
Development of RT-TDDFT for the Interaction with the Explicit Solvent and for Correct Description of Excitation Process

Kevin Koh

Publication Date

11-04-2020

License

This work is made available under a All Rights Reserved license and should only be used in accordance with that license.

Citation for this work (American Psychological Association 7th edition)

Koh, K. (2020). *Development of RT-TDDFT for the Interaction with the Explicit Solvent and for Correct Description of Excitation Process* (Version 1). University of Notre Dame. <https://doi.org/10.7274/79407w65g32>

This work was downloaded from CurateND, the University of Notre Dame's institutional repository.

For more information about this work, to report or an issue, or to preserve and share your original work, please contact the CurateND team for assistance at curate@nd.edu.

DEVELOPMENT OF RT-TDDFT FOR THE INTERACTION WITH THE
EXPLICIT SOLVENT AND FOR CORRECT DESCRIPTION OF EXCITATION
PROCESS

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Kevin J. Koh

Olaf Wiest , Co-Director

Paul Helquist , Co-Director

Graduate Program in Chemistry and Biochemistry

Notre Dame, Indiana

April 2020

This document is in the public domain.

DEVELOPMENT OF RT-TDDFT FOR THE INTERACTION WITH THE EXPLICIT SOLVENT AND FOR CORRECT DESCRIPTION OF EXCITATION PROCESS

Abstract

by

Kevin J. Koh

The electronic energy transfer processes in natural systems, such as the light harvesting pigment protein complexes, and semiconductors, such as photovoltaic devices, occur via electron dynamics in complex systems. Understanding the excited state dynamics is important in designing the biological or nanostructured materials. The computational simulation of electron dynamics in an accurate and affordable manner is crucial for making progress in this area. The time dependent open system self consistent field at second order (OSCF2) is a real-time time-dependent density functional theory (RT-TDDFT) method. The block orthogonalized partitioning method eliminates unphysical errors of the embedded mean field theory at no additional cost. This thesis focuses on the application of OSCF2 to the simulation of transient absorption spectroscopy. OSCF2 accurately reproduced transient absorption spectra and non-radiative relaxation rates of green fluorescent protein chromophore derivative 3 (GFP-3). The implementation of an embedded mean-field theory using block-orthogonalized partitioning to real-time time dependent density functional theory (RT-TDDFT) and its effect of accuracy and affordability are discussed. The block orthogonalized partitioning in RT-TDDFT accurately describes the interactions between the regions while maintaining computational affordability.

The method development for the excitation process for RT-TDDFT using the Tamm-Dancoff approximation is discussed. The effort to describe the excitation process in RT-TDDFT leads to partial success. Studies using neural networks on molecular dynamics using bullvalene and mechanistic studies on formation of oxindole derivative from squaraine dyes are discussed. The neural network force field was not successful to reproduce the conformer search through molecular dynamics simulations, presumably due to insufficient simulation times. The free energy barriers of ring opening step of squaraine ring are in line with the experimental observations. Lastly, the software implementations to PySCF, Q2MM, and automated CatVS are discussed.

For my family who encouraged and comforted me in the toughest time of my life.

CONTENTS

Figures	v
Tables	vii
Acknowledgments	viii
Chapter 1: Introduction	1
1.1 Time-Dependent Open Self-Consistent Field at Second Order (OSCF2)	1
1.2 Conclusion	3
Chapter 2: Development of Affordable Electronic Structure Method for Excited State Dynamics	4
2.1 Time-Dependent Open Self-Consistent Field at Second Order (OSCF2) Studies of Transient Absorption Spectroscopy	4
2.1.1 Introduction	4
2.1.2 Methods	5
2.1.3 Results and Discussion	7
2.1.4 Conclusions	10
2.2 Real-time Propagation Using a Block Orthogonalized Basis	13
2.2.1 Introduction	13
2.2.2 Methods	16
2.2.3 Results and Discussion	21
2.2.4 Conclusions	31
2.3 Real-time Time Dependent Tamm-Dancoff Approximation	33
2.3.1 Introduction	33
2.3.1.1 Time-Dependent Configuration-Interaction Singles	34
2.3.1.2 Tamm-Dancoff Approximation(TDA)	38
2.3.1.3 TDA Implementation to RT-TDHF	41
Chapter 3: Using Neural Networks in Chemistry Research and Education	52
3.1 Exploring Bullvalene Conformers Using Neural Network Force Field Molecular Dynamics	52
3.1.1 Introduction	52
3.1.2 Methods	54
3.1.3 Results and Discussion	55

3.1.4	Conclusion	57
3.2	Using Neural Networks to Enhance Learning in Undergraduate Organic Chemistry Courses	58
3.2.1	Introduction	58
3.2.2	Methods	58
3.2.3	Outline	59
3.2.4	Conclusion	61
Chapter 4:	Mechanistic Studies of the Reaction of Squaraine Dyes	62
4.1	Formation of Oxindole Derivatives from the Squaraine Dyes	62
4.1.1	Introduction	62
4.1.2	Methods	64
4.1.3	Result and Discussions	65
4.1.3.1	Population Analysis of Oxindole Product	65
4.1.3.2	Mechanistic Study	65
Chapter 5:	Code Implementation	75
5.1	RT-TDDFT to PySCF	75
5.2	AMBER Interface to Q2MM	78
5.3	Automated CatVS	80
Appendix A:	CUDA Implementation in C++	84
Appendix B:	Supplementary Information	89
Bibliography	90

FIGURES

2.1	Pump-probe simulation scheme in OSCF2	6
2.2	Structures of (a) pyrazole (b) (Z)-4-(biphenyl-4-ylmethylene)-1,2-diphenyl-1H-imidazol-5(4H)-one (GFP-3).	7
2.3	Zero-delay TA spectra of pyrazole	8
2.4	TA spectra of GFP-3 chromophore in acetonitrile	9
2.5	Absorption spectra of GFP-3 chromophore in acetonitrile	11
2.6	Excited state spectrum and zero-delay TA spectrum of H ₂ molecule	12
2.7	Block-Orthogonalized Manby-Miller Embedding Scheme.	17
2.8	Molecules used for benchmarking	20
2.9	Absorption spectra of one methanol with 3 water molecules	21
2.10	Dipole moments time propagation of one methanol and three water molecules	23
2.11	Electron population time propagation of one methanol and three water molecules	24
2.12	Entropy time propagation of one methanol and three water molecules	25
2.13	Speed performance of BOMME	26
2.14	Absorption spectra of benzene with water molecules	27
2.15	Absorption spectra of 3,4-Dimethylpyrazole.	28
2.16	Absorption spectra of all- <i>trans</i> -retinal	29
2.17	Absorption spectra of benzene with ten water molecules	30
2.18	Absorption spectra of a neutral GFP chromophore with ten water molecules	32
2.19	Time propagation of H ₂ molecule using TDCIS	37
2.20	Absorption spectrum of H ₂ from equation of motion under TDA	42
2.21	Time Propagation of H ₂ molecule under continuous wave at resonant frequency	43
2.22	Time Propagation of H ₂ molecule under continuous wave at resonant frequency using RT-TDHF	45

2.23	Time Propagation of H ₂ molecule under continuous wave at resonant frequency using RT-TDHF	46
2.24	Time Propagation of H ₂ molecule under continuous wave at resonant frequency using RT-TDHF	47
2.25	Time Propagation of H ₂ molecule under continuous wave at resonant frequency	49
2.26	Time Propagation of N ₂ molecule under continuous wave at resonant frequency	50
3.1	Structure of bullvalene	53
3.2	Structure of 3,4,6,10-tetramethyltricyclo(3.3.2.0 ^{2,8})deca-3,6,9-triene .	54
3.3	Simulation of 3,4,6,10-tetramethyltricyclo(3.3.2.0 ^{2,8})deca-3,6,9-triene at 2000K	55
3.4	Energy propagation of 3,4,6,10-tetramethyltricyclo(3.3.2.0 ^{2,8})deca-3,6,9-triene at 2000K	56
3.5	Website under development	60
4.1	General Structure of Squaraine	62
4.2	Nucleophilic attack of Squaraine Dyes	63
4.3	Addition of Phosphine Derivatives to Dianiline Squaraines	63
4.4	Formation of benzofuranone and oxindoles	63
4.5	Structure of Oxindole product	64
4.6	Electron potential surface	66
4.7	Mechanistic hypothesis	67
4.8	Result of Geometry optimization	68
4.9	Molecular Orbitals of Intermediate	69
4.10	Potential energy scan of bond	70
4.11	Revised mechanism	71
4.12	Energy calculated for proposed reaction	72
4.13	Orientations of aniline in product	73
5.1	CatVS Scheme	81
5.2	Automatic CatVS Scheme for Bisphosphine	82

TABLES

2.1	Position of the lowest excited state energy in Figure 2.17	31
-----	--	----

ACKNOWLEDGMENTS

First of all, I would like to appreciate Dr. John Parkhill for letting me start on this journey. He guided me in the earlier stages of my research in Notre Dame. I would like to thank Dr. Daniel Gezelter and Dr. Ken Kuno for being my committee and providing guidance. I would like to thank my former group members: Dr. Kun Yao and Dr. Jade Nguyen for their help and advice.

I wish to express my deepest gratitude to Professor Olaf Wiest and Professor Paul Helquist for accepting me into the new group and inspiring me in both research and teaching. Dr. Wiest expanded my horizon of research and made me grow as a computational chemist in new directions. The experiences in teaching Dr. Helquist shared with me helped me pursuing my career in a teaching faculty position. Without their help, I would have been lost in last two years in Notre Dame.

I would also like to thank my current group members: Dr. Anthony Rosales, Dr. Atul Bhardwaj, Dr. John Herr, Dr. Taylor Quinn, Jessica Wahlers, Himani Patel, Michael Maloney, Emily Tharnish, Allison Miller, Bozhao Nan, and Ricardo Morales. They have accepted both me and John Herr to the Wiest-Helquist group as one of them. I would like to give special thanks to John Herr. He gave me many personal advices that changed my life.

Finally, I would like to express my gratitude to my family for supporting and encouraging me during my journey. My older brother, Andrew, was my role model as a consistent pursuer of dreams. My younger brother, Chris, supported me with care emotionally. Most importantly, my parents reminded me that they are always

behind me. Without them, this journey would have been impossible.

CHAPTER 1

INTRODUCTION

1.1 Time-Dependent Open Self-Consistent Field at Second Order (OSCF2)

In studying large systems, reduced models, such as quantum master equations [1] and stochastic Schrodinger equations[2], have been used due to decrease the high computational cost of a full quantum mechanical treatment of large systems. The Aspuru-Guzik group has used quantum master equations to study the energy transfer process of photosynthetic complexes such as the Fenna-Matthew-Olson pigment-protein complex[3]. Quantum master equations rely on a bath correlation function or a bath spectral density to describe the system-bath interaction. Thus, obtaining a correlation function or a spectral density of a bath is crucial in studying reduced models. The method of obtaining the spectral density, which describes the frequency dependent coupling of the system to the bath from the correlation functions of electronic transition energy obtained from molecular dynamics (MD) simulations has been extensively studied[4, 5]. Using the correlation function of the electronic transition energy to obtain a bath spectral density, the Parkhill group studied the relaxation in the perovskite $\text{CH}_3\text{NH}_3\text{PbI}_3$ and observed picosecond timescale intra-band relaxation as seen in transient absorption experimental results[6]. Furthermore, the Parkhill group derived an equation of motion where coherences decay to zero and the electronic

population reaches Fermi-Dirac distribution population at finite temperature[7].

$$\begin{aligned} \frac{d}{dt}\gamma_{ab} = & -\frac{i}{\hbar}[F(\gamma(t)), \gamma(t)]_{ab} \\ & -\frac{i}{\hbar^2}(S_{a,f,f,c}\Gamma(\omega_{cf})V_{ae}^\nu V_{fc}^\nu \gamma_{cb}\eta_{ef} + S_{b,f,f,d}(\Gamma(\omega_{df})V_{be}^\nu V_{fd}^\nu)^\dagger \gamma_{ad}\eta_{ef} \\ & -S_{d,b,a,c}\Gamma(\omega_{ca})V_{de}^\nu V_{ac}^\nu \gamma_{cd}\eta_{eb} - S_{c,a,b,d}(\Gamma(\omega_{db})V_{ce}^\nu V_{bd}^\nu)^\dagger \gamma_{cd}\eta_{ae}) \end{aligned} \quad (1.1)$$

$$S_{a,b,c,d} = \delta_{ab}\delta_{cd} - \delta_{bc}\delta_{ad}(1 - \delta_{ab}\delta_{cd}) \quad (1.2)$$

$$J_\nu(\omega) = \sum_i \frac{\lambda_{\nu,i}}{\sqrt{\pi}} \left(\frac{\gamma_{\nu,i}}{\gamma_{\nu,i}^2 + (\omega - \Omega_{\nu,i})^2} + \frac{\gamma_{\nu,i}}{\gamma_{\nu,i}^2 + (\omega - \Omega_{\nu,i})^2} \right) \quad (1.3)$$

$$\Gamma_\nu(\omega) = \int_0^{\inf} d\tau e^{i\omega_{ab}\tau} \times \left(\frac{1}{\pi} \right) \int_0^{\inf} d\omega' J_\nu(\omega') [e^{i\omega_{ab}\tau} n_b(\omega') + e^{-i\omega_{ab}\tau} (n_b(\omega') + 1)] \quad (1.4)$$

After the time-dependent correlation term of the system-bath interaction was derived using Mukherjee's normal ordering technique[8], the equation of motion was obtained as shown in Equation 1.1, where F is the Fock matrix, γ is one-electron reduced density matrix, $V_{ab}^\nu = \sum_\nu C_{a\nu} C_{b\nu}^\dagger$, $C_{a\nu}$ is molecular orbital coefficient matrix element for atomic orbital ν to molecular orbital a , $\eta_{ab} = \delta_{ab} - \gamma_{ab}$, δ is kronecker delta, and $\hbar\omega_{ab} = \epsilon_a - \epsilon_b$. After obtaining the two-time energy correlation function $C_\nu(t) = \langle \Delta_\nu(t) \Delta_\nu(0) \rangle$ with $\Delta_\nu = F_{\nu,\nu} - \langle F_{\nu,\nu} \rangle$ from MD simulations, the spectral density (SD) for atomic orbital ν , $J_\nu(\omega)$, is obtained by Fourier transformation of $C_\nu(t)$ and written as a linear combination of Lorentzian SD peaks[5, 9], as shown in Equation 1.3, with central frequency (Ω_i), strength (λ_i) and damping (γ_i) factors. Then the relaxation rate, Γ , is computed from $J_\nu(\omega)$ as shown in Equation 1.4, where n_b is the Bose-Einstein distribution.

This implementation of a nonadiabatic relaxation model, or time-dependent open self-consistent field at second order (OSCF2), has a computational cost of $O(N^3)$ and relaxes to the ground state from the excited state at a variable rate. A broadening of the response spectra, shift in peaks, and an entropy correction for free energy calculation was obtained by this method. The OSCF2 was used to investigate transient

spectra and relaxation dynamics of the the organolead triiodide perovskite (MAPbI_3) by Triet Nguyen[10]. Using OSCF2 to reproduce the interband relaxation time scale and identify cooling channels for hot electrons and hot holes, Nguyen reported that the dynamics of the organic cation and the tetragonal-to-pseudocubic transition of the perovskite mediates relaxation by observing a 70% increase in the interband relaxation time scale when computed using only the dominant vibrational frequencies of the cation.

1.2 Conclusion

Computational simulations of electronic structure dynamics of large system using quantum embedding are common approaches to treat quantum systems because full quantum treatment of the system is out of reach due to high computational cost. The balance between accuracy and affordability of this embedding has been studied. With the improvements of modern computational resources, more accurate embedding methods are becoming more affordable. This thesis covers the OSCF2 implementation for the simulation of transient absorption spectra and its application to the GFP-3 chromophore. The implementation of an embedded mean-field theory using block-orthogonalized partitioning to real-time time dependent density functional theory (RT-TDDFT) and its effect of accuracy and affordability are discussed. Then, attempts towards developing RT-TDDFT for the correct description of the excited state during an excitation process are discussed. Furthermore, the use of a neural network on a conformer search using a neural network force field and enhancement on learning in undergraduate organic courses and mechanistic studies on formation of oxindoles from squaraine dyes are discussed. The last chapter of this thesis covers code implementation done for Q2MM, PySCF, and automated CatVS softwares.

CHAPTER 2

DEVELOPMENT OF AFFORDABLE ELECTRONIC STRUCTURE METHOD FOR EXCITED STATE DYNAMICS

2.1 Time-Dependent Open Self-Consistent Field at Second Order (OSCF2) Studies of Transient Absorption Spectroscopy

2.1.1 Introduction

Transient absorption (TA) spectroscopy is the standard tool for investigating the dynamics of UV-visible excited states[11, 12]. Interpretation of TA spectra is complicated by the fact that the energies and intensities of excited state absorptions are often uncertain[13]. Developing a method that can accurately predict TA spectra would offer insight and assist with interpretation. Electronic structure simulations can also calculate the time-dependent properties of the system, including those that are experimentally inaccessible [14, 15]. Thus, the Parkhill group was interested in developing a method that simulates TA spectra.

Real-time time-dependent density functional theory (RT-TDDFT) is an electronic structure method commonly used for absorption spectra simulation [16, 17]. However, the RT-TDDFT model lacks features that are crucial to simulate TA spectra[18]. Time scales of electronic population changes before and after the excitation as well as relaxation are the most important features in TA experiments, yet they cannot be simulated by ordinary RT-TDDFT. It also suffers from artifacts when resonant continuous electric fields are applied to a system of interest[19, 20]. Excited state energies predicted by adiabatic functionals are different in the excited state and the ground

state. This leads to spurious predictions of photo bleaching and absorption in TA spectra since the energies of the excited states shift during relaxation. In this chapter, a dissipative RT-TDDFT method, OSCF2, developed by the Parkhill group[21], is used to treat dissipation in TA spectra simulation. The correction scheme developed by Fischer et al.[22] was used to treat the excitation pump and simulate TA spectra in real time.

TA simulations require an accurate treatment of the nonradiative relaxation between the excitation pulse and the probing pulse. OSCF2 has been shown to describe the relaxation from the excited state to the thermal equilibrium[7]. OSCF2 is a non-equilibrium extension of the real-time SCF method[21]. It is an affordable method with a cost of $O(N^3)$ that incorporates an atomistic version of a standard harmonic bath model to the Liouville-like equation of TDHF or TDDFT for the one-electron reduced density[23], as shown in Equation 2.1[21], to simulate the nonradiative relaxation and dephasing in a thermal environment. It can be used to study thermalization, relaxation, localization, quantum control[24] and transport[25], and linear and nonlinear spectroscopy[26].

$$\frac{d}{dt}\gamma = -\frac{i}{\hbar}[F(\gamma(t)), \gamma(t)] - \frac{i}{\hbar^2}R(\gamma(t)) \quad (2.1)$$

2.1.2 Methods

TA simulations require the correct description of the excited state dynamics. The excitation using a resonant pump pulse has been shown to be problematic[20] since the highest possible occupation of the singly excited state in a closed shell is 50%[27]. The correction suggested by Fischer et al. avoids the problem by describing the excitation process as a propagation from the excited state calculated using LR-TDDFT gradients[28, 29] and by eliminating response of the system due to the nonstationary initial state, as labeled as P_{ref} in Figure 2.1, from the overall response as shown in

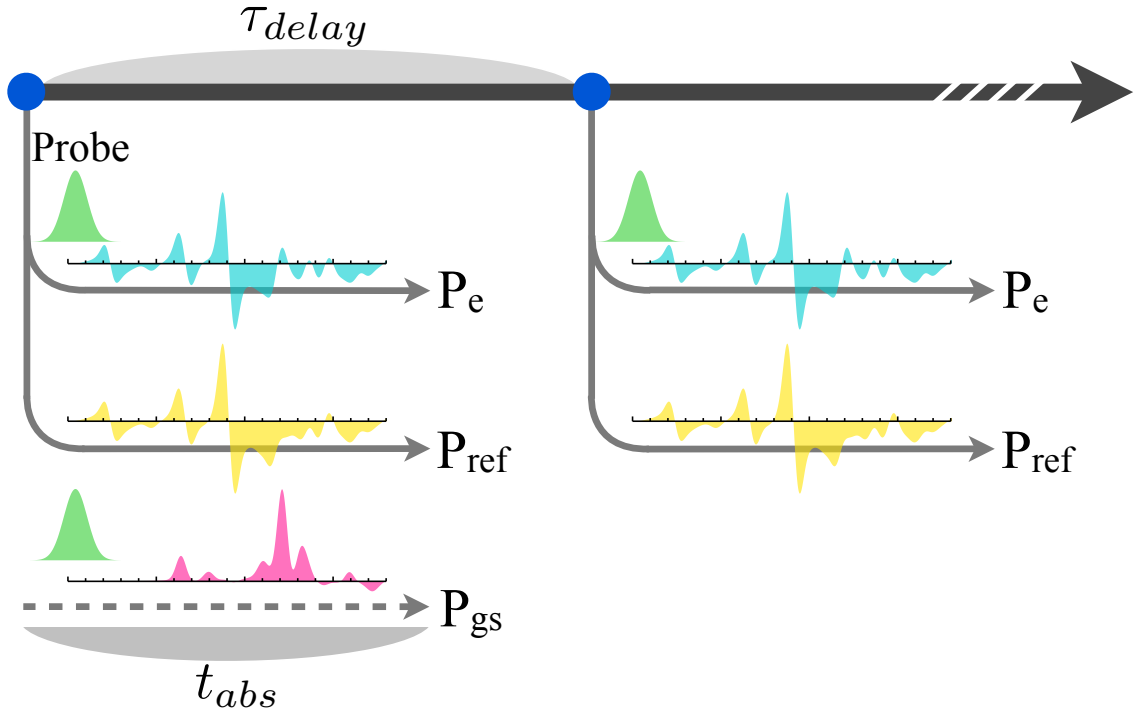


Figure 2.1. Pump-probe simulation scheme in OSCF2. Continuous black lines represent density matrices which are propagated to obtain the TA spectrum. At points where the trajectory branches, propagation splits.

Equation 2.2.

$$P(t) = P_e(t) - P_{gs}(t) - P_{ref}(t) \quad (2.2)$$

2.1.3 Results and Discussion

In this chapter, the excited states of pyrazole and the GFP-3 chromophore, shown in Figure 2.2, were studied using the linear-response(LR)-TDDFT density as shown in Equation 2.3 and the excited state density was propagated using OSCF2 using Equation 2.1.

$$\gamma_e = \sum \gamma_i \int d\omega |\mu E(\omega)|^2 \delta(\omega - e_i) \quad (2.3)$$

Using the geometry of pyrazole as shown in file `pyr.in`(<https://github.com/v3op01/DIS2020>), the pump-probe simulation was performed using the B3LYP/6-31G method. The system was pumped using frequencies of 6.32 eV and 12.6 eV. The zero-delay TA spectra was generated as shown in Figure 2.3. This demonstrated that the excitation pump was working as intended. As shown in Figure 2.3 (a), when pyrazole is excited by a pump pulse of 6.32 eV, the excited state absorption spectra

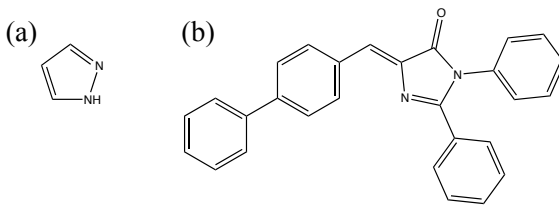


Figure 2.2. Structures of (a) pyrazole (b) (Z)-4-(biphenyl-4-ylmethylene)-1,2-diphenyl-1H-imidazol-5(4H)-one (GFP-3).

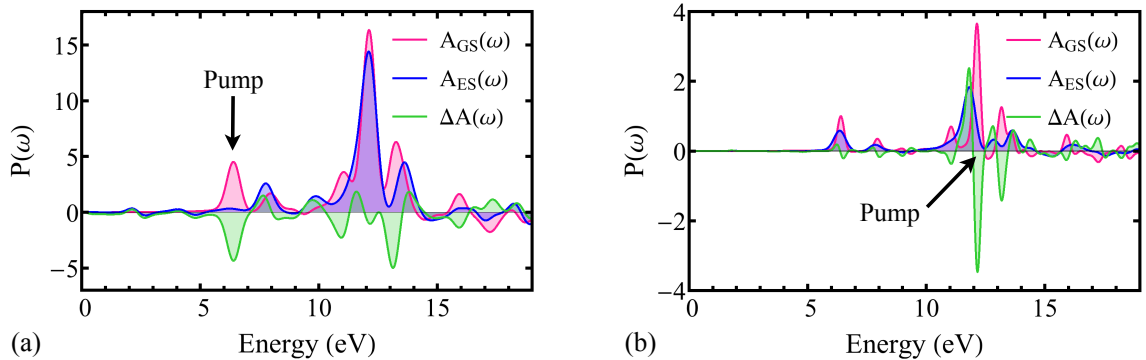


Figure 2.3. Zero-delay TA spectra of pyrazole using frequency of (a) 6.32 eV and (b) 12.6 eV. ΔA , or TA absorption spectra (green), shows bleach at the pumped energy.

(blue) showed no absorption peak at the energy of the pump pulse, leading to a photo bleach or negative absorption peak in TA spectra (green). When pyrazole is excited at 12.6 eV, as shown in Figure 2.3 (b), the excited state absorption spectra (blue) is not clearly showing a lack of absorption at 12.6 eV. However, the bleach at 12.6 eV on TA spectra (green) indicates the lack of absorption at 12.6 eV.

Using the geometry of Z-isomer of GFP-3 in an acetonitrile bath, as shown in file GFP3_PP.in(<https://bitbucket.org/jparkhill/transientabsorption>), the pump-probe simulation was performed using Q-Chem's 4.3 built-in QM/MM features. GFP-3 was treated with B3LYP/6-31G*, while the acetonitrile bath was treated with the CHARMM27 force field and the interaction between GFP-3 and the acetonitrile bath was treated using the Janus model[30]. LR-TDDFT calculations were performed in the initial step to calculate the excited state absorption using the lowest excited state energy frequency, 0.11677 a.u. and a field strength of 0.002 a.u.. The electron reduced density matrix, γ_e was calculated using Equation 2.3 and a real-time propagation was performed at a temperature of 300.0 K for 60000 iterations using a timestep of 0.8 a.u..

The probe simulations were performed at different time delays of 0 ps, 0.5 ps, 1

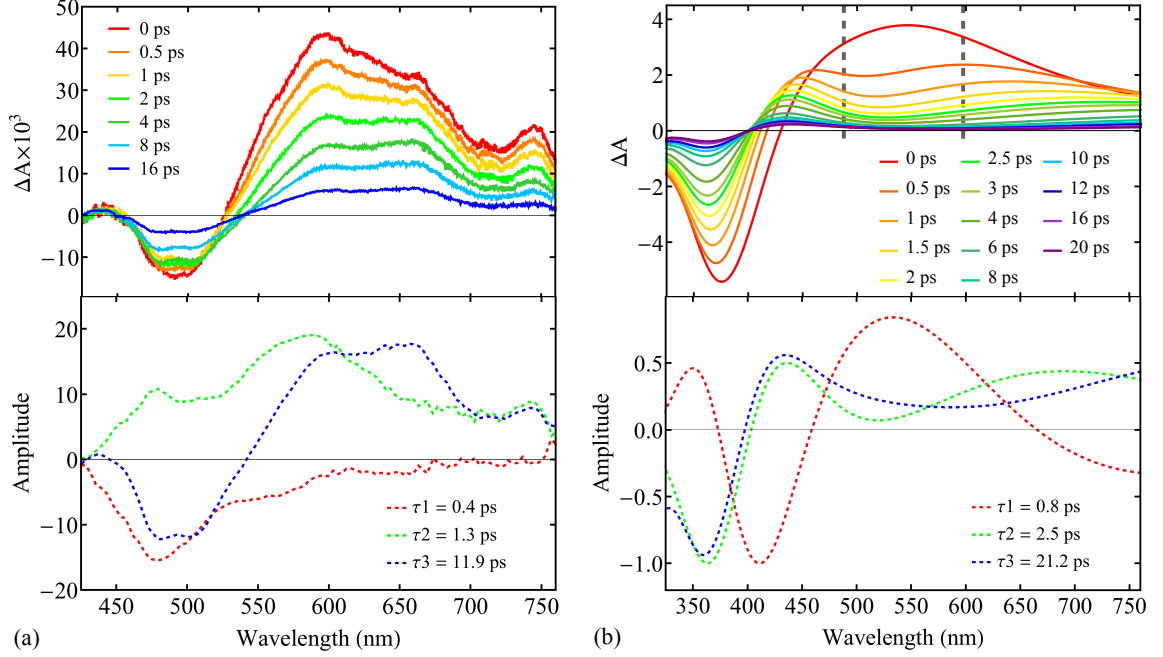


Figure 2.4. TA spectra(top) and decay-associated spectra (bottom) of GFP-3 chromophore in acetonitrile (a) simulated and (b) experimental result[31]

ps, 1.5 ps, 2 ps, 2.5 ps, 3 ps, 4 ps, 6 ps, 8 ps, 10 ps, 12 ps, 16 ps and 20 ps. The time propagation of dipole moments were Fourier transformed to the frequency domain as shown in Figure 2.4b (top). The experimental result[31] is shown in Figure 2.4a for the TA spectra (top) and the decay-associated spectra (bottom). The global analysis function in Glotaran[32] was used to generate the decay-associated spectra (DAS) observed in the experiment[31]. The software was used to generate the DAS of simulated TA spectra as shown in Figure 2.4b (bottom) to display the comparison.

The photo bleach peak and peak of experimental TA spectra are positioned around 480 nm and 600 nm, respectively. The bleach signal and the peak of the simulated TA spectra are positioned around 375 nm and 550 nm, respectively. By calculating the DAS for TA spectra, the position and time-scale of decay can be analyzed as shown in Figure 2.4b (bottom). The bleach signal around 375 nm decays to zero

as the time delay increases, but the position shifts to near 350 nm at a longer time delay as shown by a transition of the bleach signal of τ_1 at 410 nm to τ_3 at 360 nm in Figure 2.4b (bottom). The peak around 550 nm splits into two peaks positioned around 430 nm and 700 nm immediately, which is shown by peaks of τ_1 and τ_2 in Figure 2.4b (bottom). The calculated exponential decays, τ s, were 0.8 ps, 2.5 ps, and 21.2 ps, which are within a factor of 2 of their experimental counterparts. This overestimation agrees with OSCF2, where the lifetimes tend to be overestimated within an order of magnitude[21].

2.1.4 Conclusions

In this chapter, OSCF2 was combined with a scheme suggested by Fischer[22] to calculate the TA spectra of the GFP-3 chromophore in acetonitrile and to compare it with experimental data[31]. The positions of the bleach and the peaks were not in good agreement with the experimental values. However, the positions of the bleach and peaks in computed data is dependent on the geometry of the molecules and the method used. As shown in Figure 2.5, the bleach and TA peaks obtained by OSCF2 agree well with the LR-TDDFT result. Using a better bath model, instead of the QM/MM Janus model, may provide results that are in good agreement with experiment.

The problem that needs to be addressed is the peak shifting during relaxation. As shown in Figure 2.4b, bleach and TA peaks shift their positions as time increases, which is not seen in the experiment, as shown in Figure 2.4a. This is demonstrated for the case of a H_2 molecule as shown in Figure 2.6. The excited state spectrum and zero-delay TA spectrum of H_2 with the B3LYP/6-31G* method was calculated at different excited state populations. As shown in Figure 2.6a, as the population of the excited state changes, the position of the peak shifts, which leads to the problem in the TA spectrum as shown in Figure 2.6b. Addressing this problem will allow a

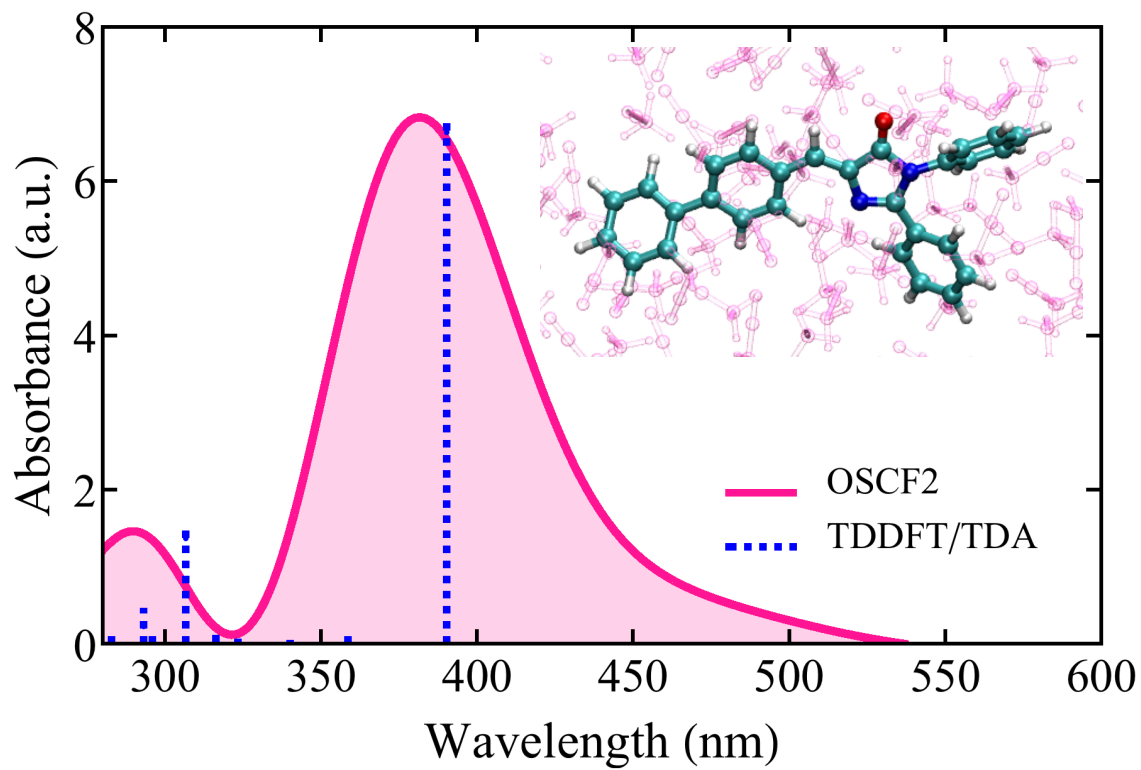


Figure 2.5. Absorption spectra of GFP-3 chromophore in acetonitrile using OSCF2 (pink) and LR-TDDFT (blue) using B3LYP/6-31G*

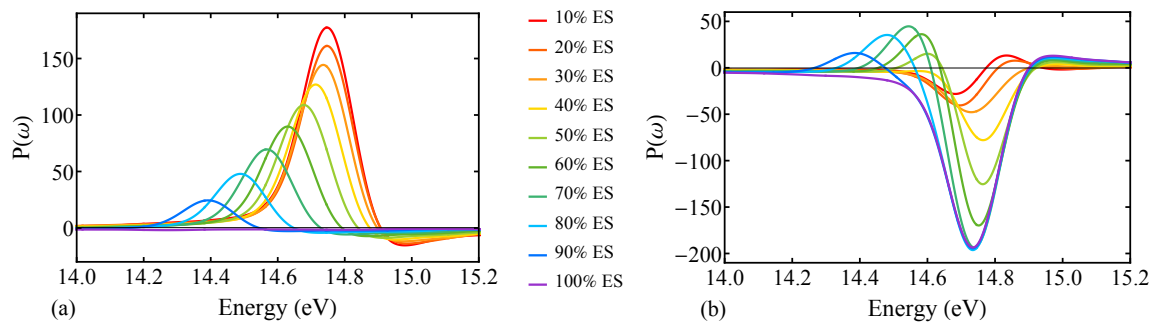


Figure 2.6. Excited state spectrum and zero-delay TA spectrum of H_2 molecule at different excited state population(B3LYP/6-31G*)

more accurate interpretation of the experimental result.

2.2 Real-time Propagation Using a Block Orthogonalized Basis

2.2.1 Introduction

While analyzing the result of pump-probe simulation, one of the problems encountered was the inaccurate absorption peak assignment due to lack of accurate system-bath interaction. Replacing the force field based solvent model with a DFT-based solvent model will allow a more accurate calculation of system-bath interaction as it includes electron coupling interaction. Quantum embedding is a powerful tool in a study where a molecule is divided into a small active system region, which is treated with higher level of theory, and a large inactive bath region, which is treated with lower level of theory.

A method where the computational cost is reduced while the accuracy is not sacrificed is needed. The solution to increase the accuracy of the calculations is to treat the whole system at a higher level of theory, such as hybrid DFT. However, this dramatically increases the cost of the computation. In the GFP-3-acetonitrile system-bath example defined in Figure 2.1, GFP-3 consisted of 51 atoms while the acetonitrile solvent consists of a total of 1560 atoms. Considering the quadratic scaling of computational cost of N basis sets in hybrid DFT, it is unrealistic to take this approach without any modifications.

The first attempt to solve this problem was the graphical processing units(GPU) in calculating the Fock matrix. It has been shown that quantum calculations can benefit from using GPU[33] when treating large systems. Using CUDA[34] in C++, I have written a code in Q-Chem that processes the Fock matrix in a GPU. As the bottleneck of the building the Fock matrix is the cost of building exchange matrix K , the two-electron integral was passed to the GPU along with the time dependent reduced density matrix to calculate the exchange matrix. After testing the code, it was found that due to the nature of application of real-time propagation of the

Fock matrix and electron reduced density matrix, this was not feasible because the increased cost of the CPU to GPU communication cost was greater than the decrease of the computational time of the Fock matrix solver as at every time step the Fock matrix had to be solved. Details on the C++ CUDA code can be found in Appendix A.

Other alternatives to solve this problem are to use different methods of QM/MM embedding [35, 36], such as density-matrix embedding theory[37], dynamical mean-field theory [38], or other density-based embedding methods[39–44]. Embedded Mean Field Theory (EMFT)[45] is a method where both regions are treated with DFT methods, but with different levels of theory and basis functions. The one particle density matrix is then divided into subsets A and B as shown in Equation (2.4) and the total energy is computed as a functional of each sub-block in the density with different level of treatments in the AO basis.

$$\gamma = \begin{bmatrix} \gamma^{AA} & \gamma^{AB} \\ \gamma^{BA} & \gamma^{BB} \end{bmatrix} \quad (2.4)$$

The energy of the EMFT is then obtained by Equation (2.5). EMFT has been shown to be comparable to ONIOM[46] in terms of energy calculation and geometry optimization, while avoiding the need to specify link atoms, link orbitals or geometry-dependent parameters and requiring only the subsystem partitioning in terms of the basis set[45]. However, EMFT with AO partitioning has failed on a range of systems, leading to unphysical collapse in energy, population of sub-blocks and dipoles. Ding et al. introduced an alternative partition scheme where the basis sets in each sub-block are block orthogonalized (BO) to the other sub-block[44].

$$E[\gamma] = E^{Low}[\gamma] + E^{High}[\gamma^{AA}] - E^{Low}[\gamma^{AA}] \quad (2.5)$$

The BO method treats a subsystem with a high-level Fock matrix and the remaining degrees of freedom with a computationally less expensive Fock matrix by reducing the quality of the basis and exchange interaction. The BO partitioning scheme uses a projected basis in place of the conventional AO partitioning to define the higher level of theory and lower level of theory in both the Fock matrix and density matrix. It was shown that this reduces errors in properties related to the embedding approximation[44]. The correct description of the population analysis on the sub-block provides an accurate interpretation and description of the system.

$$\tilde{H}_0 = U^T H_0 U, \tilde{G}^{Low}[\tilde{\gamma}] = U^T G^{Low}[\gamma] U, \text{ and } \tilde{\gamma} = U^T \gamma U \quad (2.6)$$

In Equation 2.6, U is the transformation matrix from the nonorthogonal AO basis set to the BO basis set where

$$\mathbf{U} = \begin{bmatrix} \mathbf{I}^{AA} & -\mathbf{P}^{AB} \\ \mathbf{0} & \mathbf{I}^{BB} \end{bmatrix}$$

In this blocked matrix, the identity matrices of subsystems A and B, \mathbf{I}^{AA} and \mathbf{I}^{BB} , have the dimensions of n_a and n_b which are the number of basis functions in each sub-block, and P^{AB} is the projection matrix, where $\mathbf{P}^{AB} = (\mathbf{S}^{AA})^{-1} \mathbf{S}^{AB}$, in which \mathbf{S}^{AA} and \mathbf{S}^{AB} are a sub-block of the AO overlap matrix. The effective Fock matrix in the BO basis set is then constructed as shown in Equation 2.7:

$$\tilde{F}[\tilde{\gamma}] = \tilde{H}_0 + \tilde{G}^{Low}[\tilde{\gamma}] + (\tilde{G}^{High}[\tilde{\gamma}^{AA}] - \tilde{G}^{Low}[\tilde{\gamma}^{AA}]) \quad (2.7)$$

where matrices in tildes are expressed in the BO basis. \tilde{G}^{High} and \tilde{G}^{Low} represent the two-electron contributions to the Fock matrix from the higher level and lower level of method, respectively, and $\tilde{\gamma}^{AA}$ represents the density matrix block that belongs

to the subsystem with a higher level of theory. The composition of the Fock matrix formed by the BO Manby-Miller Embedding(BOMME) is illustrated in Figure 2.7. I chose the simplest scheme to calculate $E_{EX}[\gamma^{AA}]$, where only the exact exchange interaction within the higher level of theory block (AA block) is considered as shown in Equation 2.8:

$$E_{EX}[\gamma^{AA}] = E_{EX0} = -\frac{1}{4} \sum_{\mu\nu\lambda\kappa \in A} (\mu\kappa|v\lambda) \gamma_{\mu\nu}^{AA} \gamma_{\kappa\lambda}^{AA} \quad (2.8)$$

Ding et al.[44] also discussed two more complex schemes, but determined that E_{EX0} has the best balance between accuracy and speed. I found it the best choice to reduce the computational cost for the present system.

2.2.2 Methods

Realtime TDDFT and Hartree-Fock theory lead to a Liouville equation for the one-particle density matrix[47, 48]:

$$\dot{\gamma} = -\frac{i}{\hbar} [\hat{F}(\gamma), \gamma] \quad (2.9)$$

where γ is the time-dependent one-electron density matrix, and F is the time-dependent Fock matrix of HF (or KS) theory expressed in an orthonormal basis. The Fock matrix can be written as: $F[\gamma] = H_0 + G[\gamma]$ in which H_0 is the core Hamiltonian, and G contains the two-electron contributions:

$$G[\gamma] = J[\gamma] + c_x K[\gamma] + V_{xc}[\gamma] \quad (2.10)$$

J and K are the density-dependent Coulomb and exchange operators, V_{xc} is the KS exchange-correlation contribution to the Fock matrix[49], and c_x is the coefficient of the exact exchange for hybrid KS theories. The bottleneck of integrating

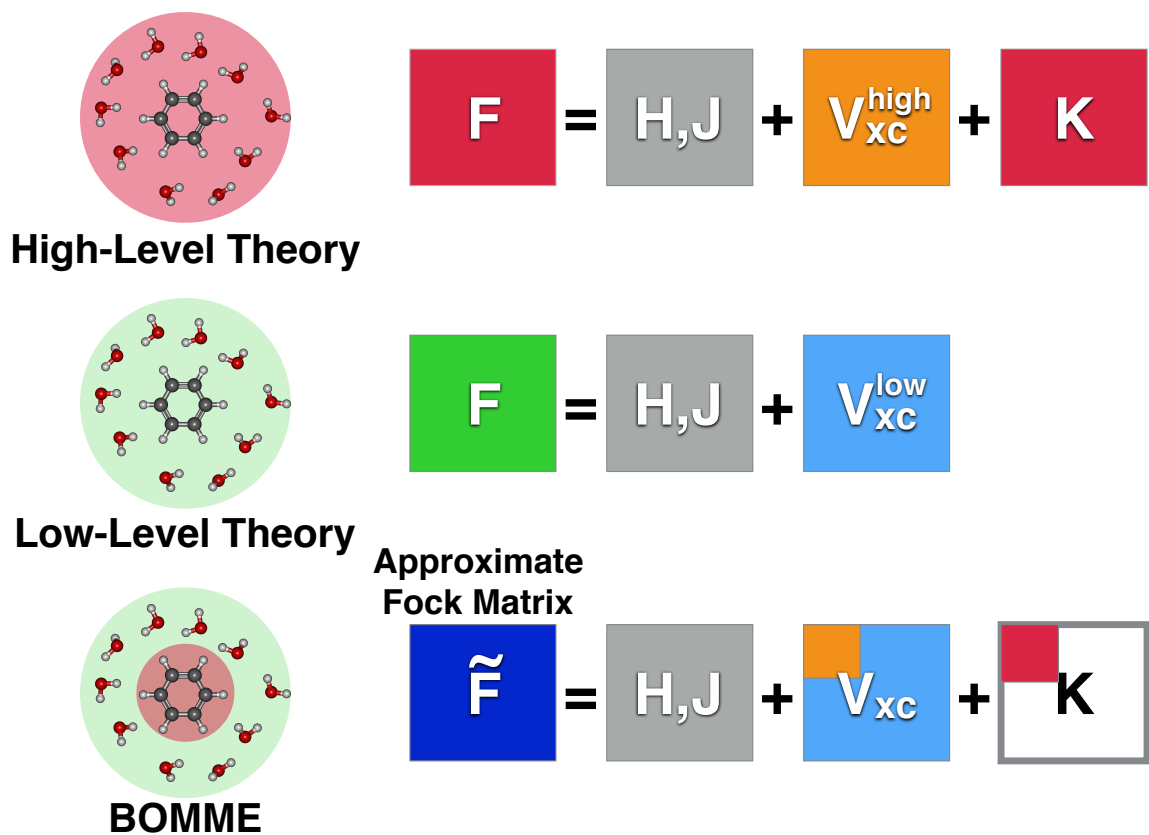


Figure 2.7. Block-Orthogonalized Manby-Miller Embedding Scheme. BO partitions Fock matrix in BOMME, where the higher level of theory is kept only within the sub-block

Equation 2.9 is the computation of J , V_{xc} and especially K , which needs to be recalculated several times for every femtosecond of propagation time. Methods and techniques that decrease the computational cost of the Fock term can save the cost of propagation[50–52].

$$\gamma[t + \Delta t] = \gamma[t] + \Delta t * \dot{\gamma} = \gamma[t] + \Delta t * -\frac{i}{\hbar}[\hat{F}(\gamma), \gamma] \quad (2.11)$$

$$\gamma[t + \Delta t] = U\gamma[t]U^* = e^{i\Delta t F[t]}\gamma[t]e^{i\Delta t F[t]} \quad (2.12)$$

Instead of using the general propagation method as shown Equation 2.11, a modified midpoint unitary transformation (MMUT) algorithm[53], shown in Equation 2.12, was used to propagate the density with a larger step size while maintaining stability of numerical noise. The MMUT method was shown to conserve energy in long-time dynamics compared to common propagation methods such as Runge-Kutta method[54]. The detailed code implementations to PySCF for the propagation are explained in Chapter 5.1.

$$F[\gamma] = H_0 + G[\gamma] + D * E[t] \quad (2.13)$$

The Fock matrix under an external field can be expressed as shown in Equation 2.12, where $E[t]$ is strength of external field at a given time t where D are the dipole moment integrals in AO basis. When treating an intramolecular embedding, applying an external field during the propagation when the Fock matrix is expressed in an AO basis exhibited unidentifiable peaks in absorption spectra. This was solved by transforming the dipole moment integrals to a BO basis, \tilde{D} , then applying the perturbation to the region with higher level of theory. This approach will be discussed

in more detail for the case of a pyrazole model system(d).

$$F[\tilde{\gamma}] = \tilde{H}_0 + G[\tilde{\gamma}] + \tilde{D}_{AA} * E[t] \quad (2.14)$$

In this study, I propagated the electron dynamics within the Born-Oppenheimer approximation. Propagation of 25000 steps with a time step of 0.02 a.u. was performed. An impulsive electric field with field strength of 0.001 a.u. , frequency of 0.9202 a.u. and τ of 0.07 in x, y, and z direction was applied. The high level of theory subsystem was calculated with PBE0/6-31G* and the lower level of theory subsystem was calculated with PBE/STO-3G.

The accuracy of the intermolecular embedding method on absorption spectra was tested for benzene and methanol solvated by water molecules. The accuracy of the intramolecular embedding method on the absorption spectra was tested for 3,4-dimethyl-1H-pyrazole and (2E,4E,6E,8E)-3,7-dimethyl-9-(2,6,6-trimethylcyclohex-1-en-1-yl)nona-2,4,6,8-tetraenal(retinal, e). Benchmarking of the computational cost of real-time propagation was done with benzene surrounded by different numbers of water molecules. Finally, a comparison of the embedding methods using (Z)-5-(4-hydroxybenzylidene)-2,3-dimethyl-3,5-dihydro-4H-imidazol-4-one (GFP neutral chromophore, c) to QM/MM and experimental value[55] was done. The QM/MM calculations were performed using PySCF’s point-charge embedding feature.

After propagation, dipole moments in the x,y,z-directions were exported and then Fourier transformed to absorption spectra, followed by summation. The population of each subsystem was monitored, and the entropy of the complete system was calculated.

All molecules were optimized using DFT calculations in their ground states before the electron propagation. In Figure 2.9 for each absorption spectrum, the atoms belonging to the higher level of theory and lower level of theory are specified by the

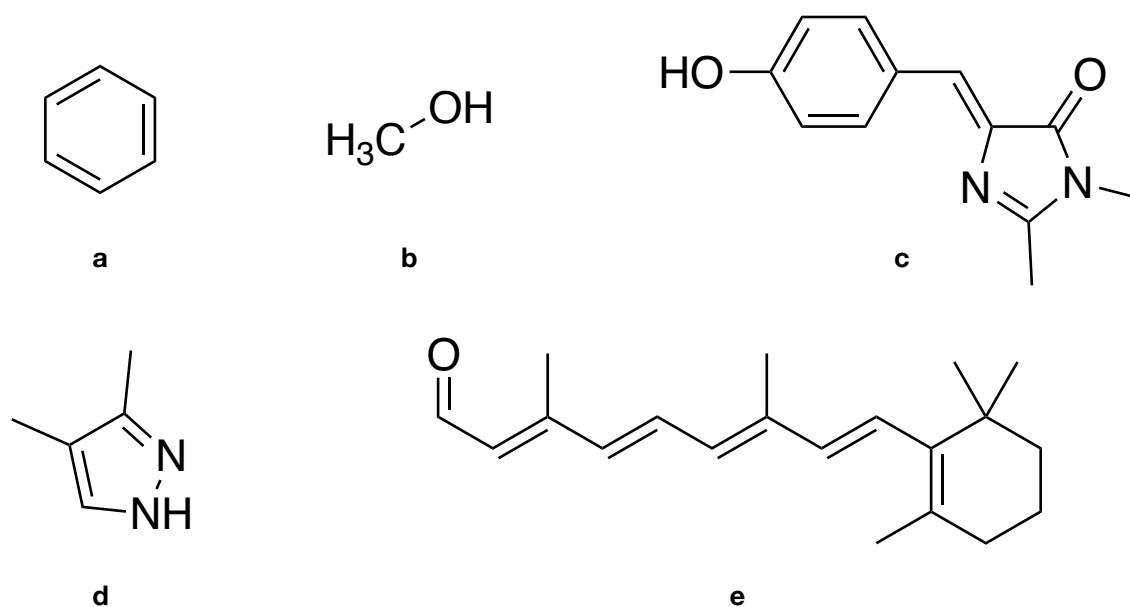


Figure 2.8. Molecules used for benchmarking. a) benzene b) methanol c) (Z)-5-(4-hydroxybenzylidene)-2,3-dimethyl-3,5-dihydro-4H-imidazol-4-one (GFP neutral chromophore) d) 3,4-dimethyl-1H-pyrazole e) (2E,4E,6E,8E)-3,7-dimethyl-9-(2,6,6-trimethylcyclohex-1-en-1-yl)nona-2,4,6,8-tetraenal(retinal)

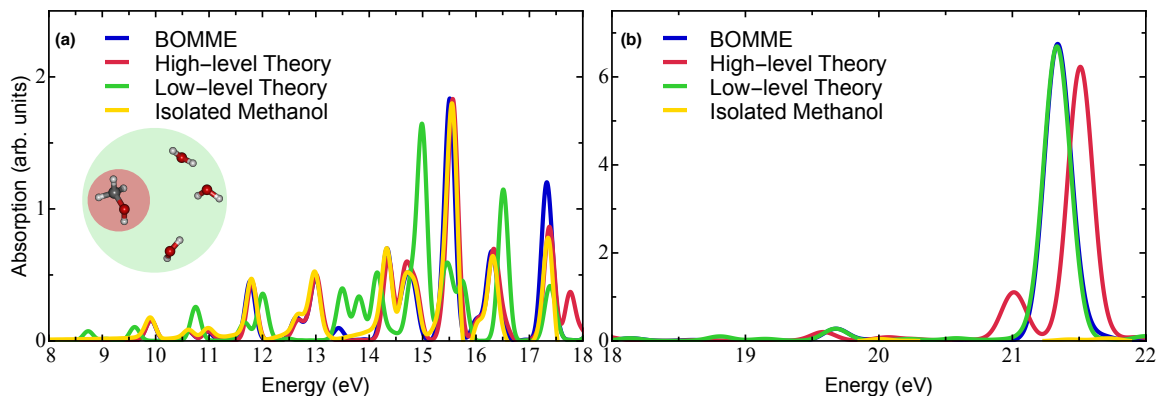


Figure 2.9. Absorption spectra of one methanol with 3 water molecules: (a) Low-energy region, and (b) high-energy region. Note that the spectra are plotted at different scales for clarity. High theory employs PBE0, and low-level theory employs PBE. Both methods used a mixed basis set of 6-31G*/STO-3G on methanol and water, respectively.

red and green shaded regions, respectively.

2.2.3 Results and Discussion

Figure 2.9 shows the absorption spectra obtained by performing RT-TDDFT on a system of one methanol molecule and three water molecules treated with PBE0/6-31G* and PBE/STO-3G, respectively. As expected for an embedding method, the absorption spectrum of the system is a combination of the PBE0/6-31G* methanol spectrum and PBE/STO-3G water spectra. The BOMME spectrum matches peaks of the PBE0/6-31G* spectrum of the system and of isolated methanol in the lower energy region up to 18 eV and matches peaks of the PBE/STO-3G spectrum in the higher energy region. All peaks of the BOMME method under 18 eV, other than peaks at 13.44 eV and 17.32 eV, correspond to the PBE0/6-31G* predicted peaks for methanol, whereas absorption above 18 eV and peaks at 13.44 eV and 17.32 eV correspond to water peaks predicted by PBE/STO-3G. For this geometry, the shift due to solvation is small. An absorption of the isolated methanol at 11.85 eV

corresponds to 11.66 eV in high-level theory and to 11.61 eV in BOMME, which are -0.19 eV and -0.24 eV from the isolated methanol peak. The effect of the bath is therefore not strong, and the shift of peaks by the low level theory of the bath is similar to that by the high level theory of the bath.

The time propagation of dipole moments in xyz directions for one methanol and three water system is shown in Figure 2.10. The amplitude of the dipole moments is stable, except for the initial oscillation that is due to initial interaction of the impulsive electric field. The trace of density in the AA block and in the BB block and total number of electrons in the whole system is shown in Figure 2.11. As was the case for the ground state population analysis[44], after the system undergoes electronic excitation, the populations of the electrons are confined within the AA and BB block of the density matrix, where the sum of traces of two blocks makes up the total number of electrons in the system. Total entropy and partial entropy in methanol are shown in Figure 2.12. The entropies of the two subsystems are not constant, since electrons are exchanged between these regions throughout the propagation, as shown in Figure 2.11. Thus, no indication of unphysical behavior, instability, or artifactual transitions is observed.

I used a simple system of benzene solvated by a different number of water molecules to benchmark the computational cost of BOMME relative to that of non-embedding methods. The time cost per ps of propagation is expressed as wall clock time. As shown in Figure 2.13, the method achieves significant relative speedup compared to simulations of purely higher level theory, and becomes more affordable as more explicit solvent molecules are used.

I then investigated the performance of BOMME method’s dependency on the size of the environment. I performed the propagation calculations and obtained the absorption spectra of a benzene with one water molecule (Figure 2.14.a) and ten water molecules (Figure 2.14.b). In Figure 2.14.a, the lowest $\pi \rightarrow \pi^*$ transition energy

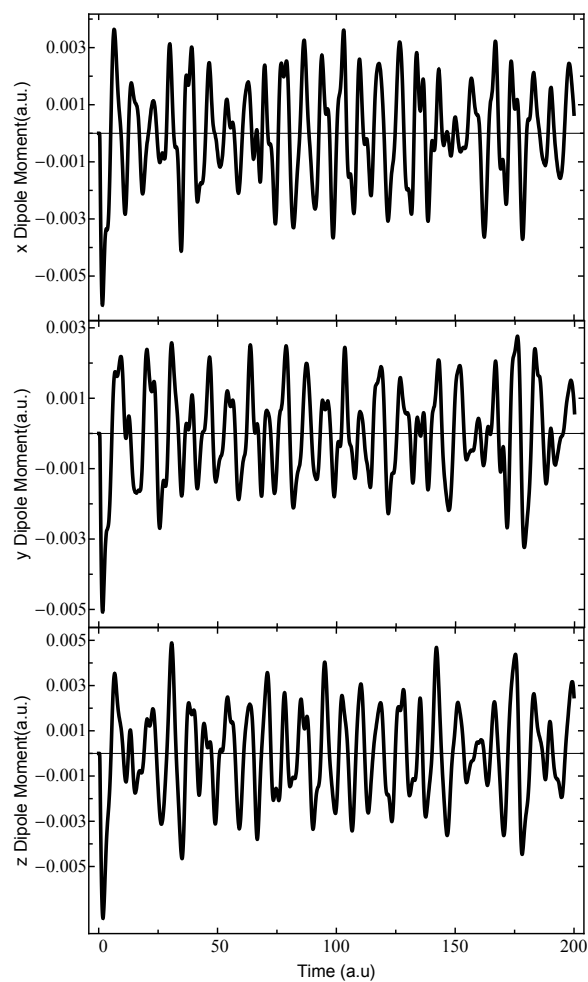


Figure 2.10. Dipole moments of a system which consists of one methanol and three water molecules. Methanol calculation employs PBE0/6-31G*; water calculation employs PBE/STO-3G.

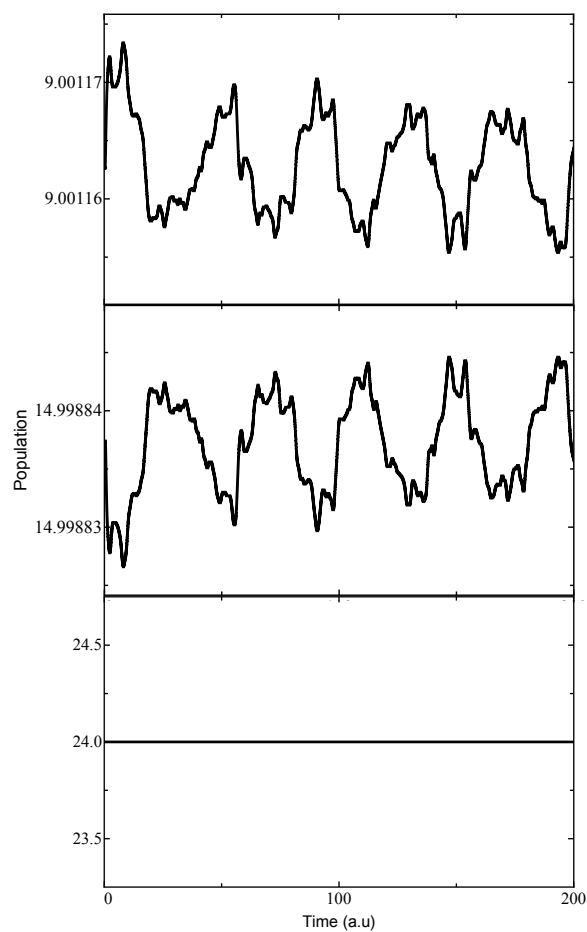


Figure 2.11. Number of electrons in subsystem A (one methanol molecule - top panel), B (three water molecules - middle panel), and the whole system (bottom panel). Methanol calculation employs PBE0/6-31G*; water calculation employs PBE/STO-3G.

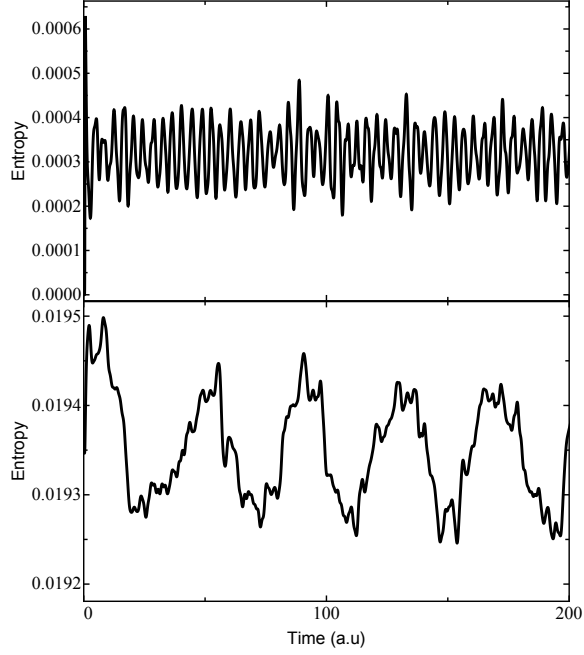


Figure 2.12. Entropy of the whole system (one methanol and three water molecules - top panel) and subsystem A (one methanol molecule - bottom panel). Entropy is calculated using

$S[\gamma] = -2 * [(\gamma)\ln(\gamma) + (1 - \gamma)\ln(1 - \gamma)]$, where γ is electron density contribution and $(1 - \gamma)$ is hole density contribution in MO basis. Factor of 2 is to account for alpha and beta electron pair.

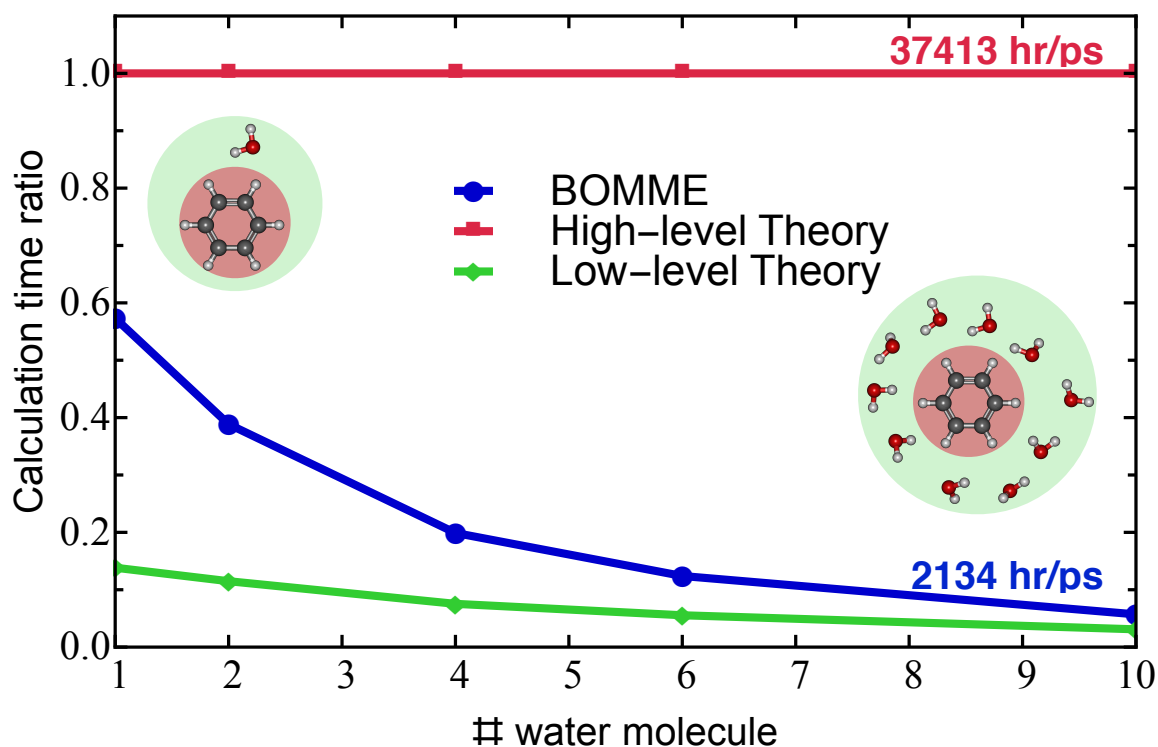


Figure 2.13. Speed performance with and without embedding is demonstrated on a benzene surrounded with an increasing number of water molecules. Ratios of simulation time are calculated with respect to the high-level theory scheme. The functional/basis set combinations are PBE0/6-31G* on high-level theory (red line and red shaded region), PBE/STO-3G on low-level theory (green line and green shaded region), and mixed bases on BOMME method (blue line and red/green shaded region).

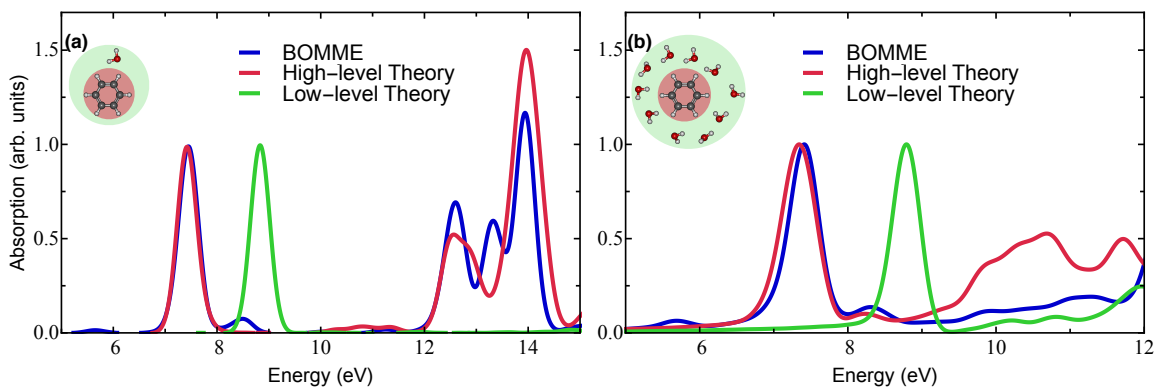


Figure 2.14. Absorption spectra of benzene with (a) one and (b) ten water molecules. High-level theory used PBE0/6-31G*, low-level theory used PBE/STO-3G, and isolated methanol used PBE0/6-31G*.

using BOMME, PBE0/6-31G*, and PBE/STO-3G theory are 7.37 eV, 7.40 eV and 8.80 eV, respectively. This fair agreement is maintained in Figure 2.14.b where the peaks are 7.24 eV, 7.35 eV and 8.74 eV, respectively. In the presence of larger number of atoms in the solvation environment, the BOMME method still performs close to the PBE0/6-31G* method as shown in Figure 2.14.b while the computational cost remains close to the PBE/STO-3G method as shown in Figure 2.13.

I tested an intramolecular embedding as well, which I expected to be significantly more challenging and possibly showing false absorption peaks due to non-physical motion of electrons between two sub-regions because the atomic orbitals between the sub-regions are overlapping closely. I chose 3,4-dimethylpyrazole as the first model system. With pyrazole’s aromaticity playing a crucial role in absorption of UV-vis spectra, it was natural to treat the ring with a higher level of theory and the methyl groups with a lower level of theory. Indeed, simply applying the same propagation described above leads to unidentifiable peaks in the BOMME method. Analyzing the time-dependent electron density in more detail, I found that characteristic DFT bias towards over-delocalization of the density from pyrazole to methyl groups caused

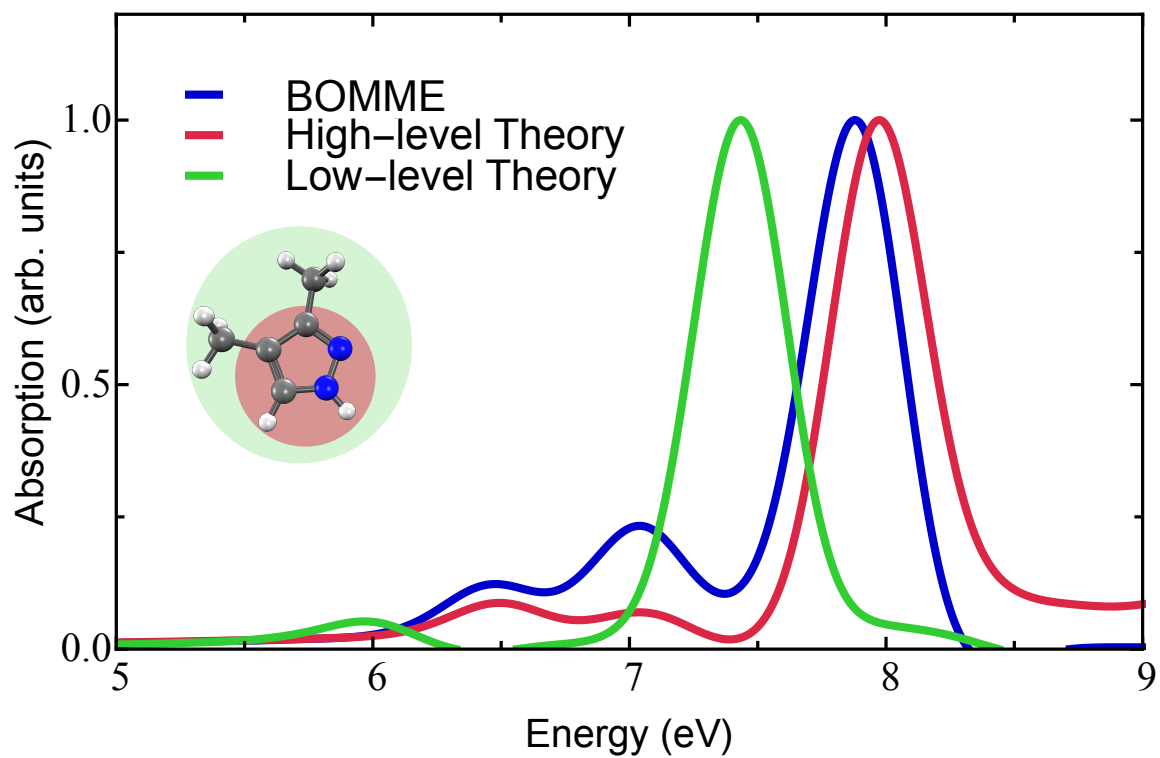


Figure 2.15. Absorption spectra of 3,4-Dimethylpyrazole. High-level theory used PBE0/6-31G*, and low-level theory used PBE/6-31G*.

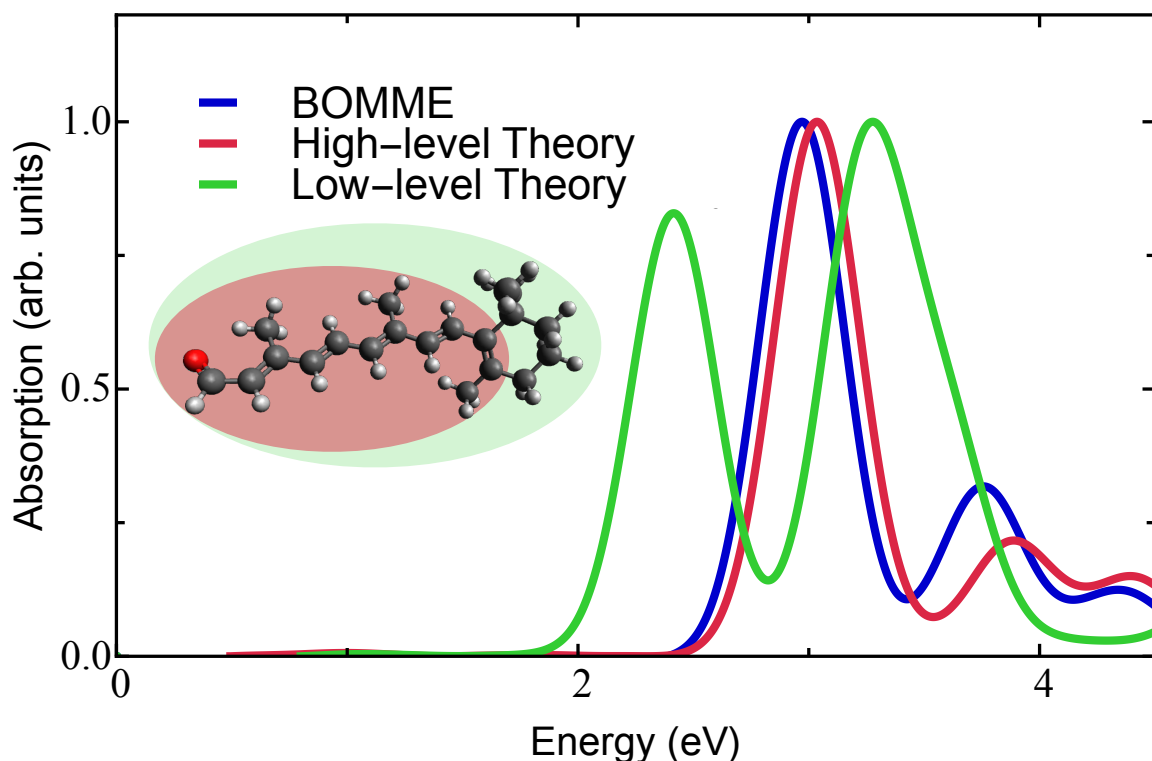


Figure 2.16. Absorption spectra of all-*trans*-retinal. High-level theory used is PBE0 and low-level theory used is PBE. All methods used a mixed basis of 6-31G*/STO-3G.

these additional peaks. The excitation pulse delocalized the electron and hole density into the two different regions[56]. This problem was solved by perturbing only the PBE0/6-31G* block, and restricting all fields to the PBE/6-31G* block. After a modified perturbation was applied to the system, the resulting spectra has no visible artifacts. The first excited-state peak is at 7.43 eV, 7.88 eV and 7.98 eV for the PBE/6-31G* method, the BOMME method, and the PBE0/6-31G* method, respectively as shown in Figure 2.15.

Applying the perturbation to only the high-level method region appears to effectively solve the problem of spurious embedding absorbances. Figure 2.16 shows the absorption spectra of all-*trans*-retinal. BOMME functions well when a reasonable definition of high- and low-level method regions are chosen such that it does not

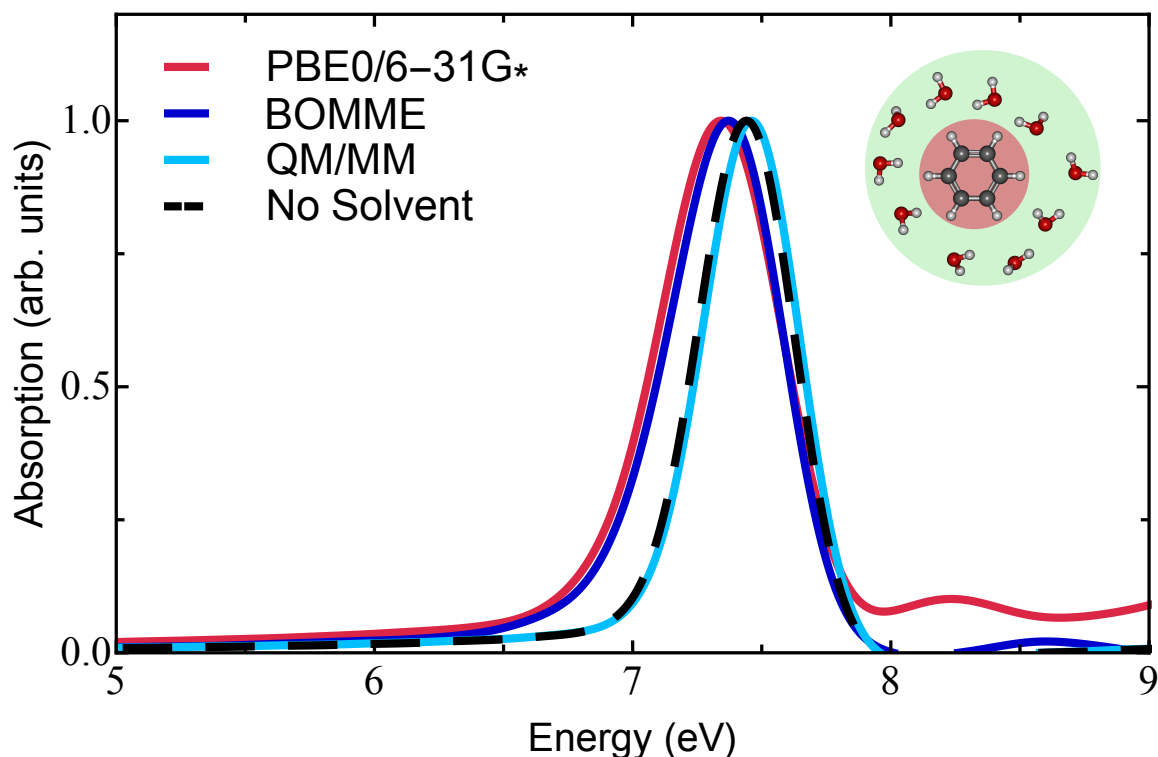


Figure 2.17. Absorption spectra of benzene with ten water molecules calculated with DFT, BOMME, and QM/MM, compared with spectrum of an benzene (PBE0/6-31G*).

disconnect the conjugate π system of the chromophore. As shown in Figure 2.16, the conjugated double bonds were chosen to be treated with a high-level method as they contribute the UV-vis absorption peak. The lowest excitation peaks were 2.41 eV, 2.97 eV and 3.04 eV for the low-level method, the BOMME method and the high-level method, respectively. The experimental values for absorption of all-trans retinal are 3.24 eV and 3.37 eV, in ethanol and hexane respectively[57].

Lastly, I compared the BOMME method with point-charge QM/MM embedding for a few systems to demonstrate the effect of explicit solvation. Figure 2.17 shows the absorption spectrum of benzene and ten water molecules generated by purely high-level theory, BOMME, and QM/MM between 5 eV and 9 eV. As shown in Table 2.1, using the absorption peak of benzene as a reference, the solva-

TABLE 2.1

POSITION OF THE LOWEST EXCITED STATE ENERGY IN FIGURE
2.17

Method	Peak (eV)	Difference
Ref	7.44	+0.00
BOMME	7.37	-0.07
PBE0/6-31G*	7.34	-0.10
QM/MM	7.46	+0.02

tion shift of the lowest excited state energy in high-level method and BOMME were -0.10 and -0.07 eV, respectively, whereas the QM/MM spectrum was +0.02 eV. This demonstrates the accuracy of the BOMME method for calculating the magnitude of solvation shift as well as the disagreement of the QM/MM method in the direction of solvation shift compared to a DFT-only method.

The disagreement of the QM/MM embedding method with the high-level reference method is further shown in Figure 2.18, where the absorption spectra of a neutral GFP chromophore obtained using the BOMME method, QM/MM method and experiment[58] are compared. Point-charge embedding produces a first bright state approximately 40 nm away from the experimental peak, which is at 370 nm, whereas BOMME is in good agreement with the experiment and is predicting solvation shift in the opposite direction compared to QM/MM embedding.

2.2.4 Conclusions

In this chapter, I investigated the potential of the BOMME method as a tool to accurately propagate the electronic system after experiencing electric field at small

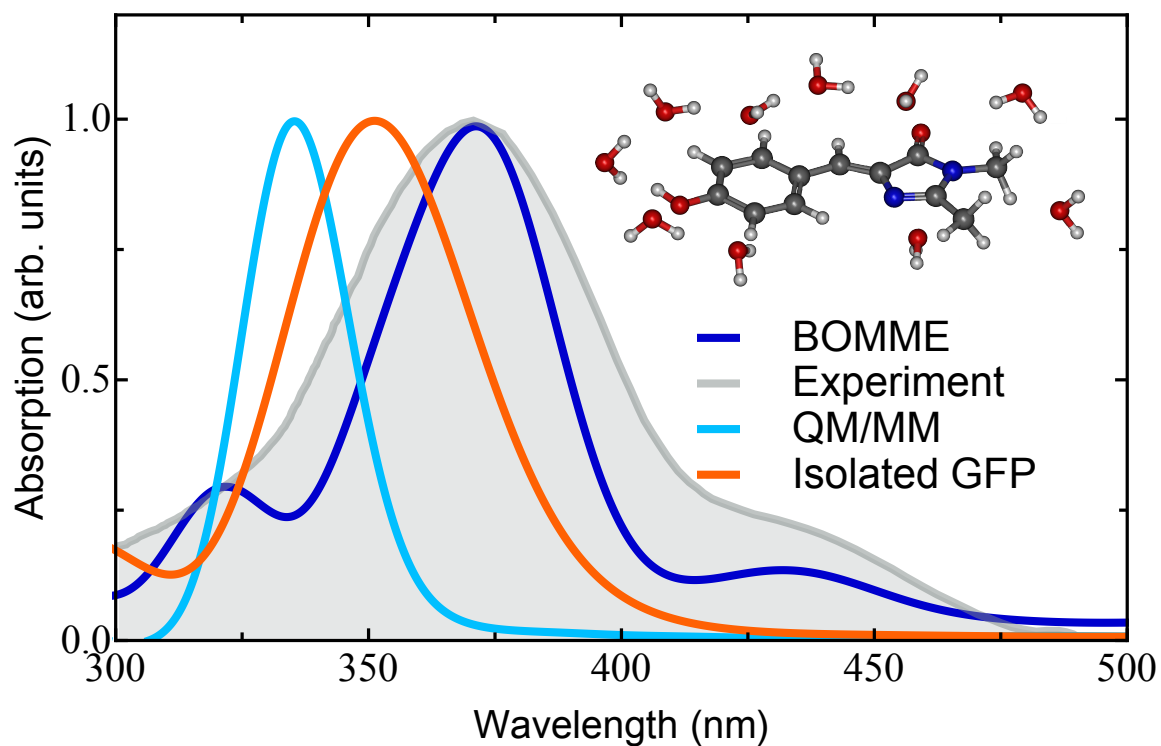


Figure 2.18. Absorption spectra of a neutral GFP chromophore with ten water molecules calculated with BOMME (6-31G/PBE0, STO-3G/PBE) and QM/MM, and of an isolated GFP neutral chromophore (PBE0/6-31G), compared with experimental spectrum from Ref. [58].

additional cost. The results shows the advantage of the BOMME method as applied to RT-TDDFT in reducing computational cost while minimizing the accuracy loss. Both intermolecular and intramolecular schemes with BOMME capture the shift of a solvated chromophore. Because of the simple framework of BOMME using the Fock matrix, the method can easily be combined with a multiple timestepping scheme to obtain further speedups[59, 60]. With the tool available to increase the accuracy at little additional computational cost to investigate the excited state system, I then sought to solve the problem of peak shift that I found to be a problem in simulating TA spectra.

2.3 Real-time Time Dependent Tamm-Dancoff Approximation

2.3.1 Introduction

The equation of motion (EOM) for the electron density is weakly non-linear. Thus, TDHF and TDDFT behave nonphysically when a large number of electrons are perturbed far from a stationary solution[61, 62]. Because TDDFT uses approximate time-independent XC functionals, it is well known that the complete excitation is not achieved when driving a TDHF or TDDFT system at a resonant frequency of the linear spectrum[53, 63, 64]. When a substantial fraction of the population is excited, the Fock operator and effective gap change, which puts the driving field out of resonance. This deviates from the textbook behavior of a closed shell system, and makes it impossible to use adiabatic RT-TDDFT to correctly model Rabi oscillations. The Isborn group investigated the peak shifting problem observed in RT-TDDFT and its dependency on the frequency and intensity of the applied field[20, 62]. As a result of these problems, electron dynamics simulated with RT-TDDFT might have problems in modeling laser-induced excitation processes.

This off-resonance problem was observed in transient absorption spectroscopy

simulations as shown in Figure 2.6, where the peaks of the excited state shifted as a function of time during the non-radiative relaxation[65]. This problem can be addressed by replacing the excitation process with excited state density calculation based on LR-TDDFT. A method that can propagate in real-time during excitation and relaxation in the presence of a solvent will allow direct comparison of calculation to experiments. I decided to investigate this problem by comparing the one-electron reduced density matrix propagation to the Time-dependent Configuration-Interaction Singles(TDCIS) method, where the full excitation and relaxation exhibits a physically correct behavior occurs[20, 66], and implement the Tamm-Dancoff approximation (TDA) to the TDSCF one-electron reduced density matrix propagator[23].

2.3.1.1 Time-Dependent Configuration-Interaction Singles

I first implemented an approximation of TDCIS for use as reference data in further implementations. In Configuration Interaction methods, the wave function, $|\Phi_0\rangle$, is defined as the summation of set of configurations, $|\Psi\rangle$, as shown in Equation 2.15:

$$|\Phi_0\rangle = \alpha_0|\Psi_0\rangle + \sum \alpha_i^a|\Psi_i^a\rangle + \sum \alpha_{ij}^{ab}|\Psi_{ij}^{ab}\rangle + \dots \quad (2.15)$$

where $|\Psi_0\rangle$ represents the ground state configuration, $|\Psi_i^a\rangle$ represents the singly excited state from occupied orbital i to virtual orbital a , $|\Psi_{ij}^{ab}\rangle$ represents the doubly excited state from occupied orbital i and j to virtual orbital a and b , and α represents the coefficient of each term such that $\langle\Phi_0|\Phi_0\rangle = 1$. In TDCIS, only the singly excited state configurations are considered and α now is dependent on the time, t , as shown

in equation 2.16.

$$|\Phi_0(t)\rangle = \alpha_0(t)|\Psi_0\rangle + \sum \alpha_i^a(t)|\Psi_i^a\rangle \quad (2.16)$$

$$i\frac{\delta}{\delta t}|\Phi_0(t)\rangle = \hat{H}(t)|\Phi_0(t)\rangle \quad (2.17)$$

The time dependent Schrodinger equation describes the time propagation of the wave function as proportional to the Hamiltonian of the wave function as shown in Equation 2.17. Combining Equation 2.16 and 2.17 allows the time propagation of the wave function as shown in Equation 2.19:

$$i\frac{\delta}{\delta t}|\Phi_0(t)\rangle = i\frac{\delta}{\delta t}\alpha_0(t)|\Psi_0\rangle + i\sum \frac{\delta}{\delta t}\alpha_i^a(t)|\Psi_i^a\rangle \quad (2.18)$$

$$\hat{H}(t)|\Phi_0(t)\rangle = i\dot{\alpha}_0(t)|\Psi_0\rangle + i\sum \dot{\alpha}_i^a(t)|\Psi_i^a\rangle \quad (2.19)$$

Using the orthonormal nature of wave functions, the time propagation of coefficients, $\dot{\alpha}(t)$ can be defined. By projecting the ground state wave function $\langle\Psi_0|$ to Equation 2.19, $\dot{\alpha}_0(t)$, the time derivative of $\alpha_0(t)$, can be defined as shown in Equation 2.23.

$$\langle\Psi_0|\hat{H}(t)|\Phi_0(t)\rangle = i\langle\Psi_0|\dot{\alpha}_0(t)|\Psi_0\rangle + i\sum \langle\Psi_0|\dot{\alpha}_i^a(t)|\Psi_i^a\rangle \quad (2.20)$$

$$\langle\Psi_0|\hat{H}(t)|\Phi_0(t)\rangle = i\dot{\alpha}_0(t)\langle\Psi_0|\Psi_0\rangle + i\sum \dot{\alpha}_i^a(t)\langle\Psi_0|\Psi_i^a\rangle \quad (2.21)$$

$$\langle\Psi_0|\hat{H}(t)|\Phi_0(t)\rangle = i\dot{\alpha}_0(t) \quad (2.22)$$

$$\alpha_0(t)\langle\Psi_0|\hat{H}(t)|\Psi_0\rangle + \sum \alpha_i^a(t)\langle\Psi_0|\hat{H}(t)|\Psi_i^a\rangle = i\dot{\alpha}_0(t) \quad (2.23)$$

$\dot{\alpha}_i^a(t)$, the time derivative of $\alpha_i^a(t)$, can be defined by projecting the singly excited state wave function $\langle \Psi_i^a |$.

$$\langle \Psi_i^a | \hat{H}(t) | \Phi_0(t) \rangle = \langle \Psi_i^a | \dot{\alpha}_0(t) | \Psi_0 \rangle + \sum \langle \Psi_i^a | \dot{\alpha}_i^a(t) | \Psi_i^a \rangle \quad (2.24)$$

$$\langle \Psi_i^a | \hat{H}(t) | \Phi_0(t) \rangle = i\dot{\alpha}_0(t) \langle \Psi_i^a | \Psi_0 \rangle + i \sum \dot{\alpha}_i^a(t) \langle \Psi_i^a | \Psi_i^a \rangle \quad (2.25)$$

$$\alpha_0(t) \langle \Psi_i^a | \hat{H}(t) | \Psi_0 \rangle + \sum \alpha_i^a(t) \langle \Psi_i^a | \hat{H}(t) | \Psi_i^a \rangle = i \sum \dot{\alpha}_i^a(t) \quad (2.26)$$

The interaction between the Hamiltonian and a laser field presented by Greenman et al.[64], as $\hat{H}(t) = \hat{H} - E_0 + E(t)\hat{z}$, where \hat{H} is sum of one-body and two-body Hamiltonian operator[67], E_0 is the Hartree-Fock ground-state energy, $E(t)$ is the electric field strength of the laser, and \hat{z} is the dipole matrix. Inserting the time-dependent Hamiltonian in an electric field in Equation 2.23 and 2.26 results in derivatives of the coefficient α_0 and α_i^a , as shown in Equation 2.27 and Equation 2.28:

$$i\dot{\alpha}_0(t) = -\sqrt{2}E(t) \sum \alpha_i^a z_{ia} \quad (2.27)$$

$$\begin{aligned} i\dot{\alpha}_i^a(t) = & (\epsilon_a - \epsilon_i)\alpha_i^a + \sum \alpha_j^b (2\langle aj|ib \rangle - \langle aj|bi \rangle) \\ & - E(t)(\sqrt{2}\alpha_0 z_{ai} + \sum \alpha_i^b z_{ab} - \sum \alpha_j^a z_{ij}) \end{aligned} \quad (2.28)$$

where a and b symbolize virtual orbitals, i and j symbolize occupied orbitals, z_{pq} is the dipole moment matrix element at orbital p and q, ϵ_p is eigenvalue of Fock matrix at orbital p, and $\langle pq|rs \rangle = \int dx_1 dx_2 \chi^*(x_1) \chi^*(x_2) r_{12}^{-1} \chi(x_1) \chi(x_2)$ for orbitals p, q, r and s. The TDCIS code is available at github.com/v3op01/DIS2020/ES/TDA/TDCIS.py.

One of the problems in TDCIS is that a smaller timestep is required to correctly propagate the system as shown in Figure 2.19. When H_2 was propagated using an 0.02 a.u. timestep under resonant frequency of 13.5617 eV, the excited electron population in LUMO did not reach the full excitation and did not relax back to the HOMO. It was only at a timestep of 0.005 a.u. that complete excitation and

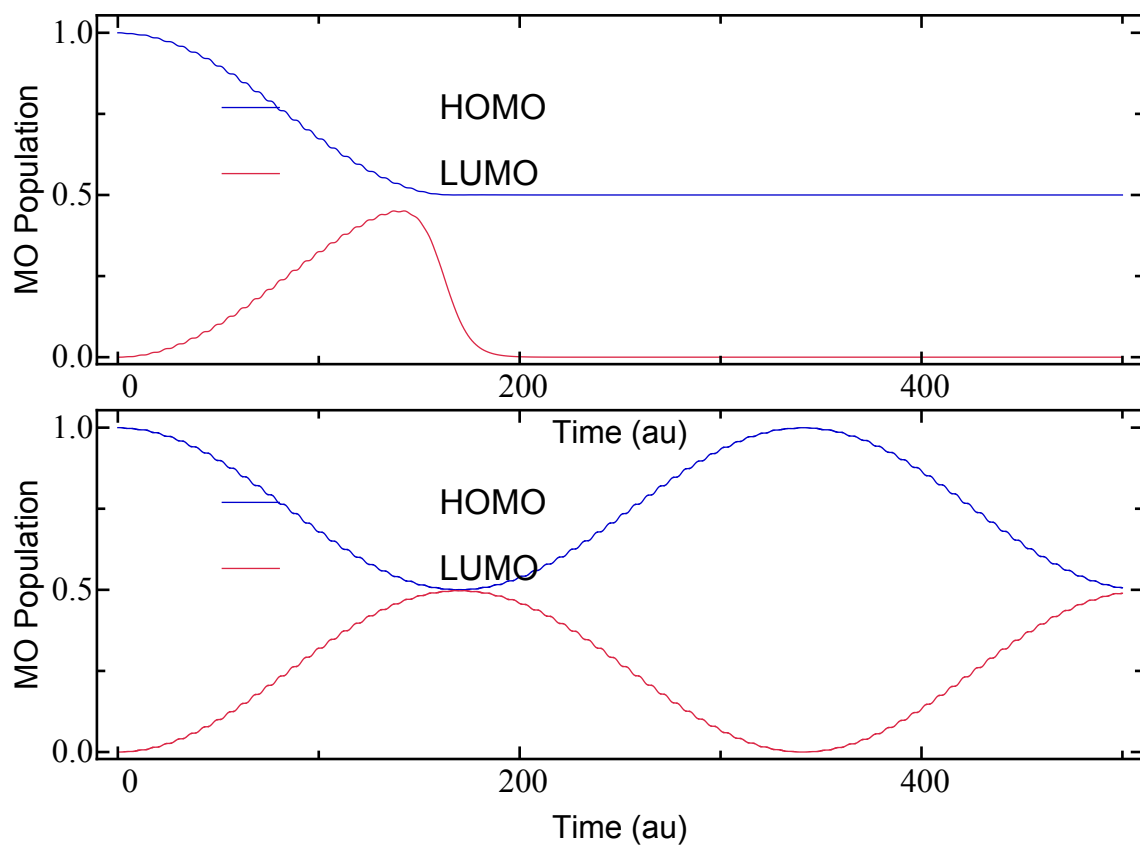


Figure 2.19. Time propagation of H_2 molecule using TDCIS under resonant electric field. 0.75 Angstrom between hydrogens. HF/cc-pvtz method. Resonant frequency of 13.5617 eV. Timestep of 0.02 a.u. (top) and 0.005 a.u. (bottom)

relaxation due to a resonant field was achieved in TDCIS. The same propagation using Equation 2.52, the latest EOM of TDA implementation, was achieved using 0.25 a.u., which is 50 times that of TDCIS. Thus, transition from TDCIS to TDA implemented EOM will allow lower computational cost if it can be implemented successfully. It should be noted that TDCIS, as opposed to RT-TDDFT, cannot have dissipation due to its equation of motion formalism. In order to study a real system, dissipation is essential. Thus, I attempted to implement TDA in RT-TDHF to keep the physically correct behavior of the Rabi oscillation while using the formulation of RT-TDDFT to allow implementation of dissipation for the later use.

2.3.1.2 Tamm-Dancoff Approximation(TDA)

For simplicity, I will be using the two-electron integral notation used in ref [67], where $\langle ab||cd \rangle = 2\langle ab|cd \rangle - \langle ab|dc \rangle$, for the restricted Hartree-Fock method. In the ground state, the Fock matrix F and density matrix γ is defined as shown below:

$$F_{pq}^0 = \delta_{pq}\epsilon_p \quad (2.29)$$

$$\gamma_{pq}^0 = \delta_{ij} \quad (2.30)$$

where p,q,r,s are general orbitals, i,j are occupied orbitals, and a,b are virtual orbitals. Under this condition, the change in density matrix γ over time is zero according to Equation 2.9. However, when an oscillatory perturbation is applied, the density matrix γ results in linear response γ' shown in Equation 2.31 and Equation 2.32, where the matrix d indicates a one-electron perturbation operator in the density matrix.

$$\gamma_{pq} = \gamma_{pq}^0 + \gamma'_{pq} \quad (2.31)$$

$$\gamma'_{pq} = \frac{1}{2}[d_{pq}e^{-i\omega t} + d_{pq}^*e^{i\omega t}] \quad (2.32)$$

First-order changes in the Fock matrix are in the one-electron part, g_{pq} , and in two-electron part induced by the linear response of the density matrix, $\frac{\delta F_{pq}}{\delta \gamma_{rs}} = \langle pq || rs \rangle$, as shown in Equation 2.33:

$$F_{pq} = F_{pq}^0 + g_{pq} + \langle pq || rs \rangle \gamma'_{rs} \quad (2.33)$$

$$g_{pq} = \frac{1}{2} [f_{pq} e^{-i\omega t} + f_{pq}^* e^{i\omega t}] \quad (2.34)$$

where f represents the one-electron perturbation operator. Substituting F_{pq} and γ_{pq} to Equation 2.9 and multiplying by $e^{-i\omega t}$ and $e^{i\omega t}$ factor results in Equation 2.35:

$$\omega d_{pr} = \sum_q [F_{pq}^0 d_{qr} - d_{pq} F_{pr}^0 + (f_{pq} + \sum_{st} \langle pq || st \rangle d_{st}) \gamma_{pr}^0 - \gamma_{pq}^0 (f_{pr} + \sum_{st} \langle qr || st \rangle d_{st})] \quad (2.35)$$

and the conjugate complex of the above equation, respectively. In TDHF, occupied-occupied and virtual-virtual elements in the d matrix are restricted to 1 and 0, respectively, due to the idem-potency condition of the density matrix[68]. Thus, only virtual-occupied and occupied-virtual elements of d matrix, x_{ai} and y_{ai} are non-zero. Using the diagonal nature of the Fock matrix and the density matrix in spin orbitals, Equation 2.35 can be converted into a pair of equations, Equation 2.38 and Equation 2.38:

$$d_{ai} = x_{ai} \quad (2.36)$$

$$d_{ia} = y_{ai} \quad (2.37)$$

$$\omega x_{ai} = F_{aa}^0 x_{ai} - x_{ai} F_{ii}^0 + (f_{ai} + \sum_{bj} (\langle aj || ib \rangle x_{bj} + \langle aj || bi \rangle y_{bj})) \gamma_{ii}^0 \quad (2.38)$$

$$\omega y_{ai} = F_{ii}^0 y_{ai} - y_{ai} F_{aa}^0 - \gamma_{ii}^0 (f_{ia} + \sum_{bj} (\langle ij || ab \rangle x_{bj} + \langle ij || ba \rangle y_{bj})) \quad (2.39)$$

Assuming a negligible perturbation in electronic transition, or $f_{ia} = f_{ai} = 0$, the eigenvalue problem is obtained as shown in Equation 2.40.

$$\begin{bmatrix} A & B \\ B^* & A^* \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \omega \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \quad (2.40)$$

In TDHF, Equation 2.40 is solved to obtain the excitation energy ω with the two-electron part in block A and B defined as in Equation 2.41 and Equation 2.42:

$$A_{ai,bj} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + \langle aj||ib \rangle \quad (2.41)$$

$$B_{ai,bj} = \langle ab||ij \rangle \quad (2.42)$$

In TDA, occupied-virtual elements of d matrix or y_{ai} are not considered. This nullifies Equation 2.39 and converts Equation 2.38 to Equation 2.43, leading to simpler eigenvalue problem as shown in Equation 2.44. :

$$\omega x_{ai} = F_{aa}^0 x_{ai} - x_{ai} F_{ii}^0 + (f_{ai} + \sum_{bj} (\langle aj||ib \rangle x_{bj})) \gamma_{ii}^0 \quad (2.43)$$

$$\omega X = AX \quad (2.44)$$

The TDA implementation of TDDFT is explained in more detail by Hirata et al.[68]. There are several assumptions made in TDHF/TDA for solving Equation 2.44 and the excitation energy ω , which has the same value from TDCIS. The most important assumption made by TDA was y_{ai} being not considered, which resulted in matrix B to be deleted from the final equation. In my TDA implementation of TDHF, TDHF removed the matrix B factor from EOM or included everything except the matrix B factor. TDA calculates the excitation energy ω in TDCIS, which has the correct excitation and relaxation process under resonant electric field. Thus, including the TDA implementation in RT-TDHF may allow the correct propagation

of electron density under resonant electric field.

2.3.1.3 TDA Implementation to RT-TDHF

The first attempt at this implementation was to apply the assumption of TDA to the EOM of TDHF by subtracting the $Y(t)$ matrix to eliminate y_{ai} contribution, as shown in Equation 2.45.

$$i\hbar\left(\frac{d\gamma}{dt}\right) = [F(t), \gamma(t)] - Y(t) \quad (2.45)$$

$$Y(t) = \begin{bmatrix} 0 & \sum_{bj} \langle ab || ij \rangle \gamma_{jb}(t) \gamma_{ij}(t) \\ \sum_{bj} \langle ab || ij \rangle \gamma_{jb}(t) \gamma_{ij}(t) & 0 \end{bmatrix} \quad (2.46)$$

The Equation 2.45 was used to propagate a hydrogen molecule with bond length of 0.75 Å for the H-H bond and the HF/cc-pvtz method. After the system undergoes the excitation by the electric field pulse, the system was propagated for 25000 iterations with a timestep of 0.02 a.u.. The dipole moment of the system over time was Fourier transformed to frequency, represented as the red curve in Figure 2.24. The system, using linear response TDDFT features in Q-Chem, has a ω of 13.4 eV and 13.5 eV for RPA and CIS excitation energy, respectively. As shown in Figure 2.24, the agreement of the red curve and CIS excitation energy peak shows TDA assumption as described in Equation 2.45 successfully removed virtual-occupied coupling, and the EOM has the frequency of TDHF/TDA.

Then the system was propagated under a continuous wave of an electric field at a resonant frequency of 13.5617 eV with an amplitude of 0.02 a.u.. The propagation of the electron density population is shown in Figure 2.21. Ideally, the population of HOMO (black) and LUMO (orange) should exhibit a Rabi oscillation, where the transition from excitation to relaxation occurs at the singly excited state and transition from relaxation to excitation occurs at the ground state. Because the method

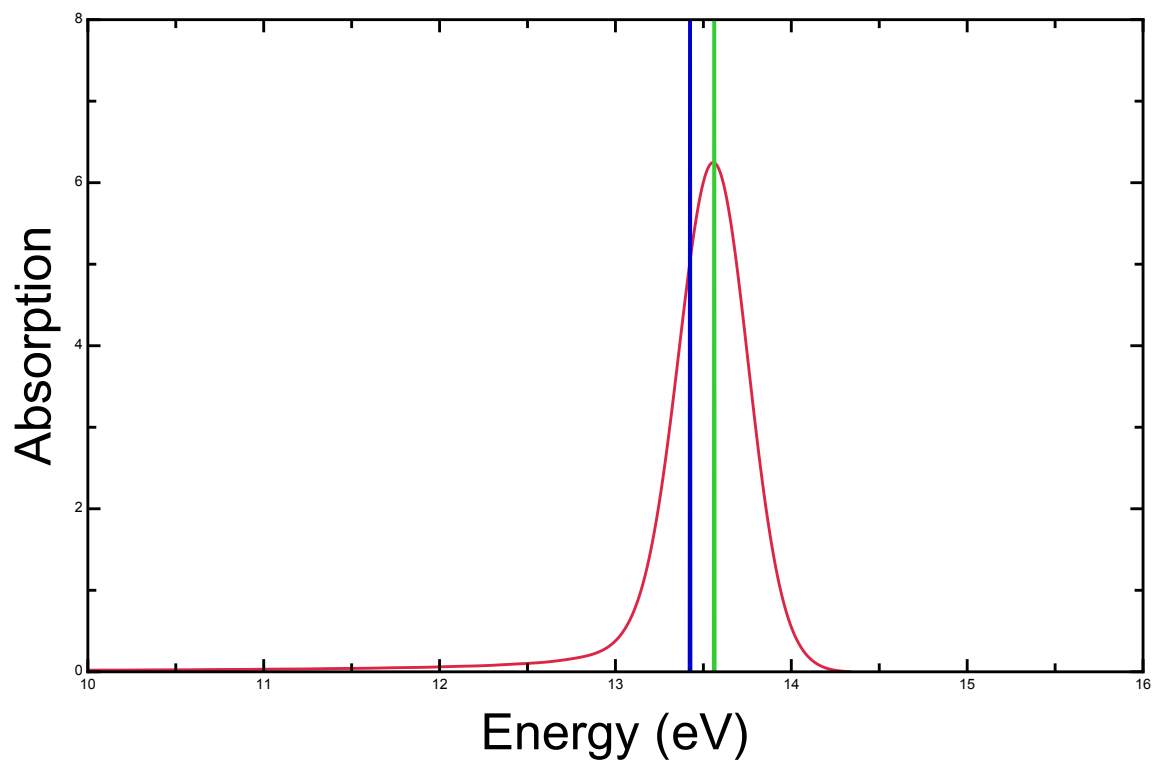


Figure 2.20. Absorption spectrum of H₂ with 0.75 angstrom H-H bond, cc-pvtz/HF method. Absorption from TDTDA (red), CIS excitation energy (green), RPA excitation energy (blue)

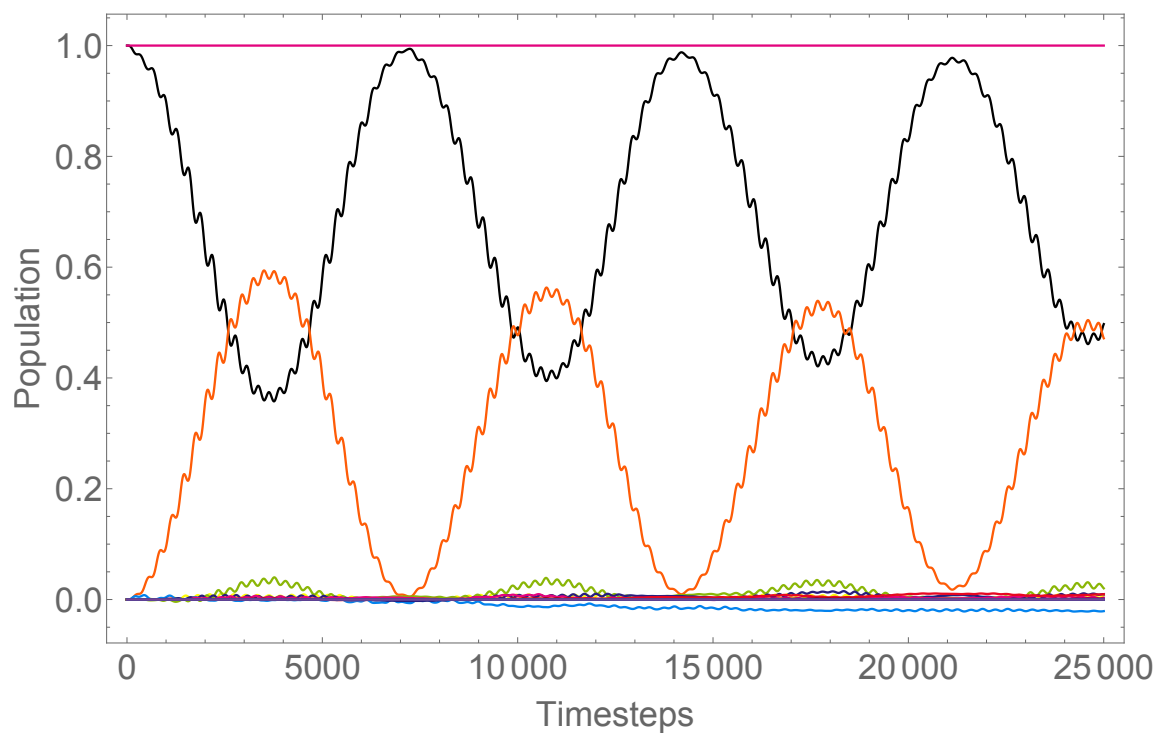


Figure 2.21. Time Propagation of H_2 molecule under continuous wave at resonant frequency. Population of LUMO(black), HOMO(orange), HOMO-5(light blue) and all orbitals (violet). (HF/cc-pvtz method. 0.02 a.u. electric field amplitude. 13.5617 eV resonant frequency. 25000 iterations with 0.02 a.u. timestep.)

used for the propagation is the restricted Hartree-Fock method, the maximum population of the HOMO is 0.5 when the molecule is in its singlet excited state. However, as shown in Figure 2.21, the population of LUMO exceeds 0.5 around timesteps of 4000, 11000, 18000 and 24000. Although the system starts the propagation from the ground state, the relaxation does not fully reach the ground state as shown in Figure 2.21 where its end of the relaxation process around timesteps of 21000 has a population of the HOMO(black) below 1. This is due to the size of the timestep of the propagation, which is too large. As shown in Figure 2.22, when the system is propagated under the identical conditions using RT-TDHF at its resonant frequency of 13.4241 eV, the system fails to relax to the ground state. When the size of the timestep is halved to 0.01 a.u., the system relaxes closer to ground state as shown in Figure 2.23, which implies that the error can be addressed using a smaller step-size. The step-size value of 0.02 a.u. did not change since the problem of relaxation starts to appear after a few Rabi oscillation cycles.

However, the negative population of HOMO-5, or the blue curve in Figure 2.21, is a problem occurring due to the $Y(t)$ term in Equation 2.45. As shown in Figure 2.22 and Figure 2.23, there is no sign of negative population using RT-TDHF. This is due to the indirect changes in the virtual-occupied and occupied-virtual block of density matrix. As shown in Equation 2.45, $Y(t)$ does not directly interfere with the diagonal element of the density matrix. The expanded equation of motion for TDHF (Equation 2.47) demonstrates why $Y(t)$ affects the density. $Y(t)$ only modifies the γ_{ov} and γ_{vo} block of the density matrix during propagation. As shown in Equation 2.48 and Equation 2.51, the modified γ_{ov} and γ_{vo} affects the γ_{oo} and γ_{vv} .

$$i\hbar\left(\frac{d\gamma}{dt}\right) = \begin{bmatrix} i\hbar\left(\frac{d\gamma_{oo}}{dt}\right) & i\hbar\left(\frac{d\gamma_{ov}}{dt}\right) \\ i\hbar\left(\frac{d\gamma_{vo}}{dt}\right) & i\hbar\left(\frac{d\gamma_{vv}}{dt}\right) \end{bmatrix} = [F(t), \gamma(t)] \quad (2.47)$$

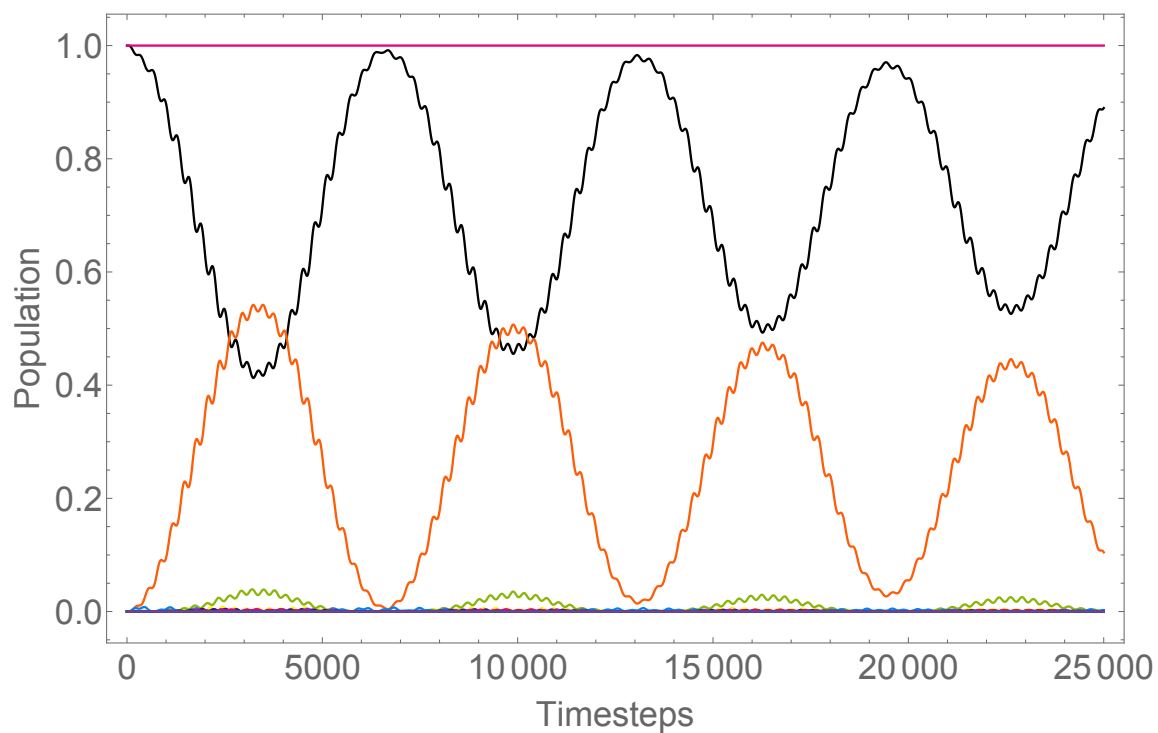


Figure 2.22. Time Propagation of H₂ molecule under continuous wave at resonant frequency using RT-TDHF. Population of LUMO(black), HOMO(orange), HOMO-5(light blue) and all orbitals (violet). (HF/cc-pvtz method. 0.02 a.u. electric field amplitude. 13.4241 eV resonant frequency. 25000 iterations with 0.02 a.u. timestep.)

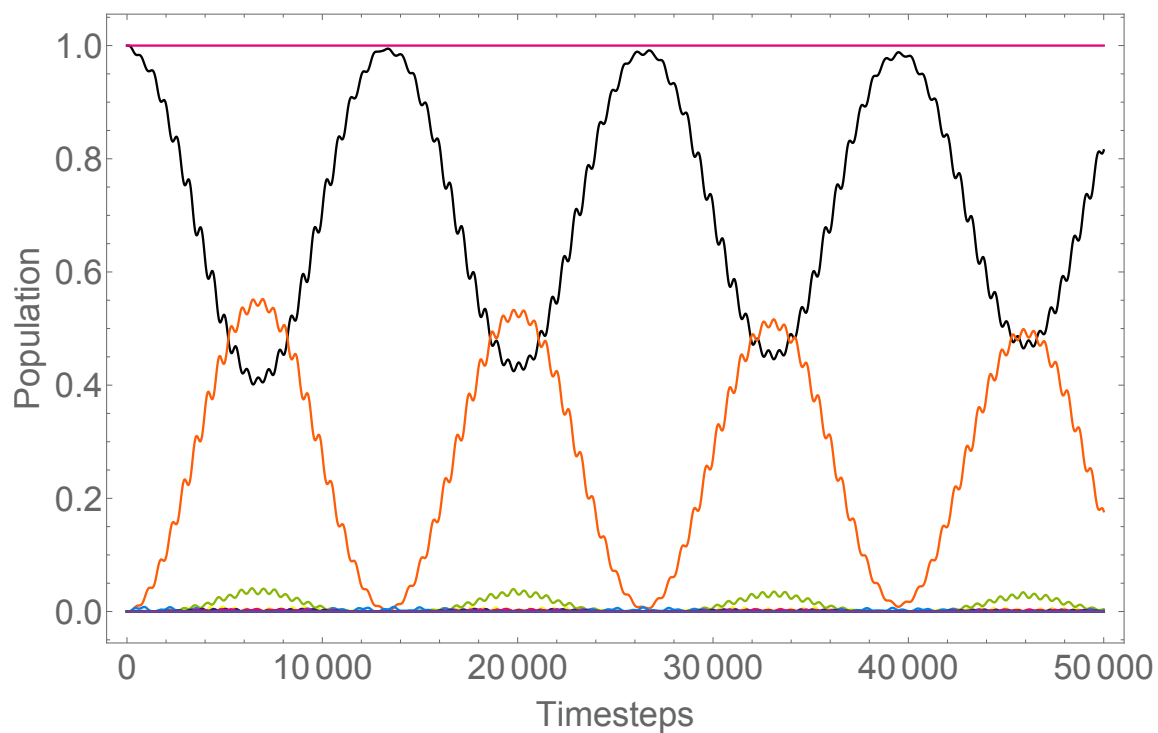


Figure 2.23. Time Propagation of H_2 molecule under continuous wave at resonant frequency. Population of LUMO(black), HOMO(orange), HOMO-5(light blue) and all orbitals (violet). (HF/cc-pvtz method. 0.02 a.u. electric field amplitude. 13.4241 eV resonant frequency. 50000 iterations with 0.01 a.u. timestep.)

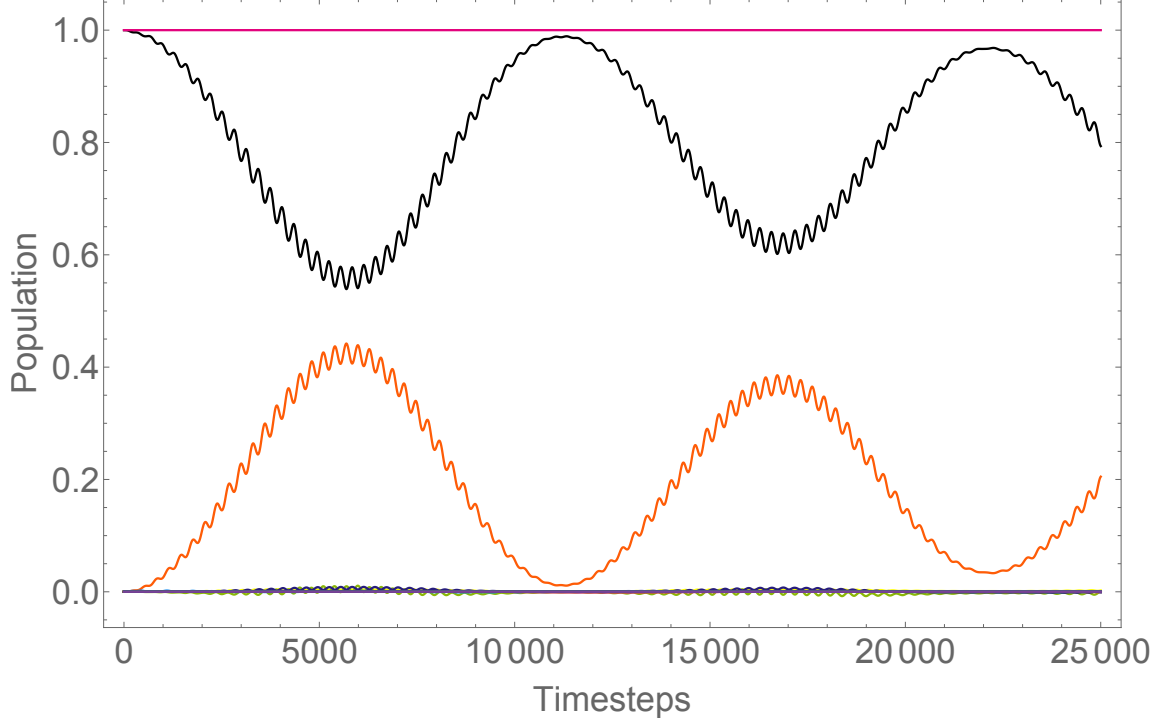


Figure 2.24. Time Propagation of H₂ molecule under continuous wave at resonant frequency. Population of LUMO(black), HOMO(orange), HOMO-5(light blue) and all orbitals (violet). (HF/cc-pvtz method. 0.01 a.u. electric field amplitude. 13.5617 eV resonant frequency. 25000 iterations with 0.02 timestep.)

$$i\hbar\left(\frac{d\gamma_{oo}}{dt}\right) = [F_{oo}(t)\gamma_{oo}(t) + F_{ov}(t)\gamma_{vo}(t)] - [\gamma_{oo}(t)F_{oo}(t) + \gamma_{ov}(t)F_{vo}(t)] \quad (2.48)$$

$$i\hbar\left(\frac{d\gamma_{ov}}{dt}\right) = [F_{oo}(t)\gamma_{ov}(t) + F_{ov}(t)\gamma_{vv}(t)] - [\gamma_{oo}(t)F_{ov}(t) + \gamma_{ov}(t)F_{vv}(t)] \quad (2.49)$$

$$i\hbar\left(\frac{d\gamma_{vo}}{dt}\right) = [F_{vo}(t)\gamma_{oo}(t) + F_{vv}(t)\gamma_{vo}(t)] - [\gamma_{vo}(t)F_{oo}(t) + \gamma_{vv}(t)F_{vo}(t)] \quad (2.50)$$

$$i\hbar\left(\frac{d\gamma_{vv}}{dt}\right) = [F_{vo}(t)\gamma_{ov}(t) + F_{vv}(t)\gamma_{vv}(t)] - [\gamma_{vo}(t)F_{ov}(t) + \gamma_{vv}(t)F_{vv}(t)] \quad (2.51)$$

Equation 2.45 also has a problem with the maximum population of HOMO being dependent on the amplitude of electric field. In TDCIS, the change in the amplitude of the electric field leads to the change in the frequency of the Rabi oscillation only[20]. As shown in Figure 2.24, when the amplitude of the electric field was halved, the

maximum population of HOMO changed as well. These observations indicate that Equation 2.45 is not sufficiently accurate to reproduce TDCIS-like propagation. The negative population in Figure 2.21 shows that $Y(t)$, which changes only in γ_{ov} and γ_{vo} , also requires changes in γ_{oo} and γ_{vv} that counteract the indirect influences such that the diagonal element of density matrix holds positive value.

The problem of HOMO population exceeding 0.5 is due to changes in the eigenvalues ϵ_p of the Fock matrix as explained in the introduction section. During the excitation process, the change in populations in the density matrix leads to the change in $F(t)$ and changes the resonant frequency[65]. TDCIS also has ϵ_p dependency in its equation of motion as shown in Equation 2.28, but ϵ is constant throughout the process as the Fock matrix is not solved during propagation. Thus, the TDA approximation cannot be applied to EOM of TDHF without significant modifications. Although TDA has the same ω as TDCIS, because TDA also assumes ground state population during the excitation process, the basic assumption is violated and the excitation frequency shifts.

$$i\hbar\dot{\gamma} = [F(0), \gamma(t)] + \begin{bmatrix} \dot{\gamma}_{ij} & \dot{\gamma}_{ia} \\ \dot{\gamma}_{ai} & \dot{\gamma}_{ab} \end{bmatrix} + [E(t), \gamma(t) - \frac{1}{2}\gamma(0)] \quad (2.52)$$

$$\begin{bmatrix} \dot{\gamma}_{ij} & \dot{\gamma}_{ia} \\ \dot{\gamma}_{ai} & \dot{\gamma}_{ab} \end{bmatrix} = \begin{bmatrix} \sum_{o,v} \langle jv || vo \rangle \gamma_{i,o} + c.c & - \sum_{o,v} \langle ao || iv \rangle \gamma_{o,v} \\ - \sum_{o,v} \langle ao || iv \rangle \gamma_{v,o} & - \sum_{o,v} \langle bo || av \rangle \gamma_{a,v} + c.c \end{bmatrix} \quad (2.53)$$

Attempts to avoid the problem with resonant frequency shifts by using $F(0)$ for the propagation while adding terms separately to sub-blocks of γ that corresponds to TDCIS were explored. The equation of motion that simulates the Rabi oscillation is shown in Equation 2.52, where TDCIS terms in Equation 2.53 are added to the EOM of ground state Fock matrix $F(0)$ and modified field contribution are added. The population propagation in orbitals is shown in Figure 2.25, where the problem

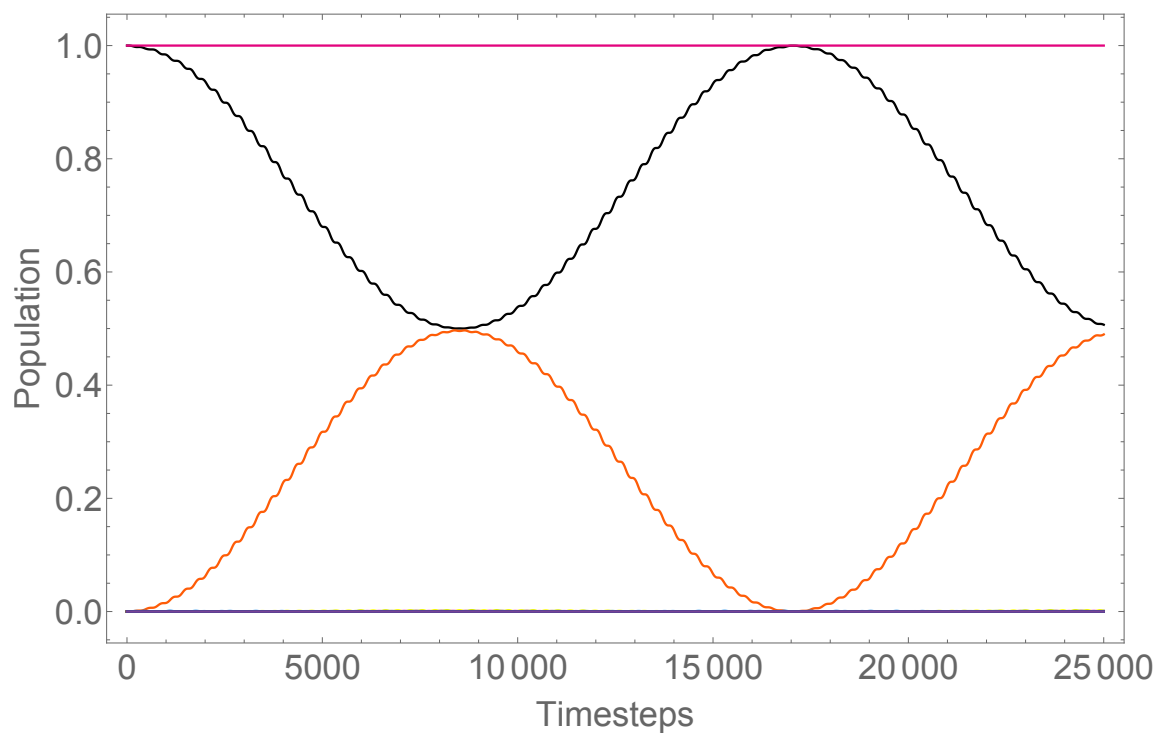


Figure 2.25. Time Propagation of H_2 molecule under continuous wave at resonant frequency. Population of LUMO(black), HOMO(orange), HOMO-5(light blue) and all orbitals (violet). (HF/cc-pvtz method. 0.01 a.u. electric field amplitude. 13.5617 eV resonant frequency. 25000 iterations with 0.02 timestep.)

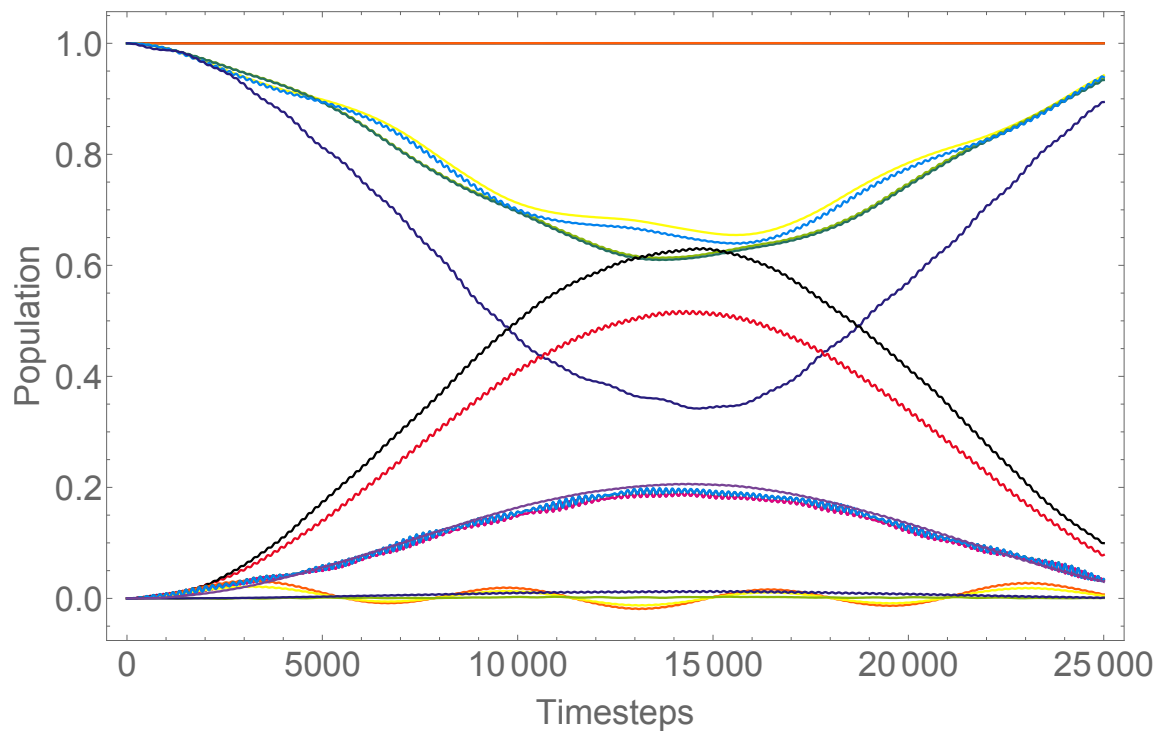


Figure 2.26. Time Propagation of N₂ molecule under continuous wave at resonant frequency. (HF/6-31G method. 0.01 a.u. electric field amplitude. 24.6330 eV resonant frequency. 25000 iterations with 0.02 timestep.)

with maximum population of virtual orbitals and negative population is not observed. However, Equation 2.52 is only applicable to a system that has one occupied or one virtual orbital. When tested on N₂ with a 6-31G basis set on its resonant frequency of 24.6330 eV, the population became negative during propagation as shown in Figure 2.26. This implies that there is a missing term in Equation 2.52 between off-diagonal elements of γ_{oo} and γ_{vv} . The work therefore was not further pursued.

CHAPTER 3

USING NEURAL NETWORKS IN CHEMISTRY RESEARCH AND EDUCATION

3.1 Exploring Bullvalene Conformers Using Neural Network Force Field Molecular Dynamics

3.1.1 Introduction

With the availability of powerful performance of modern GPUs, the application of machine learning (ML) has been expanded to the field of chemistry. Researchers have used ML to predict reaction products[69, 70], bond energies[71], partial charges[72] and other electronic structure properties[73]. The Parkhill group released a software called TensorMol[74], a Python package that trains neural networks on DFT data to predict energies and its gradients to offer simulation tools such as geometry optimization and molecular dynamics using energies and gradients predicted by a trained neural network (NN).

With the promising results of NN in predicting the energy and gradient[74], I attempted to use a neural network trained force field (NNFF) to predict a reaction via molecular dynamics. Mechanistic studies of reactions using DFT requires the chemist to predict its transition states, intermediates, and products and verify their prediction by experiment. DFT transition state structure calculation has been proven to be very accurate for studies of organic reaction mechanisms[75]. However, while the calculated results on the predicted reaction pathway using DFT are accurate, it does not predict a product when given reactant and reagents and when there is a

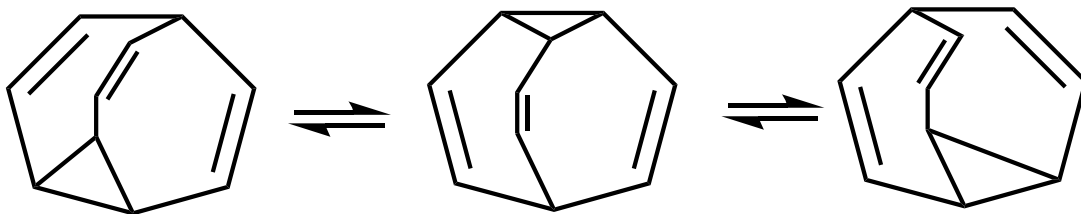


Figure 3.1. Structure of Bullvalene. Bullvalene has 1,209,600 degenerate states.[76]

possibility of a side reaction that chemist could not predict.

Ideally, a chemist can run Ab Initio Molecular Dynamics (AIMD) of all reactants and reagents using the electronic structure methods and its calculated energy and gradient combined with thermostat to predict the reaction product. This method, which is purely based on theory and not the intuition of chemists, will provide an unbiased reaction product, but the computational cost of such an application is not feasible even with the modern technology of computational chemistry. However, with the availability of a NNFF that predicts the energy and the gradients with the accuracy of the DFT method at linear scaling[74], using NNFF to study the reaction mechanism may offer an accurate and cheap tool for chemists. In this chapter, I attempt to verify NNFF’s ability to predict the reaction product using molecular dynamics simulation under thermal environment.

Bullvalene, tricyclo[3.3.2.0^{2,8}]deca-3,6,9-triene, is a molecule that rapidly undergoes Cope rearrangement[76] as shown in Figure 3.1. Because the bond formation and breakage occurs intramolecularly via Cope rearrangement around room temperature [77] with activation energy of 13.1 kcal/mol[78] and having 1.2 million degenerate states[76], bullvalene is an interesting system, in which its fluxional behavior decreases with the number of substituents[79, 80]. Thus, given a multi-substituted bullvalene, well-trained NNFF may be able to predict a most stable isomer from any

other isomers.

3.1.2 Methods

Using the dataset that contains 1,187,690 molecules of 29,552,336 atoms with atom number of 1 to 36 and 53 and its potential energies, atomic forces and Mulliken charges calculated using Q-Chem[81] with the ω B97X-D/6-311G** method, the network was trained on energy, force and charge with learning rate of 0.0001 on GTX 1080. The network that uses elemental modes generated by auto-encoder[82] was trained to perform the molecular dynamics simulation. The python script used to train the network is available at <https://github.com/v3op01/DIS2020/tree/master/NN/BV/train.py>.

Once the network was trained, the energy and force of bullvalene undergoes MD simulation for 20,000 iterations with 0.05 fs timestep using Nose thermostat[83], a built-in feature of TensorMol. The time propagation of energy and molecular coordinates was studied for bond formation and breakage. The python script used to run MD simulation is available at <https://github.com/v3op01/DIS2020/tree/master/NN/BV/sim.py>

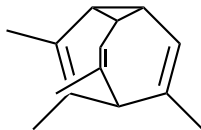


Figure 3.2. Structure of
3,4,6,10-tetramethyltricyclo[3.3.2.0^{2,8}]deca-3,6,9-triene

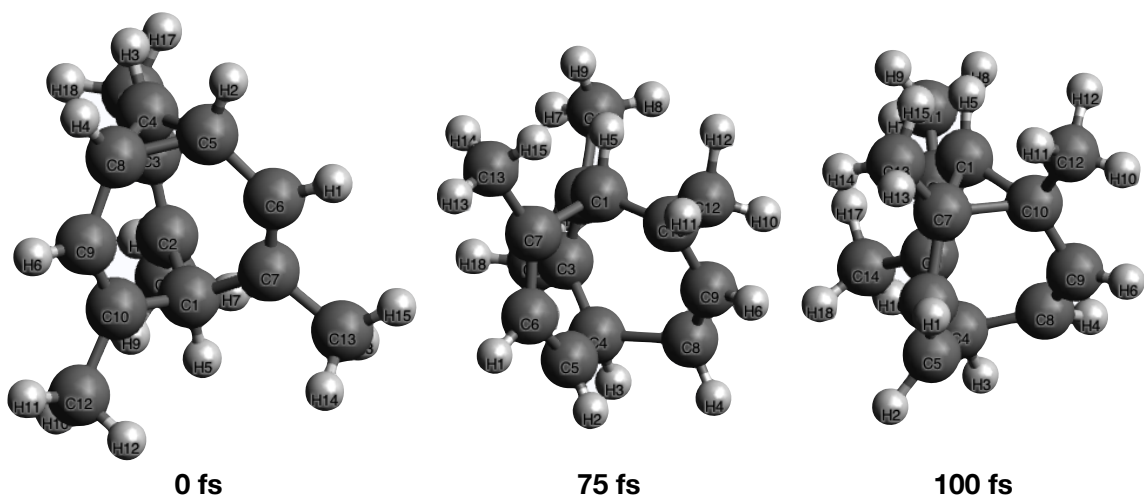


Figure 3.3. Simulation of 3,4,6,10-tetramethyltricyclo[3.3.2.0^{2,8}]deca-3,6,9-triene at 2000K

3.1.3 Results and Discussion

When the temperature was set to 300 K, the temperature that bullvalene undergoes rearrangement[77], no bond breakage or formation was observed during the simulation time of 2 ns. It was only when the temperature was set to 2000 K that the rearrangement was observed during the simulation via bond breakage and formation. Thus, using temperature setting of 2000K, the simulation was then tested on 3,4,6,10-tetramethyltricyclo[3.3.2.0^{2,8}]deca-3,6,9-triene shown in Figure 3.2.

As shown in Figure 3.3, the molecule started with the C4-C5-C8 forming 3-membered ring. Then, C5-C8 bond breaks around 1500 timesteps (75.0 fs), leading to C7-C10 bond formation around 2000 timesteps (100.0 fs), forming C1-C7-C10 3-membered ring. Then around 11440 iterations, the molecule reforms C4-C5-C8 3-membered ring, which then C4-C5 bond breaks around 15000 timesteps and C2-C7 bond forms around 16000 timesteps, resulting in C1-C2-C7 3-membered ring. During 20000 iterations of simulation, total of 3 transition state geometries and 4 ground state geometries were observed, which is expected based on the activation energy and

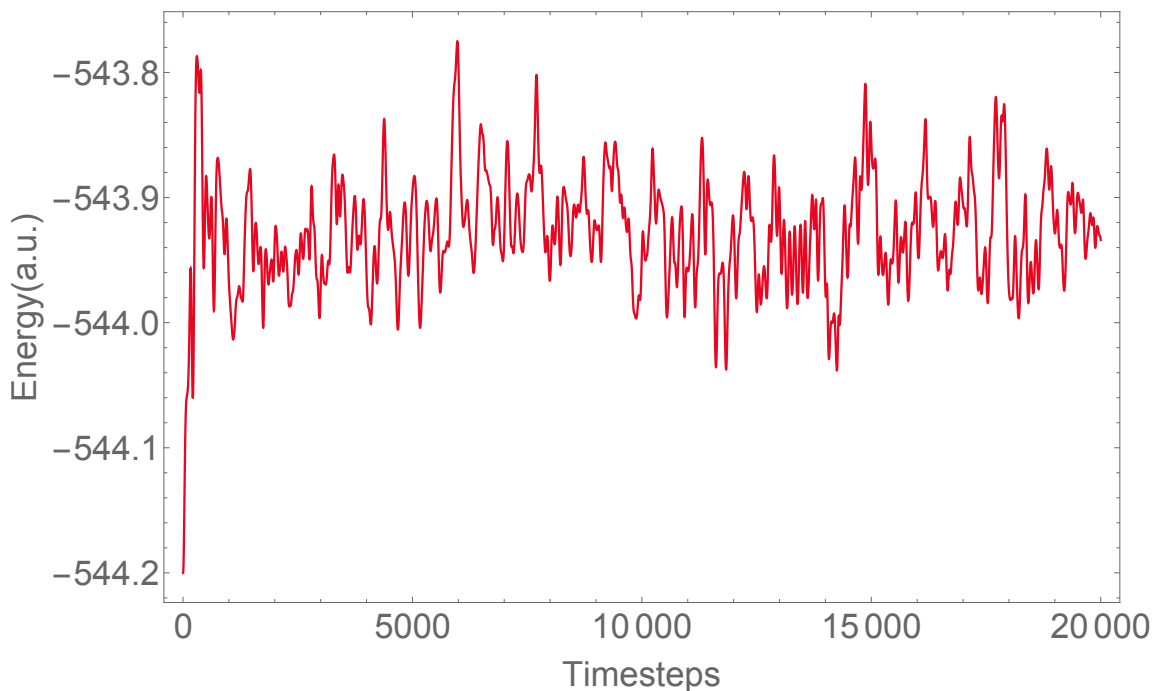


Figure 3.4. Energy propagation of 3,4,6,10-tetramethyltricyclo[3.3.2.0^{2,8}]deca-3,6,9-triene at 2000K with 0.05 fs timestep

simulation time.

However, due to the temperature setting of 2000K, the energy of a molecule did not reflect these observation nor provide any insight as shown in Figure 3.4. The potential energy when the bond breaks around 1500 timesteps is a local maximum and the potential energy when the bond forms around 2000 timesteps is a local minimum. However, there exists more than 3 local maxima and 4 local minima that are not associated with bond formation and bond breakage. This could be due to high temperature setting which distorts the molecule in various shapes. Thus, a longer simulation time at lower temperatures is needed since too much noise is observed at these temperature. In order to observe the transition at 300 K, approximately 530,000 ns of MD simulation will be required [84].

3.1.4 Conclusion

I used NNFF to investigate its ability to predict reaction mechanism and the reaction product via MD simulation. Bullvalene was selected as a test model as it breaks and forms bond continuously at room temperature. However, my trained NNFF was not able to break and form bond until the temperature was set to 2000K, which made interpreting energies of simulation difficult.

In order to improve the neural network, large amounts of data are always needed. Unfortunately, although I have used a dataset containing more than 1 million molecules, most of the molecules in the dataset used were in or close to their optimized geometry coordinates. Since the network has a low amount of data on transition states of molecules, it had difficulties breaking the bonds at room temperature. Including datasets near the transition states of molecules would be crucial in training the network for predicting the reaction mechanism and product. However, the transition state search algorithm is more expensive than the ground state optimization algorithm since transition state search requires the Hessian matrix to be calculated. Thus, it would be cheaper to randomly collect data from using meta-dynamics[85]. This work was not further pursued.

3.2 Using Neural Networks to Enhance Learning in Undergraduate Organic Chemistry Courses

3.2.1 Introduction

The flipped classroom is a new and popular instructional model, in which the classroom material is shifted to home activities, and homework and projects are shifted to the classroom[86]. Flipped classrooms are an effort to transform a traditional teacher-centered classroom, where students passively listen to information[87], to student-centered classroom, where active-learning, peer-assisted learning and collaborative learning is encouraged[86–88]. The studies with flipped classroom reported not only its advantages of improved learning performance[89], increased engagement[90] and flexible learning[91], but also its challenges of increased time consumption for the preparation of the classroom for both student[92] and teachers[93–95], difficulties of reaching out for help out of class[96], and lack of access to technologies[97].

With the improved technologies, many attempts to implement these technologies into chemistry courses have been done[98, 99]. Through computational models such as calculation of bond dissociation energies, bond distances and MO visualization, students gain enhanced understanding of the molecular and electronic structures of molecules[100]. Without the accessibility to such tools, students cannot benefit from improved technologies. Furthermore, the speed of the computations needs to be fast enough to allow interactive learning. In this chapter, I will outline a website that can be used to enhance student learning experience by providing a thermodynamics database, 3D visualization tools and prediction of reaction products with ML.

3.2.2 Methods

In the current and first phase, the goal is to build a website that is capable of building a database for further work. I used django[101], a Python-based free and

open-source web framework, to create the website. Python is used to manage the overall server, JavaScript is used to handle input, and the Html language is used to build the framework of the webpage.

3.2.3 Outline

As shown in Figure 3.5, the general web framework is composed of submission and database access. Each datapoint in the database is a molecule with a SMILES string as an ID. Once the ID is stored, 3D coordinates of the molecule can be generated using either RDKit[102] or OpenBabel[103] or be imported from existing databases such as PubChem[104]. Then, using PySCF[105], the geometry is optimized and the energy calculation results are stored in the database. The database is specifically targeted for molecules covered in undergraduate organic chemistry.

A script that automatically builds a database while allowing individual submission from the webpage will be implemented. In order to avoid duplicate data points, the SMILES strings will be used as an input format. For the generation of SMILES strings, Marvin JS[106], which allows users to draw 2D structures of molecule and export the SMILES string, will be used. For the 3D visualizer of the database, I plan to use JSMol[107], a JavaScript version of Jmol[108].

In the next phase, the database will be expanded for sets of concerted reactions such as SN2, Diels-Alder, and Claisen rearrangement. The transition states in these reactions will be calculated and collected. Using visual tools, the energy profile with reactant, energy barrier, and product will be provided, which will allow students to understand thermodynamics and the kinetic effects on the reaction path by providing the calculated result.

The last phase will be using the database to build a NNFF that is specialized in undergraduate organic chemistry, which requires additional datapoints using metadynamics [85]. The NNFF can be used to calculate energy of molecules that students

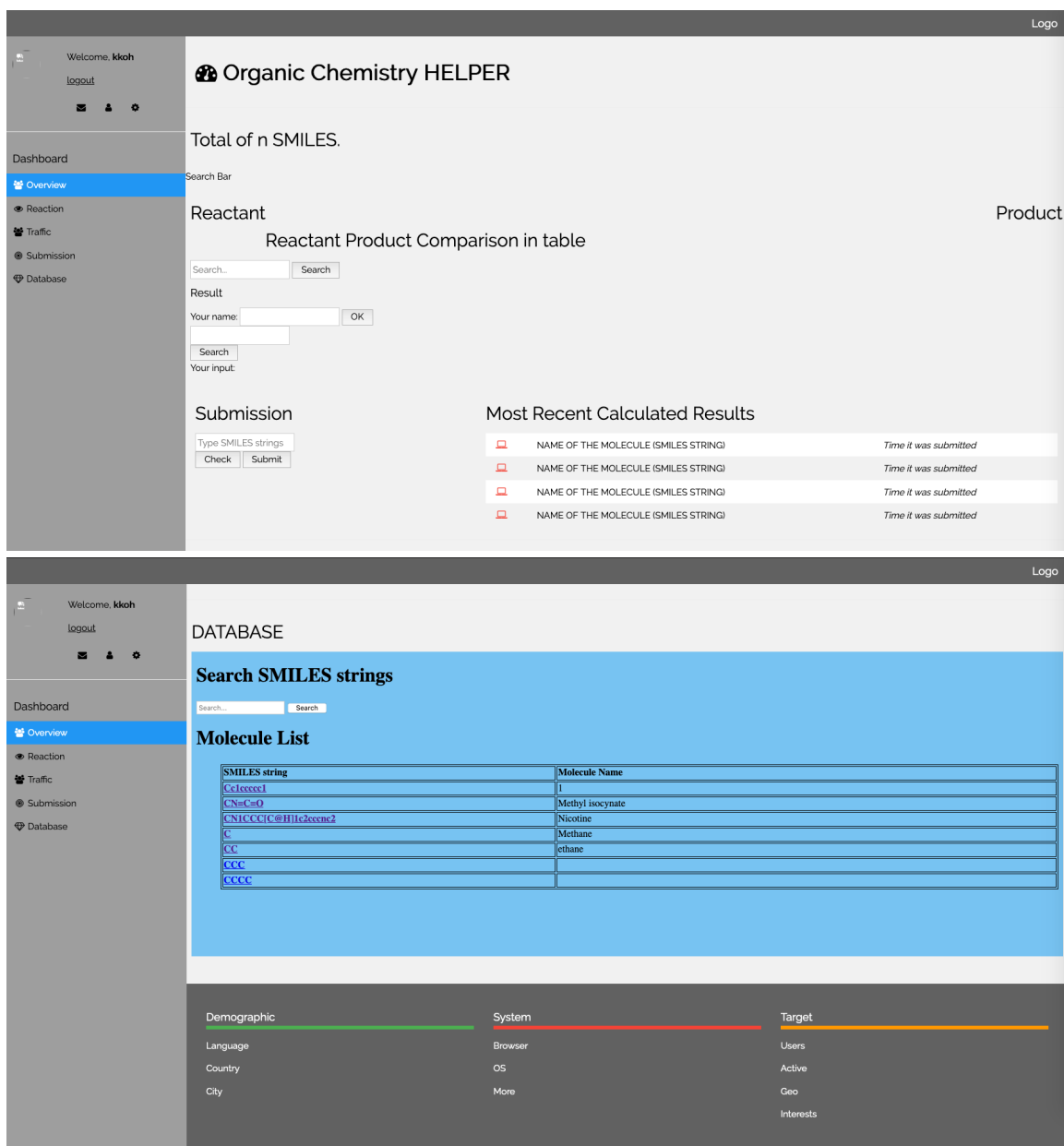


Figure 3.5. Website under development

enter, which may not be present in the database. Due to computational cost of DFT, students may not be provided with the information in an interactive manner. However, NNFF can provide students the information in time for them to use the tool in an interactive fashion.

The database of the calculated ground state energies of molecules can also be used to train a graph convolutional neural network that predicts organic outcomes[109]. When a student asks about a reaction outcome using existing reaction with a heavily modified reactant, I currently search for actual data from experiment manually to give student a feedback. By reducing the dataset to molecules of interest in undergraduate organic chemistry and by providing additional information, such as ground state energies, the trained network will perform better to predict the reaction outcome for the organic chemistry course and be useful tool for students to use to quickly check with questions regarding reaction outcomes.

3.2.4 Conclusion

The database, which will be available online, will be easily accessible to students and can be used to enhance a student’s learning experience by providing the numerical data for reaction of interest to undergraduate organic chemistry. When students are given a set of similar reactions from the database, they will recognize and locate the reaction site. By submitting a reaction and observing the feasibility of the submitted reaction, students can apply the knowledge and get evaluated quickly. Lastly, students can design a synthesis of a molecule and test it in laboratory, which will connect the theory and experiment for the students. Besides the database, simplicity and the visual aspects will be very important since no students will be willing to use tools that are not easy to use and visually captivating. The current version of the code for the website is available at <https://github.com/v3op01/DIS2020/tree/master/NN/TEACH/ochelper>.

CHAPTER 4

MECHANISTIC STUDIES OF THE REACTION OF SQUARINE DYES

4.1 Formation of Oxindole Derivatives from the Squaraine Dyes

4.1.1 Introduction

Squaraine dyes, a four-membered aromatic ring system derivative of squaric acid shown in Figure 4.1, are molecules that possess sharp and intense absorption bands in the red to near-infrared region[110]. The photo-physical properties of the dyes have been studied and find application as photo-receptors in copiers[111], photoconductors in organic solar cells[112], and sensors for metal ions[113] and for biological molecules[114]. The central ring can be attacked by nucleophiles as shown in Figure 4.2, which alters the photochemistry of the dyes. The reactivity and photochemical properties of derivatives of squaraine has been of interest to many chemists.

The Ashfeld group has investigated the nucleophilic addition of phosphorus(III) derivatives to dianiline squaraine dyes[115] and reported the structure, stability and reversible formation of zwitterionic squaraine-phosphine adducts as shown in Figure 4.3. The Ashfeld group continued their investigation on their work and

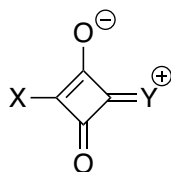


Figure 4.1. General Structure of Squaraine.

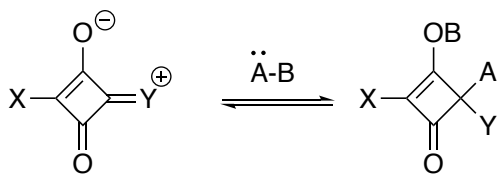


Figure 4.2. Nucleophilic attack of Squaraine Dyes

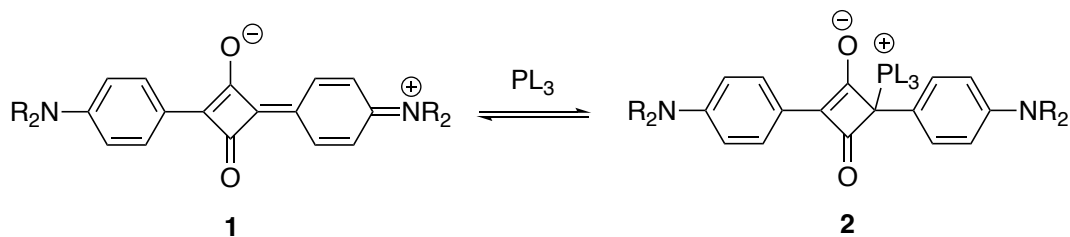


Figure 4.3. Addition of Phosphine Derivatives to Dianiline Squaraines

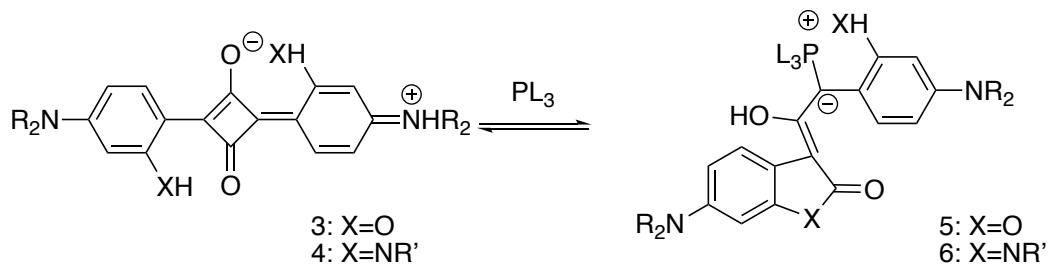


Figure 4.4. Formation of benzofuranone and oxindoles

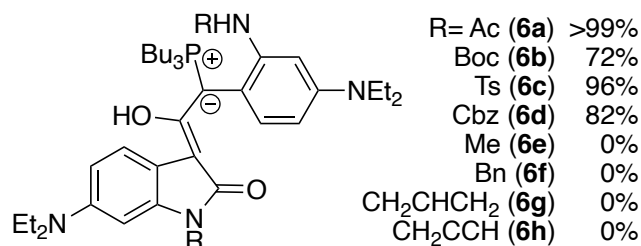


Figure 4.5. Structure of oxindole product from diaoniline squaraines and experimental yields

discovered the formation of benzofuranones and oxindoles from squaraine-phosphine adducts as shown in Figure 4.4. In the case of formation of oxindoles shown in Figure 4.5, many substituent they have studied, in case of R=acetate, the product is formed with a yield greater than 99%. They also observed reversion of **6a** to **4a** when the oxindole product was exposed to the acid, indicating that the reaction is reversible. In order to have better understanding of the reaction, I carried out a DFT study to understand the surprising stability of the zwitterionic products and determine the reaction pathway for formation of oxindole from squaraine dyes and the effect of substituents on oxindole ring on the experimental yield.

4.1.2 Methods

DFT calculations were performed with the B3LYP/6-311G+ method with a CPCM implicit solvent model and a dielectric constant of 4.8069 using Q-Chem[81]. DFT calculations for the thermodynamics studies were performed with the 6-311+G* basis set and B3LYP method, without dispersion correction, using Gaussian 16[116]. All geometry optimization was performed using a CPCM solvent model with parameters for chloroform unless noted otherwise. The transition state calculations were done using the QST2 method. The frequency calculation was performed at the transition state to ensure that the structure have only a single imaginary frequency correspond-

ing to the reaction coordinate.

4.1.3 Result and Discussions

4.1.3.1 Population Analysis of Oxindole Product

The Ashfeld group originally expected the protonation of the ylide with the exposure of acid, but instead discovered the reversion of the oxindole **6a** to squaraine dyes **4a**. This finding is surprising given the zwitterionic nature of **6a**. I performed geometry optimization and the population analysis on oxindole product **6a**. Using Avogadro[117] and population analysis output file, the electron potential surface of **6a** was visualized as shown in Figure 4.6, where the red regions represent negatively charged regions and blue regions represent positively charged regions. The result shows the negative charge across the oxindole ring, aniline ring and extended throughout the π -system, revealing that the inactivity of ylide to acid is due to lack of specific reaction site where the charge is localized. Instead, the negative charge is completely delocalized.

4.1.3.2 Mechanistic Study

I studied three distinct mechanistic hypotheses for the formation of **6a** shown in Figure 4.7. I proposed **INT 1** as the first intermediate due to dianiline squaraine dyes' reactivity with phosphine[115]. The existence of this intermediate has been confirmed experimentally by the Ashfeld group by x-ray crystallography. **INT D** is envisioned as the last intermediate before protonation that would lead to the final product **6a**. In between, we proposed three different pathways; Path A, where squaraine ring cleavage connects to formation of protonated oxindole ring via transition state **TS A**; Path B, where squaraine ring cleaves to ketene intermediate **INT 2** followed by formation of a protonated oxindole ring, and Path C, where formation of a protonated oxindole ring is followed by squaraine ring cleavage. In order to test

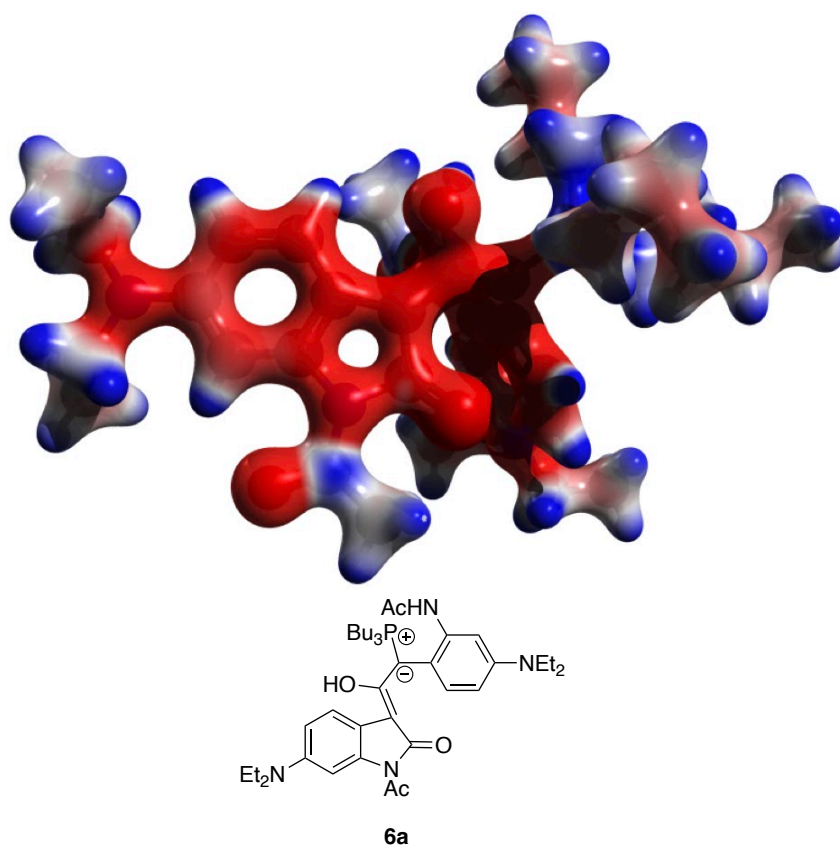


Figure 4.6. Electron potential surface of 4e. Red indicates negative charge. Blue indicates positive charge.

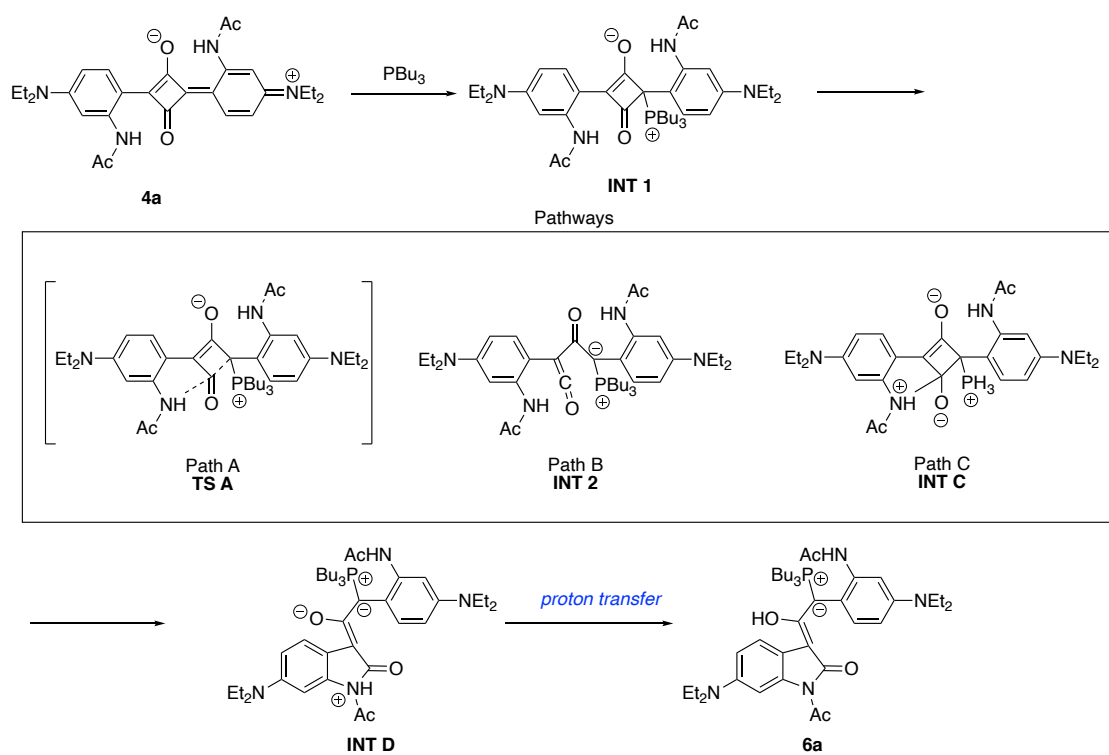


Figure 4.7. Mechanistic hypothesis for formation of **6a**

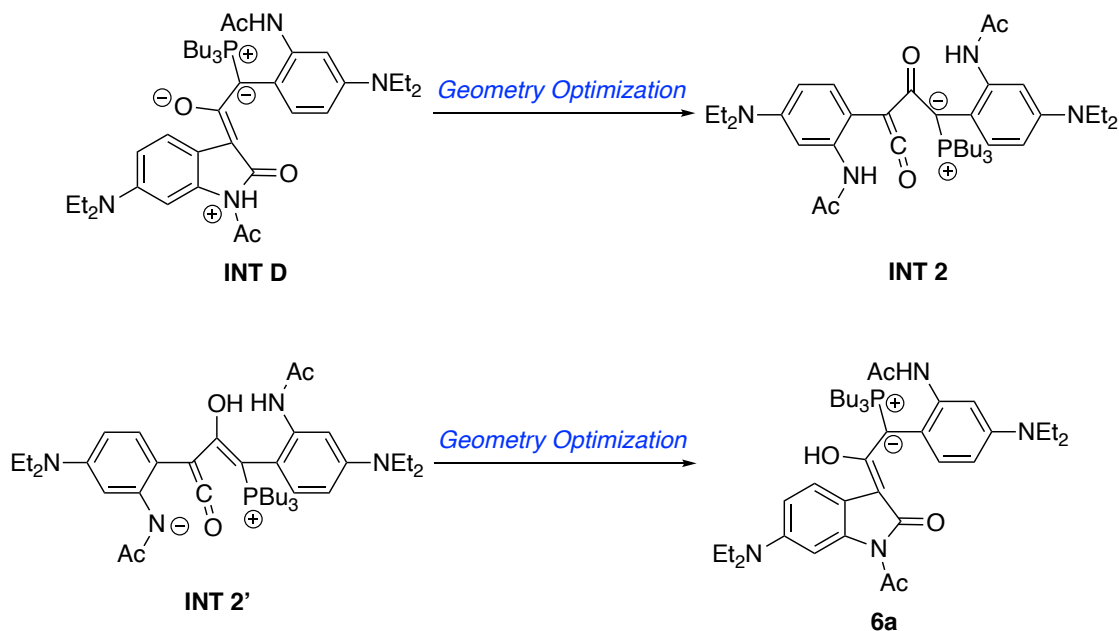


Figure 4.8. Result of Geometry optimization

path A, geometry optimization on **INT 1** and **INT D** was performed, followed by a transition state search using QST2 method for **TS A**.

However, as shown in Figure 4.8, geometry optimization on **INT D** led to **INT 2**, suggesting that path B is possible but also that the last intermediate **INT D** is highly unstable. Similarly, geometry optimization on deprotonated **INT 2**, **INT 2'** with C-N bond length of 2.76 Å, led to final product **6a**. In order to understand these observations, I visualized the molecular orbitals(MO) using Molden[118] as shown in Figure 4.9.

In Figure 4.9, MOs of **INT 2** and **INT 2'** are shown on the top and bottom, respectively. In order to form the oxindole ring, the nitrogen acts as a nucleophile

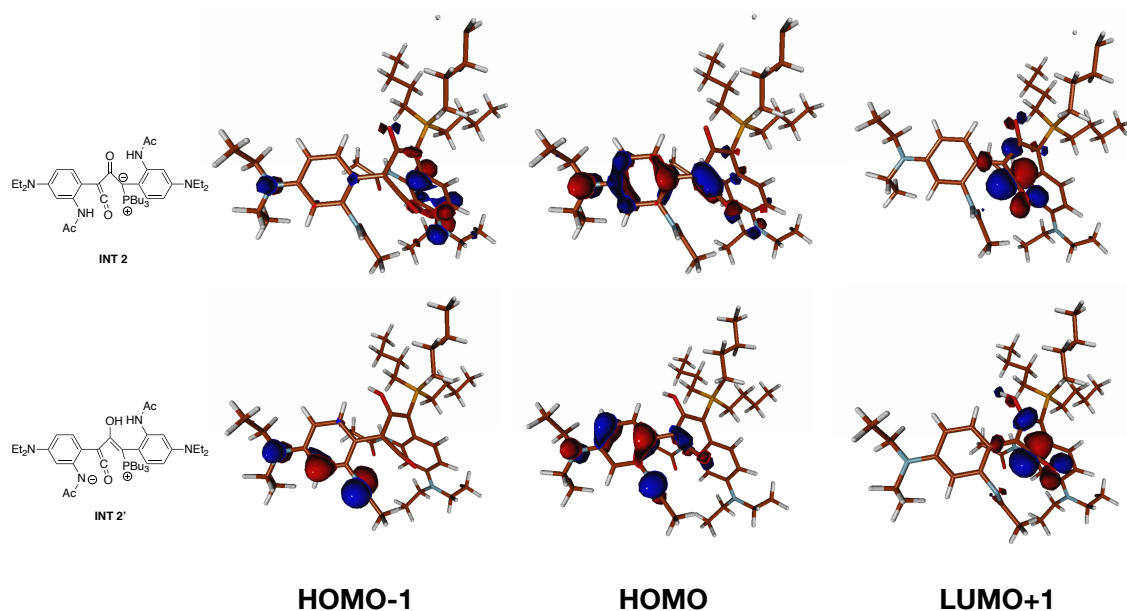


Figure 4.9. Molecular Orbitals of **INT 2** and **INT 2'**

to attack the carbonyl carbon on the ketene. In MOs of both **INT 2** and **INT 2'**, LUMO+1 indicates the electrophilic nature of the carbonyl carbon on the ketene. However, HOMOs of **INT 2** do not have orbitals around the acyl nitrogen. Thus, the protonated oxindole **INT D** is unlikely to form due to the nitrogen's lack of its nucleophilic nature. On the other hand, deprotonated nitrogen introduces HOMO and HOMO+1 around the nitrogen, which can overlap well with LUMO+1. Thus, geometry optimization will optimize **INT 2** to oxindole product **6a**.

I then performed potential energy scans (PES) of the C-N bond from 1.50 Å, corresponding to 1.47 Å C-N distance in **6a**, to 2.70 Å, corresponding to 2.76 Å C-N distance in **INT 2**, to search for local extrema, or a potential transition state. As shown in Figure 4.10, both scans shows no indication of a transition state. Thus, proton transfer on **INT 2** forms **INT 2'**, which will automatically lead to formation of **6a** as shown in Figure 4.10 with the downhill blue curve.

With these results, we proposed a new reaction pathway where squaraine dye **4a**

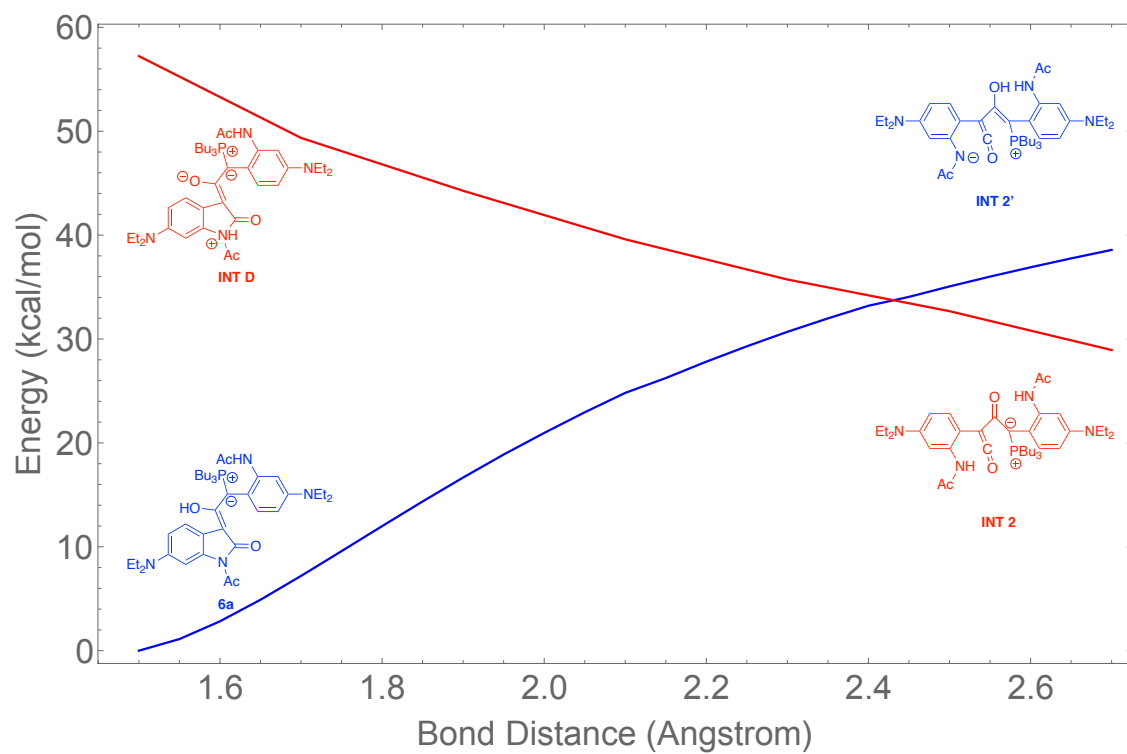


Figure 4.10. Potential energy scan of C-N bond for **INT 2**(red) and **6a**(blue)

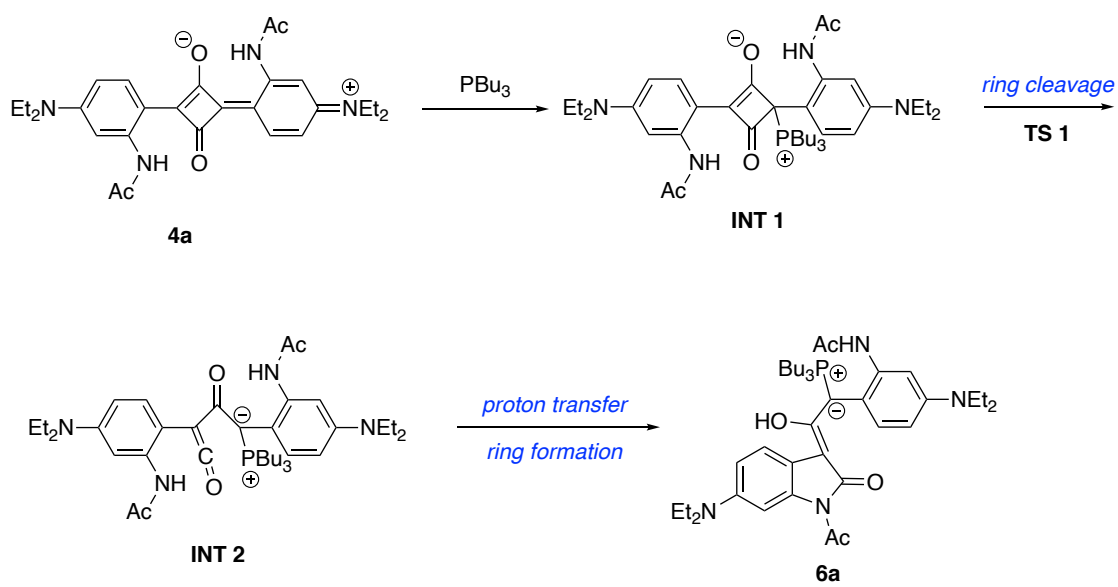


Figure 4.11. Revised mechanism

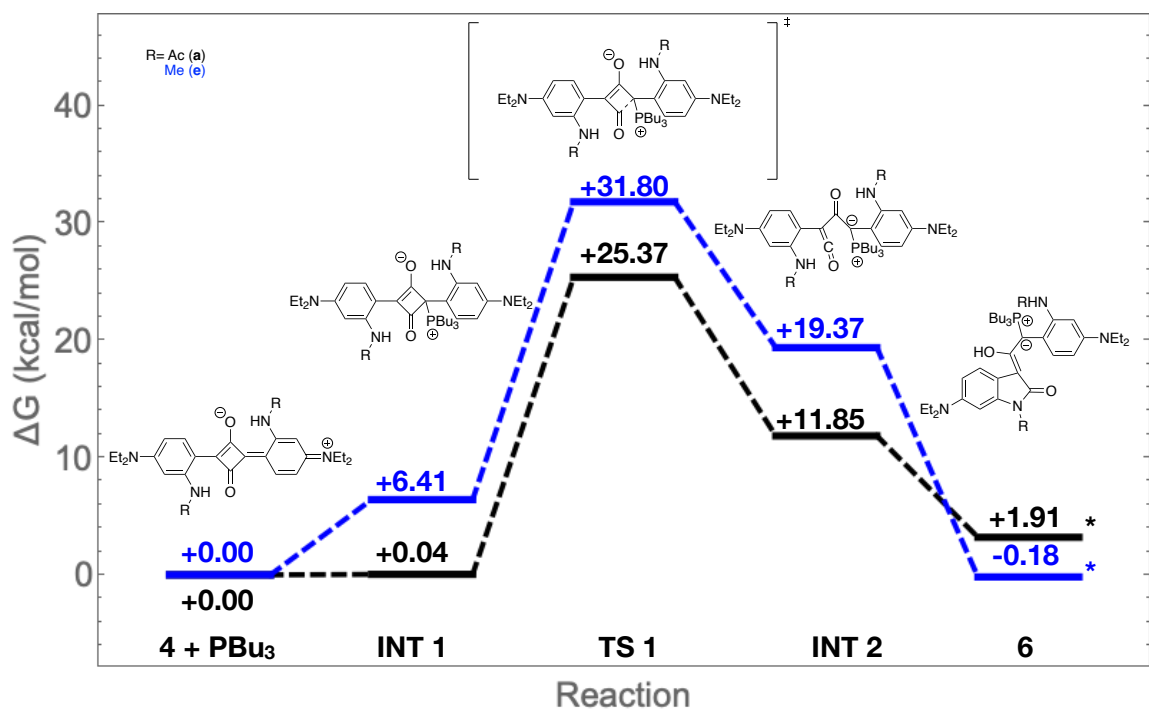


Figure 4.12. Free energy profile for oxindole **6a**, **6a***, and **6e*** formation using Gaussian16 with B3LYP/6-311+G* method.

forms phosphine adduct **INT 1**, which is followed by transition state **TS 1** forming ketene intermediate **INT 2** and where proton transfer on **INT 2** leads to formation of oxindole as shown in Figure 4.11. The result that the protonated **INT D** opens barrier-less to **INT 2** while the deprotonated **INT 2'** leads to barrier-less ring closure to **6a** suggests a concerted process where the deprotonation of the acetamide is coupled to the formation of the C-N bond. This is best understood in the framework of a general acid/base catalysis where the proton transfer is part of the rate determining step[119, 120]. Using the newly proposed reaction pathway, I performed DFT calculation on the intermediates and the transition state to understand the experimental observation.

The free energy surface of the reaction is shown in Figure 4.12. The transition state between **INT 2** and **6a** cannot be calculated due to coupled proton transfer

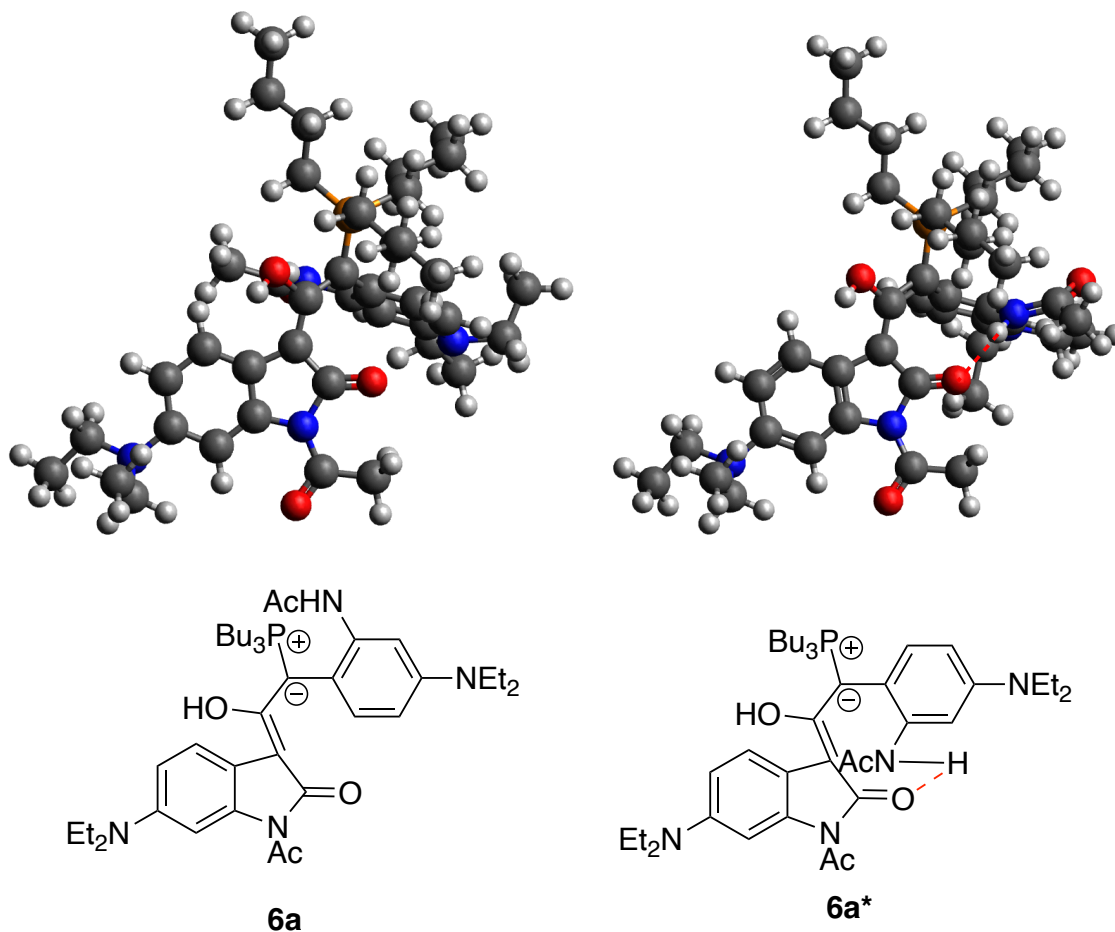


Figure 4.13. Orientations of aniline in product without hydrogen bond **6a** and with hydrogen bond **6a***

transition explained above. From squaraine dye **4a** with phosphine PBU_3 , the adduct **INT 1** only increased in energy by 0.04 kcal/mol. With the barrier of squaraine ring cleavage of 25.3 kcal/mol via **TS 1**, ketene intermediate **INT 2** is 11.8 kcal/mol higher in energy than **INT 1**. The initial conformation of oxindole product **6a** is 6.4 kcal/mol higher in energy than the squaraine, but undergoes conformational change to a lower energy structure rapidly.

The X-ray crystal structure of **6a** revealed the rotated aniline ring **6a*** as shown in Figure 4.13. The optimized structure of **6a*** is 4.5 kcal/mol lower in energy than

6a and has hydrogen bonding between oxindole’s carbonyl oxygen and amide’s hydrogen with distance of 1.90 Å, as indicated with the red dashed line in Figure 4.13. The free energy of **6a*** is 1.9 kcal/mol higher than the squaraine dye **4a** + PBu₃. Keeping in mind that implicit solvent models will give only an estimate of the relative free energy of the zwitterionic species **6a** and **INT 2**, the calculated free energy difference indicates an approximately thermochemical reaction. This is in line with the experimental observation that the reaction is reversible for sterically bulky phosphine.

The same calculation was done for formation of **6e***, which had experimental yield of 0% as shown in Figure 4.12. From squaraine dye **4e** with phosphine PBu₃, the adduct **INT 1** increased in energy by 6.4 kcal/mol. With the barrier of squaraine ring cleavage of 25.4 kcal/mol via **TS 1**, ketene intermediate **INT 2** is 13.0 kcal/mol higher in energy than **INT 1**. The oxindole product **6e*** is -0.18 kcal/mol lower in energy than the squaraine dyes. However, the barrier for formation of **6e*** is 31.8 kcal/mol, which is 6.4 kcal/mol higher than the barrier for formation of **6a***, which is in line with experimental reaction conditions that are not suitable to overcome a barrier of 31.8 kcal/mol.

CHAPTER 5

CODE IMPLEMENTATION

5.1 RT-TDDFT to PySCF

RT-TDDFT was originally developed in Q-Chem[81]. However, because the whole Q-Chem source code needed to be compiled every time the code was modified, which took hours each time, it was clear developing a method for electronic structure using Q-Chem was not time efficient. Thus, using PySCF[105], the python-based electronic structure program, I implemented RT-TDDFT code in the PySCF code to efficiently develop and test the method. The source code for PySCF is available at <https://github.com/pyscf/pyscf>.

TestTDHF.py

```
import numpy as np
import sys, re
import pyscf
import pyscf.dft
from pyscf import gto, rt
np.set_printoptions(linewidth=220, suppress = True, precision = 7)

def TestTDHF():
    """
    Tests Basic Propagation Functionality. TDDFT
```

```

"""
prm = '''
Model TDHF
Method MMUT
dt 0.02
MaxIter 100
ExDir 1.0
EyDir 1.0
EzDir 1.0
FieldAmplitude 0.01
FieldFreq 0.9202
ApplyImpulse 1
ApplyCw 0
StatusEvery 10
'''

geom = """
H 0. 0. 0.
H 0. 0. 0.9
H 2.0 0. 0
H 2.0 0.9 0
"""

output = re.sub("py","dat",sys.argv[0])
mol = gto.Mole()
mol.atom = geom
mol.basis = 'sto-3g'
mol.build()
ks = pyscf.dft.RKS(mol)

```

```

    ks.xc='HF'

    ks.kernel()

    apropr = rt.tdscf.RTTDSCF(ks,prm,output)

    return

TestTDHF()

```

In order to run the real-time code, three input sections are needed. As shown in **TestTDHF.py**, the restricted Kohn-Sham (RKS) class **ks**, which contains molecule's coordinates **geom**, basis set and DFT method, string **prm** that contains real-time propagation parameters, and the result file name **output** are sent to RTTDSCF module. Using the RKS, the Fock matrix is solved before propagation and the density matrix is then propagated using the MMUT method[53], described in Equation 2.12. The trajectory output file contains time, dipole in x,y,z direction, and energy of the molecule in atomic units, which can be used to study absorption peaks via Fourier transformation. This module in PySCF, along with additional implementation in <https://github.com/v3op01/RTMME>, was used to study the absorption of BOMME method.

The real-time propagation of the density matrix introduces one element that does not exist in non-real time propagation, imaginary value. The original exchange matrix builder for DFT in PySCF is built via a C++ extension, where the set amount of data is passed down from python to a C++ code. Because there is no usage of imaginary density matrix other than real-time propagation, the original extension only accepts real-valued density matrix. Thus, I implemented an exchange matrix builder that takes the density matrix with real and imaginary values as an input.

5.2 AMBER Interface to Q2MM

loop.in

```
DIR ./
FFLD read frcmod
PARM qparam.txt

RDAT -gabo example.log -gaao example.log -gato example.log -gh
      example.log
CDAT -abo example.in -aao example.in -ato example.in -ah example.in
COMP -o opt_start.txt
LOOP 0.01
GRAD
END
FFLD write frcmod.gaff.01
CDAT
COMP -o opt.01.txt
```

example.in

```
source leaprc.gaff
loadamberparams frcmod
mol = loadmol2 example.mol2
saveamberparm mol calc/prmtop calc/inpcrd
quit
```

The Quantum-Guided Molecular Mechanics (Q2MM) program is a software developed by the Norrby and Wiest groups and for the automated fittings of force field,

including transition state force field (TSFF), to reproduce reference data from accurate electronic structure calculations[121]. It describes the TS as a minimum by fitting FF parameters to existing DFT data. The current version of Q2MM mainly utilizes MacroModel[122] as its FF software, but links to free software such as TINKER[123] has been incorporated recently as well[124]. Due to our group’s interest of using TSFF for proteins, I have implemented a link of Amber18[125], a suite of biomolecular simulation programs, to Q2MM. It utilizes the **sander** for molecule energy calculation and geometry optimization, **cpptraj** for fitting bond, angle and dihedral angle parameters, and **nab** for fitting hessian matrix, which requires the user to compile Amber18 with modified code as explained in <https://github.com/Q2MM/q2mm>. As shown in **loop.in**, Q2MM input file, the file requires force field parameter file **fr-cmod**, which is generated using **parmchk2** from Amber18, parameter fitting file **qparam.txt**, which contains set of force field parameters that gets fitted, the Gaussian output file **example.log** with hessian matrix, the **tleap** input file **example.in** that loads force field parameters for given molecule as shown in **example.in**, and flags **bo,ao,to,h** that selects bond, angles, torsion and hessian matrix being fitted, respectively. Currently, the implementation is being tested by Taylor Quinn in the Wiest group.

5.3 Automated CatVS

CatVS is a virtual screening tool that automatically merges substrate, ligands and metal for conformational search[126]. However, not only does it requires proprietary software such as Maestro[127] and MacroModel, but also it requires the chemist to build new individual ligand geometry and manually generate a conformational search input file, using Maestro, to run calculations as indicated by arrows in Figure 5.1. This limitation allows the chemist to set up the structure manually, but it also makes the process slow, which is not feasible for large ligand libraries, and requires chemists to learn multiple computational techniques, which could limit them from using the software.

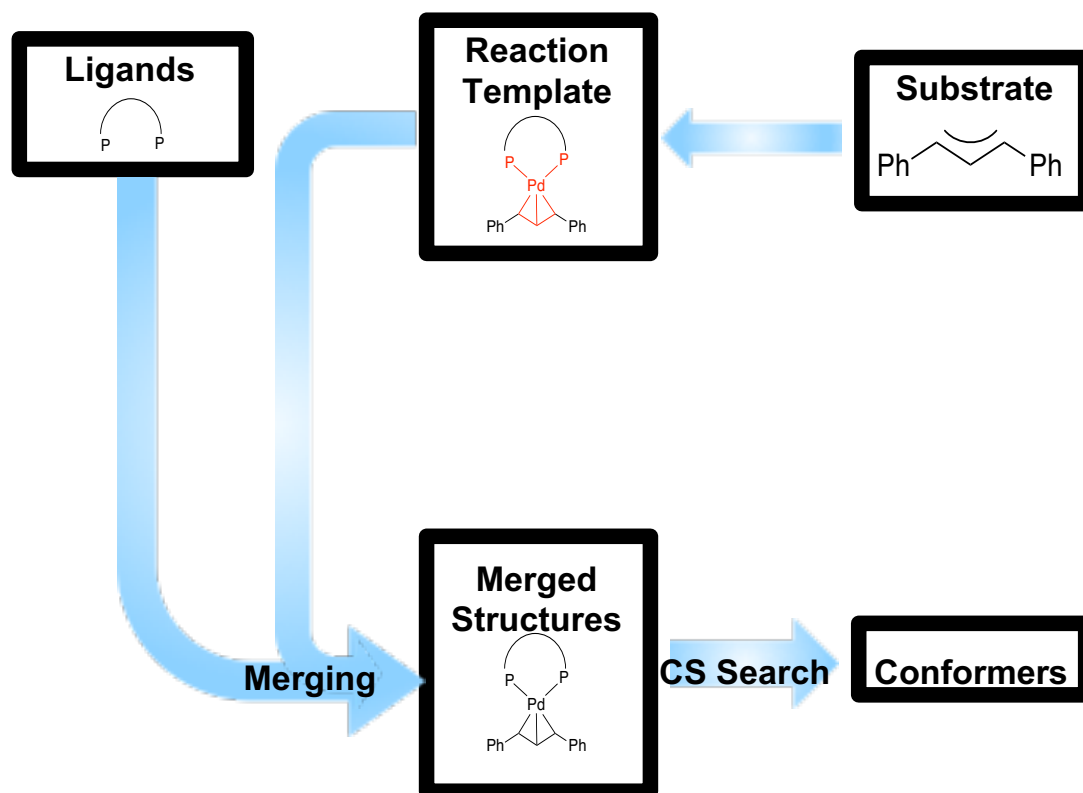


Figure 5.1. CatVS Scheme

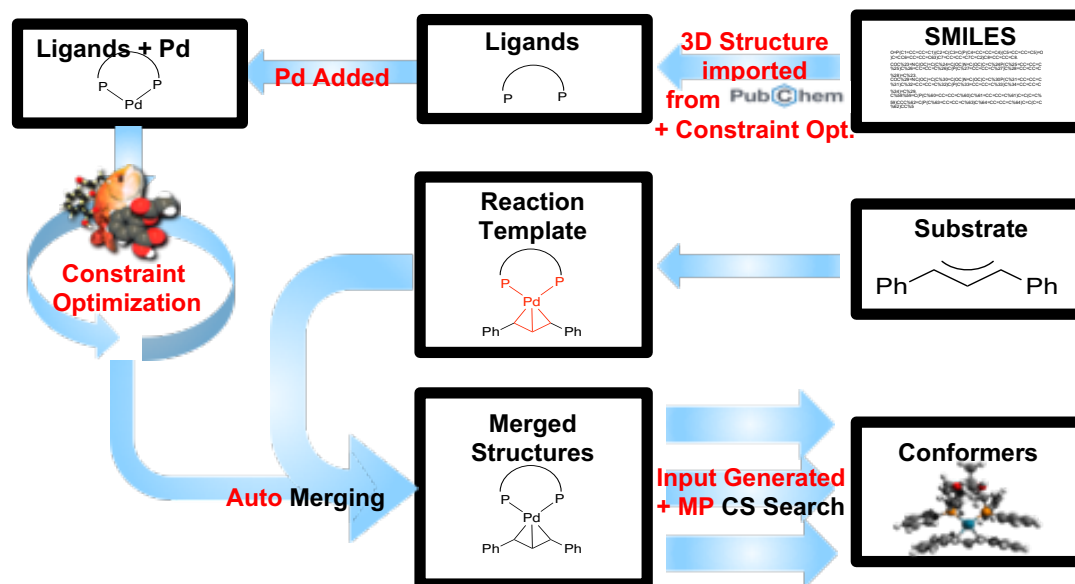


Figure 5.2. Automatic CatVS Scheme for Bisphosphine

run

```
#!/usr/bin/env python

import os

os.system("python_□$Q2MM/smiles_to_catvs/scripts/search.py")

os.system("python_□$Q2MM/smiles_to_catvs/scripts/sdftomae.py")

os.system("python_□$Q2MM/smiles_to_catvs/scripts/prepare.py")

reactions = "templatediphos.mae"
substrates = "allyl16.mae"
```

```

ligands = "ligands/*.mae"
mergedfile = "result.mae"

os.system("$SCHRODINGER/run_$$Q2MM/screen/merge.py_{}_{}_{}_{}_{}_{}-
        o_{}".format(reactions,substrates,ligands,mergedfile))

os.system("$SCHRODINGER/run_$$Q2MM/smiles_to_catvs/scripts/finish.py")

```

In an effort to make CatVS more approachable, I wrote a set of scripts that automatically generates ligand libraries from SMILES strings, generates the conformational search input file and performs the search in multi threads as shown in Figure 5.2. As shown in **run**, the ligand libraries are built using **search.py**, **sdf to mae.py**, and **prepare.py**. In **search.py**, SMILES strings are imported into 3D structure based on existing data in PubChem[128] in sdf file format, which are then optimized with bond constraints such as distance between two atoms that bind to metal using OpenBabel[103] and converted to mae file format using **sdftomae.py** script. Then, in **prepare.py**, the metal is added to the opposite side of the center of mass of the ligand, which then undergoes constraint optimization. These steps make an initial set of ligand libraries in folder **ligands**. Using **merge.py** script, the optimized ligand geometry then automatically merges with existing substrate and reaction template mae files, **templatediphos.mae** and **allyl16.mae** respectively, followed by conformational search using Maestro with **finish.py**. This was used to generate the ligand libraries of a bisphosphine ligand and is available at <https://github.com/Q2MM/q2mm> along with example files. As it only requires SMILES strings, constraints, substrate and reaction template, which will allow other chemists to easily use CatVS. As shown in the **run** file, the input variables for SMILES strings, bond constraints and number of threads are hard coded in each script, thus not visible in run file. The effort to simplify the input file is currently in progress.

APPENDIX A

CUDA IMPLEMENTATION IN C++

```
#include <cuda_runtime.h>
#include <cuda.h>
#include <stdio.h>
#include <iostream>
#include "book.h"

#define M 8

using namespace std;

__global__ void makeK(double *BpqR, double *rBRhot, double *iBRhot,
    double *rRhot, double *iRhot, double *rK, double *iK, int n_mo,
    int n_aux)
{
    //Parallelizing the GPU

    unsigned int j = threadIdx.x + blockIdx.x * blockDim.x;
    unsigned int i = threadIdx.y + blockIdx.y * blockDim.y;
    unsigned int k = threadIdx.z + blockIdx.z * blockDim.z;

    double rsum = 0;
```

```

double isum = 0;
double rKsum = 0;
double iKsum = 0;

if ( i < n_mo && j < n_mo && k < n_aux)
{
    for(int l = 0; l < n_mo; l++)
    {
        rsum += BpqR[k * n_mo * n_mo + i * n_mo + l] *
            rRhot[l * n_mo + j];
        isum += BpqR[k * n_mo * n_mo + i * n_mo + l] *
            iRhot[l * n_mo + j];
    }
    rBRhot[k * n_mo * n_mo + i * n_mo + j] = rsum;
    iBRhot[k * n_mo * n_mo + i * n_mo + j] = isum;

    for(int l = 0; l < n_mo; l++)
    {
        rKsum += rBRhot[k * n_mo * n_mo + i * n_mo + l]
            * BpqR[k * n_mo * n_mo + l * n_mo + j];
        iKsum += iBRhot[k * n_mo * n_mo + i * n_mo + l]
            * BpqR[k * n_mo * n_mo + l * n_mo + j];
    }
    rK[k * n_mo * n_mo + i * n_mo + j] = rKsum;
    iK[k * n_mo * n_mo + i * n_mo + j] = iKsum;
}
}

```

```

void CUADFock( double * BpqR, double * r_Rhot, double * i_Rhot,
               double * r_K, double * i_K, int n_mo, int n_aux)
{

    int mat_size = n_mo * n_mo;

    int size = sizeof(double) * mat_size;

    double * h_BpqR = new double[size * n_aux];
    double * h_rRhot = new double[size];
    double * h_iRhot = new double[size];
    double * h_rK = new double[size * n_aux];
    double * h_iK = new double[size * n_aux];

    double *d_BpqR, *d_rRhot, *d_iRhot, *d_rBRhot, *d_iBRhot, *
        d_rK, *d_iK;

    HANDLE_ERROR( cudaMalloc( (void **)&d_BpqR, size * n_aux ) );
    HANDLE_ERROR( cudaMalloc( (void **)&d_rRhot, size ) );
    HANDLE_ERROR( cudaMalloc( (void **)&d_iRhot, size ) );
    HANDLE_ERROR( cudaMalloc( (void **)&d_rBRhot, size * n_aux ) );
    HANDLE_ERROR( cudaMalloc( (void **)&d_iBRhot, size * n_aux ) );
    HANDLE_ERROR( cudaMalloc( (void **)&d_rK, size * n_aux ) );
    HANDLE_ERROR( cudaMalloc( (void **)&d_iK, size * n_aux ) );

    for(int k = 0; k < n_aux; k++)
        for(int i = 0; i < n_mo; i++)
            for(int j = 0; j < n_mo; j++)

```

```

        h_BpqR[k * mat_size + i * n_mo + j] =
            BpqR[ k * mat_size + j * n_mo + i];

for(int i = 0; i < n_mo; i++)
    for(int j = 0; j < n_mo; j++)
    {
        h_rRhot[i * n_mo + j] = r_Rhot[j * n_mo + i];
        h_iRhot[i * n_mo + j] = i_Rhot[j * n_mo + i];
    }

HANDLE_ERROR( cudaMemcpy( d_BpqR, h_BpqR, size * n_aux,
    cudaMemcpyHostToDevice) );
HANDLE_ERROR( cudaMemcpy( d_rRhot, h_rRhot, size,
    cudaMemcpyHostToDevice) );
HANDLE_ERROR( cudaMemcpy( d_iRhot, h_iRhot, size,
    cudaMemcpyHostToDevice) );

dim3 blocks( (n_mo + M - 1)/M, (n_mo + M - 1)/M, (n_aux + M -
    1)/M );
dim3 threads( M, M, M );

makeK <<< blocks, threads >>> (d_BpqR, d_rBRhot, d_iBRhot,
    d_rRhot, d_iRhot, d_rK, d_iK, n_mo, n_aux);

HANDLE_ERROR( cudaMemcpy( h_rK, d_rK, size * n_aux,
    cudaMemcpyDeviceToHost) );
HANDLE_ERROR( cudaMemcpy( h_iK, d_iK, size * n_aux,

```



```

        cudaMemcpyDeviceToHost) );

for(int k = 0; k < n_aux; k++)
    for(int i = 0; i < n_mo; i++)
        for(int j = 0; j < n_mo; j++)
        {
            r_K[i + j * n_mo] -= h_rK[k * n_mo * n_mo
                + i * n_mo + j];
            i_K[i + j * n_mo] -= h_iK[k * n_mo * n_mo
                + i * n_mo + j];
        }
    cudaFree(d_BpqR);
    cudaFree(d_rRhot);
    cudaFree(d_iRhot);
    cudaFree(d_rBRhot);
    cudaFree(d_iBRhot);
    cudaFree(d_rK);
    cudaFree(d_iK);

    delete[] h_BpqR;
    delete[] h_rRhot;
    delete[] h_iRhot;
    delete[] h_rK;
    delete[] h_iK;
}

```

APPENDIX B

SUPPLEMENTARY INFORMATION

Supplementary information for the data presented in mechanistic study is available as an electronic supplement. The supplement includes coordinates for DFT calculations.

BIBLIOGRAPHY

1. R. Chetrite and K. Mallick. Quantum fluctuation relations for the lindblad master equation. *J. Stat. Phys.*, 148(3):480–501, 2012.
2. I. De Vega. Non-markovian stochastic schrödinger description of transport in quantum networks. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 44(24):245501, 2011.
3. J. Zhu, S. Kais, P. Rebentrost, and A. Aspuru-Guzik. Modified scaled hierarchical equation of motion approach for the study of quantum coherence in photosynthetic complexes. *J. Phys. Chem. B*, 115(6):1531–1537, 2011.
4. S. Valleau, A. Eisfeld, and A. Aspuru-Guzik. On the alternatives for bath correlators and spectral densities from mixed quantum-classical simulations. *J. Chem. Phys.*, 137(22):224103, 2012.
5. T. Markovich, S. M. Blau, J. Parkhill, C. Kreisbeck, J. N. Sanders, X. Andrade, and A. Aspuru-Guzik. More accurate and efficient bath spectral densities from super-resolution. *arXiv preprint arXiv:1307.4407*, 2013.
6. R. Nanguneri and J. Parkhill. Relaxation between bright optical wannier excitons in perovskite solar absorber $\text{CH}_3\text{NH}_3\text{PbI}_3$. *arXiv preprint arXiv:1411.1110*, 2014.
7. T. S. Nguyen, R. Nanguneri, and J. Parkhill. How electronic dynamics with pauli exclusion produces fermi-dirac statistics. *J. Chem. Phys.*, 142(13):134113, 2015.
8. W. Kutzelnigg and D. Mukherjee. Normal order and extended wick theorem for a multiconfiguration reference wave function. *J. Chem. Phys.*, 107(2):432–449, 1997.
9. C. Meier and D. J. Tannor. Non-markovian evolution of the density operator in the presence of strong laser fields. *J. Chem. Phys.*, 111(8):3365–3376, 1999.
10. T. S. Nguyen and J. Parkhill. Nonradiative relaxation in real-time electronic dynamics of organolead triiodide perovskite. *J. Phys. Chem. A*, 120(34):6880–6887, 2016.

11. R. Berera, R. van Grondelle, and J. T. Kennis. Ultrafast transient absorption spectroscopy: principles and application to photosynthetic systems. *Photosynth. Res.*, 101(2-3):105–118, 2009.
12. Z. Zhu, J. Crochet, M. S. Arnold, M. C. Hersam, H. Ulbricht, D. Resasco, and T. Hertel. Pump-probe spectroscopy of exciton dynamics in (6, 5) carbon nanotubes. *J. Phys. Chem. C*, 111(10):3831–3835, 2007.
13. L. J. van Wilderen, C. N. Lincoln, and J. J. van Thor. Modelling multi-pulse population dynamics from ultrafast spectroscopy. *PloS one*, 6(3):e17373, 2011.
14. A. D. Dutoi and L. S. Cederbaum. Time-resolved pump-probe spectroscopy to follow valence electronic motion in molecules: Application. *Phys. Rev. A*, 90(2):023414, 2014.
15. B. Peng, D. B. Lingerfelt, F. Ding, C. M. Aikens, and X. Li. Real-time tddft studies of exciton decay and transfer in silver nanowire arrays. *J. Phys. Chem. C*, 119(11):6421–6427, 2015.
16. S. Tussupbayev, N. Govind, K. Lopata, and C. J. Cramer. Comparison of real-time and linear-response time-dependent density functional theories for molecular chromophores ranging from sparse to high densities of states. *J. Chem. Theory. Comput.*, 11(3):1102–1109, 2015.
17. C. M. Isborn, S. V. Kilina, X. Li, and O. V. Prezhdo. Generation of multiple excitons in pbse and cdse quantum dots by direct photoexcitation: first-principles calculations on small pbse and cdse clusters. *J. Phys. Chem. C*, 112(47):18291–18294, 2008.
18. J. Walkenhorst, U. De Giovannini, A. Castro, and A. Rubio. Tailored pump-probe transient spectroscopy with time-dependent density-functional theory: controlling absorption spectra. *Eur. Phys. J. B*, 89(5):128, 2016.
19. J. I. Fuks and N. T. Maitra. Challenging adiabatic time-dependent density functional theory with a hubbard dimer: the case of time-resolved long-range charge transfer. *Phys. Chem. Chem. Phys.*, 16(28):14504–14513, 2014.
20. B. F. Habenicht, N. P. Tani, M. R. Provorse, and C. M. Isborn. Two-electron rabi oscillations in real-time time-dependent density-functional theory. *J. Chem. Phys.*, 141(18):184112, 2014.
21. T. S. Nguyen and J. Parkhill. Nonadiabatic dynamics for electrons at second-order: Real-time tddft and oscf2. *J. Chem. Theory. Comput.*, 11(7):2918–2924, 2015.
22. S. A. Fischer, C. J. Cramer, and N. Govind. Excited state absorption from real-time time-dependent density functional theory. *J. Chem. Theory. Comput.*, 11(9):4294–4303, 2015.

23. A. McLachlan and M. Ball. Time-dependent hartree—fock theory for molecules. *Rev. Mod. Phys.*, 36(3):844, 1964.
24. W. Zhu, J. Botina, and H. Rabitz. Rapidly convergent iteration methods for quantum optimal control of population. *J. Chem. Phys.*, 108(5):1953–1963, 1998.
25. A.-M. Uimonen, E. Khosravi, G. Stefanucci, S. Kurth, R. van Leeuwen, and E. Gross. Real-time switching between multiple steady-states in quantum transport. In *J. Phys. Conf. Ser.*, volume 220, page 012018. IOP Publishing, 2010.
26. X. Andrade, J. Alberdi-Rodriguez, D. A. Strubbe, M. J. Oliveira, F. Nogueira, A. Castro, J. Muguerza, A. Arruabarrena, S. G. Louie, A. Aspuru-Guzik, et al. Time-dependent density-functional theory in massively parallel computer architectures: the octopus project. *J. Phys. Condens.*, 24(23):233202, 2012.
27. K. Burke, J. Werschnik, and E. Gross. Time-dependent density functional theory: Past, present, and future. *J. Chem. Phys.*, 123(6):062206, 2005.
28. F. Furche and R. Ahlrichs. Adiabatic time-dependent density functional methods for excited state properties. *J. Chem. Phys.*, 117(16):7433–7447, 2002.
29. D. W. Silverstein, N. Govind, H. J. Van Dam, and L. Jensen. Simulating one-photon absorption and resonance raman scattering spectra using analytical excited state energy gradients within time-dependent density functional theory. *J. Chem. Theory. Comput.*, 9(12):5490–5503, 2013.
30. H. M. Senn and W. Thiel. Qm/mm methods for biological systems. In *Atomistic Approaches in Modern Biology*, pages 173–290. Springer, 2006.
31. I. Petkova, G. Dobrikov, N. Banerji, G. Duvanel, R. Perez, V. Dimitrov, P. Nikolov, and E. Vauthey. Tuning the excited-state dynamics of gfp-inspired imidazolone derivatives. *J. Phys. Chem. A*, 114(1):10–20, 2010.
32. J. Snellenburg, S. Liptonok, R. Seger, K. Mullen, and I. Van Stokkum. Glotaran: A java-based graphical user interface for the r package timp. 2012.
33. N. Luehr, I. S. Ufimtsev, and T. J. Martinez. Dynamic precision for electron repulsion integral evaluation on graphical processing units (gpu). *J. Chem. Theory. Comput.*, 7(4):949–954, 2011.
34. J. Nickolls, I. Buck, M. Garland, and K. Skadron. Scalable parallel programming with cuda. *Queue*, 6(2):40–53, 2008.
35. A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 103(2):227–249, 1976.

36. H. M. Senn and W. Thiel. Qm/mm methods for biomolecular systems. *Angew. Chem., Int. Ed.*, 48(7):1198–1229, 2009.
37. G. Knizia and G. K.-L. Chan. Density matrix embedding: A simple alternative to dynamical mean-field theory. *Phys. Rev. Lett.*, 109(18):186404, 2012.
38. A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.*, 68(1):13, 1996.
39. N. Govind, Y. A. Wang, and E. A. Carter. Electronic-structure calculations by first-principles density-based embedding of explicitly correlated systems. *J. Chem. Phys.*, 110(16):7677–7688, 1999.
40. J. D. Goodpaster, N. Ananth, F. R. Manby, and T. F. Miller III. Exact non-additive kinetic potentials for embedded density functional theory. *J. Chem. Phys.*, 133(8):084103, 2010.
41. J. D. Goodpaster, T. A. Barnes, and T. F. Miller III. Embedded density functional theory for covalently bonded and strongly interacting subsystems. *J. Chem. Phys.*, 134(16):164108, 2011.
42. F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller III. A simple, exact density-functional-theory embedding scheme. *J. Chem. Theory Comput.*, 8(8):2564, 2012.
43. J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III. Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *J. Chem. Phys.*, 140(18):18A507, 2014.
44. F. Ding, F. R. Manby, and T. F. Miller III. Embedded mean-field theory with block-orthogonalized partitioning. *J. Chem. Theory Comput.*, 13(4):1605–1615, 2017.
45. M. E. Fornace, J. Lee, K. Miyamoto, F. R. Manby, and T. F. Miller III. Embedded mean-field theory. *J. Chem. Theory Comput.*, 11(2):568–580, 2015.
46. K. Morokuma. Oniom and its applications to material chemistry and catalyses. *B. Korean. Chem. Soc.*, 24(6):797–801, 2003.
47. W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133, 1965.
48. R. G. Parr. Density functional theory of atoms and molecules. In *Horizons of Quantum Chemistry*, pages 5–15. Springer, 1980.
49. J. A. Pople, P. M. Gill, and B. G. Johnson. Kohn—sham density-functional theory within a finite basis set. *Chem. Phys. Lett.*, 199(6):557–560, 1992.

50. A. F. Izmaylov, G. E. Scuseria, and M. J. Frisch. Efficient evaluation of short-range hartree-fock exchange in large molecules and periodic systems. *J. Chem. Phys.*, 125(10):104103, 2006.
51. E. Rudberg and P. Salek. Efficient implementation of the fast multipole method. *J. Chem. Phys.*, 125(8):084106, 2006.
52. W. Liang, C. T. Chapman, and X. Li. Efficient first-principles electronic dynamics. *J. Chem. Phys.*, 134(18):184102, 2011.
53. X. Li, S. M. Smith, A. N. Markevitch, D. A. Romanov, R. J. Levis, and H. B. Schlegel. A time-dependent hartree-fock approach for studying the electronic optical response of molecules in intense fields. *Phys. Chem. Chem. Phys.*, 7(2):233–239, 2005.
54. J. J. Goings, P. J. Lestrangle, and X. Li. Real-time time-dependent electronic structure theory. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 8(1):e1341, 2018.
55. S. B. Nielsen, A. Lapierre, J. U. Andersen, U. Pedersen, S. Tomita, and L. Andersen. Absorption spectrum of the green fluorescent protein chromophore anion in vacuo. *Phys. Rev. Lett.*, 87(22):228102, 2001.
56. J. M. Milanese, M. R. Provorse, E. Alameda Jr, and C. M. Isborn. Convergence of computed aqueous absorption spectra with explicit quantum mechanical solvent. *J. Chem. Theory. Comput.*, 13(5):2159–2171, 2017.
57. M. Denny, M. Chun, and R. Liu. 9-cis, 11-cis-retinal* from direct irradiation of all-trans-retinal. new geometric isomers of vitamin a and carotenoids 9. *Photochem. Photobiol.*, 33(2):267–269, 1981.
58. S. B. Nielsen, A. Lapierre, J. U. Andersen, U. V. Pedersen, S. Tomita, and L. H. Andersen. Absorption spectrum of the green fluorescent protein chromophore anion in vacuo. *Phys. Rev. Lett.*, 87:228102, Nov 2001. doi: 10.1103/PhysRevLett.87.228102. URL <https://link.aps.org/doi/10.1103/PhysRevLett.87.228102>.
59. M. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97(3):1990–2001, 1992.
60. J. R. McClean, J. A. Parkhill, and A. Aspuru-Guzik. Feynman’s clock, a new variational principle, and parallel-in-time quantum dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, 110(41):E3901–E3909, 2013.
61. J. I. Fuks, K. Luo, E. D. Sandoval, and N. T. Maitra. Time-resolved spectroscopy in time-dependent density functional theory: An exact condition. *Phys. Rev. Lett.*, 114(18):183002, 2015.

62. M. R. Provorse, B. F. Habenicht, and C. M. Isborn. Peak-shifting in real-time time-dependent density functional theory. *J. Chem. Theory. Comput.*, 11(10):4791–4802, 2015.
63. N. Rohringer, A. Gordon, and R. Santra. Configuration-interaction-based time-dependent orbital approach for ab initio treatment of electronic dynamics in a strong optical laser field. *Phys. Rev. A*, 74(4):043420, 2006.
64. L. Greenman, P. J. Ho, S. Pabst, E. Kamarchik, D. A. Mazziotti, and R. Santra. Implementation of the time-dependent configuration-interaction singles method for atomic strong-field processes. *Phys. Rev. A*, 82(2):023406, 2010.
65. T. S. Nguyen, J. H. Koh, S. Lefelhocz, and J. Parkhill. Black-box, real-time simulations of transient absorption spectroscopy. *J. Phys. Chem. Lett.*, 7(8):1590–1595, 2016.
66. P. Elliott, J. I. Fuks, A. Rubio, and N. T. Maitra. Universal dynamical steps in the exact time-dependent exchange-correlation potential. *Phys. Rev. Lett.*, 109(26):266404, 2012.
67. A. Szabo and N. S. Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.
68. S. Hirata and M. Head-Gordon. Time-dependent density functional theory within the tamm-dancoff approximation. *Chem. Phys. Lett.*, 314(3-4):291–299, 1999.
69. M. H. Segler, M. Preuss, and M. P. Waller. Planning chemical syntheses with deep neural networks and symbolic ai. *Nat.*, 555(7698):604–610, 2018.
70. P. Schwaller, T. Gaudin, D. Lanyi, C. Bekas, and T. Laino. “found in translation”: predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. *Chem. Sci.*, 9(28):6091–6098, 2018.
71. K. Yao, J. E. Herr, S. N. Brown, and J. Parkhill. Intrinsic bond energies from a bonds-in-molecules neural network. *J. Phys. Chem. Lett.*, 8(12):2689–2694, 2017.
72. B. Nebgen, N. Lubbers, J. S. Smith, A. E. Sifain, A. Lokhov, O. Isayev, A. E. Roitberg, K. Barros, and S. Tretiak. Transferable dynamic molecular charge assignment using deep neural networks. *J. Chem. Theory. Comput.*, 14(9):4687–4698, 2018.
73. K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. Von Lilienfeld, K.-R. Müller, and A. Tkatchenko. Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. *J. Phys. Chem. Lett.*, 6(12):2326–2331, 2015.

74. K. Yao, J. E. Herr, D. W. Toth, R. Mckintyre, and J. Parkhill. The tensormol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci.*, 9(8):2261–2269, 2018.
75. L. Simón and J. M. Goodman. How reliable are dft transition structures? comparison of gga, hybrid-meta-gga and meta-gga functionals. *Org. Biomol.*, 9(3): 689–700, 2011.
76. A. Ault. The bullvalene story. the conception of bullvalene, a molecule that has no permanent structure. *J. Chem. Educ.*, 78(7):924, 2001.
77. J. J. Titman, Z. Luz, and H. W. Spiess. Solid-state reactions studied by carbon-13 rotor-synchronized magic angle spinning two-dimensional exchange nmr. 2. the cope rearrangement and molecular reorientation in bullvalene. *J. Am. Chem. Soc.*, 114(10):3765–3771, 1992.
78. J. F. Oth, K. Müllen, J.-M. Gilles, and G. Schröder. Comparison of ^{13}C - and ^1H -magnetic resonance spectroscopy as techniques for the quantitative investigation of dynamic processes. the cope rearrangement in bullvalene. *Helv. Chim. Acta*, 57(5):1415–1433, 1974.
79. G. Schröder and J. Oth. Recent chemistry of bullvalene. *Angew. Chem. Int. Ed. in English*, 6(5):414–423, 1967.
80. K. Rebsamen and G. Schröder. Penta- und hexabrombullvalene. *Chem. Ber.*, 126(6):1425–1427, 1993.
81. Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, et al. Advances in molecular quantum chemistry contained in the q-chem 4 program package. *Mol. Phys.*, 113(2): 184–215, 2015.
82. J. E. Herr, K. Koh, K. Yao, and J. Parkhill. Compressing physics with an autoencoder: Creating an atomic species representation to improve machine learning models in the chemical sciences. *J. Chem. Phys.*, 151(8):084103, 2019.
83. S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1):511–519, 1984.
84. P. O. Moreno, C. Suarez, M. Tafazzoli, N. S. True, and C. B. LeMaster. Gas phase nmr study of the degenerate cope rearrangement of bullvalene. *J. Phys. Chem.*, 96(25):10206–10212, 1992.
85. J. E. Herr, K. Yao, R. McIntyre, D. W. Toth, and J. Parkhill. Metadynamics for training neural network model chemistries: A competitive assessment. *J. Chem. Phys.*, 148(24):241710, 2018.

86. B. Sohrabi and H. Iraj. Implementing flipped classroom using digital media: A comparison of two demographically different groups perceptions. *Comput. Hum. Behav.*, 60:514–524, 2016.
87. V. Betihavas, H. Bridgman, R. Kornhaber, and M. Cross. The evidence for ‘flipping out’: A systematic review of the flipped classroom in nursing education. *Nurse Educ. Today*, 38:15–21, 2016.
88. C.-L. Lai and G.-J. Hwang. A self-regulated flipped classroom approach to improving students’ learning performance in a mathematics course. *Comput. Educ.*, 100:126–140, 2016.
89. K. K. Bhagat, C.-N. Chang, and C.-Y. Chang. The impact of the flipped classroom on mathematics concept learning in high school. *J. Educ. Technol. Soc.*, 19(3):134–142, 2016.
90. Y.-N. Huang and Z.-R. Hong. The effects of a flipped english classroom intervention on students’ information and communication technology and english reading comprehension. *Educ. Technol. Res. Dev.*, 64(2):175–193, 2016.
91. B. Nguyen, X. Yu, A. Japutra, and C.-H. S. Chen. Reverse teaching: Exploring student perceptions of “flip teaching”. *Act. Learn. High. Educ.*, 17(1):51–61, 2016.
92. J. D. Smith. Student attitudes toward flipping the general chemistry classroom. *Chem. Educ. Res. Pract.*, 14(4):607–614, 2013.
93. T. Wanner and E. Palmer. Personalising learning: Exploring student and teacher perceptions about flexible learning and assessment in a flipped university course. *Comput. Educ.*, 88:354–369, 2015.
94. M. Sage and P. Sele. Reflective journaling as a flipped classroom technique to increase reading and participation with social work students. *J. Soc. Work Educ.*, 51(4):668–681, 2015.
95. B. Schneider and P. Blikstein. Flipping the flipped classroom: A study of the effectiveness of video lectures versus constructivist exploration using tangible user interfaces. *IEEE Trans. Learn. Technol.*, 9(1):5–17, 2015.
96. L. Chen, T.-L. Chen, and N.-S. Chen. Students’ perspectives of using cooperative learning in a flipped statistics classroom. *Australas. J. Educ. Technol.*, 31(6), 2015.
97. J. L. Jensen, T. A. Kummer, and P. D. d. M. Godoy. Improvements from a flipped classroom may simply be the fruits of active learning. *CBE Life Sci. Educ.*, 14(1):ar5, 2015.

98. M. M. Cooper, L. C. Williams, and S. M. Underwood. Student understanding of intermolecular forces: A multimodal study. *J. Chem. Educ.*, 92(8):1288–1298, 2015.
99. M. N. Srnec, S. Upadhyay, and J. D. Madura. A python program for solving schrödinger’s equation in undergraduate physical chemistry, 2017.
100. N. J. Hill, M. D. Bowman, B. J. Esselman, S. D. Byron, J. Kreitingner, and N. E. Leadbeater. Ligand-free suzuki–miyaura coupling reactions using an inexpensive aqueous palladium source: A synthetic and computational exercise for the undergraduate organic chemistry laboratory. *J. Chem. Educ.*, 91(7):1054–1057, 2014.
101. J. Forcier, P. Bissex, and W. J. Chun. *Python web development with Django*. Addison-Wesley Professional, 2008.
102. G. Landrum et al. Rdkit: Open-source cheminformatics. 2006.
103. N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison. Open babel: An open chemical toolbox. *J. Cheminformatics*, 3(1):33, 2011.
104. S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, et al. Pubchem substance and compound databases. *Nucleic Acids Res.*, 44(D1):D1202–D1213, 2016.
105. Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, et al. Pyscf: the python-based simulations of chemistry framework. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 8(1):e1340, 2018.
106. L. ChemAxon and J. Instant. Chem/marvinsketch. *ChemAxon Ltd., Budapest, Hungary*, 2012.
107. R. M. Hanson, J. Prilusky, Z. Renjian, T. Nakane, and J. L. Sussman. Jsmol and the next-generation web-based representation of 3d molecular structure as applied to proteopedia. *Isr. J. Chem.*, 53(3-4):207–216, 2013.
108. J. Jmol. an open-source java viewer for chemical structures in 3d. *Jmol web page: <http://www.jmol.org/>*, last accessed, 15, 2013.
109. C. W. Coley, W. Jin, L. Rogers, T. F. Jamison, T. S. Jaakkola, W. H. Green, R. Barzilay, and K. F. Jensen. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.*, 10(2):370–377, 2019.
110. K. Liang, K.-Y. Law, and D. G. Whitten. Synthesis, characterization, photophysics, and photosensitization studies of squaraine- bipyridinium diads. *J. Phys. Chem. B*, 101(4):540–546, 1997.

111. K.-Y. LAW et al. Squaraine chemistry: effect of synthetis on the morphological and xerographic properties of photoconductive squaraines. *J. Imaging Sci.*, 31 (4):172–177, 1987.
112. K. Liang, K.-Y. Law, and D. G. Whitten. Estimation of exciton sizes in squaraine monolayers by intralayer photoinduced electron transfer. *J. Phys. Chem.*, 99(45):16704–16708, 1995.
113. J. V. Ros-Lis, R. Martínez-Máñez, K. Rurack, F. Sancenón, J. Soto, and M. Spieles. Highly selective chromogenic signaling of hg2+ in aqueous media at nanomolar levels employing a squaraine-based reporter. *Inorg. Chem.*, 43 (17):5183–5185, 2004.
114. J. Xie, A. B. Comeau, and C. T. Seto. Squaric acids: a new motif for designing inhibitors of protein tyrosine phosphatases. *Org. Lett.*, 6(1):83–86, 2004.
115. E. P. Bacher, A. J. Lepore, D. Pena-Romero, B. D. Smith, and B. L. Ashfeld. Nucleophilic addition of phosphorus (iii) derivatives to squaraines: colorimetric detection of transition metal-mediated or thermal reversion. *ChemComm*, 55 (22):3286–3289, 2019.
116. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian~16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.
117. M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics*, 4(1):17, 2012.
118. G. Schaftenaar and J. H. Noordik. Molden: a pre-and post-processing program for molecular and electronic structures. *J. Comput. Aided Mol. Des.*, 14(2): 123–134, 2000.
119. E. V. Anslyn and D. A. Dougherty. *Modern physical organic chemistry*. University science books, 2006.

120. W. P. Jencks. General acid-base catalysis of complex reactions in water. *Chem. Rev.*, 72(6):705–718, 1972.
121. E. Hansen, A. R. Rosales, B. Tutkowski, P.-O. Norrby, and O. Wiest. Prediction of stereochemistry using q2mm. *Acc. Chem. Res.*, 49(5):996–1005, 2016.
122. F. Mohamadi, N. G. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Cauffield, G. Chang, T. Hendrickson, and W. C. Still. Macromodel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.*, 11(4):440–467, 1990.
123. J. W. Ponder et al. Tinker: Software tools for molecular design, 2004.
124. A. R. Rosales. *Development of Force Field Parameters for Predictions in Asymmetric Catalysis and a Foray into Mechanism Elucidation*. University of Notre Dame, 2019.
125. D. Case, I. Ben-Shalom, S. Brozell, D. Cerutti, T. Cheatham III, V. Cruzeiro, T. Darden, R. Duke, D. Ghoreishi, M. Gilson, et al. Amber 18; 2018. *University of California, San Francisco*.
126. A. R. Rosales, J. Wahlers, E. Limé, R. E. Meadows, K. W. Leslie, R. Savin, F. Bell, E. Hansen, P. Helquist, R. H. Munday, et al. Rapid virtual screening of enantioselective catalysts using catvs. *Nat. Catal.*, 2(1):41–45, 2019.
127. S. Release. 1: Maestro. *Schrödinger, LLC, New York, NY*, 2017, 2017.
128. E. E. Bolton, Y. Wang, P. A. Thiessen, and S. H. Bryant. Pubchem: integrated platform of small molecules and biological activities. In *Annu. Rep. Comput. Chem.*, volume 4, pages 217–241. Elsevier, 2008.