose deconstruction is typically done through thermally based processes that involve high temperature, pressure, or expensive catalysts.

Various pathways have been demonstrated to convert polystyrene back into the styrene monomer or to recover processable intermediates. [4, 31, 69]. One pathway of interest that can create small oxygenated molecules from polystyrene is the hydrogen atom transfer pathway (HAT). In this case, a hydrogen atom is abstracted from the PS carbon backbone resulting in a radical that can later go on to react with oxygen gas in the system, (O_2) which results in C-C bond scission via subsequent bond rearrangement [69]. HAT reactions are not only of great interest for plastic deconstruction, but also for a variety of applications that require the activation of inert C-H bonds – this includes applications such as pharmaceutical and metabolite synthesis [9]. As a result, effective routes for facilitating HATs have been developed using materials such as organic p-block radicals and transition metal centered complexes. These studies have revealed that the relative energy of the substrate BDE and the oxidant is important and correlates with performance metrics such as the reaction rate [?]. Typically, larger reaction rates are observed when there is a larger energy difference between the oxidant and substrate, as is intuited by the Evans–Polanyi relation. However, these large energy differences can potentially result in lower selectives because a higher energy oxidant also has a non-negligible driving force to react with other species in the system.

Electrochemical mediators are a class of oxidants that have showed a lot of promise in facilitating HATs for various substrates [31]. Mediators are molecules that can be activated by and external stimulus, such as an external potential, to form an activated species that can go on to react with some substrate in the system and abstract a hydrogen atom[66]. Electrochemistry takes advantage of an increasingly carbon free and powerful driving force that is readily accessible – an applied voltage. The potency of an applied voltage can replace demanding temperatures and pressures in many systems. Importantly, this driving force is not indiscriminate, as it can be finely tuned to impart the appropriate thermodynamic and kinetic driving force to the system at hand. HAT mediators can come in many forms and range from small organic molecules to metal centered complexes. Common electrochemical organic redox mediators include (2,2,6,6-Tetramethylpiperidin-1-yl)oxyl (TEMPO), N-Hydroxyphthalimide (NHPI) and their derivatives. Yan et al. even used NHPI as a PS degradation redox mediator – deconstructing PS into a variety of small oxygenated molecules[69]. In this case, the authors hypothesized that some of the issues in their system such as selectivity and mediator deactivation can be partly due to a mismatch of bond dissociation energies (BDEs) of the C-H bonds on the substrate and the NO-H bond of NHPI. As previously mentioned, the energy difference between the oxidant and substrate has already been shown to be crucial for reaction rates so for a given substrate, the ability to modulate the energy of the mediator would prove immensely beneficial.

In the case of NHPI or TEMPOH, the NO-H BDE could in principle be tuned through molecular functionalization, but only a handful of derivatives has been shown to be synthetically accessible in either case [40, 20]. Recently, N-ammonium ylides have been shown to serve as electrochemical HAT redox mediators that exhibit a high degree of synthetic tunability, with more than 80 ylide derivatives synthesized in the original work[46]. With this increased level of sythesizability, we are able to apply tools such as virtual high throughput screening (vHTP) to redox mediator design. vHTP screening has been shown to be useful in designing electroactive materials in other contexts such as redox flow battery (RFB) redox couples, but has not been applied to redox mediator design – this likely has to do with the fact that mediators must be chemistry specific and their properties cannot always be analyzed in isolation [58]. In the case of HAT mediators, the energetics of the mediator must be considered in the context of C-H bond to be activated. If a mediator with too much energy is used, it is more likely to react undesirably with either other molecules in the system or at undesirable sites on the molecule of interest.

Using N-ammonium ylides as a class of redox mediators, we demonstrate that HAT redox mediators can be computationally designed to match the bond dissociation energy of the C-H bond of interest. We generate a library of 20,000 candidate N-ammonium ylide redox mediators and calculate the relevant thermodynamic values present in the ylide HAT catalytic cycle (Fig 3-1), namely the redox potential (E^o), the hydrogen binding energy ($\Delta G_{H_{Bind}}$) and the deprotonation energy (ΔG_{DPFE}). We use density functional theory (DFT) to calculate these properties for half of our molecular library (10,000 catalytic cycles or 30,000 molecules) and use these DFT calculations to build machine learning models that are able to calculate these relevant thermodynamic values for the remaining half of the library – greatly reducing the required computational cost. We demonstrate that for PS model compounds, we can computationally screen N-ammonium ylide thermodynamics to generate candidate structures that promote the activation of the lowest energy C-H bond in the system. This means we are choosing ylide candidates that have only the required energy to activate the bond of interest, resulting in lower energy intermediates and thus reducing the likelihood of radical intermediates undergoing any unwanted side reactions. Our computational screened ylide candidate structures are then validated through electrochemical experiments.

3.2 Results and Discussion

3.2.1 Virtual Library Generation and Calculating Properties

The molecules in the virtual library were generated by first performing a substructure search in Reaxys for a carboxylic acid substructure. The details of this procedure can be found in the supporting information. N-ammonium ylides have been shown to be synthesizable from carboxylic acids, and since carboxylic acids are both diverse and easily purchasable, they are good starting motifs for the purpose of building a virtual library. The results of the Reaxys

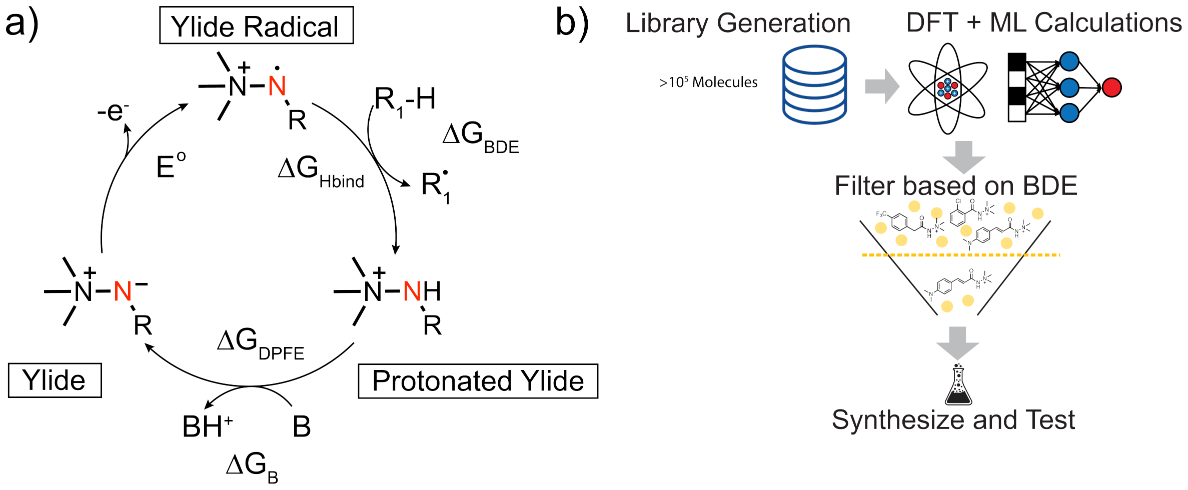


Figure 3-1: a) Catalytic cycle of ylide redox mediators for HAT reactions. b) Virtual high throughput computational workflow developed for redox mediator design in this work.

search were filtered by the criteria indicated in Table 3.1, which resulted in approximately 22,000 carboxylic acids. The screening criteria were elected to maximize the probability that each library entry was practically synthesizable. Using RDkit, the Simplified Molecular Input Line Entry System (SMILES) string for the 22,000 carboxylic acids was used to append the structure to three different cores to form the three species present in the catalytic cycles show in Figure 3-1 (Ylide = Y, Ylide Radical = Yrad, and protonated ylide = Yh).After verifying all the structures were valid and removing any duplicates, we were left with 19,097 ylides.

Physics-based modeling utilizing DFT was used to calculate key thermodynamic properties of the first 10,000 ylide catalytic cycles. Specifically, we calculated the redox potential E^o, the hydrogen binding energy ΔG_{Hbind}, and the ylide deprotonation energy ΔG_{DPFE} using the equations found in Table 3.3 at the M062X/def2-SVPD level of theory with SMD implicit solvation in acetonitrile. Because of the anionic character of the N in the ylide motif, we made sure to utilize a basis set that included diffuse functions. The results of the calculations can be found in Figures S5 and S6, where we can see that for E^o and ΔG_{DPFE} the distributions exhibit good gaussian behavior. However, for $\Delta G_{H_{bind}}$ there is a larger cluster of values around -450 kJ/mol, which could point to limitations in the ability of functionalization to modulate the energy of the hydrogen abstracting step. Of the original

10,000 catalytic cycles, we obtained 8095 cycles where Y, Yh and Yrad were all calculated successfully. The largest failure mode, was the inability to correctly optimize the geometry to a minimum, which was checked by assuring all the vibrational frequencies were positive.

We wanted to take the values of these ylide energies in the context of a specific substrate C-H bonds, and in order to do this, we considered the PS analog molecules studied by Yan et al. [69]. The structures of these molecules can be found in Figure 3-8, and include biphenyl compounds with a variable number of carbons between the two aromatic rings and some that contain oxygen functionality on the aliphatic chain. HATs for these molecules were found to occur on the aliphatic carbon backbone, so the BDEs for all possible HATs on the carbon chain were calculated with the same work flow used to calculate the ylide energetics. We observe that the energy of the benzylic position does not vary much across the molecules, occurring at around 335-340 kJ/mol. C-H bonds further away from the aromatic rings tended to be higher in energy. For example, in the case of 1,3 diphenylpropane, the two possible HAT positions differ in energy by 32 kJ/mol which suggests that there is a workable energy window that could be used to design a mediator that has selectivity toward the weaker bond.

For the purposes of exploring this concept, we chose to simplify our further analysis by focusing on a single substrate - 1,3 diphenylpropane. To design an ylide mediator to facilitate HAT reactions for this substrate, it is useful to consider the free energy landscape indicated in Figure 3-2b. The landscape shows the ylide HAT catalytic cycle up until the formation of the substrate radical $R^{\bullet}$, which goes on to react with O_2 in the system to form oxygenated products. We show the paths for both possible HATs, which diverge after step 1, and identify two important free energy differences that are crucial to screening ylides with the energy of the C-H bond in mind – ΔG_{CH} and ΔG_{DP} . These two energies are defined as follows:

$$\begin{aligned} \Delta G_{CH} &= \Delta G_{H_{bind}} + \Delta G_{BDE} \\ &= \Delta G_{Yh} + \Delta G_{\mathrm{R}^{\bullet}} - \Delta G_{\mathrm{Y}^{\bullet}} - \Delta G_{RH} \end{aligned} \tag{3.1}$$

$$\begin{aligned} \Delta G_{DP} &= \Delta G_{DPFE} + \Delta G_{B} \\ &= \Delta G_{Y} + \Delta G_{BH} - \Delta G_{Yh} - \Delta G_{B} \end{aligned} \tag{3.2}$$

and are simply the reaction energies for steps 2 and 3 on Figure 3-2b. We note that

ΔG_{CH} and ΔG_{DP} do not require knowledge of the energy of protons (H^+) or hydrogen radicals ($H^\bullet$), which can have relatively large errors associated with their calculation [54]. Using the previously mentioned DFT results, we calculated these two values for the first 10,000 structures in our ylide library, which can be found in Figure 3-2c. The figure contains 2 vertical lines, labeled BDE 1 and BDE 2, which correspond to the energies of the two BDEs of 1,3 diphenylpropane and are referenced to BDE 1, which is why it appears at $\Delta G_{CH} = 0$. We note that for an ylide to be able to activate BDE 1 or BDE 2, the value of ΔG_{DP} must be negative and the value of ΔG_{CH} must be negative of the vertical line corresponding to the BDE of interest. We indicate 3 important regions on Figure 3-2c. First, region 1 in the area left of the orange box, contains structures that have enough energy to activate both BDE 1 and BDE 2. Next, there is the orange region (region 2), with structures that have enough energy to selectively activate BDE 1. Lastly, structures in the rightmost box (region 3) have insufficient energy to activate either BDE 1 or BDE 2.

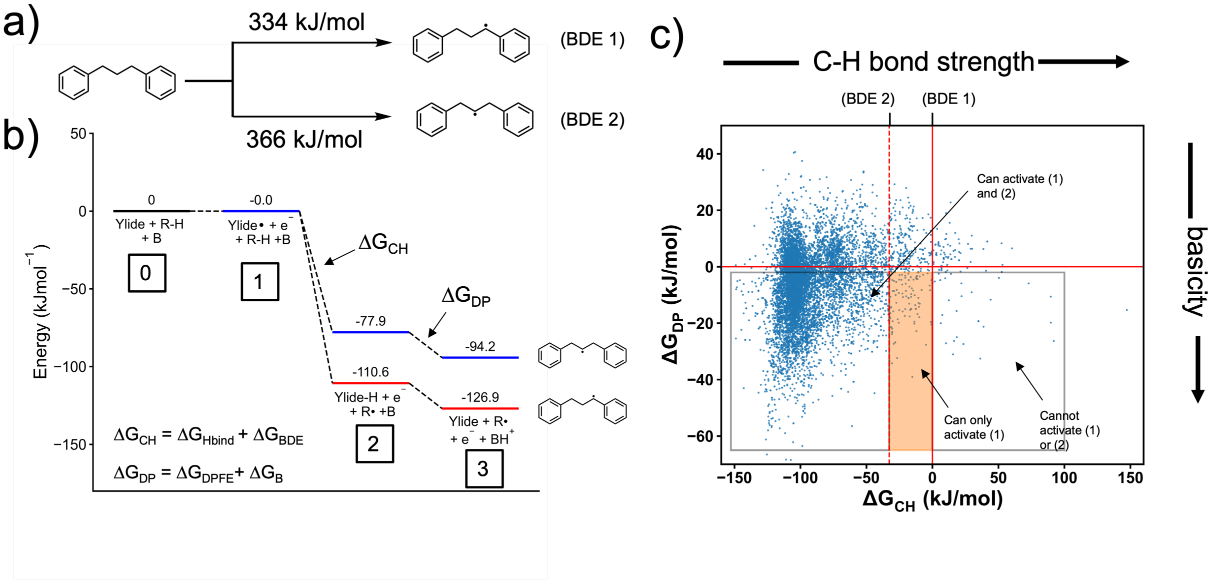


Figure 3-2: a)The two HAT reactions for 1,3-diphenylpropane, identified as (BDE 1) and (BDE 2) with their associated BDE energies. b) Free energy landscape of an ylide redox mediators showing the electron transfer, hydrogen abstraction and deprotonation. R-H = substrate, B = Base. c) Scatter plot of ΔG_{CH} and ΔG_{DP} for all DFT analyzed ylides.

Each point on Figure 3-2c corresponds to a unique ylide structure, where our analysis shows that most ylide structures are in great excess of the required energy to facilitate BDE 1 and BDE 2. The highest density of points can be found in the region $-120\text{kJ/mol} \leq \Delta G_{CH} \leq -80\text{kJ/mol}$, indicating that an ylide chosen at random, without prior consideration of the energy of the BDE, is likely to have a sufficiently large thermodynamic driving force. To see if there is any correlation between the presence of certain functional groups and the value of ΔG_{CH} , we isolated points containing halogens, amines, alcohols, ethers, phosphorous, sulfur, and only carbon containing species, the results of which can be seen in Figure 3-10. Some key insights into our library are that halogen containing ylides tend to lie in the $-120\text{kJ/mol} \leq \Delta G_{CH} \leq -80\text{kJ/mol}$ region, with no clear difference among the different halogens. Additionally, structures containing, amines, alcohols, and ethers had a larger spread in ΔG_{CH} energy, suggesting that these functional groups have more potential to modulate the energy of Yh and Yrad to result in lower ΔG_{CH} values.

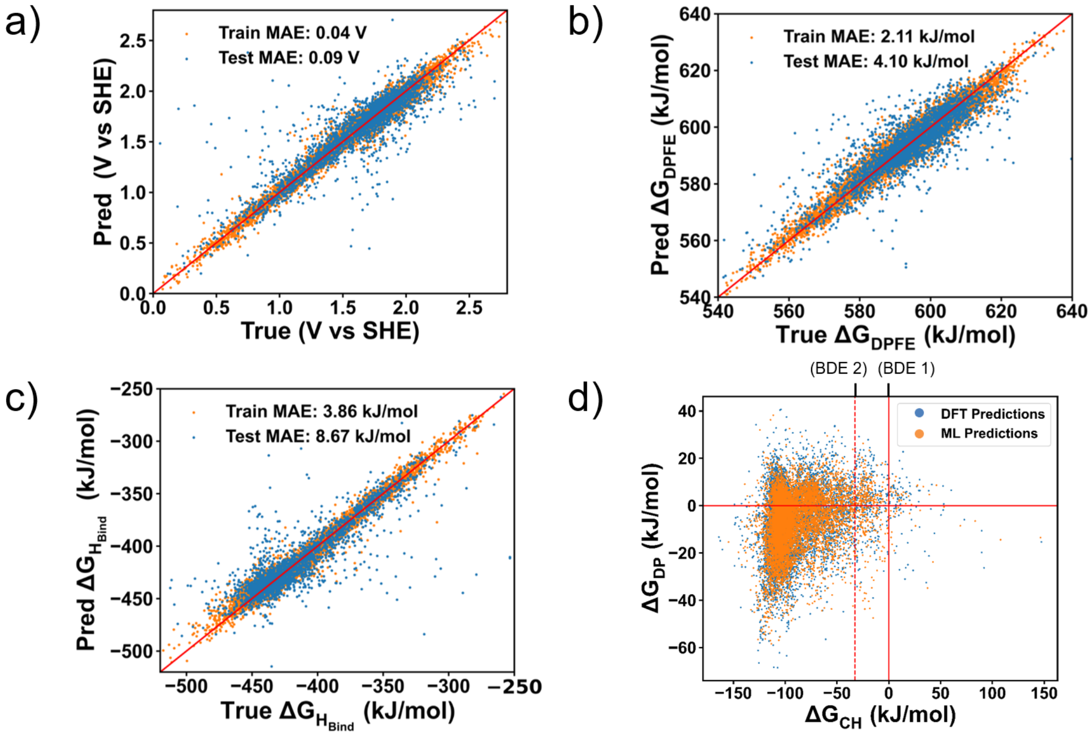


Figure 3-3: a)The two HAT reactions for 1,3-diphenylpropane, identified as (1) and (2) with their associated BDEs. b) Free energy landscape of an ylide redox mediators showing the electron transfer, hydrogen abstraction and deprotonation. c) Scatter plot of ΔG_{CH} and ΔG_{DP} for all DFT analyzed ylides.

The time requirements of screening our virtual library at our desired level of theory are elucidated in Figure 3-13 and Table 3.4, where we can see that each calculation had an average wall time of 1–3 hours. Based on our available resources, we sought to accelerate our screening for the remaining 9,097 catalytic cycles (27,291 individual calculation). To do this, we turned toward data driven modeling. In particular, we constructed a graph convolutional neural network (GNN) that is capable of simultaneously mapping the SMILES string of an ylide to E^o, ΔG_{DPFE} , and $\Delta G_{H_{bind}}$. GNNs have been shown to perform quite well in mapping chemical structure to thermochemistry, especially for datasets on the order of 10^5 and larger [63, 7, 19, 67]. The details of the methods we used to train our models can be found in the supporting information. Briefly, we used Chemprop wrapped in custom code to build and train our models, which were trained on ten 70:10:20 random splits of the data using the Adam optimizer and mean squared error loss. Bayesian hyperparameters tuning was performed using the python Optuna library, where all code for hyperparameter tuning and training can be found at `https://github.com/jmaalouf23/QMkit`. The performance of training a model using the best hyperparameters can be seen in Figure 3-3a-c. We achieve mean average errors of 0.09 V, 4.10 kJ/mol, and 8.67 kJ/mol for E^o, ΔG_{DPFE} , and $\Delta G_{H_{bind}}$ respectively on the test set. ΔG_{CH} was observed to have the largest predictive error, likely because of the higher deviation from gaussian behavior previously discussed. The non-gaussian behavior would constrain the model to map many diverse input structures to a narrow range of energy values.

We used our ML model to calculate ΔG_{CH} and ΔG_{DP} for the remaining ylides in our library, which are shown by the orange points in Figure 3-3d. This process only required seconds of computational time, orders of magnitude lower than the temporal requirements of DFT calculations. Notable, the ML calculated points occupy a similar region in ΔG_{CH} xΔG_{DP} space as the DFT calculated points, similarly demonstrating a high density of points in the low ΔG_{CH} region. The analysis we have performed with 1,3 diphenylpropane can be easily transferred to any substrate of interest without the need to recalculate any ylide properties. Only BDEs of the new substrate are required. For a new substrate, the location of the points would shift horizontally relative to the vertical line corresponding to the new BDE of interest.

3.2.2 Experimental Characterization

With values for ΔG_{CH} and ΔG_{DP} of our entire library, we sought to experimentally validate the performance of specific structures HAT reactions of 1,3 diphenylpropane. We selected 7 ylides (referred to as YN, where N is the ylide index). The ylides were chosen first because they span a wide range of values of ΔG_{CH} , from -133.95 kJ/mol to -12 kJ/mol allowing us to examine how this thermodynamic parameter effects important performance metrics. Of these 7 selected ylides, we were able to successfully synthesize 4 of them, Y1, Y2, Y3, and Y6.

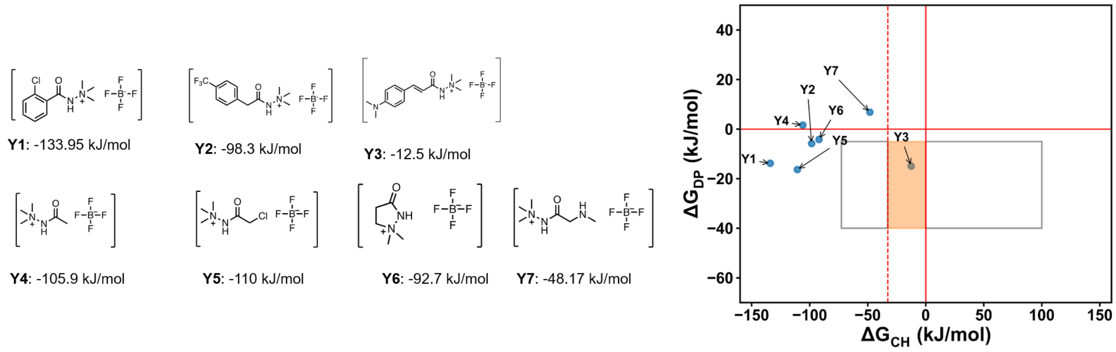
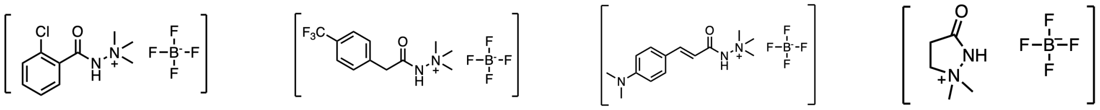


Figure 3-4: a) Selected Ylide candidates that were chosen for further examination. b) Performance of the selected candidates for HAT chemistry.

We probed their electrochemical properties by testing the materials using cyclic voltammetry. Figure 3-5 shows the results of plotting the peak potential of the CVs vs our calculated values. We observe that the trend in experimental values follows our predictions for all cases except for Y6. This suggests that conclusions that rely on relative values of different structures can be trusted, and it's likely that the error with Y6 is related to the fact that it belongs to a different structural class than Y1-Y3.

3.3 Conclusion

We have shown that for the case of 1,3 diphenylpropane we can use a virtual high throughput screening approach to select N-ammonium ylide based HAT redox mediators that have energies commensurate with the BDEs present on the molecule. By screening a library of 20,000 ylides using DFT calculations supplemented with data driven models, we were able to demonstrate that many common structures fall within a similar energy range. Only a small percentage of our library fall within a low ΔG_{CH} energy region (close to 0 kJ/mol), demonstrating the need for prior knowledge of mediator energetics in order to choose appropriate mediator structures. Our screening, allowed us to choose several ylide candidates that we synthesized and tested for their electrochemical HAT performance.



Y1: -133.95 kJ/mol **Y2**: -98.3 kJ/mol **Y3**: -12.5 kJ/mol **Y6**: -92.7 kJ/mol

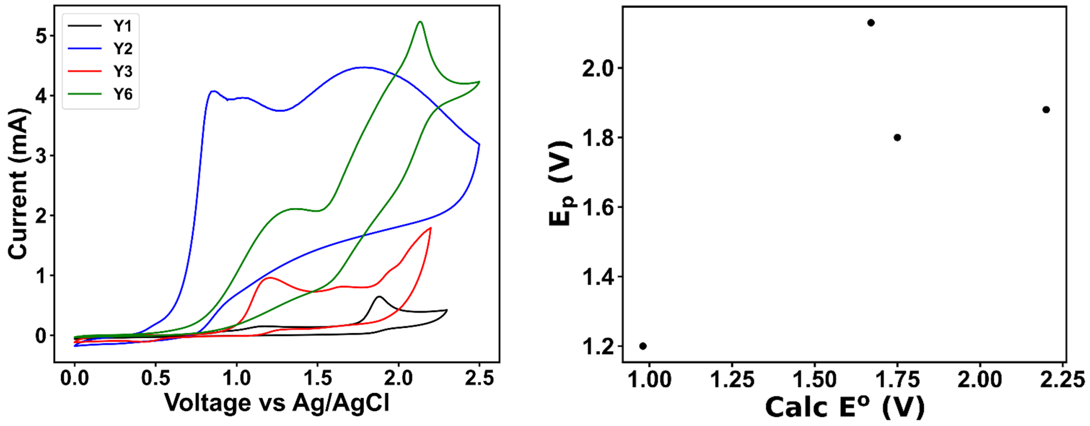


Figure 3-5: Top: structures of ylides that were successfully synthesized in this work. Bottom: Cyclic voltammetry of selected ylides. 100 mM of ylide was added to MeCN with 100 TBA BF_4 as the electrolyte.

3.4 Supporting Information

3.4.1 Methods

Virtual Chemical Library Design

Ylide Redox Mediator Library

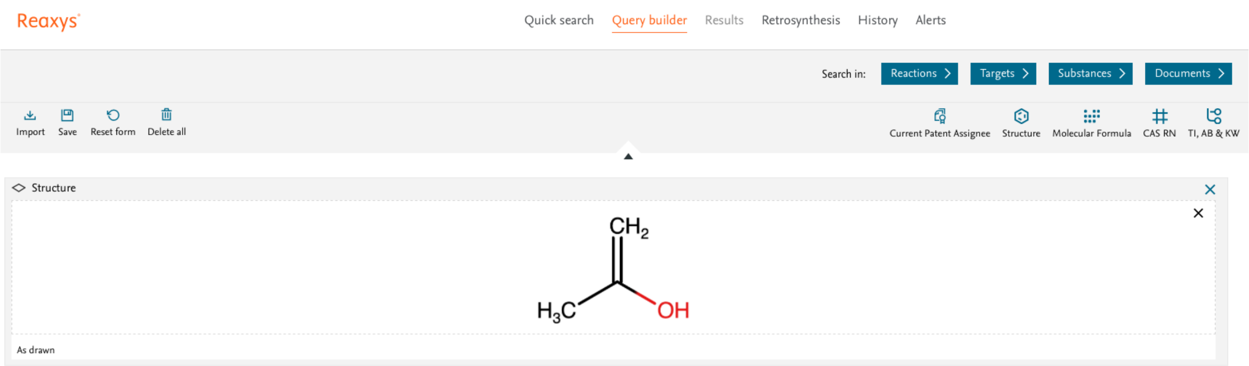

Reaxys
Quick search
Query builder
Results
Retrosynthesis
History
Alerts
Search in:
Reactions
Targets
Substances
Documents
Import
Save
Reset form
Delete all
Current Patent Assignee
Structure
Molecular Formula
CAS RN
TI, AB & KW
Structure
CH2
H3C
OH
As drawn

Figure 3-6: Original Reaxys carboxylic acid query.

N-ammonium ylides have previously been shown to be functionalizable using carboxylic acids. Thus, we built our chemical library by first performing a carboxylic acid substructure search using Reaxys. The original query was for any molecules that contained the acetic acid motif (Figure 3-6). We included constraints such as commercial availability and the existence of NMR spectra in order to increase the likelihood that the carboxylic acid was practically accessible (Table 3.1). Additionally, we removed more exotic functional groups that could potentially result in synthetic or reactive issues down the line. SMILES for the final 22.02 K molecules were exported in a csv format. Lastly, the set of carboxylic acid molecules were double-checked for duplicates and additionally any molecules containing atoms outside the set H,C,N,O,F,P,S,Cl,Br, and I were removed resulting in a final set of 17,586 carboxylic acids.

Each ylide redox mediator catalytic cycle consists of three different species: the ylide (Y), the radical cation (Yrad), and the protonated form (Yh) (Figure 1). Each carboxylic acid was

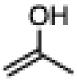
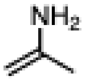
Filter	**Number of Results Post Filter**
Original Search	4.5 M
Commercially Available	1.82 M
Molecular Weight < 250	225.6 K
Remove Alkynes	98.3 K
Remove Q-Q	89.4k
Remove alkanes longer than 3 Carbons	57.2 K
Exclude 	56.52 K
Exclude 	55.42 K
Exclude Ethers	22.06 K

Table 3.1: Substructures filtered from original Reaxys query.

appended to the appropriate N-ammonium ylide core using RDKit's `ReplaceSubstruct()` function to form either Y, Yrad, or Y. As an example, Figure 3-7 shows the relevant connection to form Y. For carboxylic acids that contained more than one carboxylic acid motif, every possible attachment was considered and added to the ylide library. The original library of 17,586 carboxylic acids resulted in a library of 19,097 catalytic cycles and a total of 57,291 molecules. The SMILES strings for these molecules can be found in a csv file in the supporting information.

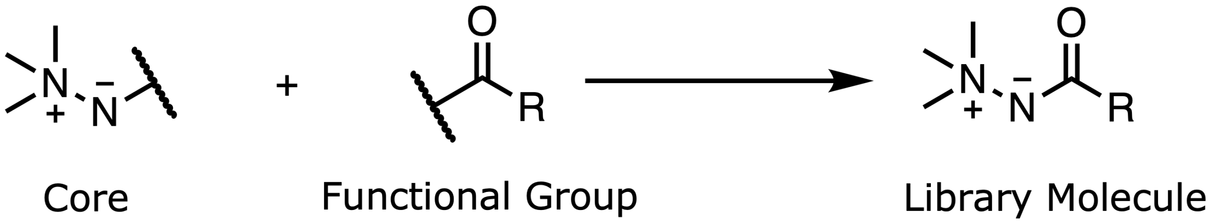


Figure 3-7: Molecular connections made using RDkit to form ylide library.

Model Substrates

Our goal was to design a HAT redox mediator for polystyrene decomposition by matching the thermodynamics of the mediator to the relevant C-H bond in polystyrene. In order to do this, we chose to first test our design capabilities on styrene model compounds shown in Figure 3-8 that were inspired by those used by Yan et al. [69]. All model compounds include 2 phenyl groups with a varying number of carbons in between, in addition to some oxygen functionality in certain cases.

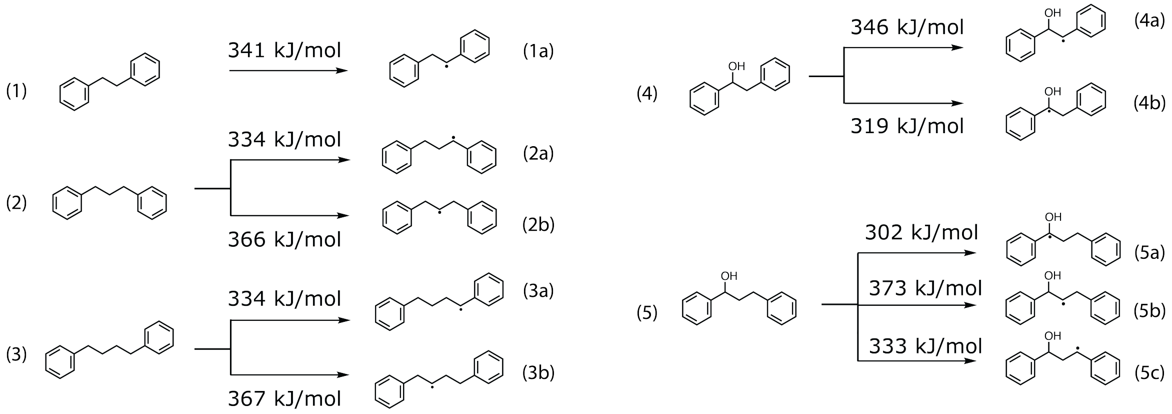

(1)
341 kJ/mol
(1a)
(2)
334 kJ/mol
(2a)
366 kJ/mol
(2b)
(3)
334 kJ/mol
(3a)
367 kJ/mol
(3b)
(4)
OH
346 kJ/mol
(4a)
319 kJ/mol
(4b)
(5)
302 kJ/mol
(5a)
373 kJ/mol
(5b)
333 kJ/mol
(5c)

Figure 3-8: Model substrates considered in this work. Energies correspond to the BDE of removing a hydrogen atom from the indicated C-H bond.

3.4.2 DFT Calculations

Due to the size of the library compared to available computational resources, DFT calculations were only performed for the first 10,000 catalytic cycles (30,000) molecules. The energetics of the remaining half of the library was screened using graph based machine learning models described later in the text. DFT calculations were run using Gaussian 16 Software and distributed across Intel Xeon Platinum 8260 nodes provided by the MIT Super Cloud [1].

For Y, Yrad, and Yh, first a gas phase calculation (geometry optimization and harmonic frequency calculation) was performed at the M06-2X/def2-TZVP level of theory, which has been previously shown to obtain good accuracy for bond dissociation calculations. For Y and Yh, the initial starting geometry was obtained by sampling 100 conformers using the RDKit library and then optimizing their structure using RDKit's implementation of the MMFF94s force field. The geometry of the lowest energy converged conformer was then taken as the starting geometry of the gas phase calculation. Since MMFF94s is not parameterized for radicals, a procedure from the literature was implemented where the initial structure of Yrad was obtained by using the DFT gas phase optimized geometry of Yh and then removing the H atom which gives Yrad [56]. A geometry optimization and frequency calculation was then performed for Yrad with this starting geometry.

The geometry optimized structure of Y, Yrad, and Yh from the gas phase calculations were then used as a starting geometry for DFT calculation that incorporated implicit solvation in acetonitrile using the SMD model. Acetonitrile was the dominant solvent used in all validation experiments. These calculations were performed at the M06-2X/def2SVPD level of theory. We chose this basis set as it includes diffuse basis functions to address the anionic character of the nitrogen atom in Y, while also being of a tractable size to allow for rapid computational screening.

All output files were parsed using cclib python library. The relevant python function used for generating conformer, 3D coordinates and Gaussian input scripts can be found at `https://github.com/jmaalouf23/QMkit`. These functions were utilized in the following set of Jupyter Notebooks (`https://github.com/jmaalouf23/ylide`) to build the molecular

Failure Mode	Y	Yrad	Yh
Geometry Optimization Failure	385	436	562
Other Failure Modes	143	358	339

Table 3.2: Summary of the number of failed calculations for the 10,000 catalytic cycles, corresponding to 30,000 DFT calculations

library, generate directories containing all the input scripts and to submit all the calculations for execution on the MIT Super Cloud. Each calculation was checked to make sure that the geometry had converged to a minimum by verifying that all vibrational frequencies were greater than zero. This resulted in 8095 converged ylide catalytic cycles where Y, Yrad, and Yh all successfully converged. A summary of all failures can be found in Table 3.2, where "other failure modes" predominantly refers to SCF convergence issues.

The converged results of the DFT calculations performed on the ylide library were stored in an PostgresSQL database. Stored data includes, the electronic energies, enthalpies, Gibbs free energy, optimized xyz coordinates, and frequencies.

3.4.3 Calculated Thermodynamic Quantities

The successful DFT calculations for the first 10,000 catalytic cycles were used to calculate key thermodynamic quantities indicated in Figure 1,namely the redox potential (E^o), the hydrogen binding energy ($\Delta G_{H_{Bind}}$) and the deprotonation energy (ΔG_{DPFE}) which are explicitly defined in Table 3.3. Free energies of the hydrogen atom and the proton, which are required to calculate the desired energies were calculated using DFT at the same level of theory as Y,Yrad, and Yh (M06-2X/Def2-SVPD) in acetonitrile using SMD. While we acknowledge there can be large errors associated with DFT calculations of protons, we note that ΔG_{DPFE} is typically subtracted from the deprotonation free energy of an explicit base, resulting in the cancellation of the proton free energy. CSV files containing $E^o, \Delta G_{DP}, \Delta G_{H_{bind}}$ for each catalytic cycle were generated from the data stored in the aforementioned PostgreSQL database.

Thermodynamic Quantity	**Definition**
Redox Potential	$E^o = \frac{\Delta G_{Yrad} - \Delta G_Y}{nF} - 4.28V$
Hydrogen Binding Free Energy	$\Delta G_{H_{Bind}} = \Delta G_{Yh} - \Delta G_{Yrad} - \Delta G_{\mathrm{H}\bullet}$
Deprotonation Free Energy	$\Delta G_{DPFE} = \Delta G_Y + \Delta G_{H+} - \Delta G_{Yh}$

Table 3.3: Definitions of ylide thermodynamic quantities used for screening.

Another key thermodynamic quantity used throughout is the free energy of the base in the system ($\Delta G_{Base} = G_{BH} - G_B - G_{H+}$). In our case, the based used in our system was predominantly $NaHCO_3$, which was shown to be a suitable base in the literature[46]. ΔG_{Base} was calculated using DFT at the same level of theory as all other solvated calculations by calculating the geometry optimized energies of HCO_3, H_2CO_3 and H^+.

3.4.4 Machine Learning Models

Model Architecture and Training

To obtain E^o, ΔG_{DPFE}, and $\Delta G_{H_{Bind}}$ for the remainder of the ylide library, we trained message passing neural networks using Chemprop to predict these properties. The original software package was wrapped in custom code to allow for more flexibility during model training and to allow us to predict all 3 energies at once. In the model, the SMILES of Y is used to form a graph representation of the molecule. Atom and bond feature vectors are built for each molecule. Chemprop by default only uses atom and bond features and does not add molecular level features to the representation after convolution. We used the Chemprop default atom and bond features. Specifically, the atom features are (i) the atomic number, (ii) the number of directly bonded neighbors, (iii) the formal charge, (iv) the chiral tag, (v) the number of connected H's, (vi) and the hybridization. These are all one hot encoded, resulting in a vector length of 130. The bond features are, (i) the bond type (single, double, triple), (ii) the aromaticity, (iii) the conjugation, (iv) and whether the atom is in a ring. The atomic and bond feature vectors are then convolved by passing them through message passing (MPNN) layers, after which the atom features were aggregated into a molecular vector representation by taking the mean of all the atom vectors. This vector is then passed into a traditional feed forward neural net (FFN) readout architecture that relates the vector representation to the properties of interest.

We used a random split of size 70:10:20 to form the training, validation and test set respectively. When training our model on this type of split, we chose to train 10 models in total with 10 different random splits in order to perform a fair evaluation. The models were trained using mean squared error as the loss function with and Adam optimizer. Bayesian hyperparameter tuning was performed using the Optuna python library. Using the best validation loss to monitor hyperparameter performance and the CMA-ES algorithm implemented in Optuna, we were able to optimize the MPN hidden size, MPN depth, dropout fraction, learning rate, batch size, FFN hidden size, and the FFN depth. All code for hyperparameter tuning and training can be found at `https://github.com/jmaalouf23/gnn`.

3.4.5 N-Ammonium Ylide Synthesis

The general synthesis procedure was adapted from Saito et al. [46] with the general steps indicated in Figure 3-9.

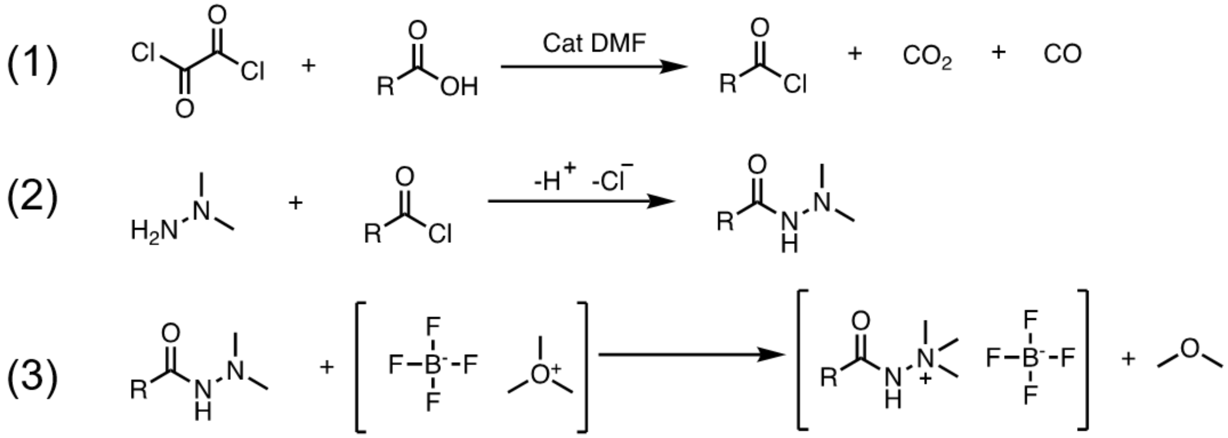


Figure 3-9: General synthesis steps to make N-ammonium ylides when starting from a carboxylic acid, as shown in 3.4.5.

Procedure 1: Starting from Carboxylic Acid

1. Step 1: Acyl Chloride Formation

   A round-bottom flask was charged with carboxylic acid (1.0 equiv.) and a magnetic stirring bar. The flask was then evacuated and backfilled with argon. Depending on the carboxylic acid solubility, either dry CH_2Cl_2 or dry THF (0.2 –0.3M) was used, and the resulting solution was cooled to 0 °C. DMF (3–5 drops) was added and oxalyl chloride (1.5–2.0 equiv.) was added dropwise over 10 min to the reaction mixture. After addition, the reaction mixture was stirred for 1 h at 0 °C and 2 h at laboratory temperature. The reaction mixture was then concentrated under reduced pressure to remove any excess oxalyl chloride and afforded the desired acyl chloride. The crude acyl chloride was used directly in the following step without further purification.

2. Step 2: dimethylhydrazide formation

   A round-bottom flask charged with a magnetic stir bar was evacuated and backfilled with argon twice. The flask was charged with N,N-dimethyl hydrazine (2.1 –3.0 equiv.) and dry CH2Cl2(0.2 –0.3 M) and the resulting solution was cooled to 0 °C. To this solution, acyl chloride (1 equiv.) was added dropwise over 15 min and the reaction was stirred for 30 min at 0 °C and 3 h at laboratory temperature. The reaction mixture was diluted with CH2Cl2(10 –50 mL) and 1 M aq. NaOH (20–50 mL), the organic phase was separated, and the aqueous phase was extracted with CH_2Cl_2 twice. Combined organic phases were dried over MgSO4, filtered, and concentrated under reduced pressure. The crude hydrazide was used directly in the following transformation, with no further purification.

3. Step 3: N-methylation of hydrazide

   A round-bottom flask charged with a hydrazide (1.0 equiv.) and a magnetic stir bar was evacuated and backfilled with argon twice. Dry MeCN (0.10 –0.25 M) was added, followed by the addition of trimethyloxonium tetrafluoroborate (1.1 equiv.) in one portion. The resulting homogeneous solution was stirred at laboratory temperature for 30 min. Then the reaction mixture was concentrated under reduced pressure to afford a solid residue. The crude ylide salt was recrystallized from refluxing EtOH then washed with Et2O and EtOH to afford pure ylide salt.

Procedure 2: Starting from Acyl Chloride

1. Step 1: N-acyl-N', N'- dimethylhydrazide formation

   A round-bottom flask charged with a magnetic stir bar was evacuated and backfilled with argon twice. The flask was charged with N, N-dimethyl hydrazine (2.1 –3.0 equiv.) and dry CH_2Cl_2(0.2 –0.3 M) and the resulting solution was cooled to 0 °C. To this solution, acyl chloride (1 equiv.) was added dropwise over 15 min and the reaction was stirred for 30 min at 0 °C and 3 h at laboratory temperature. The reaction mixture was diluted with CH2Cl2(10 –50 mL) and 1 M aq NaOH (20 –50 mL), the organic phase was separated, and the aqueous phase was extracted with CH2Cl2twice. Combined organic phases were dried over MgSO4, filtered, and concentrated under reduced pressure. The crude hydrazide was used directly in the following transformation, with no further purification.

2. Step 2: N-methylation of hydrazide

   A round-bottom flask charged with hydrazide (1.0 equiv.) and a magnetic stir bar was evacuated and backfilled with argon twice. Dry MeCN (0.10 –0.25 M) was added, followed by the addition of trimethyloxonium tetrafluoroborate (1.1 equiv.) in one portion. The resulting homogeneous solution was stirred at laboratory temperature for 30 min. Then the reaction mixture was concentrated under reduced pressure to afford a solid residue. The crude ylide salt was recrystallized from refluxing EtOH then washed with Et_2O and EtOH to afford pure ylide salt.

3.4.6 NMR

For solid samples, NMR samples were prepared by creating a stock solution of 25 mg of sample in 1 mL of d3-MeCN. Then 700 μL of this stock solution was loaded into an NMR tube.

If the sample was a liquid or oil, we took 25 μL of the ylide and diluted it in 1 mL of d3-MeCN. A 700 μL aliquot of this stock solution was then pipetted into an NMR tube and run.

3.5 Supplementary Figures

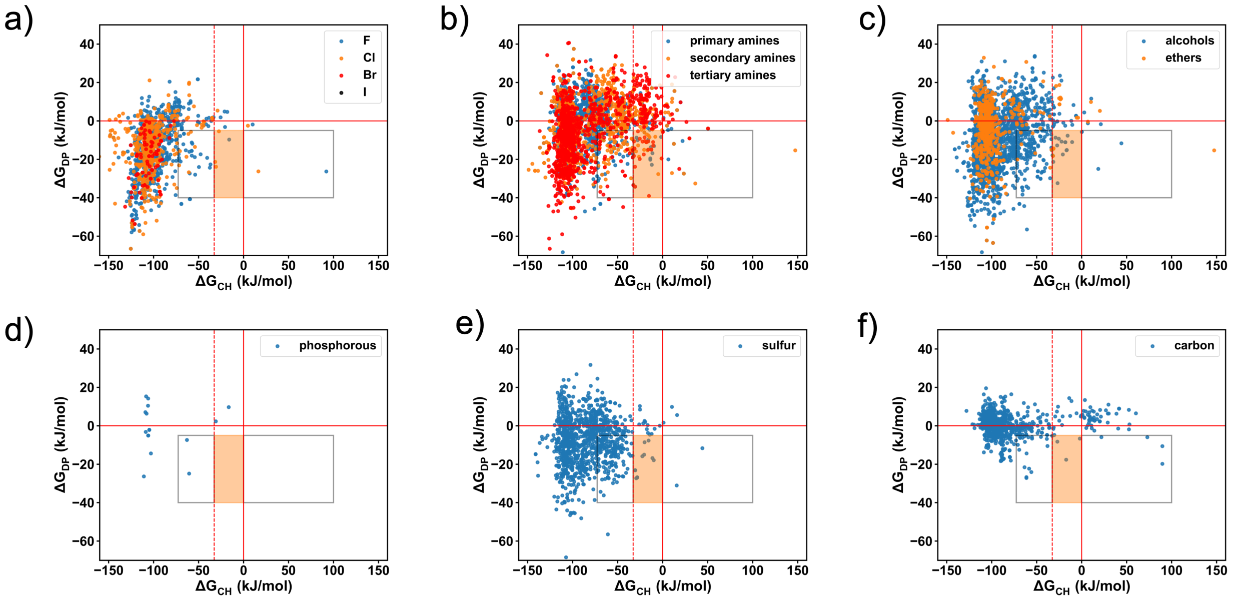


Figure 3-10: a) Ylides that contain at least one halogen in the set F, Cl, Br,I. It is possible for molecules in this plot to contain multiple halogens, in which case the ylide was plotted for each case. b) Ylides containing at least one of the amine types specified in the legend. c) Ylides containing alcohols and ethers. It is possible for other functional groups to exist on the molecule, such as amines or halogens. d) Ylides containing P atoms. e) ylides containing sulfur atoms. f) Ylides that contain only carbon atoms (excluding the ylide motif). No other functional groups are present on the molecule in this case.

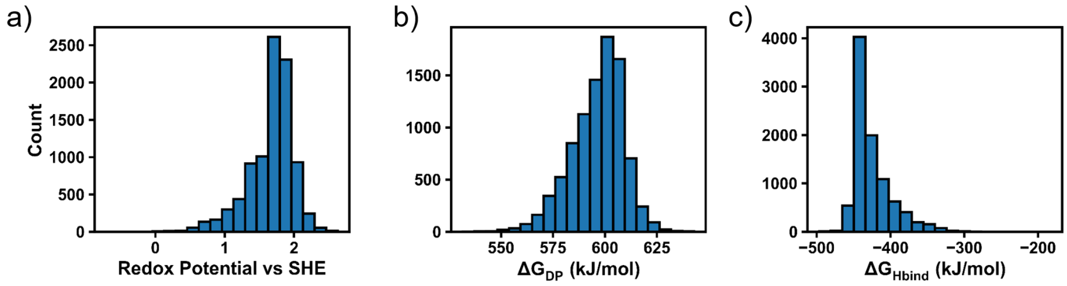


Figure 3-11: Histograms of a) E^o, b) ΔG_{DPFE}, and c) ΔG_{H_bind} from calculations performed at the M062X/def2-SVPD level of theory.

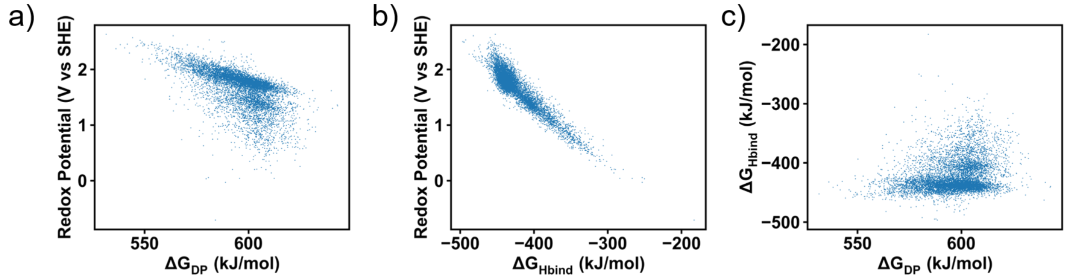


Figure 3-12: Scatter plots from calculations performed at the M062X/def2-SVPD level of theory of a) E^o vs ΔG_{DPFE}, b) E^o vs $\Delta G_{H_{bind}}$, and c) $\Delta G_{H_{bind}}$ vs ΔG_{DPFE}.

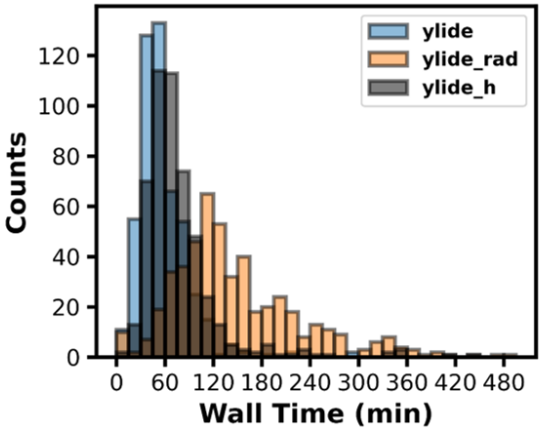


Figure 3-13: Histogram of the wall time of each ylide, ylide radical, and protonated ylide calculation.

	Ylide (Y)	Ylide Radical (Yrad)	Protonated Ylide (Yh)
Wall Time/molecule (h)	0.95	2.56	1.28

Table 3.4: Table of the average wall time per molecule of each of the three species in the ylide catalytic cycle from calculations at the M062X/def2-SVPD level of theory with SMD solvation in acetonitrile.

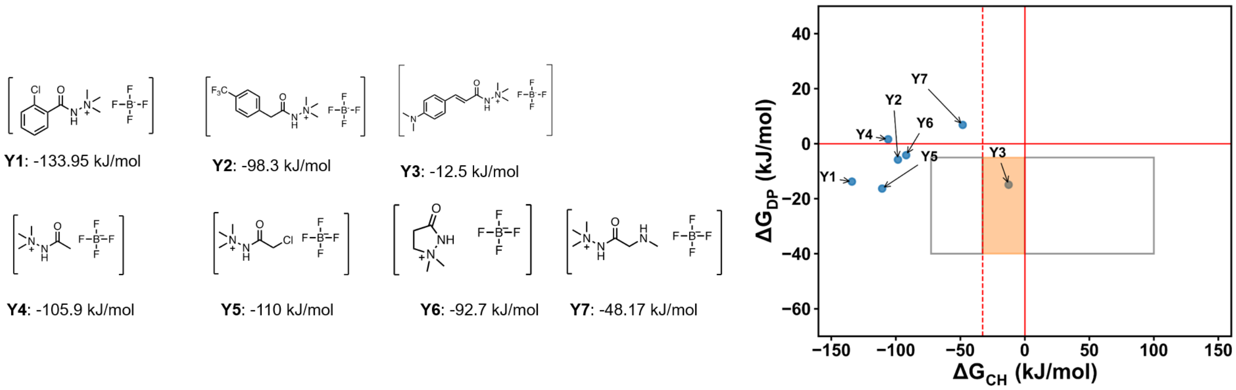


Figure 3-14: Selected candidates that were synthesized in this work.

3.5.1 NMR Spectra

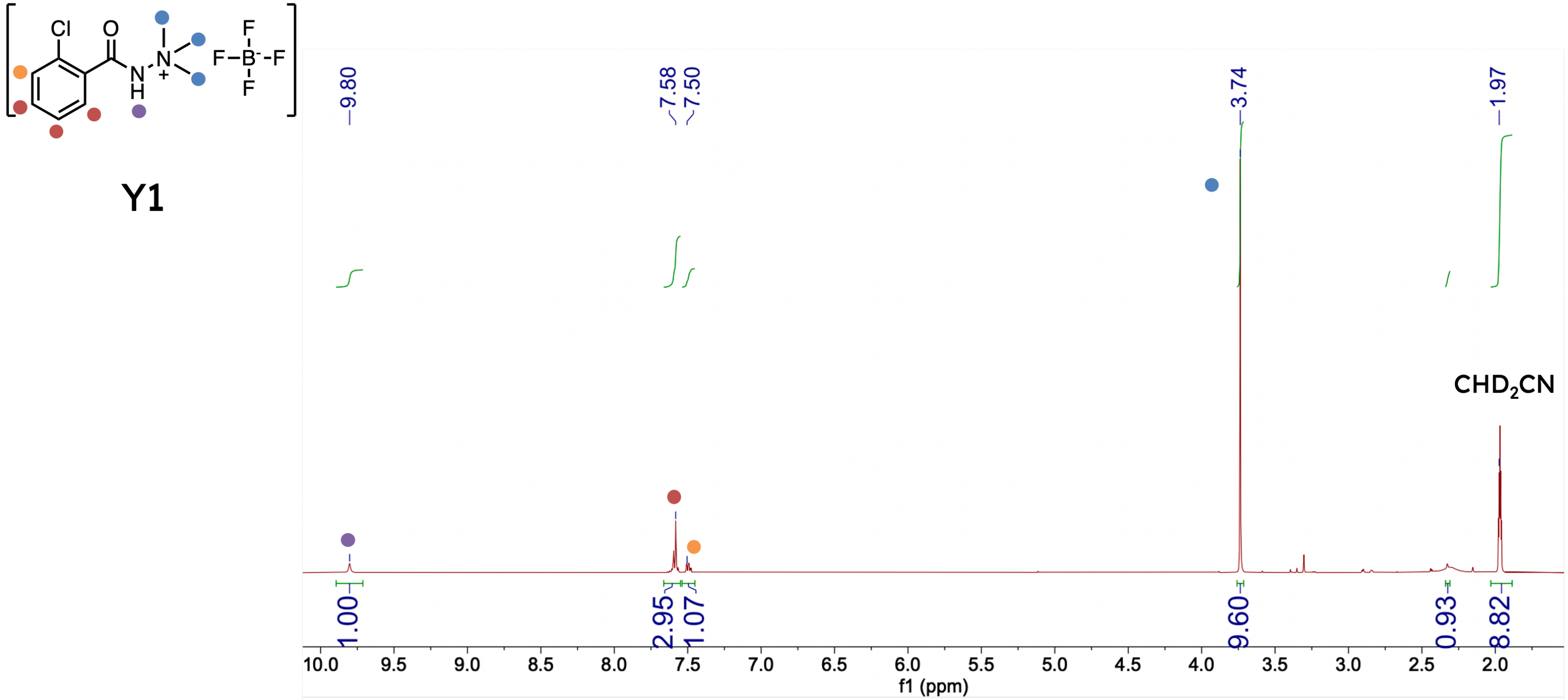


Figure 3-15: NMR Spectra of Y1 taken in deuterated MeCN.

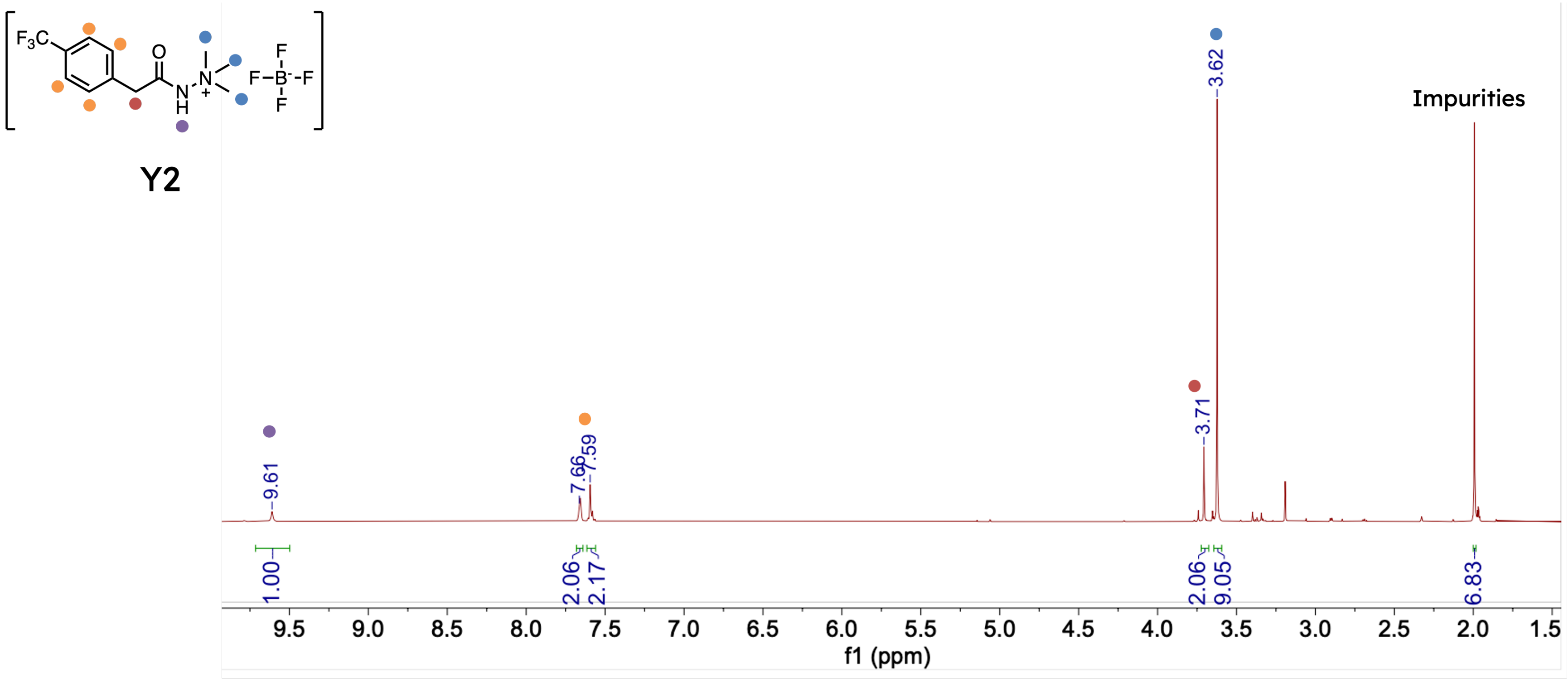


Figure 3-16: NMR Spectra of Y2 taken in deuterated MeCN.

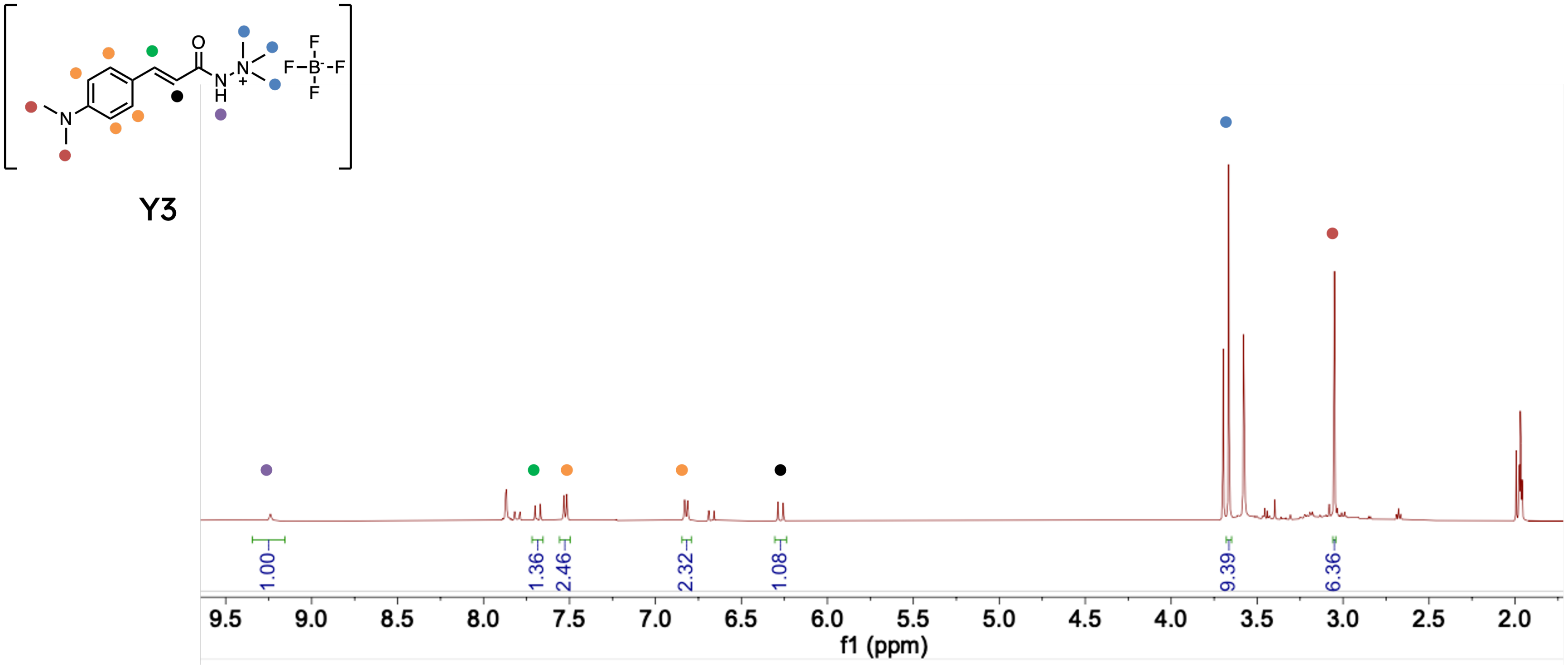


Figure 3-17: NMR Spectra of Y3 taken in deuterated MeCN.

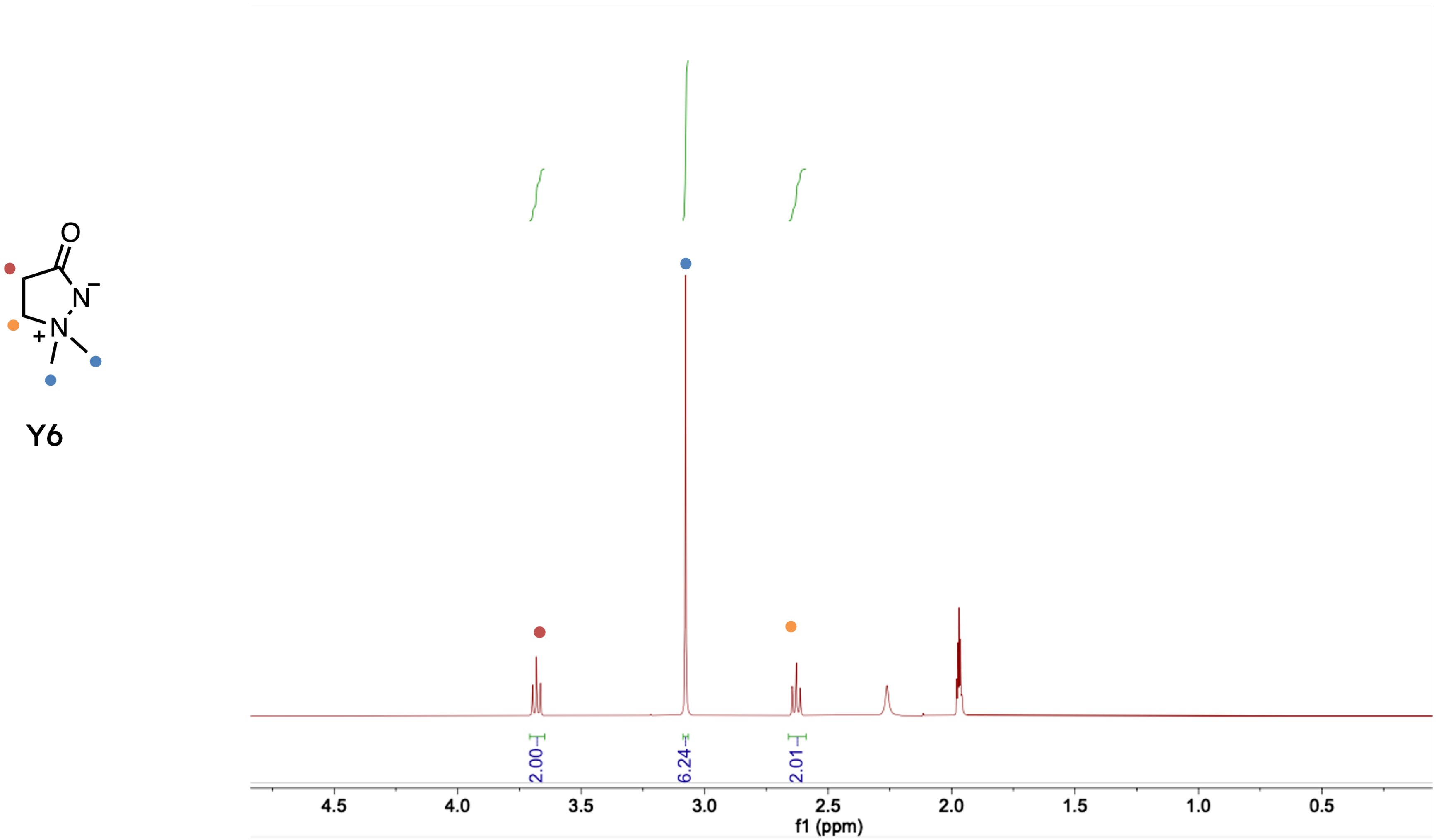


Figure 3-18: NMR Spectra of Y6 taken in deuterated MeCN.

Chapter 4

Incorporating Machine Learning into the prediction of Organic PCET Redox Potentials in Multiple Solvents

Abstract

Redox potentials of proton coupled electron transfers (PCETs) are important reaction properties that dictate capacities in energy storage systems such as redox flow batteries. They can be calculated using techniques such as DFT, accompanied by implicit solvation to handle environmental effects, but recently researches have turned toward fast and accurate data driven models as a way to help bypass the need for repeated large-scale QM calculations during high throughput computational screening of electroactive materials. Herein, we explore the degree to which machine learning models can be used to calculate the redox potentials of PCETs referenced to H_2/H^+. We compiled 75 experimental PCET redox potentials from various literature sources and compared a baseline DFT + SMD approach with another that replaces SMD implicit solvation with a graph convolutional neural network (GNN) trained on 10^6 COSMO-RS calculation. We show that at the B3LYP/6-31G(2df,p) level of theory, this new approach results in an average improvement of 186 mV. We also show that GNNs that predict gas phase values for ΔH_f and ΔS_o can also be used in certain cases. Specifically, for 2 electron PCETs, we obtain a MAE of 189 mV, in line with DFT calculations at the same level of theory. However, for 1 electron PCETs the approach is not valid, as MAEs of over 1V are obtained. These finding demonstrate the degree to which ML models that predict gas phase molecular thermochemistry can be incorporated into workflows for the large scale calculations of redox potentials.

4.1 Introduction

Electrochemical equilibrium potentials are fundamental thermodynamic quantities that determine the propensity of a molecule to accept or donate an electron. Redox reactions as a general class of reactions are prevalent in areas such as photovoltaics, electrocatalysis, energy storage, and biological processes. However, each application tends to involve different classes of molecules that undergo electron transfers in drastically varying solvent environments (water, acetonitrile, THF, DME, IPA, etc.). The ability to ascertain the equilibrium potential of a redox reaction for an arbitrary redox couple in an arbitrary solvent would give great insight to the design of optimal redox couples in a myriad of fields.

Two specific applications that would benefit from information on the redox potential include redox mediated electrocatalysis and non-aqueous organic redox flow batteries. In electroorganic synthesis, redox mediators are species that are activated at an electrode and then go on to subsequently react with the intended substrate via a chemical reaction. Redox mediators are typically used because they can be oxidized at much lower potentials than what is usually required to directly activate the substrate. Additionally, finer control over the redox potential and thus the thermodynamics of the system is available through means such as chemical functionalization of the redox mediator. In principle, this functionalization results in a rich design space, but it cannot be adequately investigated experimentally. For this reason, being able to rapidly calculate the redox potential of redox mediators could help in thorough searches of chemical space for candidate redox mediators for a given reaction.

Redox flow batteries are an energy storage technology based on the oxidative and reductive cycling of a redox couple and offer key benefits when compared to conventional electrochemical energy storage technologies such as batteries. In particular, they allow for the decoupling of energy and power density. RFBs based on organic redox couples are cheaper and more environmentally friendly than their traditional metal ion counterparts. Additionally, their solubility in various organic solvents in addition to water is also desirable, since these solvents enable larger cell voltages due to their wider electrochemical stability windows.

The difference in standard redox potential (operation voltage), along with solubility of the active species and the number of electrons involved in the redox reaction, are the key parameters that can dictate the energy density of RFBs. Some examples of typical classes of molecules used in organic RFBs include quinones, viologens, nitroxides, methoxybenzenes, and phenazines, both in aqueous and non-aqueous organic systems. These classes of molecules offer a high degree of tunability of over their standard redox potentials through the selective addition of functional groups at various positions in the molecule. This tunability can result in thousands of potential redox couple candidates, and even more system configurations when different solvents are considered. This makes it infeasible in some cases to experimentally investigate a meaningful portion of this chemical space if one is in search of promising candidates for energy dense RFBs. Instead, computational modeling is often used to screen candidates in a high throughput fashion by using molecular properties such as redox potentials as a screening parameter. Typical computational approaches calculate the standard redox potential by exploiting the relationship to the Gibbs free energy of reaction:

$$E^o = \frac{-\Delta_r G}{nF} \tag{4.1}$$

Here is the free energy change of the redox reaction and can be computed via the relationship:

$$\Delta_r G = \sum_i \nu_i \Delta G_i \tag{4.2}$$

where ν_i is the stoichiometric coefficient, n is the number of electrons transferred during the reaction, and F is Faraday's constant. When calculating ΔG_i for a given species i , the free energy is typically divided into the gas phase free energy and solvation free energy:

$$\Delta G_i = \Delta G_{i,gas} + \Delta G_{i,solv} \tag{4.3}$$

Each of these free energy terms can be calculated from any one of many techniques such as DFT, MD, wave function methods, etc. with the incorporation of some implicit or explicit solvation method when calculating the solvation free energy (PCM, SMD, COSMO-RS, etc). The overall relationship between the previously mentioned free energies is typically summarized in the following Born-Haber cycle (**Figure 4-1**). For a redox reaction

$$A(s) + ne^- \rightleftharpoons B(s) \tag{4.4}$$

we can draw:

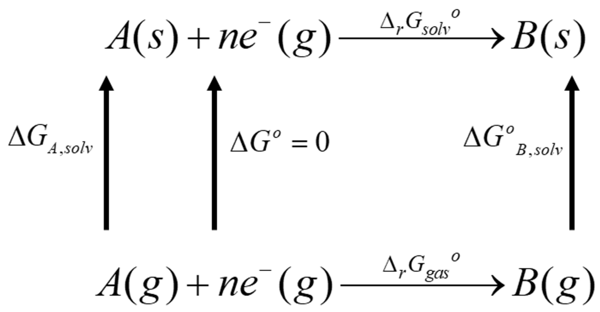


Figure 4-1: Born-Haber cycle, demonstrating how the free energy of a redox reaction can be broken up into its constituent gas phase and solvation free energies. Note that charges on A and B are implicit, such that conservation of charge is maintained.

and decompose the relevant free energy into

$$\Delta_r G_{solv} = \Delta_r G_{gas} + \Delta G_{B,solv} - \Delta G_{A,solv} \tag{4.5}$$

where,

$$\Delta_r G_{gas} = \Delta G_{B,gas} - \Delta G_{A,gas} - \Delta G_{e^-,gas} \tag{4.6}$$

ΔG_{gas} is the gas phase free energy of formation and ΔG_{solv} is the solvation free energy. $\Delta_r G$ indicates a quantity associated with a reaction. It is also useful to recall that the free energy can be decomposed into enthalpy and entropy terms via the equation:

$$\Delta_r G = \Delta_r H - T \Delta_r S \tag{4.7}$$

where expressions similar to equation (4.2) can be used for $\Delta_r H$ and $\Delta_r S$

An important note is that $\Delta_r G$ is often taken in conjunction with a reference electrode half-reaction. Some popularly used reference electrodes include the standard hydrogen electrode (SHE) in water and the Ferrocene/Ferrocenium redox couple in non-aqueous solvents. When chosen properly, referencing to a reference electrode reaction can avoid the need to calculate the free energies of species such as electrons and other ions. In particular, this is the case for proton coupled electron transfers (PCETs), a class of redox reaction prevalent throughout electrochemistry. PCETs can be generally described by the following:

$$XH_n(s) \rightleftharpoons X(s) + nH^+(s) + ne^-(g) \tag{4.8}$$

In this case, the same number of electrons and protons are transferred in the reaction. As an example of a well-chosen reference reaction, for a PCET, if one uses the popular normal hydrogen electrode (NHE) as a reference shown in (**4.9**).

$$XH_2(s) \rightleftharpoons X(s) + 2H^+(s) + 2e^-(g) \tag{4.9}$$

Where the proton is solvated in the same solvent in which the PCET occurs, resulting in the following overall reaction:

$$XH_2(s) \rightleftharpoons X(s) + H_2 \tag{4.10}$$

From (**4.10**) one can see it is no longer necessary to calculate any thermodynamic values for charged species such as protons or electrons to obtain a redox potential referenced to NHE(s). Specifically, when referenced to NHE(s), the reaction above can be written as:

$$E_{NHE}(s) = -\frac{\Delta G^o_X(s) + \Delta G^o_{H_2}(g) - \Delta G^o_{XH_2}(s)}{nF} \tag{4.11}$$

This avoids a common assumption that must be made when computationally determining redox potentials vs SHE. Typically, the value of E_{abs} vs SHE , which is the value of SHE referenced to vacuum, is assumed to be 4.422 volts based on computational studies. However, others have reported values varying by a few hundred mV. Thus, the value of E_{abs} vs SHE can be a source of error for any method that relies on its precise value, and it is noteworthy that it is not required to calculate V vs NHE(s) for PCETs.

While there have been many efforts validating the use of cheaper computational methods to compute the desired energies required to calculate a redox potential, which can help speed up large scale computational screening efforts for potential molecules to serve as redox couples in RFBs, there is still a push to lower the computation time while maintaining accuracy and generality. Rapid protocols for computing reasonably good free energies and solvation free energies could be used in high-throughput screening studies to determine molecular and solvent contributions to stabilizing the oxidized or reduced state of a given reactant, which has been shown to be a key contribution in obtaining accurate predictions. In this vein, machine learning has been shown to be a useful tool in the development of rapid and general models to predict various molecular properties. Building machine learning models trained on redox potential data to directly predict is likely not possible as large, ($N > 10^5$), diverse, and curated datasets for this property do not exist. Instead, an abundance of thermochemical data of organic molecules exists which can be used to build models and predict the necessary values shown in equations (**4.1**) and (**4.3**) to obtain E . Incorporating ML into the prediction of redox potentials has been demonstrated in a few different ways, but these studies are either limited to very specific classes of molecule or never when consider redox reactions in multiple solvents [55, 51, 2, 17].

Here in, we present and validate two approaches for incorporating thermochemical machine learning models into the prediction of 72 organic PCET reactions in 5 solvents (**4-2**). In the first proposed method, we demonstrate a mixed DFT-machine learning approach where gas phase free energies are predicted using DFT calculations (B3LYP or M062X) and the solvation energy is determined using a graph convolutional network that has been trained on 10^6 computational solvation free energies at the COSMO-RS level of theory. Secondly, we present the incorporation of fast and accurate machine learning models that can predict ΔH_f, ΔS^o , ΔG_{solv} for diverse sets of organic molecules and solvents. These thermochemical values are then used to calculate vs NHE(s) of PCETs as is shown in equation (**4.1**) resulting in a pure machine learning based approach to calculating redox potentials. In this second approach, ΔH_f and ΔS^o are predicted using graph convolutional networks (GNNs) trained on 10^5 gas phase DFT calculations while is determined using the same model as before (GNN that has been trained on 10^6 computational solvation free energies at the COSMO-RS level of theory). We also compare the accuracy and temporal requirements of our approaches to traditional computational methods for calculating standard redox potentials (DFT with implicit solvation such as SMD). The described new approaches are validated on 72 experimental PCET reactions from the literature. We demonstrate that when comparing our machine learning models to separate DFT calculations at the same level of theory that the ML model was trained on, we are able to recover similar accuracy on our experimental data set, validating our approach.

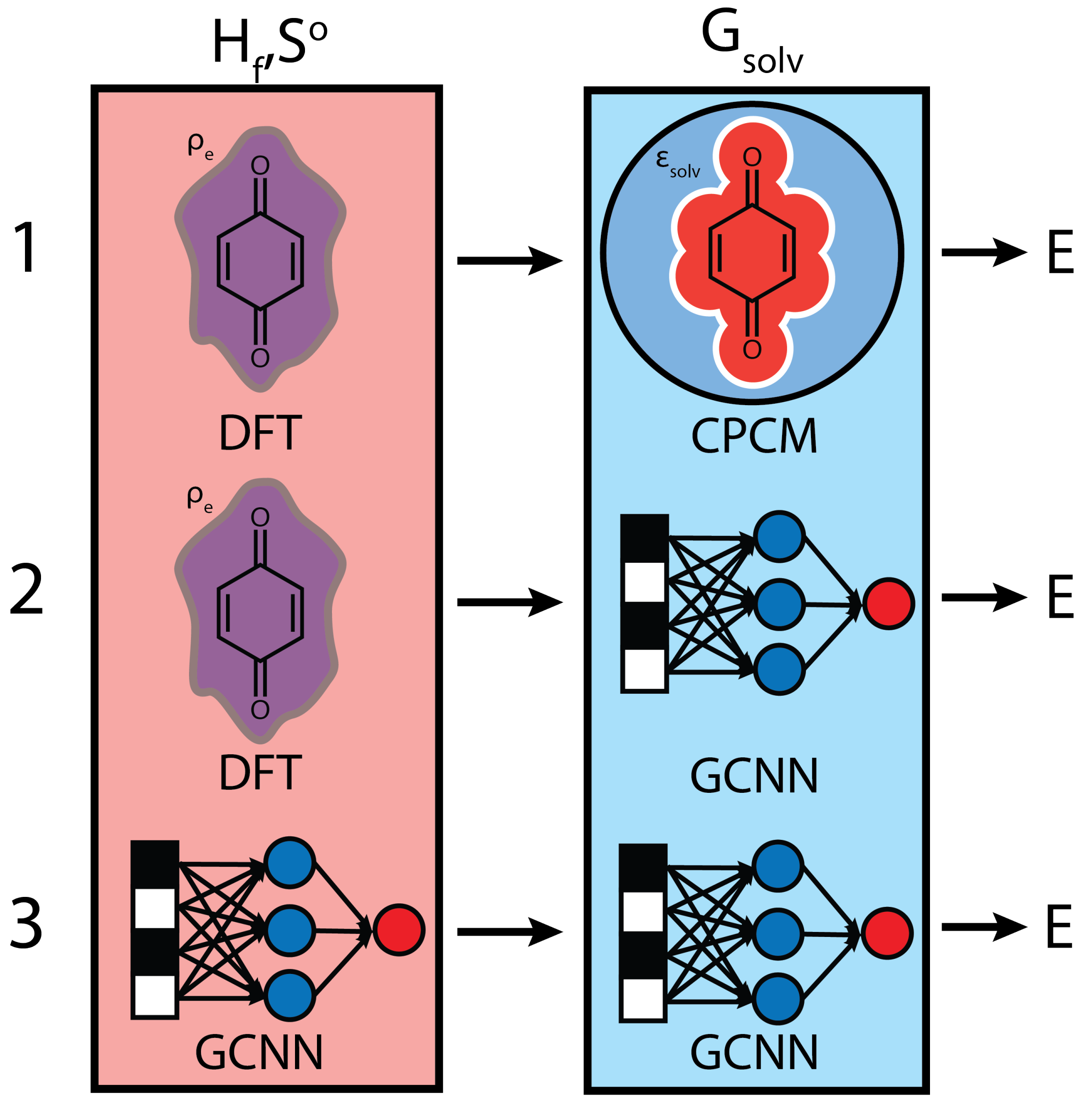


Figure 4-2: Summary of all approaches used to calculate redox potentials. Approach 1 uses DFT is used to calculate gas phase thermodynamics along with SMD to account for solvation. Approach 2 uses DFT to calculate gas phase thermodynamics and uses a machine learning model to obtain the solvation free energy (ΔG_{solv}). Approach 3 uses a machine learning model to predict gas phase free energies ($\Delta H_f, \Delta S^o$) and a separate machine model (same as Approach 2) to predict ΔG_{solv}

4.2 Datasets

4.2.1 E^o

When validating the accuracy of our computational approaches, we compared our predictions to two classes of experimentally obtained PCET redox potentials, summarized in Table 4.1:

Class	Reaction	n	N
1	$XH_2 \rightleftharpoons X + 2H^+ + 2e^-$	2	60
2	$XH \rightleftharpoons X + H^+ + e^-$	1	12

Table 4.1: Classes of electron transfer reactions considered in this work. We considered both 1 and 2 electron proton coupled electron transfers.

Values for redox potentials in each class were aggregated from various literature sources [68, 21]. The data set consists of experimental measurements in 5 different solvents commonly used for electrochemistry (water, acetonitrile, tetrahydrofuran, dimethylformamide, and isopropyl alcohol). Two techniques were predominantly used to experimentally determine redox potentials. The first was cyclic voltammetry, a more common technique where E^o is the midpoint of the oxidative and reductive peaks in a potential sweep[21]. The second technique was based on open circuit techniques, showing high accuracy for non-aqueous proton coupled electron transfers[68].

4.2.2 ΔH_f, ΔS^o, and ΔG_{solv}

Data sets for each of the 3 given thermochemical properties were taken from previous literature sources. For ΔH_f and ΔS^o the QM9 database was adapted to remove all F containing species resulting 130,000 DFT calculations at the B3LPY/6-31G(2df,p) level of theory [41, 45, 19]. The data sets consisted of H, C, O, and N containing molecules up to nine heavy atoms. For entropy, only the most stable conformer was considered. For ΔG_{solv} the CombiSolv-QM database was used, which is a dataset of 10^6 computational solvation free energies calculated at COSMO-RS level of theory[63]. CombiSolv-QM considers 11029 solutes with 284 common solvents and contains the elements H, B, C, N, O, F, P, S, Cl, Br, and I.

Quantity	N	Elements	Level of Theory	Reference
E^o	72	H, C, N, O, S, Cl	Experiment	[68, 21]
ΔH_f	$1.3*10^5$	H, C, N, O	B3LPY/6-31G(2df,p)	[41, 45, 19]
ΔS^o	$1.3*10^5$	H, C, N, O	B3LPY/6-31G(2df,p)	[41, 45, 19]
ΔG_{solv}	$1.3*10^5$	H, B, C, N, O, F, P, S, Cl, Br, and I	COSMO-RS	[63]

Table 4.2: Summary of datasets for various thermochemical quantities used to build data driven models in this work.

4.3 Methods

4.3.1 DFT Calculations

DFT calculations were performed using Gaussian 16 with the B3LYP functional and a 6-31++G** basis set. Initial 3D coordinates were obtained from SMILES strings using Open Babel software to generate the starting conformer. Solvation energies were included using continuum solvation models (SMD) with Bondi radii. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) Expanse SDSC.

4.4 Results and Discussion

4.4.1 Machine Learning Models

Machine learning models were built using Pytorch and Chemprop software [71]. Training was performed using the MIT Supercloud on Nvidia Volta V100 GPUs [1]. In order to implement approach 2 and 3, we developed 3 separate machine learning models trained on distinct data sets to predict ΔH_f,ΔS^o, and ΔG_{solv}. Depending on the prediction task, each model takes in one or two SMILES strings as inputs and outputs the corresponding thermochemical value. We chose to implement a graph neural network architecture for each model, as they are highly nonlinear models that have been shown to accurately relate various chemical properties to their molecular structure. Briefly, GNNs represent molecules as a graph with atoms as nodes and bonds as edges, and can importantly capture the connectivity of atoms within a molecule. They do however miss out on discrete 3-dimensional information that can be important for the prediction of certain quantities, such as ΔS^o where conformational information can greatly affect values. In these models, each atom and bond are assigned a given set of features that are convolved multiple times through message passing steps that use parameters specific to the training data, resulting in unique molecular representations for each prediction task.

Machine Learning Models for Solvation

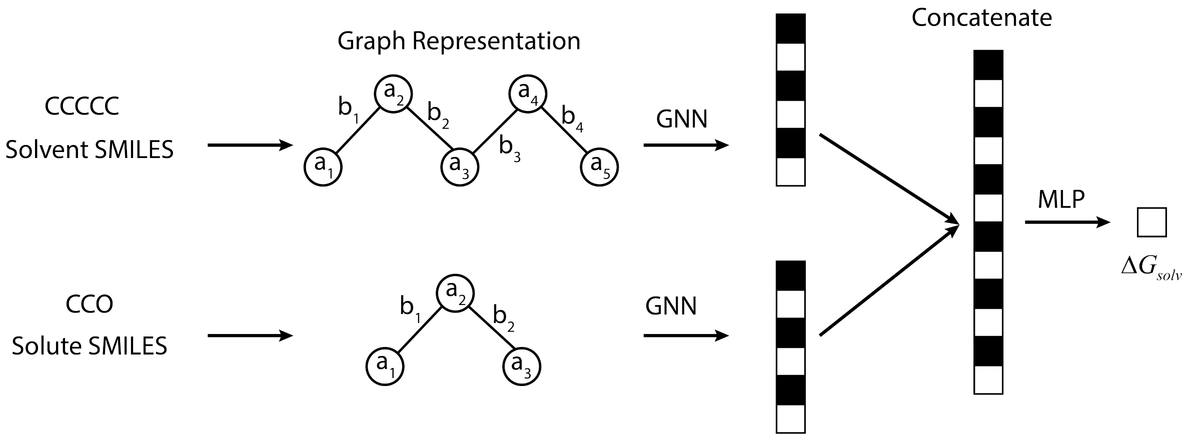


Figure 4-3: Architecture of our model to predict solvation free energies. SMILES strings of the solute and solvent are transformed into graph representation where atom and bond features are assigned. Convolutions are performed with a GNN to form embedding that are concatenated and then passed through an MLP for property prediction.

The specific architecture we used to build our model to predict solvation free energies can be summarized in Figure 4-3. We used a state-of-the-art directed message passing neural network implemented with Chemprop software, build on top of Pytorch7. The solvent SMILES and solute SMILES are each passed through their own distinct GNN to form a graph representation using the open source cheminformatics software RDkit. Atom and bond feature vectors are built for each molecule. Chemprop by default only uses atom and bond features and does not add molecular level features to the representation after convolution. We used the Chemprop default atom and bond features. Specifically, the atom features are (i) the atomic number, (ii) the number of directly bonded neighbors, (iii) the formal charge, (iv) the chiral tag, (v) the number of connected H's, (vi) and the hybridization. These are all one hot encoded, resulting in a vector length of 130. The bond features are, (i) the bond type (single, double, triple), (ii) the aromaticity, (iii) the conjugation, (iv) and whether the atom is in a ring. The atomic and bond feature vectors are passed through MPNNs to form distinct latent molecular representations for both the solvent and solute. In particular, for each GNN/MPN we implemented 4 rounds of convolutions and each message passing MLP for the convolution steps consisted of a single hidden layer with 300 nodes. After performing all the convolutions, the atom features were aggregated into a molecular vector representation by taking the mean of all the atom vectors. These vectors are then concatenated into a single vector embedding that can be passed into a traditional MLP readout architecture that relates the vector representation to the solvation free energy. The final dimension of this embedding was 600. This concatenated vector was then fed into a readout MLP that consisted of two hidden layers of size 300, with the final output being the solvation free energy.

We used a random split of size 70:10:20 to form the training, validation and test set respectively. When training our ΔG_{solv} model on this type of split, we chose to train 10 models in total with 10 different random splits in order to perform a fair evaluation. The models were trained using mean squared error as the loss function with and Adam optimizer. The models were trained for 100 epochs with a batch size of 50 and an initial learning rate of 10^{-3} which was coupled with a patience of 5 to reduce the learning rate by a factor of 2 if no improvement was observed on the validation set over 5 epochs. Hyperparameters were chosen based on similar models[63]. We chose not to optimize our hyperparameters grid search or other more sophisticated methods due to constraints on the time required to train our models.

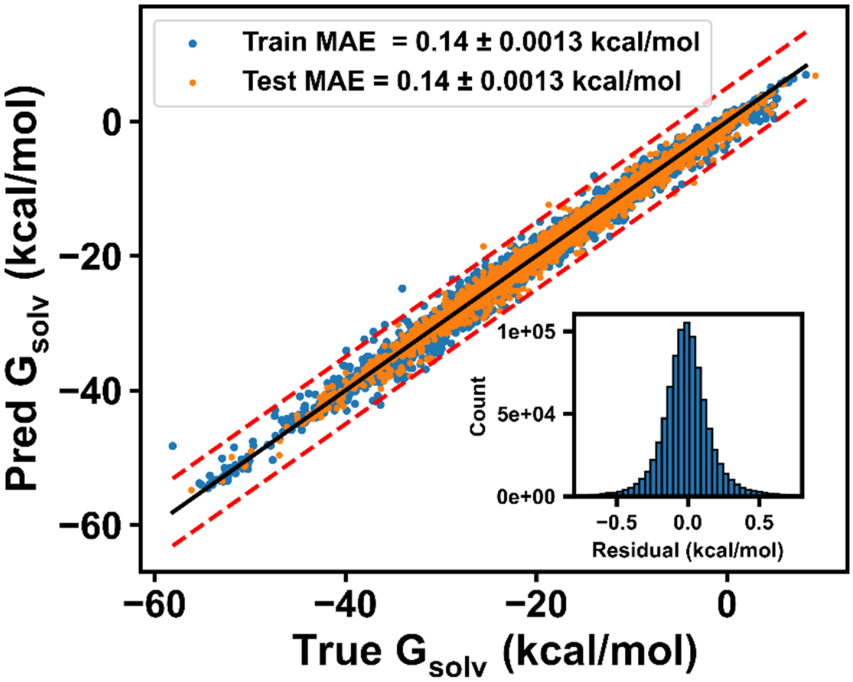


Figure 4-4: Parity plot of GNN used to fit ΔG_{solv} .RMSE in kcal/mol.

Model	MAE (kcal/mol)	RMSE (kcal/mol)
ΔG_{solv}	0.140 ±0.013	0.216 ± 0.0

Table 4.3: Performance metrics of solvation free energy GNN models.

We find that for we are able to predict ΔG of with an average MAE of 0.071 kcal/mol (Table 4.3) which is on par with some of the best ML models reported in the literature [63]. The MAE for the training and test sets in this case are of equal value, suggesting that our model has good generality and will have trustable predictive power when used on the E^o dataset. Using equation 4.1 the MAE would correspond to an error of 8.5 mV for n=1 and 4.25 mV for n=2, although this is likely just a lower bound on the error as the molecules in the E^o differ in structure from the G_{solv} dataset.

The main caution to consider for this model is that because each data point consists of a solvent and solute pair, a random split of the train and test set may result in an overly optimistic model. This is because there is some information leakage between the training and test sets, as there could be solvents in the training set that also exist in the test set, although each solute-solvent pair is unique.

Machine Learning Models for ΔH_f and ΔS^o

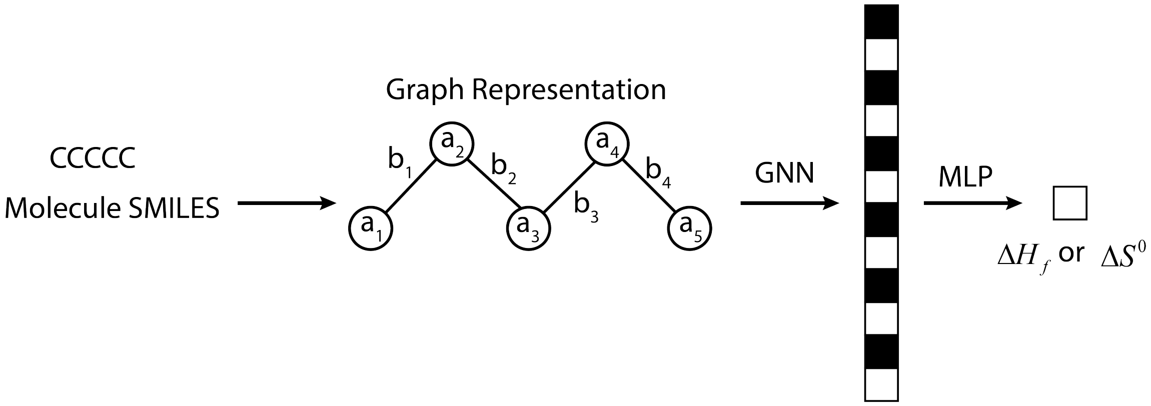


Figure 4-5: Architecture of our model to predict solvation free energies. SMILES strings of the solute and solvent are transformed into graph representation where atom and bond features are assigned. Convolutions are performed with a GNN to form embeddings that are concatenated and then passed through an MLP for property prediction.

The models used to predict ΔH_f and ΔS^o were very similar to the one used to predict, ΔG_{solv} with the main difference being that for each model there is only one molecular input instead of two. For each GNN/MPN we implemented 4 rounds of convolutions and each message passing MLP for the convolution steps consisted of a single hidden layer with 300 nodes. The convolutions incorporated both bond and atom features (the same features as the ΔG_{solv} model). Additionally, after performing all the convolutions, the atom features were aggregated into a molecular vector representation by taking the mean of all the atom vectors. The final dimension of this vector was 600. This vector was then fed into a readout MLP that consisted of two hidden layers of size 300, with the final output being either the enthalpy of formation or the standard entropy.

Model	**MAE**	**RMSE**
ΔH_f (kcal/mol)	0.450 ±0.026	0.604 ± 0.025
ΔS^o (cal/molK)	0.297 ± 0.031	0.404± 0.025

Table 4.4: Results of GNN models to predict gas phase thermochemistry. Error bars are the standard deviation of 10 random splits and trained models.

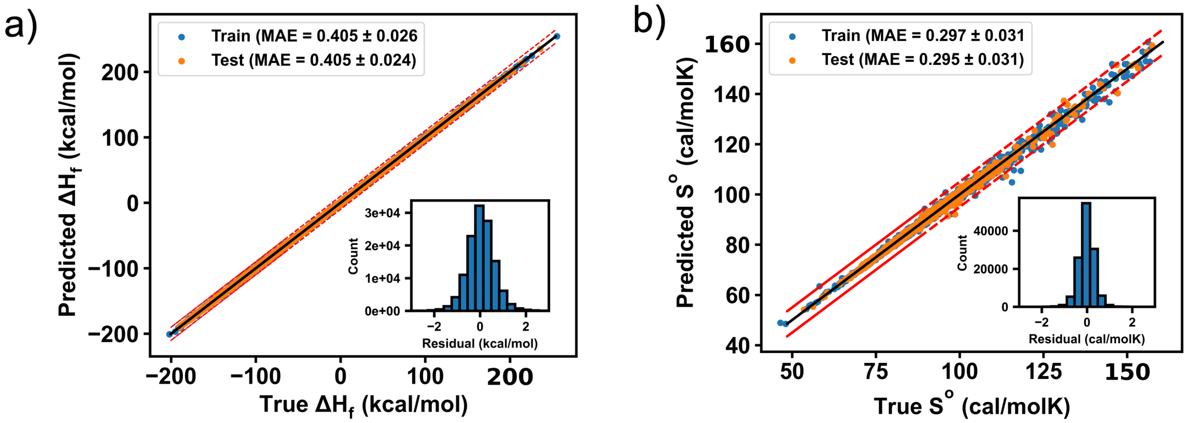


Figure 4-6: Parity plot of GNN used to fit ΔH_f (a) and S^o (b).

The models were each trained three times using random 70:10:20 splits of the corresponding datasets using the mean squared error as the loss function with an Adam optimizer. The models were trained for 150–200 epochs with a batch size of 50 and an initial learning rate of 10^{-3} which was coupled with a patience of 5 to reduce the learning rate by a factor of 2 if no improvement was observed on the validation set over 5 epochs.

The results of the ΔH_fmodel and S^omodels are summarized in Table 4.4. For ΔH_f, the MAE of 0.195 kcal/mol corresponds to a redox potential error of 19.5 mV for n=1 and 9.75 mV for n=2. On the other hand, for $S^o$the MAE of 0.297 cal/molK corresponds to an error of 3.8 mV for n=1 and 1.9 mV for n=2. Fitting S^oproved to be more difficult, this is likely because atom connectivity is likely not the most dominant factor in mapping molecules to their entropies as conformational information is a crucial factor is determining entropy, but no 3D positions were given for the ML model. Overall, the result of the 3 separate models for ΔG_{solv}proved accurate enough to proceed with verifying whether they could be used to obtain values of E^o.

4.4.2 Prediction of Redox Potential

Performance of DFT and SMD

E^o was obtained from the approaches in Figure 4-2, where molecular thermochemistry was related to E^o using equations 4.1, 4.3, and 4.2. The results of all approaches at the B3LYP/6-

31(2df,p) level of theory for calculating the redox potential are summarized in Table 4.5. The other levels of theory investigated in this work can be found in the supporting information.

Approach	**MAE(mV)**	**$\mathbf{MAE_n}$(mV)**	**RMSE(mV)**	**$\mathbf{RMSE_n}$(mV)**
DFT-SMD	216	294 (n=1)	229	308 (n=1)
		206 (n=2)		217 (n=2)
DFT-ML	137	234 (n=1)	156	247 (n=1)
		125 (n=2)		141 (n=2)
ML-ML	285	1037 (n=1)	428	1039 (n=1)
		189 (n=2)		264 (n=2)

Table 4.5: Performance of different approaches depending on the number of electrons passed during the reaction (n). DFT calculations are at the **B3LYP/6-31(2df,p) level of theory.**

In order to set a baseline for the performance of approaches 2 and 3 to calculating redox potentials using DFT, we first investigated the results of using DFT at various levels of theory complemented with SMD implicit solvation. In the literature, B3LYP is one of the most commonly used functional for articles that report calculating redox potential, often accompanied by some type of Pople basis set [10, 44, 33]. However, to test the importance of functional choice, we utilized both the B3LYP and M062X functionals, along with the 6-31G(2df,p) (same as QM9 dataset) and 6-31++G** basis sets. We find that using M062X over B3LYP resulted in a 250 mV improved accuracy on average. The greatest increase in accuracy came when n=1, implying that M062X is able to more accurately model radical thermochemistry. The accuracy of the B3LYP/6-31(2df,p) calculation are of note since they are the only ones directly comparable to ML models for $\Delta H_f$and S^o. In this case, the DFT-SMD approach achieves an MAE of 216 mV, with the n=2 redox reactions having an 88 mV lower error.

Approach 2 DFT + ML

Results of Approach 2 (DFT+ML) are shown in Figure 4-7a,b and Table 4.5. They indicate that using the B3LYP/6-31(2df,p) level of theory for gas phase thermochemistry and an ML model to predict solvation free energy, we recover an accuracy of 137 mV for the 75 calculated experimental data points. Including solvation results in an improved accuracy of approximately 90 mV when compared to redox potentials calculated from gas phase DFT calculations, demonstrating how crucial information about the solvation environment is to accurate predictions. The results of using the B3LYP/6-31++G**, M062X/6-31(2df,p), and M062X/6-31++G** levels of theory can be found in the supporting information. When using M062X, we observe that the accuracy that approaches 1 and 2 are on par with each other. Specifically, approach 2 has a 37 mV higher MAE at the M062X/6-31++G** level of theory and a 10 mV higher MAE at the M062X/6-31(2df,p) level of theory than approach 1.

We note that the ML model for ΔG_{solv} is trained on COSMO-RS calculations, which is known to be one of the most accurate methods for obtaining solvation free energies and in general should give better result than using a technique such as SMD. This would imply that in principle the approach 2 should give better results than approach 1, but this is only observed when using B3LYP as the functional. From approach 1 we learned that M062X gives more accurate results overall and when compared to B3LYP, but both functional have their own systematic errors. Error cancelation is a concept that can lead to more accurate results for a given technique, even though the way the physics is modeled is not superior in any way. So in this case, it is possible that for M062X, there is fortuitous error cancellation with the SMD solvation results that give more accurate MAEs when compared to using the COSMO-RS trained ΔG_{solv} ML model. As a result, even though the COSMO-RS results could give more accurate solvation free energies, the overall value for E^o has a higher error. Additionally, the ML models have their own error that is compounded onto the error inherent to COSMO-RS, so it is possible the assumption that approach 2 should perform better is flawed.

Approach 3 ML + ML

Approach 3, which involves replacing both the gas phase DFT and the SMD solvation calculations used in Approach 1 with ML models, shows poorer performance overall. This approach achieves an overall MAE of 285 mV which is 148 mV higher than approach 2 and 69 mV higher than approach 1. When considering accuracy as a function of n we see that approach 3 is in par with approaches 1 and 3 when n=2 (Table 4.6), but when n=1 approach

3 is not predictive at all with an MAE over 1V. This indicates that the error in ML models that predict gas phase thermochemistry is still too large to accurately predict redox potentials in general, but if one is only considering 2 electron transfer PCETs then approach 3 is usable for initial screening. While this class of redox reaction seems limiting, it applies to a large quantity of redox couple candidates for applications such as redox flow batteries. The reason approach 3 does not preform well for n=1 PCETs is likely because 1 electron processes produce radicals, meaning the only difference between the reduced and oxidized species is the loss of a hydrogen atom ($H^\bullet$) and since our GNN ML models only consider the molecular graph of the heavy atoms on the molecules it may not be able to accurately differentiate the energies of the reduced and oxidized species. This effect is diminished when n=2 because we effectively have a C−OH group becoming a C=O functional group, which is a bigger difference and is easier for the ML model to capture.

As previously mentioned, we observed greater accuracy with the Minnesota functionals, but comparing approach 3 to these calculations would not be a fair comparison since the ML models were trained with data at the B3LYP/6-31G(2df,p) level of theory. With this said, the disparity between 1 and 2 electron reactions can be mitigated through the proper choice of functionals for training data or by including hydrogen atoms in the molecular graphs of the GNNs used for ΔH_f and S^o. It is noteworthy that the machine learning models were all trained on a diverse set of molecules that don't include the those in the E^o data set, and even so we are still able to obtain reasonable accuracy. This speaks to the generalizability of our ML models and implies that they are learning reasonable representations of molecules for the purpose of predicting redox potentials.

Solvent Dependent Performance

We attempted to determine how the accuracy of our approached varied based on the solvent. Our E^o data set contained 5 different solvents, but the data density within each solvent varied drastically at times. Water, MeCN and THF are represented in relatively equal amounts, but DMF and IPA only have one data point, making conclusions drawn for these two solvents tough to extrapolate. Focusing on water, THF, and MeCN we see that at the B3LYP/(2df,p) level of theory there is no significant difference between the solvents for approach

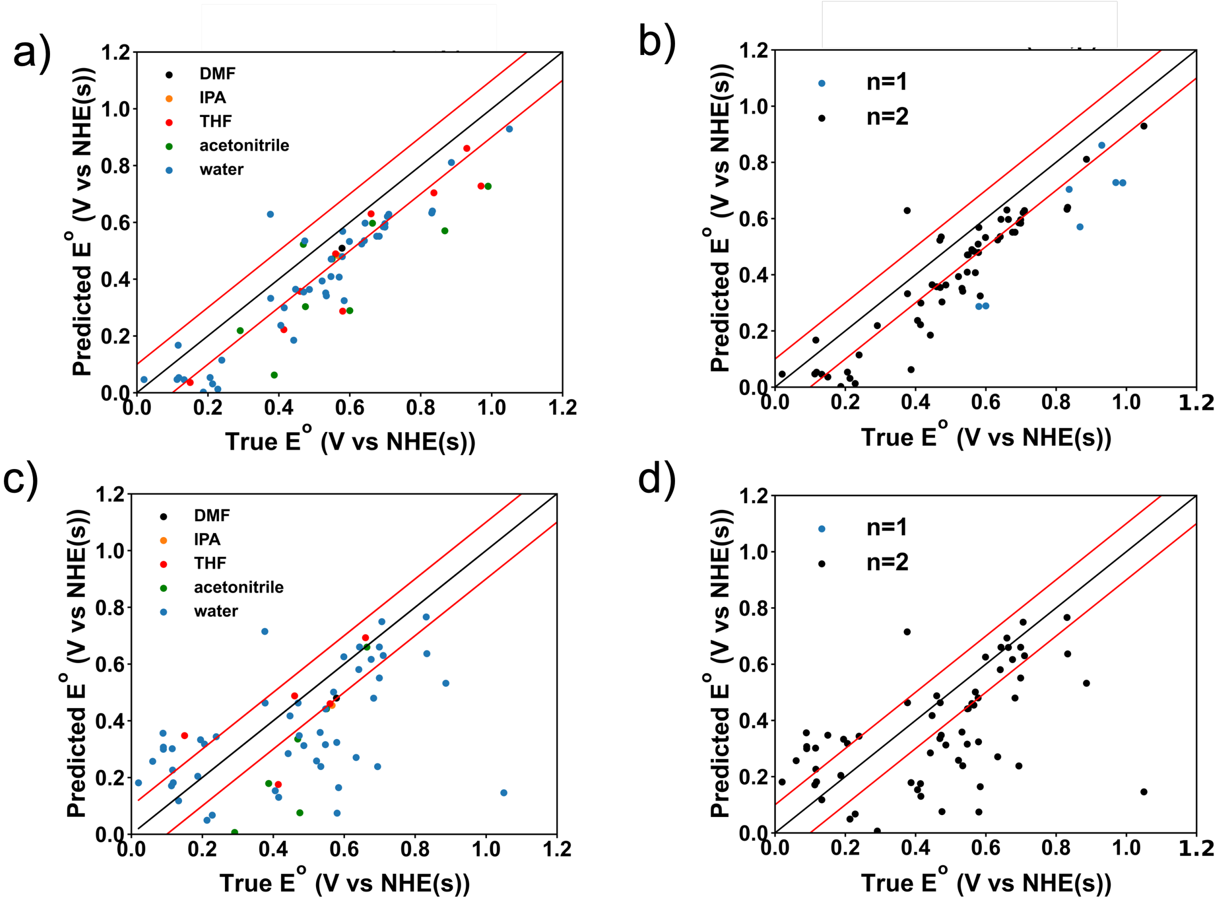


Figure 4-7: a,c: Parity plots of approach 2 where gas phase free energies are predicted using DFT and solvation free energies are predicted using a machine learning model. DFT calculations are at the **B3LYP/6-31(2df,p)** level of theory. b,d: Parity plots of approach 2 where gas phase free energies are predicted using an ML model and solvation free energies are predicted using a machine learning model.

1 and a max difference of 60 mV between the three solvents for approach 2 (Table 4.6. It appears as though for approach 3, THF and MeCN perform significantly worse, but this is likely convoluted with the poor performance of approach 3 on n=1 PCETs. In the E^o dataset when n=1 the solvent is either MeCN or THF, for water n is always 2, when we compare the n=2 case for THF and MeCN we see that we recover similar accuracy across the three solvents. The roughly constant accuracy among these three solvents could point to the robustness of the ΔG_{solv} ML model that we built and its ability to generalize across different solvent. However, we should note that the redox potentials in the E^odataset are referenced to V vs NHE which is given by equation (**4.11**), this equation indicates that the

species required to calculate V vs NHE are the reduced and oxidized forms of the redox couple along with H_2 (potentially radicals in the case of n=1). It's likely that the difference between in ΔG_{solv} between the reduced and oxidized species is not large, meaning the contribution of ΔG_{solv} to E^oeffectively cancels out.

Solvent	MAE			RMSE		
	DFT + SMD (mV)	**DFT + ML (mV)**	**ML + ML (mV)**	**DFT + SMD (mV)**	**DFT + ML (mV)**	**ML + ML (mV)**
DMF	194	69	97	194	69	97
IPA	183	83	111	183	83	111
THF	233	139	515	246	161	683
MeCN	235	190	539	262	218	697
Water	211	129	200	220	143	281

Table 4.6: Performance of different approaches on specific solvents. DFT calculations are at the **B3LYP/6-31(2dfp) level of theory.**

4.5 Conclusion

We have presented two new approaches to incorporate ML into the calculation of organic PCET redox potentials. First, we investigated replacing solvation free energy calculations with graph based ML models instead of using implicit solvation techniques. Next, we determined if the gas phase DFT calculations could be replaced by ML models for $\Delta H_f$and S^oin approach 3. In general, we found that replacing using ML models to predict the solvation free energy recovers similar accuracy to the standard DFT + ML regardless of the solvent, but the accuracy is improved when considering only PCETs where n=2. Furthermore, when using a pure machine learning based approach, the performance recovers accuracy on par to approach 1 only when n=2, and we also observe that there is little solvent dependence. In this case, when n=1 this approach cannot be used with predictive power. To address these issues, it's likely that new ML models would have to be trained with data sets that include species such as organic radicals both in the gas phase and in various solvents. We expect these validated approaches can be used in the future to support high throughput screening efforts of electroactive materials.

4.6 Supporting Information

4.7 Figures

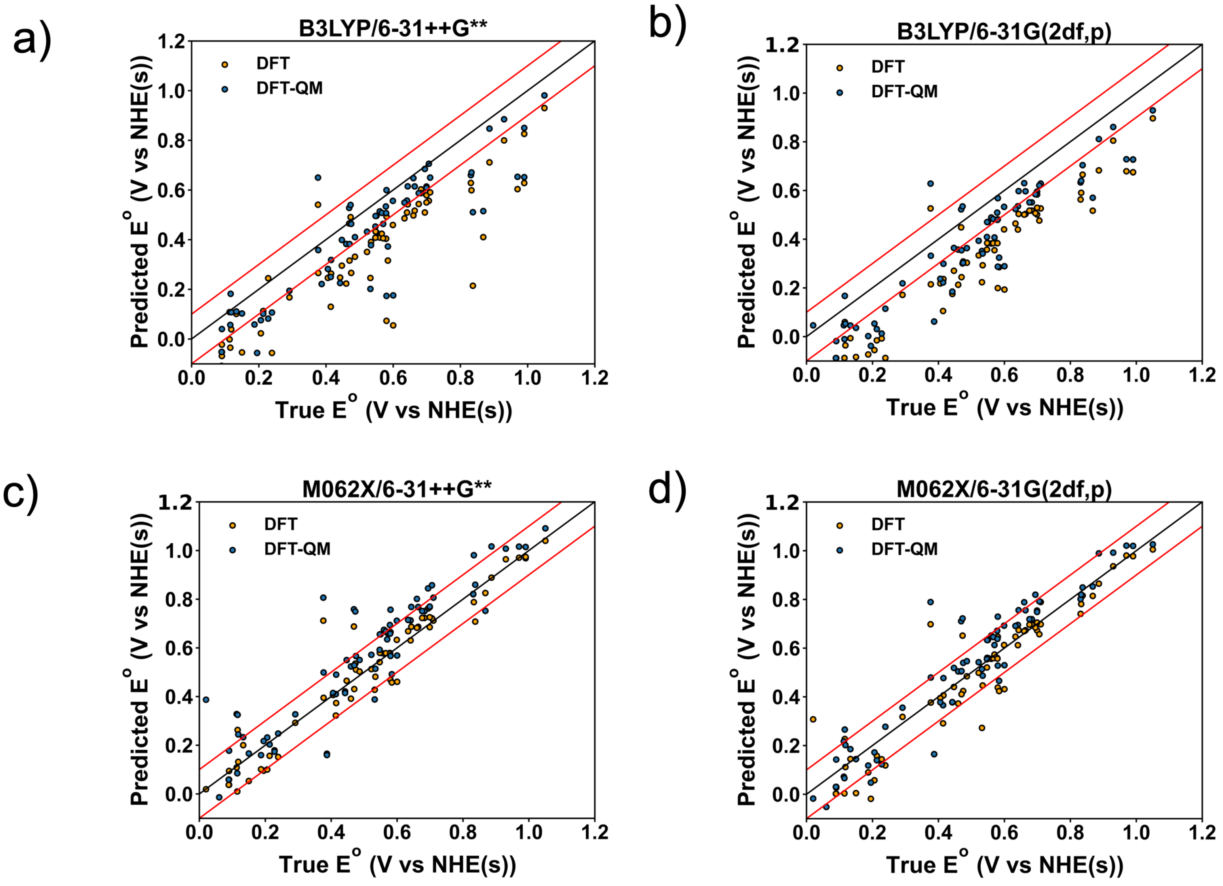


Figure 4-8: Comparison of performance of the DFT and DFT + ML approaches. Results for all possible combinations of the functionals B3LYP and M062X and the basis sets 6-31++G** and 6-31G(2df,p) are shown.

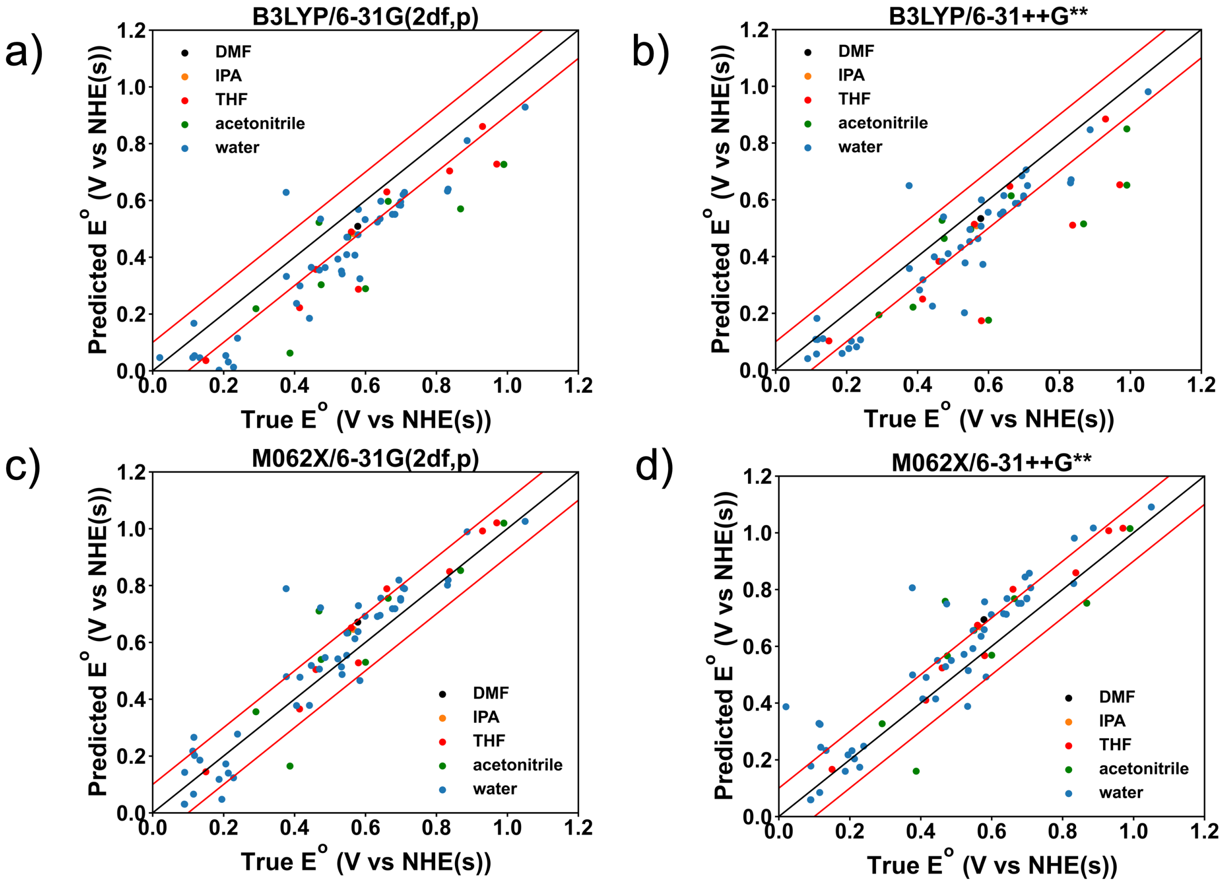


Figure 4-9: Comparison of performance of the DFT + ML approach (approach 1) based on solvent. Results for all possible combinations of the functionals B3LYP and M062X and the basis sets 6-31++G** and 6-31G(2df,p) are shown.

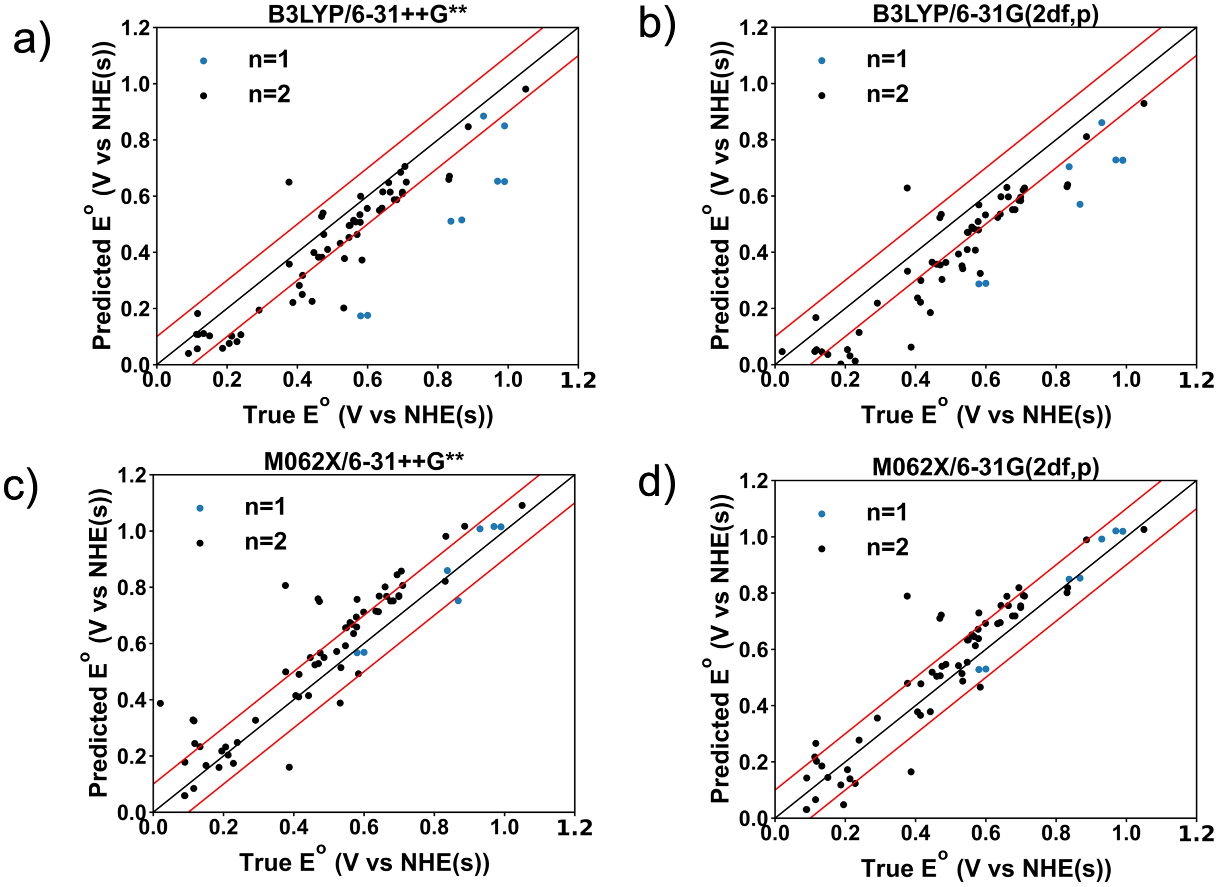
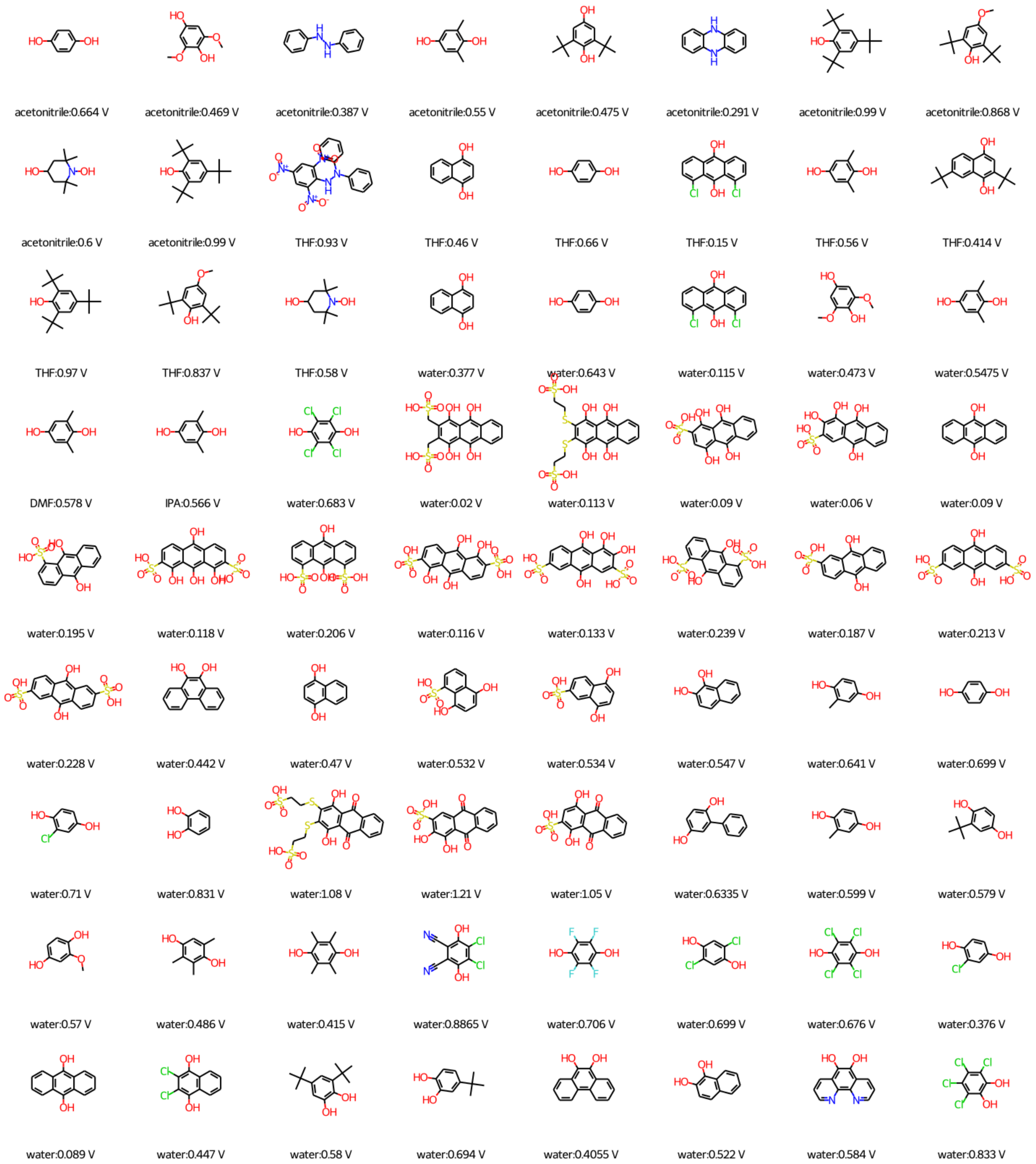


Figure 4-10: Comparison of performance of the DFT + ML approach (approach 1) based on the number of electrons transferred during the reaction (n). Results for all possible combinations of the functionals B3LYP and M062X and the basis sets 6-31++G** and 6-31G(2df,p) are shown.



acetonitrile:0.664 V
acetonitrile:0.469 V
acetonitrile:0.387 V
acetonitrile:0.55 V
acetonitrile:0.475 V
acetonitrile:0.291 V
acetonitrile:0.99 V
acetonitrile:0.868 V

acetonitrile:0.6 V
acetonitrile:0.99 V
THF:0.93 V
THF:0.46 V
THF:0.66 V
THF:0.15 V
THF:0.56 V
THF:0.414 V

THF:0.97 V
THF:0.837 V
THF:0.58 V
water:0.377 V
water:0.643 V
water:0.115 V
water:0.473 V
water:0.5475 V

DMF:0.578 V
IPA:0.566 V
water:0.683 V
water:0.02 V
water:0.113 V
water:0.09 V
water:0.06 V
water:0.09 V

water:0.195 V
water:0.118 V
water:0.206 V
water:0.116 V
water:0.133 V
water:0.239 V
water:0.187 V
water:0.213 V

water:0.228 V
water:0.442 V
water:0.47 V
water:0.532 V
water:0.534 V
water:0.547 V
water:0.641 V
water:0.699 V

water:0.71 V
water:0.831 V
water:1.08 V
water:1.21 V
water:1.05 V
water:0.6335 V
water:0.599 V
water:0.579 V

water:0.57 V
water:0.486 V
water:0.415 V
water:0.8865 V
water:0.706 V
water:0.699 V
water:0.676 V
water:0.376 V

water:0.089 V
water:0.447 V
water:0.58 V
water:0.694 V
water:0.4055 V
water:0.522 V
water:0.584 V
water:0.833 V

Figure 4-11: All molecules used in the E^o dataset with the solvent and redox potential vs NHE indicated below.

4.8 Tables

Approach	**MAE(mV)**	**MAE$_n$(mV)**	**RMSE(mV)**	**RMSE$_n$(mV)**
DFT-SMD	411	394 (n=1)	1931	427 (n=1)
		413 (n=2)		2045(n=2)
DFT-ML	117	293 (n=1)	153	318 (n=1)
		95 (n=2)		117(n=2)
ML-ML	285	1037 (n=1)	428	1039 (n=1)
		189 (n=2)		264 (n=2)

Table 4.7: Performance of different approaches depending on the number of electrons passed during the reaction (n). MAE= mean average error. MAE$_n$ = mean average error for a specific n value. RMSE = root mean squared error.DFT calculations are at the **B3LYP/6-31++G**** level of theory.

Approach	**MAE(mV)**	**MAE$_n$(mV)**	**RMSE(mV)**	**RMSE$_n$(mV)**
DFT-SMD	55	61 (n=1)	82	80 (n=1)
		55 (n=2)		83 (n=2)
DFT-ML	92	44 (n=1)	122	55 (n=1)
		98 (n=2)		128 (n=2)
ML-ML	285	1037 (n=1)	428	1039 (n=1)
		189 (n=2)		264 (n=2)

Table 4.8: Performance of different approaches depending on the number of electrons passed during the reaction (n). MAE= mean average error. MAE$_n$ = mean average error for a specific n value. RMSE = root mean squared error.DFT calculations are at the **M062X/6-31++G**** level of theory.

Approach	**MAE(mV)**	**MAE$_n$(mV)**	**RMSE(mV)**	**RMSE$_n$(mV)**
DFT-SMD	66	53 (n=1)	100	80 (n=1)
		68 (n=2)		102 (n=2)
DFT-ML	76	40 (n=1)	99	45 (n=1)
		81 (n=2)		104 (n=2)
ML-ML	285	1037 (n=1)	428	1039 (n=1)
		189 (n=2)		264 (n=2)

Table 4.9: Performance of different approaches depending on the number of electrons passed during the reaction (n). MAE= mean average error. MAE$_n$ = mean average error for a specific n value. RMSE = root mean squared error. DFT calculations are at the **M062X/6-31(2dfp)** level of theory.

Solvent	MAE			RMSE		
	DFT + SMD (mV)	**DFT + ML (mV)**	**ML + ML (mV)**	**DFT + SMD (mV)**	**DFT + ML (mV)**	**ML + ML (mV)**
DMF	174	44	97	174	44	97
IPA	142	58	111	142	59	111
THF	296	160	515	337	213	683
MeCN	229	169	539	280	220	697
Water	478	102	200	2293	124	281

Table 4.10: Performance of different approaches on specific solvents. DFT calculations are at the **B3LYP/6-31++G****level of theory.

Solvent	MAE			RMSE		
	DFT + SMD (mV)	**DFT + ML (mV)**	**ML + ML (mV)**	**DFT + SMD (mV)**	**DFT + ML (mV)**	**ML + ML (mV)**
DMF	1.3	116	97	1.3	116	97
IPA	12	102	111	12	102	111
THF	63	55	515	77	72	683
MeCN	75	105	539	110	135	697
Water	53	96	200	79	127	281

Table 4.11: Performance of different approaches on specific solvents. DFT calculations are at the **M062X/6-31++G****level of theory.

Solvent	MAE			RMSE		
	DFT + SMD (mV)	**DFT + ML (mV)**	**ML + ML (mV)**	**DFT + SMD (mV)**	**DFT + ML (mV)**	**ML + ML (mV)**
DMF	18	93	97	18	18	97
IPA	7.1	80	111	7.1	7.1	111
THF	61	55	515	85	65	683
MeCN	60	91	539	92	117	697
Water	72	77	200	106	100	281

Table 4.12: Performance of different approaches on specific solvents. DFT calculations are at the **M062X/6-31G(2df,p)**level of theory.

Chapter 5

Conclusions and Future Directions

5.1 Conclusions

5.2 Future Directions

This thesis work broadly aimed to address the challenges and complexities of increasing plastic circularity. It is likely that many different technologies will need to be developed in order to achieve this, and that fact is reflected in the diversity of tools and techniques studied in this thesis. Specifically, we focused on 1) using electrochemistry to synthesize lactones, which have the potential to serve as circular monomers 2) using high throughput computational screening to design redox mediators that are specific to C-H bond energies which can potentially deconstruct polystyrene and 3) using ML models to accelerate the prediction of important properties of electroactive material which can aid in high throughput screening projects.

In the case of lactonization, we have investigated a method for using electricity derived from renewable sources to convert cyclohexanone into 5 and 6 membered substituted lactones with an FE of approximately 20% using Pt electrodes and water as the sole oxygen atom source. These products expand and complement the set of possible lactone products synthesized from cyclohexanone via the Baeyer-Villiger reaction. Electrochemical kinetic studies allowed us to propose a mechanism that involves the initial inner sphere activation of water to form reactive oxygen surface species that can then react with cyclohexanone.

With our studies of HAT redox mediator design, we demonstrated that we were able to build a library of 20,000 ylide redox mediators and calculate their relevant catalytic thermodynamic values and then match these quantities to different substrates to propose structures that are tailored to different bond energies. We then experimentally validated our proposed candidates for 1,3 diphenylpropane. Our calculations indicate that most ylide structures have more than enough energy to activate the bonds in 1,3 diphenylpropane with a much lower number of structures that for example can only activate the lowest energy C-H bond on 1,3 diphenylpropane. Our ability to identify these candidates apriori could allow us to access unique selectivities and FEs that would not be available by using a "random" ylide structure.

Lastly, upon completing the project in chapter 3 we found that the screening could be further accelerated using ML. We investigated two new approaches to incorporate ML into the calculation of organic PCET redox potentials. First, we investigated replacing solvation free energy calculations with graph based ML models instead of using implicit solvation techniques. Next, we determined if gas phase DFT calculations could be replaced by ML models for $\Delta H_f$and S^o. In general, we found that replacing using ML models to predict the solvation free energy recovers similar accuracy to the standard DFT + ML regardless of the solvent. Furthermore, when using a pure machine learning based approach, the performance recovers accuracy on par to approach 1 only when n=2, but when n=1 this approach cannot be used with any predictive power. We expect these validated approaches can be used in the future to support high throughput screening efforts of electroactive materials.

5.2.1 Lactonization

In chapter 2 we were able to demonstrate that we could electrochemically transform cyclic ketones into lactones, mimicking aspects of the Baeyer-Villiger reaction, but with products that are isomers of the Baeyer-Villiger reaction. Due to the polymerization thermodynamics of the BV products, they tend to be more desirable. For this reason, future directions of lactonization work include targeting the BV products, namely ϵ caprolactone. This will involve the same experimental set up but require new catalyst design. The approach will involve mimicking BV homogeneous and heterogeneous catalysts that use hydrogen peroxide as their

O atom source. The crucial reactive intermediate to mimic will be the Criegee intermediate that is proposed in the BV mechanism. This intermediate allows for the migration of a carbon on the starting ketone, resulting in the ring expanded lactone that is observed from the BV reaction. The new BV electrocatalysts will need to activate the ketone substrate in a Lewis acidic way and likely have to form a peroxo surface oxygen species. With oxidative electrochemistry, the -OOH group is tougher to observe on a catalyst surface (single oxygen -OH and =O species are more common), but it is a common intermediate during O_2 reduction, so future directions could attempt to synthesize lactones at the cathode using oxygen reduction intermediates.

5.2.2 vHTP screening of redox mediators for plastic deconstruction

The main future directions for projects 2 and 3 involve applying the tools that we built to actually polystyrene. Plastic deconstruction was the motivation of the work, but the analysis was done as a proof of concept with polystyrene model compounds such as 1,3, diphenylpropane. To apply our work to polystyrene deconstruction, we will need to calculate its properties and then screen ylide candidates based on the new polystyrene energies. These candidates will then need to be synthesized in the lab and tested for their capabilities for polystyrene deconstruction. It is likely that trying to use our designed HAT mediators for polystyrene deconstruction will result in a lot of practical issues such as solubility of the plastic and the synthesizability of the proposed mediators. For this reason, we will likely need to propose various ylide structures for each desired energy so that we have various options to choose from.

Upon completing these two projects, we also have developed a computational workflow that will make any screening in the future much faster. There are various other applications that would benefit from the computational screening performed in chapters 3 and 4. For example, we could expand our scope to consider molecules outside the domain of HAT redox mediators and consider redox mediators such as quinones, derivatives of NHPI, and TEMPOH, although in these cases we may run into issues with being able to synthesize any candidates that we are able to propose.

Bibliography

[1] Interactive supercomputing on 40,000 cores for machine learning and data analysis. In *2018 IEEE High Performance extreme Computing Conference (HPEC)}, 2018.*

[2] James Barker, Laura-Sophie Berg, Jan Hamaekers, and Astrid Maass. Rapid Prescreening of Organic Compounds for Redox Flow Batteries: A Graph Convolutional Network for Predicting Reaction Enthalpies from SMILES. Batteries & Supercaps*, pages 1–10, 2021.*

[3] Stacie S. Canan Koch and A. Richard Chamberlin. Modified conditions for efficient baeyer-villiger oxidation with m-CPBA. Synthetic Communications*, 19(5-6):829–833, 1989.*

[4] Ruochen Cao, Mei-Qi Zhang, Chaoquan Hu, Dequan Xiao, Meng Wang, and Ding Ma. Catalytic oxidation of polystyrene to aromatic oxygenates over a graphitic carbon nitride catalyst. Nature Communications*, 13(1):4809, 8 2022.*

[5] Lei Cheng, Rajeev S. Assary, Xiaohui Qu, Anubhav Jain, Shyue Ping Ong, Nav Nidhi Rajput, Kristin Persson, and Larry A. Curtiss. Accelerating Electrolyte Discovery for Energy Storage with High-Throughput Screening. Journal of Physical Chemistry Letters*, 6(2):283–291, 2015.*

[6] Geoffrey W Coates and Yutan Getzler. Chemical recycling to monomer for an ideal, circular polymer economy. Nature Reviews Materials*, 2020.*

[7] Connor W. Coley, Regina Barzilay, William H. Green, Tommi S. Jaakkola, and Klavs F. Jensen. Convolutional Embedding of Attributed Molecular Graphs for Physical Property Prediction. Journal of Chemical Information and Modeling*, 57(8):1757–1772, 2017.*

[8] Christopher J Cramer. Essentials of Computational Chemistry Theories and Models Second Edition. Technical report.

[9] Julia W. Darcy, Brian Koronkiewicz, Giovanny A. Parada, and James M. Mayer. A Continuum of Proton-Coupled Electron Transfer Reactivity. Accounts of Chemical Research*, 51(10):2391–2399, 10 2018.*

[10] Carlos De La Cruz, Antonio Molina, Nagaraj Patil, Edgar Ventosa, Rebeca Marcilla, and Andreas Mavrandonakis. New insights into phenazine-based organic redox flow batteries by using high-throughput DFT modelling. Sustainable Energy and Fuels*, 4(11):5513–5521, 2020.*

[11] *Marino Del, Todesco Frisone, Francesco Pinna, and Giorgio Strukul. Baeyer-Villiger Oxidation of Cyclic Ketones with Hydrogen Peroxide Catalyzed by Cationic Complexes of Platinum(II): Selectivity Properties and Mechanistic Studies.* Organometallics*, 12:148–156, 1993.*

[12] *M. Favaro, C. Valero-Vidal, J. Eichhorn, F. M. Toma, P. N. Ross, J. Yano, Z. Liu, and E. J. Crumlin. Elucidating the alkaline oxygen evolution reaction mechanism on platinum.* Journal of Materials Chemistry A*, 5(23):11634–11643, 2017.*

[13] *K. Fujimoto, N. Yamashita, Y. Tokuda, Y. Matsubara, H. Maekawa, T. Mizuno, and I. Nishiguchi. Unusual rearrangement by anodic oxidation of cycloalkanones in the presence of trifluoroacetic acid.* Electrochimica Acta*, 42(13-14):2265–2266, 1997.*

[14] *Mariacristina Gagliardi and Angelo Bifone. Ring-opening copolymerization thermodynamics and kinetics of γ-valerolactone/-caprolactone.* PLoS ONE*, 13(6), 2018.*

[15] *Bill Gates.* How to avoid a climate disaster: the solutions we have and the breakthroughs we need. *Random House Large Print, New York, first edition, 2021.*

[16] *Roland Geyer, Jenna R Jambeck, and Kara Lavender Law. Production, use, and fate of all plastics ever made. 2017.*

[17] *Siddharth Ghule, Sayan Bagchi, and Kumar Vanka. Machine Learning the Redox Potentials of Phenazine Derivatives: A Comparative Study on Molecular Features. 2021.*

[18] *Reid A. Gilsdorf, Matthew A. Nicki, and Eugene Y.X. Chen. High chemical recyclability of vinyl lactone acrylic bioplastics.* Polymer Chemistry*, 11(30):4942–4950, 8 2020.*

[19] *Colin A. Grambow, Yi Pei Li, and William H. Green. Accurate Thermochemistry with Small Data Sets: A Bond Additivity Correction and Transfer Learning Approach.* Journal of Physical Chemistry A*, 123(27):5826–5835, 2019.*

[20] *Evan J. Horn, Brandon Rosen, Yong Chen, Jiaze Tang, Ke Chen, Martin D. Eastgate, and Phil S. Baran. Scalable and Sustainable Electrochemical Allylic C–H Oxidation.* Physiology & behavior*, 176(5):139–148, 2017.*

[21] *Mioy T. Huynh, Colin W. Anson, Andrew C. Cavell, Shannon S. Stahl, and Sharon Hammes-Schiffer. Quinone 1 e- and 2 e-/2 H+ Reduction Potentials: Identification and Analysis of Deviations from Systematic Scaling Relationships.* Journal of the American Chemical Society*, 138(49):15903–15910, 2016.*

[22] *Edward L. Jeffery, Rajinder K. Mann, Graham J. Hutchings, Stuart H. Taylor, and David J. Willock. A density functional theory study of the adsorption of acetone to the (1 1 1) surface of Pt: Implications for hydrogenation catalysis.* Catalysis Today*, 105(1 SPEC. ISS.):85–92, 2005.*

[23] *Coralie Jehanno, Jill W. Alty, Martijn Roosen, Steven De Meester, Andrew P. Dove, Eugene Y.X. Chen, Frank A. Leibfarth, and Haritz Sardon. Critical advances and future opportunities in upcycling commodity polymers.* Nature*, 603(7903):803–814, 3 2022.*

[24] *Kyoungsuk Jin, Joseph Maalouf, Nik Lazouski, Nathan Corbin, Dengtao Yang, and Karthish Manthiram. Epoxidation of Cyclooctene Using Water as Oxygen-atom Source at Manganese Oxide Electrocatalysts.* Journal of the American Chemical Society*, 141:6413, 2019.*

[25] *Yu Kawamata, Ming Yan, Zhiqing Liu, Deng-Hui Bao, Jinshan Chen, Jeremy T Starr, and Phil S Baran. Scalable, Electrochemical Oxidation of Unactivated CH Bonds. 2017.*

[26] *George A. Kraus and Bruce Roth. Synthesis of Quinone Pyrano- γ-lactone Antibiotics. 1. Synthesis of 9-Deoxykalafungin.* Journal of Organic Chemistry*, 43(26):4923–4924, 1978.*

[27] *Y. L. Hu, Y. B. Xie, and D. J. Li. Efficient Baeyer-Villiger electro-oxidation of ketones with molecular oxygen using an activated carbon fiber electrode in ionic liquid [bmim][OTf].* Bulletin of the Chemical Society of Ethiopia*, 30(2):297, 2016.*

[28] *Marianne Labet and Wim Thielemans. Synthesis of polycaprolactone: A review.* Chemical Society Reviews*, 38(12):3484–3504, 2009.*

[29] *Arnold Lambert, J. A. Elings, Duncan J. Macquarrie, Graham Carr, and James H. Clark. The Baeyer-Villiger oxidation of ketones using HMS supported peroxycarboxylic acids.* Synlett*, (7):1052–1054, 2000.*

[30] *A M Lawson, F A J M Leemans, and James A Mccloskey. Oxygen-18 exchange reactions in steroidal ketones. Determination of relative rates of incorporation by gas chromatography-mass spectrometry.* Steroids*, page 603, 1969.*

[31] *Tengfei Li, Arjun Vijeta, Carla Casadevall, Alexander S. Gentleman, Tijmen Euser, and Erwin Reisner. Bridging Plastic Recycling and Organic Catalysis: Photocatalytic Deconstruction of Polystyrene via a C-H Oxidation Pathway.* ACS Catalysis*, 12(14):8155–8163, 7 2022.*

[32] *Marcelo Lopez-Segura, Fructuoso Barba, and Antonio Aldaz. Anodic oxidation of 2- and 3-methylcyclohexanones.* Electrochemica Acta*, 31:83–84, 1985.*

[33] *Alexandra R. McNeill, Samantha E. Bodman, Amy M. Burney, Chris D. Hughes, and Deborah L. Crittenden. Experimental validation of a computational screening approach to predict redox potentials for a diverse variety of redox-active organic molecules.* Journal of Physical Chemistry C*, 124(44):24105–24114, 2020.*

[34] *T. Patrick Montgomery, Abbas Hassan, Boyoung Y. Park, and Michael J. Krische. Enantioselective conversion of primary alcohols to α- Exo -methylene γ-butyrolactones via iridium-catalyzed C-C bond-forming transfer hydrogenation: 2-(Alkoxycarbonyl)allylation.* Journal of the American Chemical Society*, 134(27):11100–11103, 2012.*

[35] *Kenta Motobayashi, Líney Árnadóttir, Chikako Matsumoto, Eric M. Stuve, Hannes Jónsson, Yousoo Kim, and Maki Kawai. Adsorption of water dimer on platinum(111): Identification of the -OHPt hydrogen bond.* ACS Nano*, 8(11):11583–11590, 2014.*

[36] *Michael Musser. Cyclohexanol and Cyclohexanone.* Ullmanns Encyclopedia of Industrial chemistry*, 11:673–710, 2012.*

[37] *Scott R Nicholson, Julie E Rorrer, Avantika Singh, Mikhail O Konev, Nicholas A Rorrer, Alberta C Carpenter, Alan J Jacobsen, Yuriy Román-Leshkov, and Gregg T Beckham. The Critical Role of Process Analysis in Chemical Recycling and Upcycling of Waste Plastics. 2022.*

[38] *Kenley M. Pelzer, Lei Cheng, and Larry A. Curtiss. Effects of functional groups in redox-active organic molecules: A high-throughput screening approach.* Journal of Physical Chemistry C*, 121(1):237–245, 2017.*

[39] *Alessia Petti, Matthew C. Leech, Anthony D. Garcia, Iain C.A. Goodall, Adrian P. Dobbs, and Kevin Lam. Economical, Green, and Safe Route Towards Substituted Lactones by Anodic Generation of Oxycarbonyl Radicals.* Angewandte Chemie - International Edition*, 58(45):16115–16118, 2019.*

[40] *Mohammad Rafiee, Kelsey C. Miles, and Shannon S. Stahl. Electrocatalytic Alcohol Oxidation with TEMPO and Bicyclic Nitroxyl Derivatives: Driving Force Trumps Steric Effects.* Journal of the American Chemical Society*, 137(46):14751–14757, 2015.*

[41] *Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules.* Scientific Data*, 1, 8 2014.*

[42] *Tobias Reier, Mehtap Oezaslan, and Peter Strasser. Electrocatalytic oxygen evolution reaction (OER) on Ru, Ir, and pt catalysts: A comparative study of nanoparticles and bulk materials.* ACS Catalysis*, 2(8):1765–1772, 2012.*

[43] *Michael Renz and Bernard Meunier. MICROREVIEW 100 Years of Baeyer–Villiger Oxidations.* European Journal of Organic Chemistry*, 1999(4):737–750, 1999.*

[44] *Hudson G. Roth, Nathan A. Romero, and David A. Nicewicz. Experimental and Calculated Electrochemical Potentials of Common Organic Molecules for Applications to Single-Electron Redox Chemistry.* Synlett*, 27(5):714–723, 2016.*

[45] *Lars Ruddigkeit, Ruud Van Deursen, Lorenz C. Blum, and Jean Louis Reymond. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17.* Journal of Chemical Information and Modeling*, 52(11):2864–2875, 11 2012.*

[46] *Masato Saito, Yu Kawamata, Michael Meanwell, Rafael Navratil, Debora Chiodi, Ethan Carlson, Pengfei Hu, Longrui Chen, Sagar Udyavara, Cian Kingston, Mayank Tanwar, Sameer Tyagi, Bruce P. McKillican, Moses G. Gichinga, Michael A. Schmidt, Martin D. Eastgate, Massimiliano Lamberto, Chi He, Tianhua Tang, Christian A. Malapit, Matthew S. Sigman, Shelley D. Minteer, Matthew Neurock, and Phil S. Baran. N-Ammonium Ylide Mediators for Electrochemical C-H Oxidation.* Journal of the American Chemical Society*, 143(20):7859–7867, 2021.*

[47] *Deborah K. Schneiderman and Marc A. Hillmyer. Aliphatic Polyester Block Polymer Design.* Macromolecules*, 49(7):2419–2428, 4 2016.*

[48] *Changxia Shi, Ryan W. Clarke, Michael L. McGraw, and Eugene Y.-X. Chen. Closing the "One Monomer–Two Polymers–One Monomer" Loop via Orthogonal (De)polymerization of a Lactone/Olefin Hybrid.* Journal of the American Chemical Society*, 144(5):2264–2275, 2 2022.*

[49] *Changxia Shi, Zi Chen Li, Lucia Caporaso, Luigi Cavallo, Laura Falivene, and Eugene Y.X. Chen. Hybrid monomer design for unifying conflicting polymerizability, recyclability, and performance properties.* Chem*, 7(3):670–685, 3 2021.*

[50] *Changxia Shi, Liam T. Reilly, V. Sai Phani Kumar, Matthew W. Coile, Scott R. Nicholson, Linda J. Broadbelt, Gregg T. Beckham, and Eugene Y.X. Chen. Design principles for intrinsically circular polymers with tunable properties, 11 2021.*

[51] *Piotr De Silva. A Computational Protocol Combining DFT and Cheminformatics. 2021.*

[52] *Dharmender Singh, Nisha Devi, Vipin Kumar, Chandi C. Malakar, Saloni Mehra, Sunita Rattan, Ravindra K. Rawal, and Virender Singh. Natural product inspired design and synthesis of β-carboline and γ-lactone based molecular hybrids.* Organic and Biomolecular Chemistry*, 14(34):8154–8166, 2016.*

[53] *V. A. Sirotkin, B. N. Solomonov, D. A. Faizullin, and V. D. Fedotov. IR spectroscopic study of the state of water in dioxane and acetonitrile: Relationship with the thermodynamic activity of water.* Journal of Structural Chemistry*, 41(6):997–1003, 2000.*

[54] *Brian H. Solis and Sharon Hammes-Schiffer. Proton-coupled electron transfer in molecular electrocatalysis: Theoretical methods and design principles.* Inorganic Chemistry*, 53(13):6427–6443, 2014.*

[55] *Elif Sorkun, Qi Zhang, Abhishek Khetan, Murat Cihan Sorkun, and Suleyman Er. RedDB , a computational database of electroactive molecules for redox flow batteries. (2):https://doi.org/10.26434/chemrxiv.14398067.v1., 2021.*

[56] *Peter C. St. John, Yanfei Guan, Yeonjoon Kim, Seonah Kim, and Robert S. Paton. Prediction of organic homolytic bond dissociation enthalpies at near chemical accuracy with sub-second computational cost.* Nature Communications*, 11(1), 12 2020.*

[57] *Felix Strieth-kalthoff, Frederik Sandfort, Marwin H S Segler, and Frank Glorius. Machine learning the ropes: principles, applications and directions in synthetic chemistry.* Chem Soc Rev*, 2020.*

[58] *Daniel P. Tabor, Rafael Gómez-Bombarelli, Liuchuan Tong, Roy G. Gordon, Michael J. Aziz, and Alán Aspuru-Guzik. Mapping the frontiers of quinone stability in aqueous media: Implications for organic aqueous redox flow batteries.* Journal of Materials Chemistry A*, 7(20):12833–12841, 2019.*

[59] *Xiaoyan Tang and Eugene Y.X. Chen. Toward Infinitely Recyclable Plastics Derived from Renewable Cyclic Esters, 2 2019.*

[60] *Xiang Zhang Tao, Jian Jun Dai, Jie Zhou, Jun Xu, and Hua Jian Xu. Electrochemical CO Bond Formation: Facile Access to Aromatic Lactones.* Chemistry - A European Journal*, 24(27):6932–6935, 2018.*

[61] *Ierasia Triandafillidi, Maroula G. Kokotou, and Christoforos G. Kokotos. Photocatalytic Synthesis of γ-Lactones from Alkenes: High-Resolution Mass Spectrometry as a Tool to Study Photoredox Reactions.* Organic Letters*, 20(1):36–39, 2018.*

[62] *Angelo Vargas, Thomas Bürgi, and Alfons Baiker. Adsorption of activated ketones on platinum and their reactivity to hydrogenation: A DFT study.* Journal of Catalysis*, 222(2):439–449, 2004.*

[63] *Florence H. Vermeire and William H. Green. Transfer learning for solvation free energies: From quantum chemistry to experiments.* Chemical Engineering Journal*, 418(March), 2021.*

[64] *Zhongyu Wan, Quan De Wang, and Jinhu Liang. Accurate prediction of standard enthalpy of formation based on semiempirical quantum chemistry methods with artificial neural network and molecular descriptors.* International Journal of Quantum Chemistry*, 121(2):1–16, 2021.*

[65] *Fei Wang and Shannon S. Stahl. Electrochemical Oxidation of Organic Molecules at Lower Overpotential: Accessing Broader Functional Group Compatibility with Electron-Proton Transfer Mediators.* Accounts of Chemical Research*, 53(3):561–574, 2020.*

[66] *Hao Wang, Sayed Youssef Sayed, Erik J. Luber, Brian C. Olsen, Shubham M. Shirurkar, Sankaranarayanan Venkatakrishnan, Ushula M. Tefashe, Anna K. Farquhar, Eugene S. Smotkin, Richard L. McCreery, and Jillian M. Buriak. Redox Flow Batteries: How to Determine Electrochemical Kinetic Parameters.* ACS Nano*, 14(3):2575–2584, 2020.*

[67] *Logan Ward, Naveen Dandu, Ben Blaiszik, Badri Narayanan, Rajeev S. Assary, Paul C. Redfern, Ian Foster, and Larry A. Curtiss. Graph-Based Approaches for Predicting Solvation Energy in Multiple Solvents: Open Datasets and Machine Learning Models.* The Journal of Physical Chemistry A*, 125(27):5990–5998, 2021.*

[68] *Catherine F. Wise, Rishi G. Agarwal, and James M. Mayer. Determining Proton-Coupled Standard Potentials and X-H Bond Dissociation Free Energies in Nonaqueous Solvents Using Open-Circuit Potential Measurements.* Journal of the American Chemical Society*, 142(24):10681–10691, 2020.*

[69] *Bing Yan, Changxia Shi, Gregg T. Beckham, Eugene Y.X. Chen, and Yuriy Román-Leshkov. Electrochemical Activation of CC Bonds through Mediated Hydrogen Atom Transfer Reactions.* ChemSusChem*, 15(6), 3 2022.*

[70] *Ming Yan, Yu Kawamata, and Phil S Baran. Synthetic Organic Electrochemical Methods Since 2000: On the Verge of a Renaissance. 2017.*

[71] *Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hua Gao, Angel Guzman-Perez, Timothy Hopper, Brian Kelley, Miriam Mathea, Andrew Palmer, Volker Settels, Tommi Jaakkola, Klavs Jensen, and Regina Barzilay. Analyzing Learned Molecular Representations for Property Prediction.* Journal of Chemical Information and Modeling*, 59(8):3370–3388, 2019.*

[72] *Qi Zhang, Abhishek Khetan, and Süleyman Er. A quantitative evaluation of computational methods to accelerate the study of alloxazine-derived electroactive compounds for energy storage.* Scientific Reports*, 11(1):1–11, 2021.*