

Development and Application of Efficient Data Analytic Methods in Quantum Chemistry

Matteus Tanha

September 2015

Mellon College of Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:

David J. Yaron, Advisor

Maria Kurnikova

Bruce A. Armitage

Geoffrey Hutchison

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.*

Copyright © 2015 Matteus Tanha

Dedicated to my sister, Marieh Tanha

Abstract

This work presents systematical approaches to improve both the application and theory of quantum chemical methods. Semi-empirical and *ab initio* quantum chemical methods are used to model fluorescent dyes for bio-imaging. A new approach to development of semiempirical quantum chemical methods is explored, and imitation learning is used to accelerate convergence of self consistent field calculations.

The fluorescence of the SKC-513 dye is highly sensitive to the binding of K^+ ion. Computations reveal that, in the absence of K^+ , excitation is to two nearly-degenerate states, a neutral (N) excited state with high oscillator strength and a charge-transfer (CT) state with lower oscillator strength. Binding of K^+ raises the CT state far above the N state, shutting down a non-radiative pathway mediated by the CT state. This rationalizes the high sensitivity of the quantum yield to ion binding. Computations on a series of thiazole orange derivatives are used to successfully account for the observed spectral shifts through frontier orbital analysis.

A means to take advantage of molecular similarity to lower the computational cost of electronic structure theory is explored, in which parameters are embedded into a low-cost, low-level (LL) *ab initio* model and adjusted to obtain agreement with results from a higher-level (HL) *ab initio* model. In the parametrized LL (pLL) model, selected matrix elements of the Hamiltonian are scaled by factors that depend on element types. Various approaches to scaling, including making parameters sensitive to atomic charges and bond orders, are explored. The models are trained on ethane and ethylene, substituted with $-NH_2$, $-OH$ and $-F$, and tested on substituted propane, propylene and t-butane. The molecules are distorted and placed in electrostatic fields.

Fitted properties include total and decomposed energies, frontier orbital energies, and interactions with test charges. The best-performing model forms reduce the root mean square (RMS) difference between the HL and LL energy predictions by over 85% on the training data and over 75% on the test data.

Many computational methods, including self consistent field calculations in quantum chemistry, work by fixed-point iteration—repeatedly applying a given update function until convergence is achieved. A means to accelerate fixed-point iteration via imitation learning is explored. The approach is simple and needs only black-box access to the original update function. Experiments show that the approach successfully accelerates Hartree-Fock convergence, and that policies trained on one set of molecules transfer successfully to other molecules of the same general class. Imitation learning also leads to more-robust transfer than alternative methods that do not take into account the distribution of states induced by the learned policies.

Acknowledgments

Conducting good research is not a simple task. Throughout these five years working with my advisor, David Yaron, I have learned to dare to explore areas of research that are far beyond the low hanging fruit and always keeping in mind the importance of honest presentation of results. I am really grateful for what I have learned and all the interesting discussions we have had. I appreciate all that my advisor has thought me and all the support he has given.

David Yaron introduced me to the field of machine learning which I have found to be an exceptionally interesting field. He also introduced me to our collaborator Geoff Gordon who we had many meetings and great conversations. I truly appreciate all that I have learned from Geoff Gordon who has also been a great mentor. Working with Shiva Kaul and Tse-Han Huang was a great way to get familiar with machine learning and they guided me to get hands on experience on how to implementing different algorithms. Being able to attend the select group meetings gave me a chance to learn more about functional analysis and in more detail about the research conducted in Geoff Gordon's group.

I also had the privilege to work with Bruce Armitage, Guy Salama and Alan Waggoner on characterization of different organic dyes. I truly appreciate everything that I have learned from all the different dye meetings we have had. Collaborating with my friend Lisa Rastede, as well as Subhasish Chakraborty and Beth Gabris gave me a great insight in the actual experimental work. I want to thank all my undergraduates for their hard work and cooperation, especially Sukjin Jang who might not really be my undergraduate but still a good friend.

I want thank Bruce Armitage, Maria Kurnikova, Tomasz Kowalewski and Geoffrey Hutchi-

son for serving on my committee and providing me with great suggestions and feedback. Apart from my committee there are many other people in the department that I am thankful for. I appreciate all the great conversations I have had with faculty members such as Roberto Gil, Alexandr Ryabov (Sasha) and Gayathri Withers. Gizelle Sherwood, Carolyn Neiderlander, Robert Dalmasse and Lilian Crawford made teaching assisting a fun yet rich experience. Rea Freeland had a key role in my motivation for staying and proceeding my graduate studies. Brenda Chambers has been an amazing person and always very helpful. I also want to thank Patsey Haddock and Sara Wainer.

One thing that has really made me happy is the way that the theory suite has bloomed into a more social environment where people have great discussions about science and life at a daily basis. I want to thank first of all my group members starting with Christian Legaspi (LeGospen), who I go way back with, then Haichen Li (HC) and Christopher Collins (Tophee) for all the great discussions and everything that you have taught me, I really believe that the Yaron group has amazing people now, that includes Michael Karabinos and Eric Gottlieb (the goat), although he is more of a Tomek group member. I also want to thank my friends in the Kim group, especially Jon Wilcox and Nilesh Dhumal (Nils), and the people of the Kurnikova group.

I want to thank my friends and family who have supported me through the ups and downs. I will start with Anthony Brooks (Ant) who taught me so much about American culture. I, Ant and John Pettersson, who is not only a good friend but my only Swedish friend in Pittsburgh, together with Riley Ohlson (big smooth) are the proud members of ÅÄÖ. I want to thank Rafael Stern (Lusken), who supported me a lot, especially in the end, and we have had so many amazing discussions and I have learned much from him. I really did enjoy the good times with Tony Tran, who introduced me to all the Heinz people. I want to thank Alen Knapic (the knapkin) for his awesomeness, also Alen, Cynthia Chau Knapic and Jobin for everything (especially community Sunday). Mayank Mittal, probably the most carefree person who ever attended Carnegie Mellon University, always showed up at the perfect time. I got to bond best with Alex Beutel (el cambodiano) on the top of Mount Fuji along with my good friends Xavi Gonzalez Cam-

pos and Emil Ilysaov. I am utterly grateful to my close friend Sebastian Monadjemzadeh who not only supported and I had great conversations with but also helped me practice my Swedish language. Also all my other friends in Sweden who supported me such as Joakim trädgårdh (Jaypen), Martin Koskela (Korre), Stefan Yu, Ali Rabieh, Retta Abraham Tsehaye, Sam Solten and many others. Marina Masaki (Masago) has been great at giving me support and motivation. I am happy that I had the chance to meet so many good people and make such good friends: Antonina Simakova, Jake Mohin, Lorenzo Vasini, Melissa Lamson, Sourav Dey, Adam Ahern, Pritesh Patel, Ambarish Joshi, Mayank Mamtani, Ankush Kochhar, Marilia Cabral, Prabhat KC, Melih Özbeck, Taylor Canady, Carolyn Commer, Chirag Dhull, Joel Kranich, Pizza Paul, and so many more. I do not want to miss to acknowledge my friends in the taekwondo group as well as the walnut crew. I want to thank my entire family, my cousins, Daniel Wikström amongst others, for motivating me with their hard work, as well as my aunt and my uncles for their support.

Lastly, I want to thank my mother, Shirin Lalrezaie who always was persistent on motivating me and my sister on understanding the importance of studies, my father, Peyman Tanha who always challenged me to learn and want to teach me so much of his knowledge, and finally my sister, Marieh Tanha who has not only been a great support but also an excellent role model.

Contents

1	Overview	1
1.1	Problem Statement and Thesis Outline	1
1.2	Fluorescent Dyes	7
1.3	New Semiempirical Models	10
1.4	Using Machine Learning to Accelerate Convergence of Self Consistent Field Iterations	17
2	Characterization of a fluorescent dye for detection of potassium ion concentration	19
2.1	Introduction	19
2.2	Experimental Characterization	20
2.3	Computational Methods	21
2.4	Computational Results	22
2.4.1	Effects of Torsional Motion	22
2.4.2	Effects of a Point Charge	23
2.4.3	Origin of the Degenerate Excited States	24
2.4.4	Effects of a Potassium Ion	27
2.4.5	Excited States in the Presence of K^+	29
2.4.6	Semiempirical Calculations	30
2.4.7	Effects of Electronic Substituents	32
2.4.8	Effects of a Bridging Methylene	34

2.5	Conclusions	35
3	Computational characterization of fluorogenic dyes	37
3.1	Introduction	37
3.2	Computational Methods	38
3.3	Comparison of Computational Methods	39
3.4	Results and Discussion	40
3.5	Conclusion	42
4	Embedding parameters in <i>ab initio</i> theory to develop well-controlled approximations based on molecular similarity	43
4.1	Introduction	43
4.2	Methods	46
4.2.1	Chemical Data	46
4.2.2	Effective Hamiltonian	47
4.2.3	Context Sensitive Parameters	49
4.3	Results	50
4.4	Conclusions	58
5	An Embedded Parameters Model of CNOFH Containing Molecules	61
5.1	Introduction	61
5.2	Chemical Data	62
5.3	Embedding Parameters in the Low-Level Hamiltonian	66
5.4	Context Sensitive Scaling Factors	69
5.5	Training the pLL Model	70
5.6	Results	72
5.7	Discussion	82

6	Accelerating Hartree Fock convergence	87
6.1	Introduction	87
6.2	Background	89
6.2.1	Hartree-Fock	89
6.2.2	Imitation learning	90
6.3	Learning the Policy	91
6.4	Exprimental Design	94
6.5	Results	96
6.6	Conclusion	97

List of Figures

1.1	Chemical structure of SKC-513	8
2.1	Experimental data for SKC-513 showing the fluorescence intensity (in arbitrary units) as a function of the concentration of various ions in the range from 0-1000 mM. SR refers to experiments done in the presence of a sarcoplasmic reticulum.	20
2.2	As in Figure 2.1, but over an ion concentration range of 0-40 mM.	20
2.3	Structure of the simplified SKC-513 dye (SSKC-513).	21
2.4	Torsional potential for SSKC-513 obtained from DFT (B3LYP/6-21G) calculations and from various semiempirical model Hamiltonians.	22
2.5	Populations at room temperature as a function of torsional angle in the SSKC-513 dye, obtained from DFT (B3LYP/6-21G) calculations and from various semiempirical model Hamiltonians.	22
2.6	Two lowest excited-state energies of SSKC-513 as a function of torsional angle. Structures were optimized in ground state using SAM1 and vertical excitations were obtained with TD-DFT (TD CAM-B3LYP/6-31G**). The radii of the circles are proportional to the oscillator strengths.	23
2.7	Effects of a point charge on the excited states of SSKC-513. (The radii of the circles are proportional to the oscillator strength.)	24

2.8	Frontier orbitals of SSKC-513. There are two nearly degenerate HOMOs, one located on the heterocycle and the other on the phenylene ring. The LUMO is non-degenerate and located on the heterocycle. Optical intensity relies on good overlap between orbitals, such that only the transition from HOMO-N to the LUMO carries large optical intensity.	25
2.9	Frontier molecular orbitals of the three-fold symmetric parent of SSKC-513. . . .	25
2.10	Effects of a point charge on the excited states of the three-fold symmetric parent of SSKC-513. (The radii of the circles are proportional to oscillator strength.) . .	26
2.11	Electronic states of SSKC-513 (Figure 2.7) with (right) and without (left) a K^+ ion at a location corresponding to the center of the crown ether in SKC-513. . . .	28
2.12	Frontier orbitals of SSKC-513 in the presence of K^+ ion along with the orbital contributions to the three lowest-energy excited states. The vertical excitation energies and oscillator strengths, f , are also shown.	29
2.13	Excited states of the SSKC-513 dye as a function of torsional angle, obtained using semiempirical methods.	31
2.14	Effects of a point charge of magnitude q/ϵ on the excited states of SSKC-513 obtained using semiempirical methods.	32
2.15	Effects of a point charge on the excited states of the three-fold parent symmetric molecule, obtained using semiempirical methods	32
2.16	Effects of Flourine substituents on the two lowest excited states of SSKC-513. Adding F substituents to the phenyl-ring (left) raises the energy of the CT excited state, while adding F substituents to the heterocycle has the opposite effect. . . .	33
2.17	The two lowest excited states of a modified SSKC-513, in which a methylene bridges between the phenylene group and the heterocycle. (The radii of the circles are proportional to the oscillator strength.)	34

3.1	Correlation plots of predicted values from different calculations of the TO derivative dyes and experimental results	39
3.2	Frontier orbital analysis for TO and its donor/acceptor substituted analogues, including HOMO and LUMO orbitals of TO (far left), HOMO and LUMO energies (black), HOMO-LUMO gaps (blue arrows) and TDDFT excitation energies (red text).	40
3.3	(A) Normalized absorbance spectra of TO dyes ($1\mu\text{M}$) in methanol. (B) Normalized fluorescence spectra of TO dyes ($1\mu\text{M}$) in 90% glycerol in water.	41
4.1	Training of the pLL model on an ethane data set (solid lines), while monitoring performance on an ethane test set (dotted lines). The objective is the RMS error, Eq. 4.2, summed over all operators including E_{tot} (black lines). The RMS error, Eq. 4.2, for just the total energy, E_{tot} , (red lines) is also shown. The vertical lines show addition of context variables to the model (Table 4.2).	52
4.2	Error in the individual operators, excluding E_{tot} , for the optimization of Figure 4.1. The RMS error is obtained using Eq. 4.2, with summations limited to a particular operator type.	53
4.3	RMS error, Eq. 4.2, in the total energy, E_{tot} , for a variety of molecules, using the pLL model of Figure 4.1.	54
4.4	RMS error, Eq. 4.4, in the total energy, E_{tot} , for a variety of molecules, using the pLL model of Figure 4.1. This error reflects performance on the change in total energy arising from changes in geometry or environment.	55
4.5	RMS error associated with individual operators for the fits of Figure 4.4, computed using Eq. 4.4 with summation over all operators except E_{tot}	55
4.6	RMS error, Eq. 4.4, in total energy of various molecules, obtained from fitting the pLL model to ethane, with the total energy multiplied by a weighted factor. .	56

4.7	RMS error associated with individual operators for the fits of Figure 4.6, computed using Eq. 4.4 with summation over all operators except E_{tot}	56
5.1	The objective of Eq. 5.5 versus iteration during training of a model on the <i>combined</i> dataset. The pLL model uses the <i>JK</i> policy and decomposed energies are included in the objective. Vertical lines indicate addition of the contexts of Table 5.3.	73
5.2	Residual error in E_{tot} as a function of context level for pLL models trained on the <i>ethane</i> dataset, without inclusion of decomposed energies in the objective of Eq. 5.5. Residual error is RMS disagreement between the pLL and HL models, quoted relative to initial disagreement between the LL and HL models. Panels refer to the policies of Section 5.3.	74
5.3	Residual error in E_{tot} as a function of context level for pLL models trained on the <i>ethane</i> dataset, with inclusion of decomposed energies in the objective of Eq. 5.5. Conventions are as in Fig. 5.2.	75
5.4	Residual error in E_{tot} as a function of context level for pLL models trained on the <i>combined</i> dataset, without inclusion of decomposed energies in the objective of Eq. 5.5. Conventions are as in Fig. 5.2.	76
5.5	Residual error in E_{tot} as a function of context level for pLL models trained on the <i>combined</i> dataset, with inclusion of decomposed energies in the objective of Eq. 5.5. Conventions are as in Fig. 5.2.	77
5.6	Boxplots of disagreements in E_{tot} between HL and the best performing pLL model. The pLL model uses the <i>2elec</i> policy and is trained on the <i>combined</i> dataset without inclusion of decomposed energies. Initial refers to the unparametrized LL model.	78
6.1	The chemical structure of molecules in the datasets, R = H, F, OH or NH ₂	94
6.2	Average error in energy as a function of iteration tested on <i>pent2ene</i> test dataset .	96

6.3	Average error in energy as a function of iteration tested on <i>cycloPentene</i> dataset .	97
6.4	Average error in energy as a function of iteration tested on <i>fluoroMethylHeptene</i> dataset	97

List of Tables

1.1	Computational cost for a set of different methods for including electron correlation (columns) and with varying basis sets (rows), on a dataset of fluoromethyl compounds. [18]	3
3.1	Extinction coefficients, wavelength maxima (nm) and fluorescence quantum yields for TO dyes in homogeneous solution.	42
4.1	List of parameter types for ethane	49
4.2	hoc context variables	50
4.3	Detailed errors (in kcal/mol) from the fit in Figure 4.6 corresponding to $w=10$	57
5.1	RMS differences between the HL model and SEQC methods for the <i>combined</i> dataset. E_{tot} refers to the change in energy associated with geometric distortions and applied fields (see Section 5.5).	66
5.2	Amount of training data in the datasets used to develop pLL models.	66
5.3	Context variables, c_i of Eq.5.4, for the three levels of context added sequentially to the model during training.	70
5.4	RMS errors for models trained on the <i>combined</i> dataset without inclusion of decomposed energies. n_{par} is the number of parameters in the model. Large RMS errors reflect failures of SCF iterations to converge (Section 5.6). Units are eV for E_{orb} and kcal mol ⁻¹ for all other quantities.	79

5.5	RMS errors for models trained on the <i>combined</i> dataset with inclusion of decomposed energies. Conventions are as in Table 5.4.	80
5.6	Effects of increasing the weighting of E_{chg} in the objective of Eq. 5.5, for pLL models using the <i>2elec</i> policy with full context trained without inclusion of decomposed energies.	81
6.1	Example: applying DAgger on expert demonstrations with step size = 2	92

Chapter 1

Overview

1.1 Problem Statement and Thesis Outline

With the discovery of the atom a new world of science was opened up and quantum mechanics was born.[1] Schrödingers equation along with laws of quantum mechanics established a basis for understanding the small particles that construct the world that we perceive. This gave rise to the field of quantum chemistry which can be defined as the application of quantum mechanics to chemical systems.[2, 3, 4, 5] Although the equations governing the motion of electrons in molecular systems were known, solving them for systems beyond the two-body problem was challenging. While this challenge still remains, a variety of *ab initio* approaches have been developed that provide solutions to the many-body problem with accuracy that is chemically useful. Ultimately the goal is to develop methods that give as accurate of results as possible, at a reasonable computational cost. The methods that exist today for calculating the electronic structure of molecules give us valuable information that helps us understand molecular and chemical behavior. The most reliable methods are for the ground electronic state of stable molecules.[6, 7, 8, 9, 10] Methods for excited states[11, 12, 13] and transition states[14, 15, 16] are also steadily improving. A remaining fundamental challenge is lowering the computational cost such that larger molecular systems such as polymers or proteins may be addressed.

The computational cost of highly accurate quantum chemical methods increases rapidly with system size. *Ab initio* electronic structure calculations vary along two general dimensions (Table 1.1). The first is the basis set, which for the methods considered in this thesis, consists of atomic orbitals centered on each atom. A minimal basis has just enough atomic orbitals to hold the electrons, e.g. $1s$ on H and $1s, 2s, 2p$ on second row elements. To get a reasonable result, a much larger basis set is typically needed. A common choice for organic molecules is the 6-311++G** basis set. In this basis, the valence atomic orbitals are expanded to include a narrow and a diffuse function, such that the charge density can expand and contract. In addition, the basis includes polarization functions (p on H, d on second-row elements), as indicated by the “**” in the name of the basis, and diffuse functions, as indicated by the “++”.

The second dimension along which *ab initio* methods vary is the approach used to handle electron correlation, i.e. the tendency of electrons to move in concerted ways in order to avoid repulsive Coulombic interactions. Hartree Fock (HF) theory, also referred to as Self Consistent Field (SCF) theory, is the most fundamental form and does not include electron correlation. A large variety of correlated methods have been developed, with computational costs that increase steeply with the expected accuracy of the results. Density Functional Theory (DFT) is a popular method that includes the effects of electron correlation yet has computational costs similar to that of Hartree Fock.[17]

Table 1.1 illustrates the rapid increase in computational cost along the above two dimensions. The cost increases as the size of the basis raised to a power that depends on the method used to include correlation. This increase in cost grows with system size, such that highly accurate methods become impractical on large systems.

Finding ways to lower the cost of computations on large molecular systems is an active area of current research. Most current approaches attack the issue of scaling by taking advantage of *nearsightedness*. [19] Nearsightedness implies that interactions become simpler at long-range, and can be replaced by increasingly coarse-grained multi-polar interactions. Algorithms that take full advantage of nearsightedness, while evaluating Coulomb interactions, can achieve a

Basis set	HF	B3LYP	MP2
6-31G	1	1	1
6-311G	1.6	1.4	2.9
6-31+G	1.3	1.5	1.8
6-311+G	2.0	2.4	5.8
6-31G(d,p)	1.9	2.0	3.3
6-311G(d,p)	2.9	3.1	7.9
6-31+G(d,p)	3.5	3.2	8.5
6-311+G(d,p)	7.2	5.6	14.2

Table 1.1: Computational cost for a set of different methods for including electron correlation (columns) and with varying basis sets (rows), on a dataset of fluoromethyl compounds. [18]

computational cost that scales linearly with system size in the limit of large systems. Such linear-scaling algorithms have been developed for many *ab initio* methods.[20, 21, 22, 23, 24]

Fragment based methods use a different approach to take advantage of nearsightedness. The molecular system is divided into smaller partitions (fragments), that may have overlap. These fragments are calculated separately and the results are combined in various ways to predict the energy of the entire molecule. A recent review by Gordon *et al* [25] provides a thorough review of such methods. In addition, Hebert[26] has proposed a unified view of such methods based on a generalized many-body expansion. These approaches have the additional advantage of allowing the fragment calculations to be run in parallel, leading to substantially enhanced performance on modern computer architectures.

An alternative approach to enabling calculations on large molecules is the use of semiempirical quantum chemical (SEQC) models. As the name suggests, these methods are parameterized and fitted to empirical data. Early examples include the Parsier-Parr-Pople (PPP) model of π -electrons[27, 28], and the Complete Neglect of Differential Overlap (CNDO/2) method[29, 30] which extended PPP to include σ electrons. The Intermediate Neglect of Differential Overlap (INDO) method is closely related to CNDO/2, and is still in use for computing electronic excitations in organic systems.[31, 32, 33] For the structure and properties of ground electronic states, the Neglect of Diatomic Differential Overlap (NDDO) methods remain in common use. NDDO includes Modified Neglect of Diatomic Overlap (MNDO)[34], Austin Model 1 (AM1)[35],

Semiempirical *ab initio* Model 1 (SAM1)[36] and Parameterized Model 3 (PM3). [37] Other more recent NDDO methods are PM6 [38, 39] and RM1 [40], which are reparameterized versions of PM3 and AM1 trained on a larger dataset, as well as orthogonalization-corrected methods (OM1, OM2, OM3). [41, 42] These semiempirical models perform quite well, especially given the simplicity of the model and the low computational cost. A more recent commonly used semiempirical model is Density Functional Tight Binding (DFTB) and its derivative models.[43, 44, 45, 46, 47] DFTB derives the electronic parameters of a tight binding model from DFT calculations on isolated atoms. The empirical portion is a set of repulsive interactions between atoms that are parameterized to accurate calculations on representative molecules.

Semiempirical models use minimal basis sets and make simplifications to the Coulomb interactions that go beyond what may be justified via nearsightedness. The errors introduced by these approximations are minimized by fitting the empirical parameters to data on molecules that are similar to those on which the model will be used. These methods can therefore be viewed as taking advantage of *molecular similarity*, i.e. the tendency of molecular fragments to behave similarly in similar environments. A central theme of this thesis is exploring alternative means to take advantage of molecular similarity in quantum chemical computations.

Approximations based on nearsightedness benefit from having a clear physical basis that inspires the form of the approximation. For example, when computing Coulomb interactions at long range, charge distributions may be replaced with multipoles. This physical basis suggests specific forms for the approximations of linear scaling [20, 21, 22, 23, 24] and fragment-based methods. [48] The resulting algorithms may have parameters that define, for example, the degree to which interactions are simplified at long range. These parameters can be adjusted to find an optimal balance between computational savings and errors introduced by the approximation. This high level of control is related to the direct physical basis underlying the approximation.

Achieving a similar level of control in approximations based on molecular similarity is more challenging. Ideally, the resulting methods would provide a means to smoothly improve the accuracy of the approximation such that an appropriate balance between accuracy and computational

cost could be found for the target application.

The flexible model forms of machine learning, such as neural nets, enable construction of models that are trained to data on representative molecules and then used to make predictions for molecules not included in the training set. Such models have been successfully used to develop corrections to predictions of *ab initio* theory that are applicable across broad classes of molecules and that substantially reduce the errors in heats of formation and other properties. [49, 50, 51, 52, 53, 54, 55, 56, 57, 58] Models have also been developed that generate predictions using descriptors, such as the Coulomb matrix, [59] that do not require solution of the Schrödinger equation. The Coulomb matrix is a square matrix with dimension equal to the number of atoms in the molecule. The diagonal elements hold $0.5Z_i^{2.4}$ where Z_i is the nuclear charge of the atom. The off-diagonal elements hold the Coulomb energy between the associated nuclei, $Z_iZ_j/R_{i,j}$, where $R_{i,j}$ is the distance between the nuclei. Various means of sorting the Coulomb matrix to obtain a unique map between the Coulomb matrix and the molecular structure have been explored. The ability of machine learning methods to predict the atomization energy of molecules from the Coulomb matrix has been extensively studied, with results that compete with quantum chemistry at substantially reduced computational cost. [59, 60] Extensions to other properties, such as polarizability and frontier orbital energies, have also led to promising results. [61] Recently, Guzik *et al* have obtained highly accurate predictions of frontier orbital energies using features based on Morgan chemical fingerprints as inputs to neural nets.[62]

The current work explores the use of model Hamiltonians as flexible functional forms for model development, and so bridges between the flexible model forms of machine learning, [63] such as neural nets or reproducing-kernel Hilbert space methods, and the model Hamiltonians of SEQC, such as Intermediate Neglect of Differential Overlap (INDO) [31] and Neglect of Diatomic Differential Overlap (NDDO). [35] As in machine learning, the goal is to create model forms that are sufficiently flexible that patterns in the training data may be discovered. As in SEQC, the strategy is to use *ab initio* quantum chemistry as inspiration for model forms that can describe chemical phenomena well.

The approach used in the investigations is adapted from machine learning. First, a set of molecular systems across which similarity may be reasonably expected to hold is specified. This is done by selecting a group of related molecules and creating a set of molecular “instances” by distorting the molecular geometry and applying external electrostatic environments. Predictions of a high-level (HL) *ab initio* model are then generated to serve as target training data (also referred to as “labels” in supervised learning). The parameters of a lower-cost model are adjusted to reproduce the target training data and the performance of the model is evaluated on instances not included in the training data.

The approach used to create the model is adapted from traditional SEQC: we embed empirical parameters into a quantum chemical Hamiltonian. First, matrix elements of the operators that appear in a low-level (LL) *ab initio* Hamiltonian are generated, including kinetic energy, attraction between the electrons and each nucleus, and the electron-electron repulsion. Selected matrix elements of these operators are then multiplied by scaling factors that serve as the parameters of the model, leading to a parametrized low-level (pLL) model. Matrix elements that are not scaled retain the values of the LL *ab initio* Hamiltonian. This general approach is flexible because a variety of LL Hamiltonians may be envisioned into which parameters may be embedded. For these initial studies, the LL model is restricted to self-consistent-field (SCF) solutions within a minimal basis set. A variety of schemes for embedding parameters are also explored.

This thesis begins with the application of quantum chemical models to the design of efficient biological dyes for monitoring cellular activity. These dyes are chromophores whose fluorescence is sensitive to changes in their local environment induced by binding to a biomolecule or by a change in local ion concentration. The general strategy in the investigation of these fluorescent dyes is to take advantage of computationally inexpensive semiempirical models to explore the general behavior and properties of the dyes and their derivatives. Using methods such as SAM1 for ground state optimization and INDO for vertical excitation provides us with the properties of the molecule in both the ground and excited state. Once we have a general understanding of the dye and its properties, DFT methods are used to confirm and refine the semiempirical results.

This strategy is quite effective since the low computational cost of SAM1 and INDO provide almost instantaneous results, enabling extensive explorations of the photophysics. The more costly *ab initio* computations can then focus on making the results more quantitative and refining the general picture that emerged from the semiempirical investigations.

From the application of quantum chemical models we will move onto the exploration and development of novel quantum chemical methods. The focus of this work is to develop a model inspired by semiempirical methods and to explore the limits of molecular similarity. Unlike nearsightedness, molecular similarity is more difficult to define precisely. However, the study of molecular similarity maps onto machine learning approaches.[63] Machine learning is the study of algorithms that can learn from data and so lead to predictive models. Current *ab initio* calculations allow us to generate massive amounts of data on molecular fragments. The challenge is to develop models that can utilize patterns in this data to make useful predictions. The models explored here are inspired by past semiempirical models, but adopt a quite different means for both embedding parameters in a model Hamiltonian and for adjusting these to find agreement with data.

Through the application of existing models and the development and exploration of new methods, this thesis aims to lay useful groundwork for continued development and use of semiempirical quantum chemical methods.

1.2 Fluorescent Dyes

Fluorescence imaging provides a set of powerful techniques for monitoring biological processes in living organisms. [64, 65, 66, 67, 68, 69] These techniques rely on dye molecules that report on the location and the environment of biomolecules. Fluorogenic dyes are a special class of fluorophores that show distinctly different emission intensities depending on their local environment. [70, 71, 72, 73, 74, 75, 76] Two types of fluorogenic dyes are studied here.

For neural processes, it is useful to have dyes whose fluorescence tracks the concentration

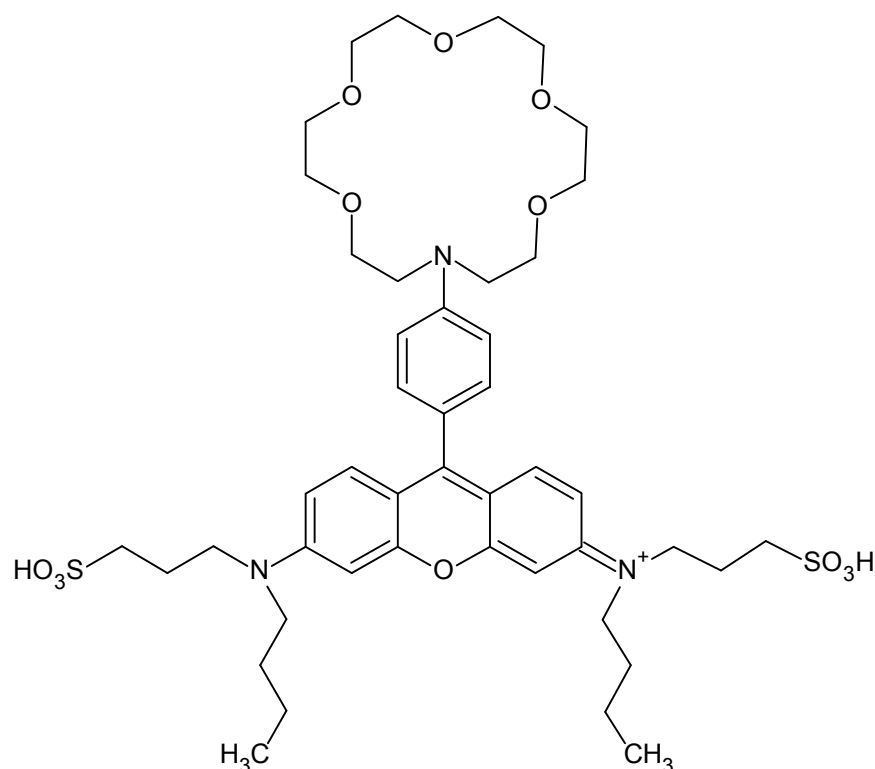


Figure 1.1: Chemical structure of SKC-513

of K^+ ion. SKC-513 ((E)-N-(9-(4-(1,4,7,10,13-pentaoxa-16-azacyclooctadecan-16-yl)phenyl)-6-(butyl(3-sulfopropyl)amino)-3H-xanthen-3-ylidene)-N-(3-sulfopropyl)butan-1-aminium) is a promising candidate for ion detection (Figure 1.1). SKC513 has a crown ether portion that can bind a potassium ion, and a chromophore that signals the presence of the ion. Experimental data on SKC513, from collaborators in the Waggoner and Salama groups, shows a substantial change in fluorescence quantum yield as the K^+ concentration varies from 0 to 1000 mM. Chapter 2 reports quantum chemical calculations that explore the mechanism through which the ion alters the fluorescence of the chromophore. These computations indicate that SKC-513 is a photoinduced electron transfer (PET) sensor, a class of dyes whose fluorescence quantum yield is known to be sensitive to ion binding. [67, 68, 69]

The change in quantum yield for SKC-513 arises from the competition between a radiative pathway, emission from a neutral excited state (N) created on photoexcitation, and a non-radiative pathway, which is mediated by a charge-transfer (CT) excited state. Binding of the

ion raises the relative energy of the CT state and this alters the branching ratio between the radiative and non-radiative pathways. The computations presented in this work find that the lowest-energy excitations of SKC-513 correspond to a fluorescent state with predominantly N character and a less optically intense state with predominantly CT character. The potassium ion preferentially destabilizes the CT state, thereby altering the quantum yield, consistent with the mechanism implicated in other PET sensors. A feature of the excited states of SKC-513 that may be contributing to the high sensitivity of the fluorescence quantum yield to ion binding is a near degeneracy of the N state, which mediates the radiative pathway, and the CT state, which mediates a nonradiative pathway.

The high sensitivity of SKC-513 to ion binding is related to the near degeneracy of the N and CT states. This near degeneracy may appear to be accidental, given that SKC-513 has only a two-fold symmetry axis. However, comparison to a three-fold symmetric parent reveals a systematic origin. In the parent molecule, the lowest excited state is doubly degenerate. In addition, the molecular orbitals involved in these degenerate excitations have low density in the regions of the molecule that must be altered to convert the parent molecule to SKC-513. The degeneracy of excited states in SKC-513 is therefore not accidental, but can be traced back to degeneracies expected for this three-fold symmetric parent.

The calculations also find that replacing the potassium ion with a point charge has little effect on the nature of the relevant excited states. This suggests that changes in the excited states of the chromophore in SKC-513 arise from field effects of the potassium ion, as opposed to more specific interactions such as ligation with the nitrogen atom that connects the crown ether to the chromophore.

Chapter 2 concludes by suggesting modifications to SKC-513 that may improve, leading to larger changes in fluorescence upon ion binding.

Chapter 3 explores another class of fluorogenic dyes, that of unsymmetrical cyanine dyes. These dyes may exhibit enhanced fluorescence when placed in an environment that constrains the conformation, such as when dissolved in a viscous solution or when bound to a protein.

For fluorescence imaging, the goal is to find dyes for which the quantum yield is low in water but high when bound to a protein. This gives the dye a switch property which can be used to detect protein rich environments.[77, 78] For the dyes studied here, the excited state has a nonradiative path which requires the molecule to twist along the central methine bridge. In a constrained environment this pathway is inhibited, thus increasing the branching ratio to the radiative pathway.[77, 79]

Experiments on derivatives of Thiazole Orange, carried out by collaborators in the Armitage group, show that the addition of electron donating and withdrawing substituents provides a means to fine tune the wavelength of the light emitted, while maintaining fluorogenic properties. Chapter 3 reports computations on these dyes and a comparison with experimental data. The results indicate that the changes in wavelength can be understood in terms of the effects of the substituents on the frontier orbitals. The Highest Occupied Molecular Orbital (HOMO) resides mainly on the benzothiazole while the Lowest Occupied Molecular Orbital (LUMO) is spread across the entire molecule. This causes electron withdrawing and donating groups to shift the HOMO and LUMO energies in predictable ways, with the predicted effects on the HOMO-LUMO gap agreeing with the experimentally observed spectral shifts.

The predictions of TDDFT, obtained using various functionals, and the predictions of INDO computations were compared to each other and to the experimental data. The results indicate that INDO is nearly as reliable as TDDFT, with both approaches giving good agreement with experimental results in most cases. This supports the use, in this thesis, of INDO theory for the initial explorations of the dye properties.

1.3 New Semiempirical Models

The general approach of semiempirical methods is to construct a Hamiltonian that is a simplified version of the *ab initio* Hamiltonian, and embed parameters in this simplified form. These parameters are then fit to training data, consisting of either experimental or *ab initio* data on

molecules related to those for which the model is being developed. Past semiempirical methods share a few general approximations as listed below:

MinimalBasis All assume a minimal basis set.

IgnoreCore All ignore the core orbitals.

UnitOverlap All ignore overlap between basis functions on different atoms.

SimplifiedCoulomb All make substantial approximations to two electron integrals between atoms as discussed further below.

Uncorrelated Most are parameterized to agree at the SCF level. An exception is INDO/S and other methods that target excitation energies. These models are parameterized to excitation energies obtained at the Singles Configuration-Interaction (SCI) level.[31, 32, 33] Note that the resulting model Hamiltonians are sometimes used for correlated calculations, even though the Hamiltonians were parameterized at the SCF level.

The one-electron terms of the electronic Hamiltonian including the kinetic energy of the electrons, KE , and the Coulomb interaction between the electrons and the nuclei, V_{nuc} . In *ab initio* theory, the KE and V_{nuc} appear through separate operators. In semiempirical theory, the effects of these operators are combined into two types of terms in the Hamiltonian. The first class of such terms are the energies of the atomic orbitals, which are analogous to the α terms of Hückel theory. These terms set the energies of the s and p orbitals of the isolated atoms, including the effects of both the kinetic energy of the electron in that atomic orbital and the interactions with that atom's nucleus. The second class of one-electron terms are the interaction energies between atomic orbitals, which are analogous to the β terms of Hückel theory. These terms model the bonding interaction between atoms and again include the effects of both kinetic energy and electron-nuclear interaction. However, rather than base β on the kinetic energy or electron-nuclear operators, semiempirical methods typically take the one-electron interaction between atomic orbitals as being proportional to the overlap between these orbitals, $S_{i,j}$. The proportionality constants serve as the semiempirical parameters adjusted in fitting the model.

(Note that DFTB does not treat the one-electron terms empirically and instead computes them from the electron density of reference atoms.[43, 44, 45, 46, 47])

The principle feature that distinguishes semiempirical model Hamiltonians is the treatment of Coulomb interactions between electrons, and this is reflected in the names given to many of these methods. The simplified treatment of electron-electron interactions, along with the use of a minimal basis, leads to significant reductions in computational cost. The Coulomb interactions between electrons appear in the quantum chemical Hamiltonian through two-electron integrals

$$(ij|kl) = \int dr_1 dr_2 \phi_i(r_1) \phi_j(r_2) \frac{1}{r_{12}} \phi_k(r_1) \phi_l(r_2) \quad (1.1)$$

where i, j, k and l refer to atomic basis functions while r_1 and r_2 refer to positions of two interacting electrons. Eq. 1.1 captures the Coulomb interaction between an electron in a charge distribution corresponding to the product of basis functions, $\phi_i(r_1)\phi_j(r_1)$, and an electron in a charge distribution, $\phi_k(r_2)\phi_l(r_2)$. If one assumes the atomic orbitals are localized to an individual atom, $\phi_i(r)\phi_j(r)$ is zero unless i and j are orbitals on the same atom, or “center”. The term “differential overlap” refers to products $\phi_i(r)\phi_j(r)$ where i is not equal to j . Another common terminology is that of “n-center” integrals, where n is the total number of centers across which i, j, k and l are spread. Past semiempirical models set all 3- and 4-center integrals to zero and keep only a restricted set of the remaining integrals.

1-center integrals involve only atomic orbitals on the same atom, and describe the self-energy due to interactions between electrons on that atom. The Slater-Condon parameters (F0, G1 and F2) can be used to construct the two-electron integrals from a set of just three parameters per element, in a manner that retains rotational symmetry[80]. This approach is used in the INDO and NDDO methods, with the values being taken from spectroscopic analysis of atomic spectra for most methods. Notable exceptions are PM3 and PM6, in which these parameters are included in the empirical parameterization.

The only 2-center integrals retained in past semiempirical models are those where i and j are on one atom and k and l are on another atom. These integrals describe the interactions between the electron densities of these two atoms. In INDO and DFTB, the following approximation is applied,

$$(ij|kl) = (ii|kk)\delta_{ij}\delta_{kl}$$

which leads to the Coulomb interactions between atoms being treated as that between point charges. NDDO methods extend this to include interactions between multipoles on the atoms. NDDO assumes $\phi_i(r_1)\phi_j(r_1)$ of Eq. 1.1 is zero if i and j are not on the same atom. When i and j are on the same atom, $\phi_i(r_1)\phi_j(r_1)$ is replaced with: a point charge, if i and j are both s orbitals; a dipole, if i and j are an s and p orbital; and a quadrupole, if i and j are both p orbitals. The same treatment is applied to $\phi_k(r_2)\phi_l(r_2)$. The integral is evaluated as the interaction between these multipoles. SAM1 is similar to NDDO, except that the two-center integrals are explicitly evaluated and then multiplied by factors that are chosen to mimic the results of NDDO. [36]

Another aspect of past semiempirical models is the inclusion of empirical potentials between atoms, that are rationalized as including effects from the core electrons, which are neglected in the overall Hamiltonian. In density functional tight binding (DFTB), such core-core interactions are the only terms treated empirically, with the electronic Hamiltonian being computed from DFT. [43, 44, 45, 46, 47] In all cases, the inclusion of such terms is essential for obtaining reasonable predictions for both the molecular structure and heats of formation.

This thesis explores an approach that is inspired by the above semiempirical models, in that parameters are embedded in an *ab initio* Hamiltonian and adjusted to obtain agreement with external data. However, rather than make substantial approximations regarding the form of the Hamiltonian, such as the above treatments of the two-electron integrals, the complete form of an *ab initio* Hamiltonian is retained. The parameters are then used only to bring a low-level (LL) *ab initio* model into agreement with a high-level (HL) *ab initio* model. In the studies below, both the LL and HL model use Hartree-Fock theory. The difference is in the basis set, with the LL

model using a minimal basis and the HL model using a much larger basis.

The LL model with embedded parameters will be called the pLL model, for parameterized-LL model. When all parameters are set to zero, the original LL *ab initio* model is recovered. As detailed in Chapter 4, the disagreement between the LL and HL model is relatively small compared to the errors introduced by the substantial approximations of traditional semiempirical theory. The parameters are therefore used to make a relatively small correction to the LL model, as opposed to compensating for large approximations regarding two-electron integrals. Retaining the form of an *ab initio* Hamiltonian also makes the pLL model parallel to the HL model. This allows, for instance, the fits to include bringing various decompositions of the Hamiltonian operator into agreement between the pLL and HL models. The pLL model form is also quite flexible, because a number of different schemes may be devised for embedding the parameters.

A number of factors support the use of minimal-basis Hartree Fock theory for the pLL model. This level of theory is qualitatively able to describe a broad range of chemical phenomena, including chemical valency, hybridization, and resonance. In addition, it is possible to transform, via Quambo[81], a self-consistent-field solutions in an arbitrary basis to a minimal-basis form, with no loss in accuracy. The parameters of the pLL model may therefore be viewed as an attempt to bring the Hamiltonian matrix elements of the LL model into agreement with those predicted by Quambo.

For the model to be useful, it must be possible to train the model on small molecules and then apply the model to larger molecules. This is promoted by training to both the energy and the charge distribution of small molecules. By training to the energy of small molecules, the model incorporates information on short-range interactions. By training to the charge distribution of small molecules, the model may incorporate information on longer-range interactions. The pLL models used here support this by using the LL form for those portions of the Hamiltonian that describe long-range interactions. At sufficiently long range in SCF calculations, the interaction between regions of electron density becomes equal to the Coulomb interaction between those regions. In the pLL models, the electron density, $\rho(\mathbf{r})$, is described by the density matrix expressed

in the LL basis set,

$$\rho(\mathbf{r}) = \sum_{i,j} \rho_{i,j} \phi_i(\mathbf{r}) \phi_j(\mathbf{r}), \quad (1.2)$$

where $\rho_{i,j}$ is the density matrix and $\phi_i(\mathbf{r})$ are LL basis functions. Because integrals describing longer-range interactions retain their LL *ab initio* values, the longer-range interactions in the pLL models have the form

$$E_{long\ range} = \sum_{\substack{i,j \in region_1 \\ k,l \in region_2}} \rho_{i,j} (ij|kl) \rho_{k,l} \quad (1.3)$$

where $(ij|kl)$ are two-electron integrals in chemist’s notation. Eq. 1.3 is an accurate expression for the interaction between the charge distributions of Eq. 1.2. The degree to which the pLL model provides an accurate description of longer-range interactions is then related to the degree to which the density of Eq. 1.2 accurately describes the HL density. This accuracy is limited by both the use of a LL basis and by the degree of success obtained in training the model parameters to best reproduce the HL charge density. The studies in Chapters 4 and 5 test the ability of models trained on small molecules to make predictions for larger systems.

The following list of features is parallel to that given above, to aid comparison with past semiempirical models:

MinimalBasis The pLL model continues to use a minimal basis set.

IgnoreCore Unlike past semiempirical models, the pLL model includes the core orbitals such that the pLL model reduces to the LL model when all parameters are set to zero.

UnitOverlap The pLL model uses the overlap matrix of the LL model, as opposed to the unit matrix used in past semiempirical models.

SimplifiedCoulomb The pLL model retains all of the two-electron matrix elements of the LL model, such that the LL model is recovered when the parameters are set to zero.

Uncorrelated The pLL model, and the HL model to which it is parameterized, use Hartree-Fock theory and so ignore correlation.

The way in which parameters are embedded in the pLL model also leads to some substantial differences from past semiempirical approaches. In past models, the one-electron portion of the Hamiltonian is described with parameters that are analogous to Hückel theory: the energy levels of atomic orbitals are given terms analogous to α and the bonding parameters are analogous to β . In the pLL model, on the other hand, the matrix elements corresponding to kinetic energy and electron-nuclear interaction are multiplied by scaling factors. In past models, the two-electron portion of the Hamiltonian invokes the strong approximations detailed above. In the pLL model, all two-electron integrals are retained and parameters are embedded through either of the following two schemes:

2elec The 2-electron integrals are modified directly by the embedded parameters.

JK The coulomb, J , and exchange, K , matrices of Hartree Fock theory are modified by the embedded parameters.

A number of different schemes, or “policies” are explored for embedding the parameters into the operators of the LL Hamiltonian. In addition, the parameters are made functions of the local “context”, by including dependence on bond lengths, atomic charges and bond orders. This context-sensitivity is found to be important for achieving good performance.

The pLL model uses substantially more parameters than past semiempirical models, with the goal of creating a model that is sufficiently flexible to describe chemical phenomena. This makes the approach analogous to the flexible model forms of machine learning, such as neural nets, which use orders of magnitude more parameters than the pLL models explored here. Two general strategies are used to prevent the large number of parameters from leading to models that do not transfer well across chemical systems. The first strategy is fitting the data to a large amount of detailed data. For each molecule type included in the fits, a number of strongly perturbed geometries are included, each geometry is placed in a set of external electrostatic environments, and a broad set of properties is computed including decomposition of the energy into subcomponents. The second strategy is adopting methodologies from machine learning to

prevent “over fitting”. Over fitting occurs when the model does well on the data used to train the parameters, but poorly on data not included in the fits. One means used to prevent over fitting is the use of a validation set to decide when to terminate the fitting process. The model parameters are fit to a training data set and, as the fit progresses, the performance on the validation set is monitored. When the performance on the validation set begins to degrade, the fit is terminated. Another means used to prevent over fittings is “regularization”, in which a penalty is applied to large values for the parameters. This helps keep the pLL model close to the LL model, and so helps prevent the pLL model from giving spurious results on new molecules.

Chapter 4 considers hydrocarbons in strongly perturbing electrostatic environments. The results show that models trained on small systems can transfer to larger systems, while reducing the error between the HL and LL by 85%, to ~ 2 kcal/mol. A number of different policies for embedding the parameters are explored. This chapter also finds that including decomposed energies, such as kinetic energy and electron-nuclear attractions, in the fits does not improve the model performance.

Chapter 5 considers hydrocarbons substituted with amine, hydroxyl and fluoro groups. The results again show that models trained on small systems can transfer to larger systems, while reducing the error between the HL and LL by 75%, to ~ 2 kcal/mol. Various policies for embedding parameters in the pLL model are compared. As in Chapter 4, the inclusion of decomposed energies did not improve model performance. This points toward the need for future work to include alternative decompositions of the energy.

1.4 Using Machine Learning to Accelerate Convergence of Self Consistent Field Iterations

Self consistent field (SCF) calculations are among the most commonly used approaches in quantum chemistry, either as an end in themselves or as a starting point for more advanced methods.

Such calculations start with a guess for the electronic density matrix and generate a sequence that, ideally, converges to a fixed point. The mostly commonly used approaches are variants of the Direct Inversion in the Iterative Subspace (DIIS) method. [82] DIIS uses a linear combination of previous density matrices to construct a new input guess density at each iteration of SCF. The coefficients for the linear combination are found by minimizing an error vector. A number of different DIIS methods have been developed, each of which uses a different error vector. [83, 84, 85]

Chapter 6 explores the use of imitation learning to accelerate SCF convergence. Imitation learning is a branch of machine learning which usually focuses on multistep processes, such as driving of an automobile. Such multistep processes generate a series of states, referred to as a trajectory. Imitation learning algorithms are trained to trajectories generated by experts, such as humans driving a car. The goal of the algorithm is to, from a given starting point, generate a trajectory that mimics that of the expert. Here, the trajectory is the series of density matrices that DIIS generates on its trajectory from the initial guess density matrix to the final converged density matrix. From the DIIS trajectory, expert trajectories with accelerated convergence can be created. For example, deleting every second density matrix from the DIIS trajectory leads to an expert trajectory that converges twice as fast as the original. In Chapter 6, the DAgger [86] algorithm is used to train an algorithm that can mimic such accelerated trajectories. The algorithm is trained on a set of organic molecules and then tested on similar molecules. The results on the test molecules are promising, suggesting that imitation learning has the potential to help accelerate SCF convergence.

Chapter 2

Characterization of a fluorescent dye for detection of potassium ion concentration

2.1 Introduction

Florescence imaging provides a set of powerful techniques for monitoring biological processes in living organisms [64, 65, 66, 67, 68, 69]. These techniques rely on dye molecules that change their fluorescence behavior under varying environments. For neural processes, it is useful to have dyes whose fluorescence tracks the concentration of K^+ ion. The experimental data presented below for the dye SKC-513 shows a substantial change in quantum yield as K^+ concentration varies between 0 and 1000 mM. The dye has a crown ether portion that binds the potassium ion, and a chromophore that signals the presence of the ion. The experimental data was gathered by our collaborators Subhasish K. Chakraborty, Beth Gabris, Alan S. Waggoner and Guy Salama. The contribution of this thesis is computations that give insight into the high sensitivity of the fluorescence quantum yield upon ion binding, and provide suggestions for improving the sensitivity of the quantum yield to ion binding. The experimental data is included below to place the computational work in context. Portions of this chapter is based on published work. [87]

2.2 Experimental Characterization

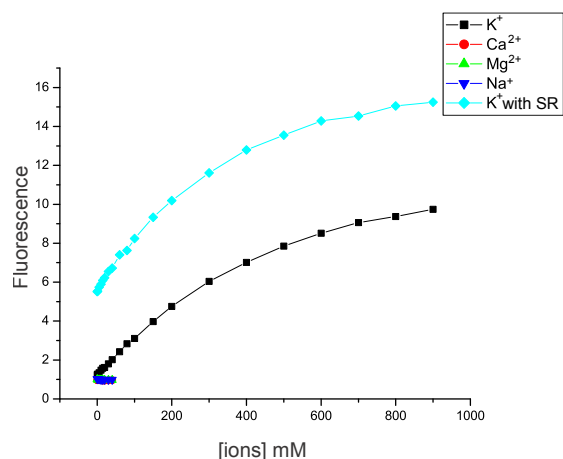


Figure 2.1: Experimental data for SKC-513 showing the fluorescence intensity (in arbitrary units) as a function of the concentration of various ions in the range from 0-1000 mM. SR refers to experiments done in the presence of a sarcoplasmic reticulum.

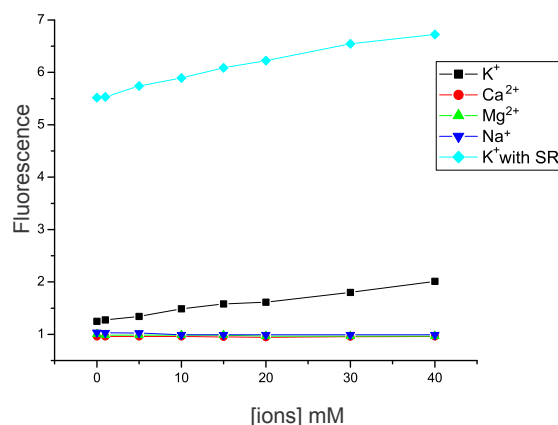


Figure 2.2: As in Figure 2.1, but over an ion concentration range of 0-40 mM.

The fluorescence intensity of SKC-513 as a function of ion concentration is shown in Figures 2.1 and 2.2. Figure 2.1 highlights how fluorescence intensity increases with K^+ concentration, saturating at about 1000 mM, giving an enhancement of nearly an order of magnitude in the fluorescence intensity. Figure 2.2 shows that the binding is selective over the biologically-relevant range of concentrations, with no change in fluorescence intensity found for Ca^{+2} , Mg^{+2} , or Na^+ . Also shown are the effects of K^+ on fluorescence intensity in the presence of the sarcoplasmic reticulum (SR), isolated via the procedure described in Salama et al [88]. The experiment was conducted in a solution composed of the isolated SR with $5\mu M$ ion (KCl , $NaCl$, $CaCl_2$ or $MgCl_2$) in 100 mM Sucrose, 20 mM HEPES and 1mM Gluconic Acid.

2.3 Computational Methods

Unless otherwise indicated, all calculations used Density Functional Theory (DFT) for the ground electronic state and Time Dependent DFT (TDDFT) for excited states, with the CAM-B3LYP functional [89] and a 6-31G** basis. CAM-B3LYP has previously been shown to provide good results for excitation energies of conjugated dyes [90]. For all calculations reported below, the Polarizable Continuum Model (PCM) [91], as implemented in Gaussian 09 [92], was used with water as solvent.

The structure of the SKC-513 dye is shown in Figure 1.1. The ion binds to the crown ether at the top of the dye. To facilitate the calculations, the structure was simplified to contain only the optical chromophore shown in Figure 2.3, which will be referred to as Simplified-SKC-513 (SSKC-513). Effects of ion binding were modeled either by explicit inclusion of a potassium ion or by including the electrical potential arising for a point charge placed at a location corresponding to the center of the crown ether. The charge was positioned at a distance of 3 Å from the nitrogen of the crown ether, along an axis connecting the nitrogen to the oxygen of the heterocycle. The distance of 3 Å was taken from experimental and theoretical studies on potassium crown ether complexes. [93, 94, 95].

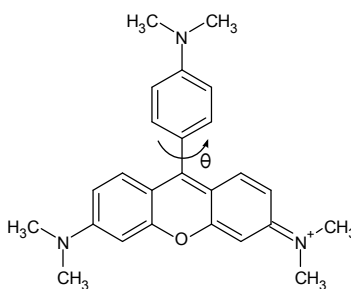


Figure 2.3: Structure of the simplified SKC-513 dye (SSKC-513).

Torsion about the angle θ of Figure 2.3 is relatively facile with a minimum at 52° and the range 30° to 80° being populated at room temperature. Since the energy and oscillator strengths of the lowest-energy excited states are only weakly dependent on θ (see section 2.4.1), the results shown below are for either the geometry-optimized ground or excited state.

2.4 Computational Results

2.4.1 Effects of Torsional Motion

Here, we examine the energy as a function of torsion about the bond connecting the heterocycle to the phenylene of SSKC-513 (θ of Figure 2.3). The molecular structure was optimized using DFT with the B3LYP functional and 6-21G basis set. A constraint was applied to the torsional angle to allow a scan of energy versus θ . The thick black line of Figure 2.4 shows that DFT theory predicts a torsional angle of 52° with a barrier of approximately 2 kcal/mol for the rotation through 90° . The corresponding population at room temperature is shown in Figure 2.5. From these results, we conclude that angles between 30° and 80° are well-sampled by the dye at room temperature.

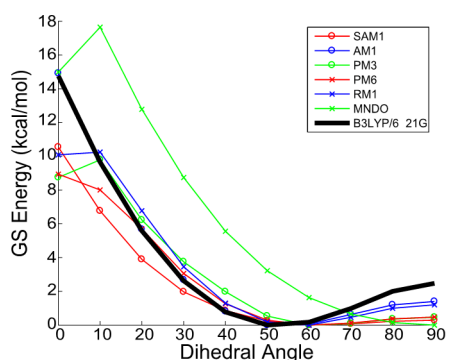


Figure 2.4: Torsional potential for SSKC-513 obtained from DFT (B3LYP/6-21G) calculations and from various semiempirical model Hamiltonians.

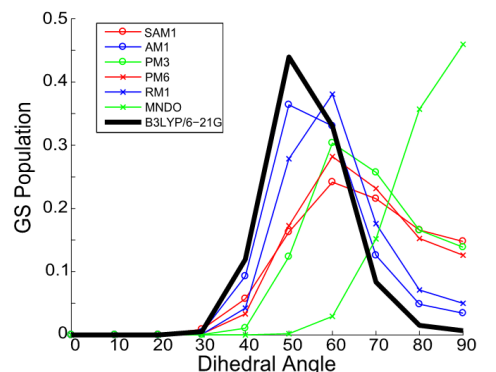


Figure 2.5: Populations at room temperature as a function of torsional angle in the SSKC-513 dye, obtained from DFT (B3LYP/6-21G) calculations and from various semiempirical model Hamiltonians.

The energy and optical intensity of the two lowest excited states depend only weakly on torsional angle. Figure 2.6 shows that these states are nearly degenerate and only weakly dependent on angle over the populated range of 30° to 80° .

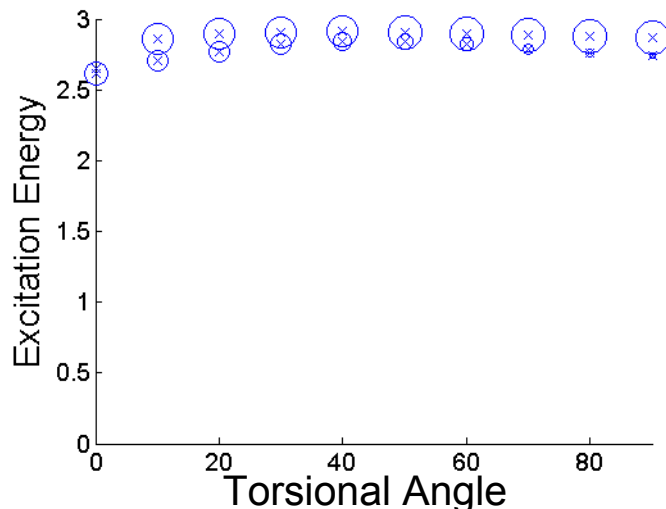


Figure 2.6: Two lowest excited-state energies of SSKC-513 as a function of torsional angle. Structures were optimized in ground state using SAM1 and vertical excitations were obtained with TD-DFT (TD CAM-B3LYP/6-31G**). The radii of the circles are proportional to the oscillator strengths.

2.4.2 Effects of a Point Charge

Figure 2.7 shows the effect of a point charge placed at a position corresponding to the center of the crown ether in SKC-513. In the absence of a point charge, there are two nearly degenerate excited states. These states have different oscillator strengths and relative positions that are strongly dependent on the sign and magnitude of the charge. Following Kasha's rule [96], we expect the molecule to fluoresce strongly only when the lowest excited electronic state carries significant optical intensity. This rule assumes that excitation to the optically intense excited state is followed by relaxation to the lowest excited electronic state on a time scale that is much faster than the fluorescence lifetime. If the lowest excited state has a strong optical transition to the ground state, fluorescence will occur. As the optical intensity of the lowest excited-state decreases, more population is lost to non-radiative pathways and the fluorescence quantum yield decreases.

In the absence of a point charge, the lowest-energy state is the less intense state, which implies the molecule should be only weakly fluorescent. As the magnitude of the point charge is

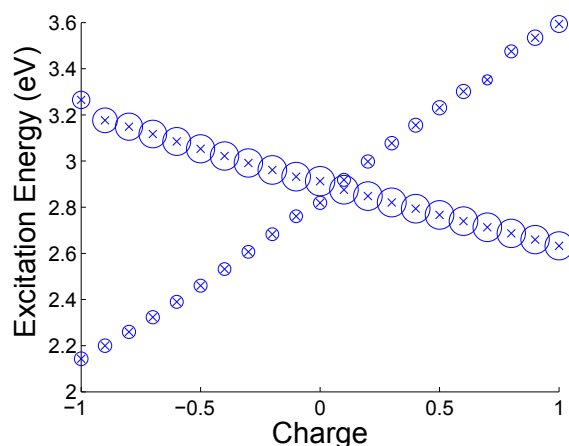


Figure 2.7: Effects of a point charge on the excited states of SSKC-513. (The radii of the circles are proportional to the oscillator strength.)

increased in the positive direction, the bright state drops below the less intense state, suggesting that the presence of a cation such as potassium should substantially enhance the fluorescence quantum yield. We next explore the origin of the excited-state degeneracy, and then consider calculations that use an explicit K^+ ion instead of a point charge.

2.4.3 Origin of the Degenerate Excited States

Examination of the frontier orbitals allows assignment of neutral (N) and charge-transfer (CT) character to the excited states of Figure 2.7. These states are essentially pure excitations between the orbitals shown in Figure 2.8. For vertical excitation in the absence of a point charge, the first excited state has 95% HOMO-CT $\rightarrow$ LUMO character and the second excited state has 97% HOMO-N $\rightarrow$ LUMO character. The lowest unoccupied molecular orbital (LUMO) resides on the heterocycle. HOMO-CT resides primarily on the phenylene ring, such that the HOMO-CT $\rightarrow$ LUMO corresponds to transfer of charge from the phenyl ring to the heterocycle, indicating a CT state. Since there is little overlap between HOMO-CT and the LUMO, this CT state carries little optical intensity. HOMO-N resides on the heterocycle, such that HOMO-N $\rightarrow$ LUMO corresponds to a neutral excitation, with substantially more oscillator strength than the CT state.

The effects of the point charge in Figure 2.7 can also be understood in terms of the frontier

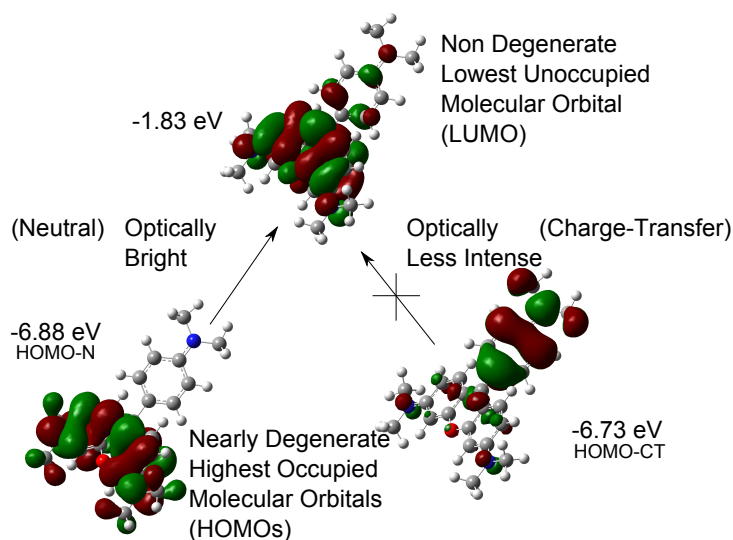


Figure 2.8: Frontier orbitals of SSKC-513. There are two nearly degenerate HOMOs, one located on the heterocycle and the other on the phenylene ring. The LUMO is non-degenerate and located on the heterocycle. Optical intensity relies on good overlap between orbitals, such that only the transition from HOMO-N to the LUMO carries large optical intensity.

orbitals. Since a positive charge bound to the crown ether is closer to the phenylene ring than to the heterocycle, it preferentially lowers the energy of the HOMO-CT. This raises the energy of the HOMO-CT→LUMO transition that dominates the CT state. This is in agreement with Figure 2.7, which shows that a positive charge causes the less intense CT state to rise significantly above the bright N state.

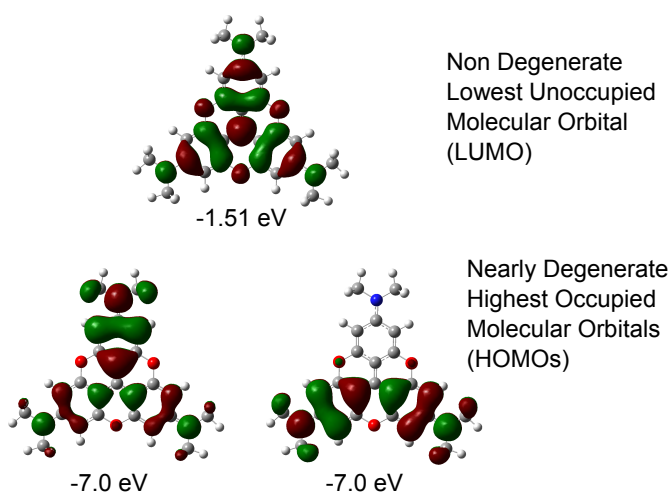


Figure 2.9: Frontier molecular orbitals of the three-fold symmetric parent of SSKC-513.

The near degeneracy of the two lowest-excited states observed in SSKC-513 may be unexpected, since the excitations have a largely different character, one being a N excitation on the heterocycle and the other being a CT excitation from the heterocycle to the phenylene ring. Furthermore, SSKC-513 has at most two-fold symmetry, and degeneracies are not expected for C_2 symmetry groups. However, doubly degenerate states are expected for molecules with three-fold symmetry, since the C_3 symmetry groups have doubly degenerate representations. The molecule of Figure 2.10 uses bridging oxygen atoms to convert SSKC-513 to a three-fold symmetric system. The orbitals of this symmetric parent molecule are shown in Figure 2.9. The HOMO is again doubly degenerate, while the LUMO is non-degenerate. The nodal pattern of the HOMOs also reveals why removal of oxygen and rotation of the upper phenylene ring to form SSKC-513 does not substantially lift the degeneracy: the HOMOs have little amplitude in the regions of the additional oxygen atoms. The nearly degenerate HOMOs of SSKC-513 are therefore not an accidental degeneracy, but rather a result of the three-fold symmetry of this parent molecule.

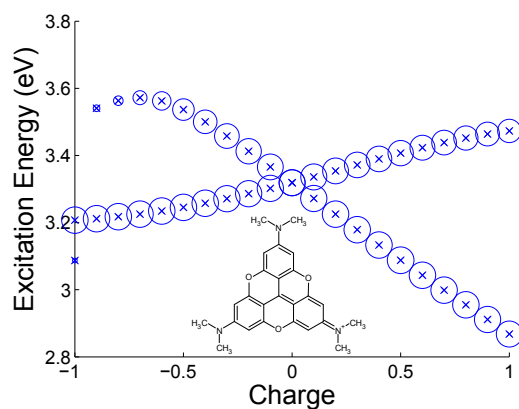


Figure 2.10: Effects of a point charge on the excited states of the three-fold symmetric parent of SSKC-513. (The radii of the circles are proportional to oscillator strength.)

The effects of a point charge on the three-fold symmetric parent molecule are shown in Figure 2.10. The point charge lifts the degeneracy, similar to the behavior of SSKC-513 in Figure 2.7. However, both states retain optical intensity such that binding of the ion would not be expected to have a strong effect on the fluorescent behavior. The three-fold symmetric molecule

is therefore useful for understanding the origin of the degenerate excited states of SSKC-513, but is not expected to be a useful chromophore for ion detection.

2.4.4 Effects of a Potassium Ion

Above, the effects of ion binding were explored by examining the effects of a point charge located at a position corresponding to the center of the crown ether. The results suggest that a positive charge preferentially stabilizes HOMO-CT, which resides primarily on the phenylene ring (Figure 2.8). This causes the less intense CT state to rise above the bright N state (Figure 2.7), which rationalizes the observed increase in fluorescence quantum yield upon binding of a cation.

Results from DFT calculations that explicitly include a K^+ ion (Figure 2.11) lead to similar conclusions. Figure 2.11 shows the lowest two excited states, S_1 and S_2 , with CT and N character assigned based on the predicted optical intensity and examination of the orbitals that contribute to the excitation. Results are shown for both the ground electronic-state geometry, corresponding to vertical excitation, and the relaxed S_1 geometry, corresponding to fluorescence.

In the presence of K^+ (right side of Figure 2.11), the gap between S_1 and S_2 is large, with S_1 having strong N character and so carrying substantial oscillator strength. We therefore expect high fluorescence quantum yield in the presence of a K^+ ion, for reasons that are consistent with those obtained above based on the effects of a point charge.

In the absence of K^+ (left side of Figure 2.11), the gap between S_1 and S_2 is small, being 0.05 eV for the ground-state geometry and 0.15 eV for the relaxed S_1 geometry. In the ground-state geometry, the S_1 and S_2 states have the same character as seen in the presence of a point charge: S_1 has CT character and carries 2.5 times less intensity than the S_2 state, which has N character. Upon excited-state relaxation, the S_1 and S_2 states are predicted to cross: S_1 has N character and carries 2.5 times more intensity than the S_2 state, which now has CT character. The weaker fluorescence seen in the absence of a K^+ ion can then be attributed to the near degeneracy

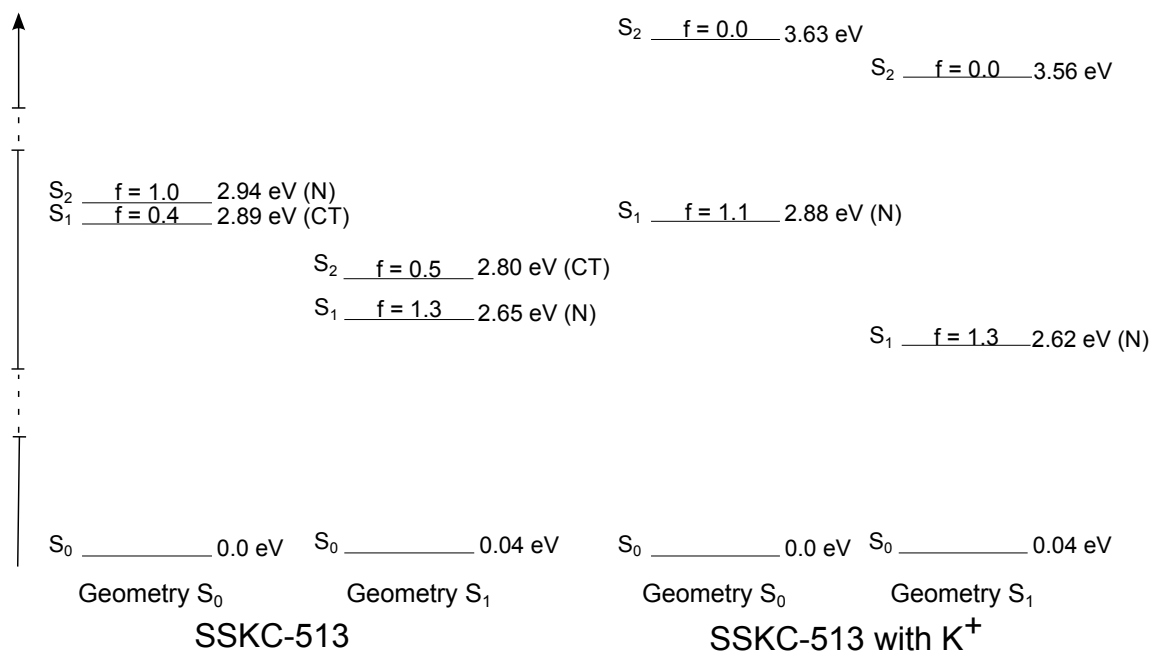


Figure 2.11: Electronic states of SSKC-513 (Figure 2.7) with (right) and without (left) a K^+ ion at a location corresponding to the center of the crown ether in SKC-513.

of S_1 and S_2 , with the energy separation being so close that they cross as the geometry relaxes from that of the vertical excitation to that of the S_1 state.

Since torsions associated with θ of Figure 2.3 are low-frequency, it is plausible that the excited state relaxation is dominated by motion along this coordinate. However, the change in torsional angles are relatively small. In the absence of a K^+ ion, the relaxation of the torsional angle is from 54° to 57.5° and in the presence of a K^+ ion, the relaxation is from 63° to 60.7° .

Also of interest, is the degree to which the interaction with K^+ goes beyond field effects to include more specific chemical interactions. The molecular orbitals that participate in the excitation have negligible amplitude on the potassium ion. This, along with the similar behavior observed for a point charge, suggests that the effects of the ion on the excited state arise primarily from the electric field of the ion.

2.4.5 Excited States in the Presence of K^+

Figure 2.8 shows the orbital contributions to the excited states of the SSKC-513 dye in the absence of K^+ . In that case, the near degeneracy of the HOMO-CT and HOMO-N orbitals leads to a near degeneracy of the two lowest-energy excited states, which correspond to a less-intense HOMO-CT $\rightarrow$ LUMO transition and a bright HOMO-N $\rightarrow$ LUMO transition.

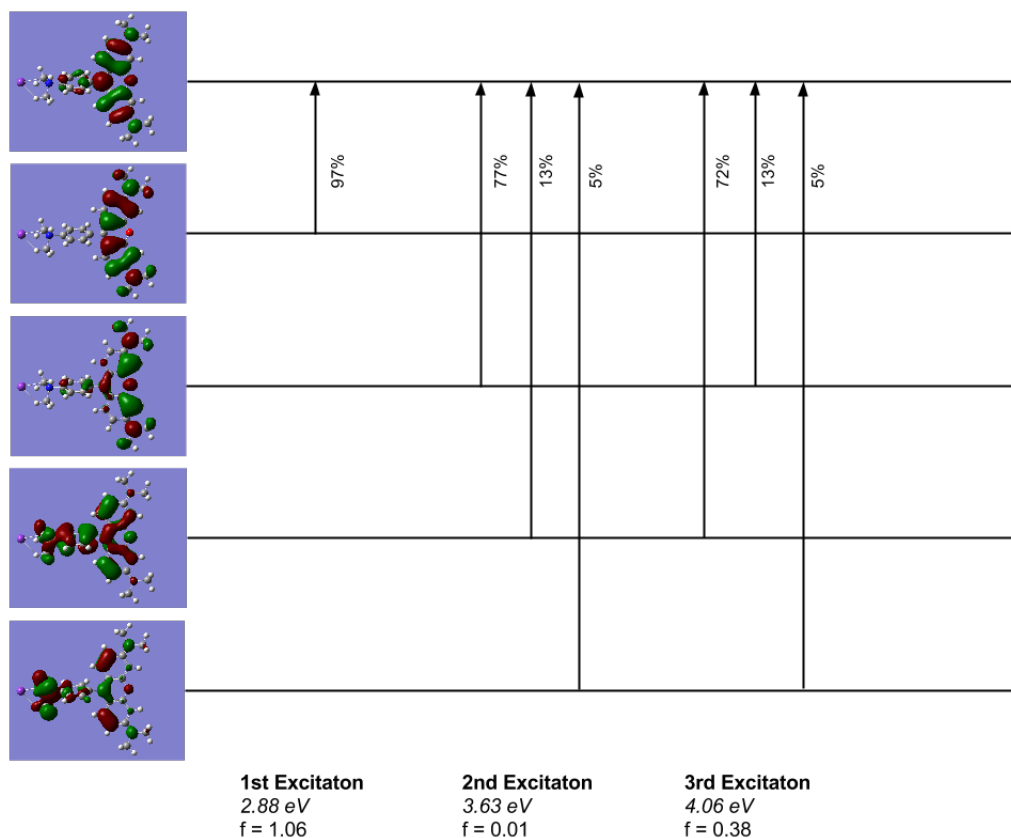


Figure 2.12: Frontier orbitals of SSKC-513 in the presence of K^+ ion along with the orbital contributions to the three lowest-energy excited states. The vertical excitation energies and oscillator strengths, f , are also shown.

Figure 2.12 shows the orbitals that contribute to the lowest-energy excited states in the presence of K^+ . The lowest-energy excited state is bright and predominantly a HOMO $\rightarrow$ LUMO transition, with the HOMO and LUMO residing primarily on the large heterocycle. This state therefore retains the character of the N state observed in the absence of K^+ . The 2nd and 3rd excited state lie more than 0.7 eV higher in energy and involve promotions to the LUMO from

the HOMO-1, HOMO-2, and HOMO-3. This suggests that binding of K^+ substantially raises the energy of the less-intense state and leads to state mixing with other high-lying excited states. The result that is of most relevance to the use of SSKC-513 as a K^+ detector is that the lowest-energy state in the presence of K^+ is both optically bright and well-separated from other excited states. The results therefore predict binding of K^+ to substantially enhance fluorescence.

2.4.6 Semiempirical Calculations

Initial exploration of the photophysical properties of the SSKC-513 was done using semiempirical methods, with the results informing the DFT computations discussed. Partial results from the initial exploratory work are presented here to illustrate that the main conclusions are robust with respect to the level of theory. As discussed earlier, the relevant excited states result from breaking the symmetry of the parent molecule (Figure 2.9). Since the main conclusions are a result of symmetry and symmetry breaking, it is not surprising that the conclusions are insensitive to the level of theory used.

Since accurate predictions of low torsional barriers are challenging for quantum chemical methods, we compared the results of various models to estimate the likely uncertainty (Figure 2.4). Semiempirical methods place the minimum energy structure near 60° and predict a low barrier to rotation through 90° , although they differ considerably regarding the magnitude of this barrier. These disagreements regarding the magnitude of the torsional barrier do not, however, impact relevant predictions.

The remaining semiempirical calculations summarized here use the SAM1 method to obtain optimized ground state structures [36]. Excited states are obtained with the INDO Hamiltonian and direct singles configuration-interaction (S-CI) calculations that include excitations between all filled and empty orbitals[97]. Only vertical excitations are considered here, with effects of geometry relaxation being addressed only in the DFT computations.

Figure 2.13 examines the lowest two excited states as a function of torsional angle. Similar

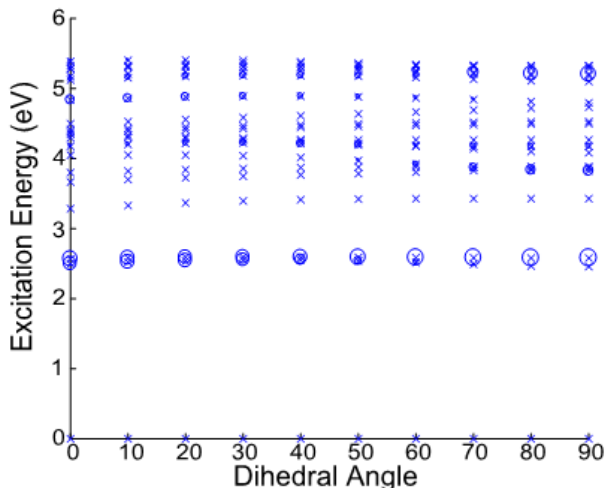


Figure 2.13: Excited states of the SSKC-513 dye as a function of torsional angle, obtained using semiempirical methods.

to DFT, the two degenerate states are only weakly affected by the torsional rotation. The effects of the potassium ion are modeled by including the potential of a point charge placed at a location corresponding to the center of the crown ether. The charge lies 2.8 Å from the nitrogen of the crown ether, somewhat shorter than the 3 Å used in the DFT calculations, but within the experimental range [93, 94, 95]. Screening of the ionic charge due to solvent was included by dividing the charge by the dielectric constant, q/ϵ . For pure water, $\epsilon \approx 80$, while for organic media $\epsilon \approx 2$. Since the ionic charge is partially screened by water, values of q/ϵ between -1/3 and 1/3 were considered. Figure 2.14 shows the dependence of the two lowest excited states of SSKC-513 on the magnitude of the point charge. For a positive charge, the lowest state is bright, and the molecule is predicted to be highly emissive. In the absence of a charge or in the presence of a negative charge, the lowest state is less intense and so the molecule is predicted to be less fluorescent.

The effects of a point charge on the parent three-fold symmetric molecule are shown in Figure 2.15. The effects of the point charge on the energies of the two lowest states is similar to that of SSKC-513 (Figure 2.14) however, in the parent, both states carry significant optical intensity. This is the same behavior as seen in the DFT calculations.

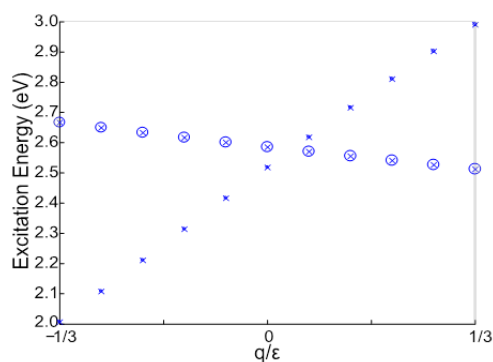


Figure 2.14: Effects of a point charge of magnitude q/ϵ on the excited states of SSKC-513 obtained using semiempirical methods.

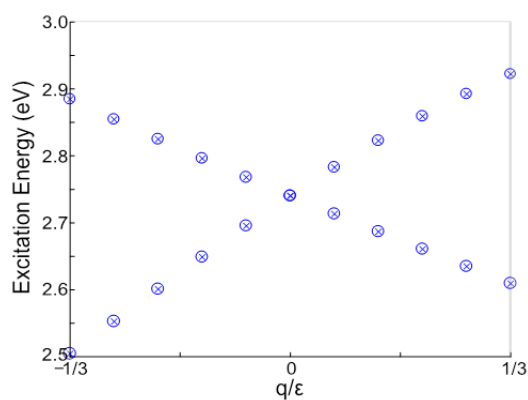


Figure 2.15: Effects of a point charge on the excited states of the three-fold parent symmetric molecule, obtained using semiempirical methods

The agreement between DFT and semiempirical methods regarding the nature of the lowest excited states and the effects of a point charge on these states, provides strong support for the mechanism proposed here regarding sensitivity of fluorescence to the presence of the ion.

2.4.7 Effects of Electronic Substituents

The suggested mechanism for the effects of K^+ on the fluorescence quantum yield is that the electric field of the ion causes the molecule to move from the less-emissive regime on the left of Figure 2.7 to the more emissive regime on the right. The computations reported here are approximate, not only in the methods used for the electronic structure of the chromophore, but also regarding the use of a continuum dielectric model for the solvent and the use of a simpli-

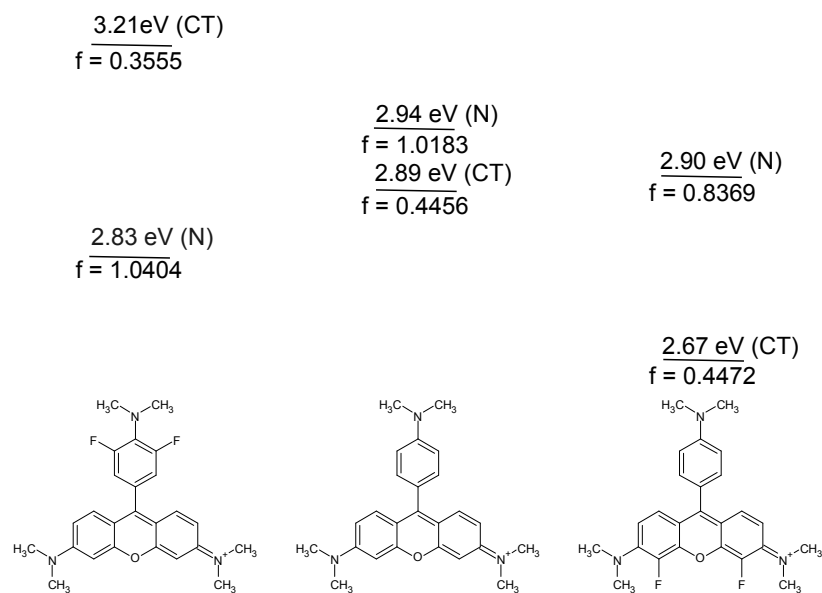


Figure 2.16: Effects of Fluorine substituents on the two lowest excited states of SSKC-513. Adding F substituents to the phenyl-ring (left) raises the energy of the CT excited state, while adding F substituents to the heterocycle has the opposite effect.

fied structure for the dye (Figure 2.3). While such approximations likely do not invalidate the proposed mechanism, they do suggest that the predicted location of the crossing point between the less emissive and more emissive regimes may not be highly reliable. A reasonable target for synthetic modification to SKC-513 is adding substituents that alter the crossing point. This can be done by adding electronic acceptors or donors that alter the relative energies of the HOMOs in Figure 2.8.

Two derivatives of SSKC-513 are compared with SSKC-513 in Figure 2.16. The results can be understood in terms of the effects of the substituents on the HOMOs of Figure 2.8. Addition of fluorine atoms to the phenylene group stabilizes the HOMO-CT located on the phenylene ring (the HOMO on the right in Figure 2.8), thereby raising the energy of the CT state relative to that of the N state. Similarly, addition of fluorine atoms to the heterocycle stabilizes the HOMO-N shown on the left of Figure 2.8, thereby raising the N state relative to the CT state.

These results suggest that electronic substituents provide a handle that may be used to optimize the sensitivity of SSKC-513 fluorescence to ion binding, by altering the relative energies of

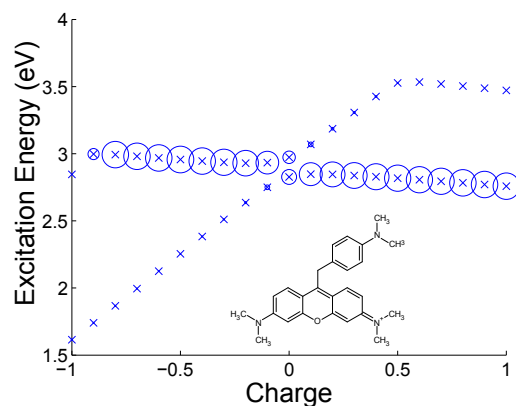


Figure 2.17: The two lowest excited states of a modified SSKC-513, in which a methylene bridges between the phenylene group and the heterocycle. (The radii of the circles are proportional to the oscillator strength.)

the N and CT states.

2.4.8 Effects of a Bridging Methylene

Figure 2.17 shows the effects of a point charge on the excited states when an extra methylene group is placed between the heterocycle and the phenylene group. The calculations are done for the optimized geometry, in which the phenylene ring is nearly perpendicular to the heterocycle. The avoided crossing between N and CT states suggests a stronger electronic coupling between these states than is seen in Figure 2.7 for SSKC-513. A more detailed examination of the curve crossings suggests a coupling of 0.15 eV for this system, compared to about 0.01 eV for SSKC-513. The proposed mechanism for ion sensitivity assumes rapid relaxation between N and CT states. A stronger coupling between N and CT states may enhance this relaxation, such that the addition of a bridge methylene group may enhance the sensitivity of fluorescence to ion binding [98]. The introduction of a bridging methylene also leads to a substantial decrease in the oscillator strength of the CT state. This increased contrast between the N and CT states may also enhance the sensitivity of fluorescence to ion binding.

2.5 Conclusions

The quantum chemical calculations presented here suggest a mechanism for the sensitivity of SKC-513 to ion binding, and a possible means for enhancing this sensitivity. The mechanism relates to the near degeneracy of a bright excitation, corresponding to a N excitation on the heterocycle of SKC-513, and a CT excitation corresponding to charge transfer from the phenylene group to the heterocycle. In the absence of K^+ , these two states lie close in energy. Binding of K^+ destabilizes the CT excited state, raising its energy far above the N state. In the presence of K^+ , the lowest-energy excited state has high oscillator strength and is well separated from other electronic states. This rationalizes the increase in fluorescence intensity seen experimentally upon binding of K^+ . Computations suggest that electronic substituents may be used to alter the relative location of the N and CT states, while the introduction of a methylene group as a bridge between the heterocycle and the phenylene group alters the electronic coupling between these states. Such modifications may therefore provide synthetic handles with which to optimize the sensitivity of the fluorescence to ion binding.

This chapter opens up a number of avenues for future work. One such avenue is better understanding why the electronic coupling between the nearly degenerate states is increased when a non-conjugated group is placed between the heterocycle and the phenylene (Figure 2.17). More detailed computational studies may shed light on the origin of this increase, and may suggest other routes to increasing the coupling. Another possible avenue of research is exploring the degree to which dye molecules may be designed by breaking the symmetry of a 3-fold symmetric parent. The lifting of the three-fold symmetry nearly retains the two-fold degeneracy of the HOMO because of the nodes present at the linking oxygen atoms (Figure 2.9). A more detailed symmetry analysis may reveal the conditions under which these nodes will be present. Given those conditions, it may be possible to identify other classes of dyes that can serve as PET sensors.

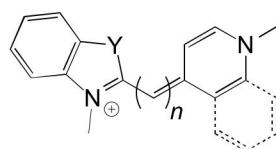
Chapter 3

Computational characterization of fluorogenic dyes

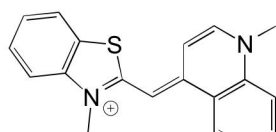
3.1 Introduction

This chapter reports computational studies of the spectral properties of a series of Thiazole Orange derivatives. These computations were motivated by experimental work conducted by our collaborators Elizabeth E. Rastede, Simon C. Watkins, Alan S. Waggoner and Bruce A. Armitage. The experimental work is included in this chapter to place the computations in context. This chapter is based on published work. [99]

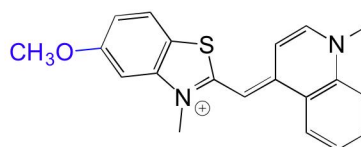
Unsymmetrical cyanine dyes exhibit low fluorescence quantum yields in fluid solution but strongly enhanced emission in viscous solution or other environments that conformationally constrain the dyes [77, 78]. This phenomenon arises from a nonradiative twisting pathway about the central methine bridge that is inhibited when the dye is constrained [77, 78, 79]. To enable simultaneous monitoring of multiple biomolecules, fluorgens that span a range of wavelengths are highly desirable. Electron donating and withdrawing substituents provide a means to shift the spectral peaks of dyes and so design dyes that emit at different wavelengths. This chapter uses quantum chemical calculations to understand the origins of the shifts seen experimentally.



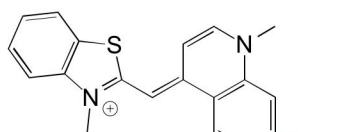
**Unsymmetrical
Cyanine Dye**



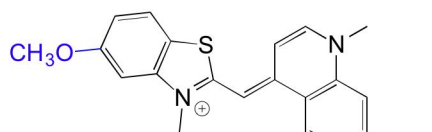
TO



CH₃O-TO



TO-CF₃



CH₃O-TO-CF₃

A number of different computational approaches are compared, including TDDFT with different functionals and semiempirical INDO calculations. The results indicate that INDO is nearly as reliable as TDDFT, with both approaches giving good agreement with experiment in most cases.

3.2 Computational Methods

The molecular geometries of the ground electronic state were optimized using DFT with the B3LYP functional [100] and 6-31G** basis set. Vertical excitations were obtained from single-point TDDFT calculations with the CAM-B3LYP/6-31G** [89], since this functional has been shown to perform well on cyanine dyes [90, 101, 102]. In addition, the correlation between the predicted and observed λ_{max} is compared for CAM-B3LYP, M06HF and PBE0 functionals as well as ZINDO [97]. All calculations were performed using GAUSSIAN09 [103] and included the effects of the methanol solvent through the Polarization Continuum Model (PCM) [91]. INDO calculations were performed using a direct singles configuration interaction method that

includes excitations between all molecular orbitals [97]. Solvent was not included in the INDO computations

3.3 Comparison of Computational Methods

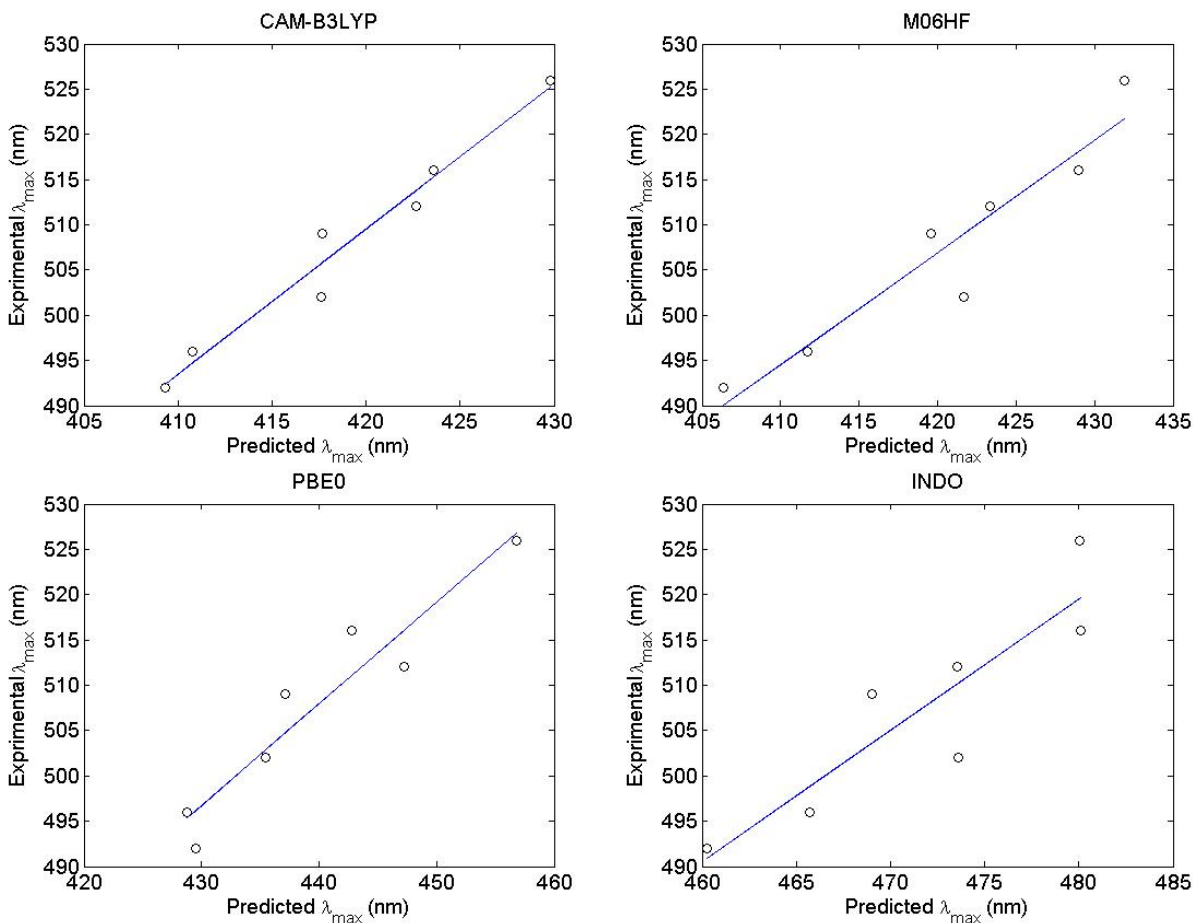


Figure 3.1: Correlation plots of predicted values from different calculations of the TO derivative dyes and experimental results

The predicted absorption wavelengths (Table 3.1) are plotted against the experimental observations in Figure 3.1. The correlations are significantly better for CAM-B3LYP than for the other methods. We see that there is in general a good correlation between predicted and experimentally observed absorption λ_{max} values.

Method (nm)	TO	TO-1F	TO-p2F	TO-4F	TO-CF ₃	TO-OMe	MeO-TOCF ₃
CAM-B3LYP	417.6	417.66	410.8	409.3	423.6	422.6	429.8
PBE0	435.5	437.2	428.8	429.5	442.8	447.2	456.8
M06HF	421.7	419.6	411.7	406.4	429	423.4	431.9
INDO	473.6	469	465.7	460.3	480.1	473.6	480
Experiment	502	516	512	509	492	496	526

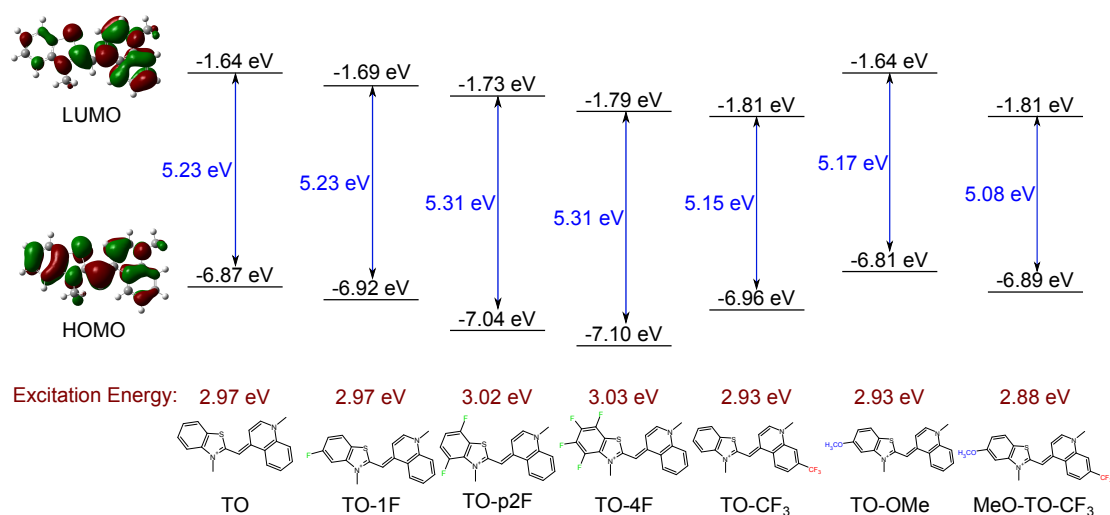


Figure 3.2: Frontier orbital analysis for TO and its donor/acceptor substituted analogues, including HOMO and LUMO orbitals of TO (far left), HOMO and LUMO energies (black), HOMO-LUMO gaps (blue arrows) and TDDFT excitation energies (red text).

3.4 Results and Discussion

Figure 3.2 shows the calculated frontier orbitals for TO, along with the effects of the substituents on the HOMO and LUMO energies and on the excitations energies predicted by TDDFT. The trends in the HOMO-LUMO gaps correlate well with the excitation energies, such that the spectral shifts can be understood in terms of the effects of the substituents on the frontier orbitals. These effects are consistent with EWGs (F and CF₃) preferentially stabilizing, and EDGs (methoxy) preferentially destabilizing, the frontier orbital whose density is largest on the heterocycle to which the group is attached. Fluorine substitution on the benzothiazole thus preferentially stabilizes the HOMO and leads to blue-shifted absorption, while CF₃ substitution on the quinoline preferentially stabilizes the LUMO and leads to red-shifted absorption. Methoxy

substitution on the benzothiazole preferentially destabilizes the HOMO, leading to red-shifted absorption, as observed previously by Ivanov and co-workers. [104] The substituent effects reinforce one another in CH₃O-TO-CF₃, leading to the largest red shift (Table 3.1).

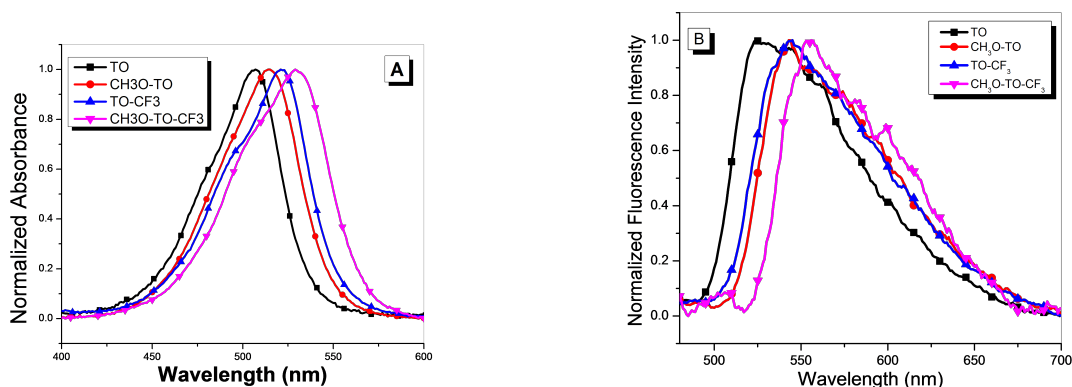


Figure 3.3: (A) Normalized absorbance spectra of TO dyes (1 μ M) in methanol. (B) Normalized fluorescence spectra of TO dyes (1 μ M) in 90% glycerol in water.

Normalized UV-vis spectra for the new TO derivatives in methanol solution are shown in Figure 3.3A, along with the spectra for TO and the previously reported TO-CF₃. Trifluoromethylation of the quinoline group causes a 14 nm red shift in the absorption spectrum relative to TO. Introduction of a methoxy group at position 5 on the benzothiazole heterocycle causes a similar red shift of 10 nm, and the combination of both substituents, methoxy on the benzothiazole and trifluoromethyl on the quinoline, shows an additive effect with a red shift of 24 nm (Table 3.1).

Similar results were observed in fluorescence emission spectra recorded for the dyes (Figure 3.3B and Table 3.1). Since TO and its analogues exhibit very low fluorescence in fluid solution, we measured emission spectra in a viscous solvent consisting of 90% glycerol in water. The spectra are relatively broad, but the emission clearly red-shifts in the order TO < TO-CF₃ $\approx$ CH₃O-TO < CH₃O-TO-CF₃.

The fluorescence quantum yields were also determined for TO analogs in 90% glycerol in water (Table 3.1; the values for TO and TO-CF₃ are lower than we previously reported; this could be due to slight variation in the glycerol content.) As previously reported the addition of

Dye	ϵ^a ($\text{M}^{-1}\text{cm}^{-1}$)	Abs λ_{max}^a	Em λ_{max}^b	ϕ_f^b
TO	68,000	502	529	0.013
CH ₃ O-TO	68,600	512	545	0.009
TO-CF ₃	65,000	516	547	0.009
CH ₃ O-TO-CF ₃	58,000	526	560	0.007

Table 3.1: Extinction coefficients, wavelength maxima (nm) and fluorescence quantum yields for TO dyes in homogeneous solution.

^a in methanol

^b in 90% glycerol in water

the trifluoromethyl group to the quinoline side decreases the quantum yield due to reduction in the activation barrier for twisting about the central methine in the excited state. Addition of a methoxy substituent to the benzothiazole group also decreases the quantum yield, and when both substituents are present the quantum yield decreases further.

3.5 Conclusion

This chapter reports calculations on a family of fluorogenic cyanine dyes in which strategic placement of electron donating and withdrawing groups was used in order to finely tune the absorbance and fluorescence spectra. The computed shifts in wavelength induced by the substituents are in good agreement with experiment. In addition, a frontier orbital analysis successfully accounts for the observed spectral shifts. The relative shifts of the HOMO and LUMO orbitals is as expected based on the amplitude of these orbitals at the atom where the substitution occurs, and the electron donating versus withdrawing character of the substituent.

Chapter 4

Embedding parameters in *ab initio* theory to develop well-controlled approximations based on molecular similarity

4.1 Introduction

Section 1.3 motivated and introduced a new approach to developing semiempirical quantum chemical models. The approach embeds parameters into a low level (LL) *ab initio* theory and adjusts these to obtain agreement with results of a high level (HL) *ab initio* theory. We will refer to such models as parameterized-LL (pLL) models, to reflect that the LL model is being used as a functional form to be trained through machine learning techniques. The embedded parameters are associated with molecular fragments, and agreement between pLL and HL models is sought on data that spans a range of chemical systems over which molecular similarity can be plausibly assumed to hold. To achieve a predictable level of accuracy, we divide the data used to parameterize the model into a training and a test set. The model parameters are adjusted to obtain agreement on the training set, while monitoring the performance on the test set. The accuracy achieved on the test set thereby provides an estimate of the accuracy of the pLL model on systems

that are sufficiently similar to those used to train the model. The model can be systematically improved by increasing either the sophistication of the LL *ab initio* theory, by increasing the size of the basis or level of correlation, or by improving the means through which the parameters are embedded to form the pLL model, as is done below through the use of context sensitive parameters.

The goal is explore the degree to which a pLL model can take advantage of molecular similarity to incorporate aspects of the electronic structure. Here, we use minimal-basis Hartree-Fock theory (HF/STO-3G) as the basis for the pLL model and a split-valence Hartree-Fock theory (HF/6-31G) as the HL theory. The HL theory differs from the LL by allowing the charge distribution to expand and contract in response to changes in the molecular geometry and the electrostatic environment. Our hypothesis is that the effects of this expansion and contraction behave similarly within a restricted class of molecules. If this holds, then we should be able to train a pLL model on a subset of molecules within that class and create a model that applies across the entire class. In addition, we embed the parameters in a manner that builds on intuitions regarding the effect the pLL model is attempting to capture. For instance, we invoke the intuition that the expansion and contraction is influenced by the charge on the atom. The training of the pLL model on HL data, as opposed to experimental data, has the advantage of allowing the computational experiments to explore the degree to which the pLL model can capture the differences between the LL and HL model. These differences are much better understood than the differences between the LL model and experiment.

Our approach is similar to traditional semiempirical quantum chemistry (SEQC) [31, 105] in that parameters are embedded into a quantum mechanical Hamiltonian and adjusted to obtain agreement with available data. There are however a number of important distinctions. First is the use of a training and test set to obtain an estimate of the error introduced by the approximation. A second distinction relates to the nature of the data used for the parameterization. Here, the model is parameterized to a large set of data generated from HL *ab initio* theory, including the expectation values of all operators that appear in the electronic Hamiltonian. Traditional

SEQC attempts to obtain agreement with either a small set of experimental data, or with sparse data (geometries, heats of formation) obtained from HL *ab initio* theory. A third distinction is the meaning attributed to the embedded parameters. The parameters of SEQC are typically assumed to have some meaning, such as atomic ionization potentials or screened Coulombic interactions. Here, the parameters are viewed in a manner similar to the parameters embedded into a neural net, with little meaning being attached to the parameters themselves. The pLL model serves as a convenient functional form for an approximating function that is assumed to hold over some limited range. Finally, the model is assumed to be valid only for molecules that are sufficiently similar to those included in the training of the model. The intent is that models will be developed for different classes of molecules, making the approach more analogous to molecular mechanics, where parameters are often specific to particular fragments such as amino acids or polymer subunits.

The SCC-DFTB method [43, 46] has some similarities with the current approach, including the use of detailed data from a higher-level theory to extract parameters for a LL model and the use of parameters tuned to specific classes of systems. However, our approach differs in the use of a training and test set to estimate error and in the general approach of using a pLL model as an approximating function that is trained to data with little regard for the meaning of the embedded parameters.

The choice of LL and HL theory for the current work is partly motivated by our previous work on empirical models. Features extracted from the results of a LL calculation were used to predict the results of a HL calculation. For instance, such a model was successful at predicting the two-electron density matrix, and thus the correlation energy, from the one-electron density matrix obtained from Hartree-Fock theory [106]. Another study considered the collinear reaction $\text{H}_2 + \text{F} \rightarrow \text{HF} + \text{H}$ in environments that strongly perturb the reaction energy profile [58]. A linear regression obtained chemical accuracy (<0.6 kcal/mol) in predicting the results of a HL calculation (QCISD/6-31G++**) using only inputs (energy and distributed multipoles) generated from a LL calculation (HF/3-21G). The error in such models was dominated by the extrapolation from

small to large basis sets (3-21G $\rightarrow$ 6-31G++**), with much less error resulting from extrapolating to correlated theories (HF $\rightarrow$ QCISD). These results suggest that a key challenge in the development of empirical models is the extrapolation across basis set. This past work was empirical, in the sense that simple linear models were used to predict HL results from LL inputs. The current work is semi-empirical, in the sense that a modified Hamiltonian is used as the functional form in which to embed parameters. This model form may lead to substantially improved performance, especially given the success of traditional semiempirical quantum chemistry (SEQC), which used only a handful of embedded parameters fit to a handful of experimental data. Our choice of LL and HL model is meant to explore the degree to which a semiempirical model can address the aspect that was most difficult in our past empirical models, the extrapolation across basis set.

The chemical systems considered here are saturated hydrocarbons. Section 4.2.1 describes the data used to train and test the model. The form of the model is then described, including the embedding of parameters into the LL Hamiltonian (Section 4.2.2), and addition of context sensitivity to these parameters (Section 4.2.3). The training and performance of the model are then described in Section 4.3.

4.2 Methods

4.2.1 Chemical Data

The data consists of the electronic structure of a hydrocarbon with varying geometries in a range of environments. Both LL (HF/STO-3G) and HL (HF/6-31G) data is generated. Random geometries are obtained by perturbing the equilibrium geometry with random distortions. These distortions are generated using a z-matrix that defines the molecular geometry in terms of bond lengths, angles, and dihedrals. A random number is then added to each value in the z-matrix, using a uniform random distribution with a width of ± 0.15 Å for bond lengths, $\pm 6^\circ$ for bond

angles, and $\pm 7^\circ$ for dihedral angles. Bond angles that are not explicitly included in the z-matrix thus have a range that is up to twice that of the angles included in the z-matrix. For ethane, the random variable is chosen to span the full range for the internal rotation angle.

The environment perturbs the electronic structure of the fragment in a manner that explores the types of perturbations that will be present in large molecules. This includes perturbations from external electrostatic potentials, due to other portions of the molecule or from solvent, and inductive effects from acceptors and donors. The environments consist of a cube, each corner of which holds a point charge. The length of each side of the cube is 12 Å for methane and ethane and 14 Å for propane and butane, with the molecule placed at its center. The magnitudes of the point charges are randomly generated using a uniform distribution between -25 and 25 amu, chosen to induce variations in the Mulliken charges on the C and H in methane and ethane that are similar to the charges induced on the methyl group in CF_3CH_3 (~ 0.2 amu).

For each pairing of a molecular configuration with an environment, we generate expectation values of each operator that appears in the Hamiltonian (total kinetic energy, interaction of the electron density with each nucleus, and total two-electron repulsion energy). Each data set consists of 10 configurations in 10 environments, corresponding to 100 calculations. Ethane has eight nuclei which, along with kinetic energy and two-electron energy, leads to 1000 data points. The goal of the model fitting is to get the 1000 values generated from the pLL model to agree with those from the HL theory.

4.2.2 Effective Hamiltonian

Rather than use the ad hoc functional forms of traditional SEQC, we embed parameters into a LL (STO-3G) *ab initio* model. A systematic comparison of different embedding schemes will be presented in next chapter. Here, we choose a scheme that performs well and focus on issues related to the training and testing of the model.

For one-electron operators, we embed parameters by multiplying the matrix elements, $\langle i|h_1|j\rangle$,

by a multiplicative constant, $(1+x)$, with x constrained to be greater than -1. For diagonal blocks, where i and j are atomic orbitals on the same atom, a different parameter is used for each atom type (C, H) and for each shell (1s, 2s, and 2p). For off-diagonal blocks, where i and j are on different atoms, modifications are included only between bonded atoms. For each bond, the atomic orbitals of C are transformed to create an sp^3 hybrid orbital directed along the bond, and the matrix element between the orbitals participating in this bond (1s for H, sp^3 for C) are multiplied by $(1+x)$. (Note that only singly-bonded molecules are included here.) The parameter, x , used between bonded atoms is a function of the atom types participating in the bond. Integrals that are not modified according to the above rules retain their LL values, as opposed to being set to zero.

For the two electron operators, we use a multiplicative constant, $(1+x)$, to modify the following classes of two electron integrals:

Diagonal: $(ij|kl)$ with i j k and l all on the same atom
 Off-diagonal: $(ij|kl)$ with i j on one atom and k l on another atom

The diagonal element in which all orbitals are 1s has a single parameter for each atom type. For the (2s, 2p) shells of carbon, we use the form introduced by Slater and utilized in INDO theory [80] to express the on-atom integrals in terms of three parameters, F0, G1 and F2. For the off-diagonal integrals, the values from the LL STO-3G theory are multiplied by a constant that depends on the two atom types. Between C and H, and between C and C, the integrals are modified only if there is a bond between the atoms. For integrals between H and H, the integrals are always modified. This approach to the two-electron operator has some similarities with the Zero Differential Overlap (ZDO) approximation, since the two-electron integrals being modified are those that are included in ZDO theories. However, unlike ZDO, the remaining two-electron integrals are retained, and set to the value they have in the unmodified LL theory.

To further illustrate the embedding scheme, Table 4.1 lists the parameters used for ethane.

Matrix element type	Atom types	Number of Parameters
Kinetic energy (KE)	Diagonal	C 3 (1s, 2s, 2p)
		H 1 (1s)
	Off-diagonal	C-H 1 (sp^3 1s)
		C-C 1 (sp^3 sp^3)
Elec-nuclear interaction	Diagonal	C 3 (1s, 2s, 2p)
		H 1 (1s)
	Off-diagonal	C-H 1 (sp^3 1s)
		C-C 1 (sp^3 sp^3)
Two-electron	On-atom	C 4 (1s, F0, G0, G2)
		H 1 (1s)
	Between atoms	C-H, C-C, H-H 3
TOTAL		20

Table 4.1: List of parameter types for ethane

4.2.3 Context Sensitive Parameters

We expect that a given set of parameters will be valid only over some limited range of molecules. We can extend this range by making the parameters functions of the current context of the molecule, where the context is extracted from the geometry and electronic density matrix. The current work considers the ad hoc context variables shown in Table 4.2. (The use of feature extraction methods to develop machine-learning derived contexts will be presented in a future work.) Below, these contexts are added sequentially. The first context is bond length, so we can determine the improvement obtained from including only geometry dependence in the model. The second context is the aspect of the electron density that seems most likely to be of relevance, with regards to the expansion and contraction of charge density present only in the HL model. For diagonal blocks, as charge is pushed onto (or pulled from) an atom, we expect the charge density to expand (or contract). For the off-diagonal blocks, the bond order may play a similar role of influencing the expansion and contraction of the electron density. The third context crosses these, using bond order for diagonal blocks and atomic charges for off-diagonal blocks.

Since the density matrix is updated on each iteration of the Hartree-Fock (HF) solution process, the context can also be updated on each iteration and so integrate smoothly into the HF algorithm. However, for the fits shown here, the context variables are derived from HF/STO-3G

and are not updated as the model is trained. In addition, the charges are those induced by the environment, as opposed to absolute charges.

Contexts for diagonal blocks

- 1) r: Average bond length to bonded atoms
- 2) q: Mulliken charge on the atom
- 3) bo: Average bond order to bonded atoms

Contexts for off-diagonal blocks

- 1) r: Bond length
- 2) bo: Bond order
- 3) q: Difference in charges on bonded atoms

Table 4.2: hoc context variables

Each of the parameters embedded in the Hamiltonian (Table 4.1) is made a linear function of the three context variables listed in Table 4.2. An exception is the parameter that modifies the two-electron integrals between hydrogen atoms: it becomes a linear function of only the bond-order. The restriction to bond order is motivated by the fact that these interactions are

present between non-bonded atoms. The inclusion of bond order allows the model to make a small distinction between adjacent versus distant hydrogens. The resulting model has 78 embedded parameters.

4.3 Results

As discussed in Section 4.2.1, each data set consists of 10 molecular configurations coupled with 10 electrostatic environments, for a total of 100 molecule/environment pairs. The model is trained on such a data set for ethane. During training, performance is monitored on a test set, also for ethane, which has the same size as the training set. The model parameters are optimized using the trust-region reflexive algorithm [107, 108], as implemented in MATLAB 2012a [109], with the objective being the RMS error of the training set. The error is computed for each operator,

and each molecular environment pair,

$$Err_{mol,env}^{\hat{O}_i} = \langle \hat{O}_i \rangle_{mol,env}^{HL} - \langle \hat{O}_i \rangle_{mol,env}^{pLL}, \quad (4.1)$$

where i ranges over kinetic energy (KE), the electron-nuclear interaction for each atom in the molecule, and the two-electron energy (E_2). We also consider the operator that sums all of these terms to give E_{tot} . Note that the energy of interaction with the environment is not included in E_{tot} . The environment is used to perturb the electronic distribution, and the model is adjusted to the self-energy of the molecule in the presence of such perturbations. The RMS error sums this over all environment/molecule pairs and over each operator,

$$RMSError = \sqrt{\frac{1}{N} \sum_{i,mol,env} (Err_{mol,env}^{\hat{O}_i})^2}, \quad (4.2)$$

where N is the total number of terms in the summation. In the fits shown below, the summation of Eq. 4.2 includes E_{tot} , with an optional weighting factor, w .

The errors of Eqs. 4.1 and 4.2 reflect disagreement between the pLL and HL models with regards to total energy. However, the absolute energy from a quantum chemical model has little meaning since it is only energy differences that can be measured experimentally. To capture the error associated with energy differences, we subtract the mean of the error,

$$Err'_{mol,env}^{\hat{O}_i} = Err_{mol,env}^{\hat{O}_i} - \overline{Err_{mol,env}^{\hat{O}_i}}, \quad (4.3)$$

where the mean is taken over all molecule/environment pairs, for each operator type. The RMS error is then,

$$RMSError' = \sqrt{\frac{1}{N} \sum_{i,mol,env} (Err'_{mol,env}^{\hat{O}_i})^2}, \quad (4.4)$$

The error of Eq. 4.4 thus reflects the disagreement between the pLL model and HL model regarding changes in the operator expectation values arising from changes in either the geometry

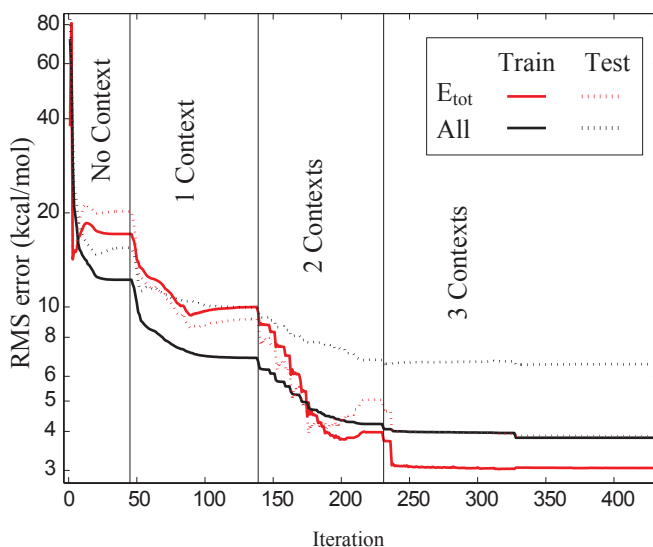


Figure 4.1: Training of the pLL model on an ethane data set (solid lines), while monitoring performance on an ethane test set (dotted lines). The objective is the RMS error, Eq. 4.2, summed over all operators including E_{tot} (black lines). The RMS error, Eq. 4.2, for just the total energy, E_{tot} , (red lines) is also shown. The vertical lines show addition of context variables to the model (Table 4.2).

or environment of the molecule. Since there are sufficient parameters in the pLL model for the fitting algorithm to adjust the mean of the predictions to that of the HL model, Eqs. 4.2 and 4.4 are expected to be the same for the training set. However, Eq. 4.4 provides a more relevant measure of the performance on test sets. Figure 4.1 shows the error as a function of iteration number for a weighting factor of $w=1$ on E_{tot} . The model is first optimized without the inclusion of context dependence in the parameters. Once convergence is achieved, the first level of context dependence is added to the model (contexts labeled 1 in Table 4.2), and so on. The fit shown here used tight convergence criteria (10-10 relative change in either the parameters or the RMS error, or 200 maximum iterations at each stage). The objective tends to decrease smoothly, such that in later fits, we stop the optimization when there are 4 successive steps in which the performance on the test set drops by less than 0.01 kcal/mol. In most cases, this stopping criteria is met when performance on the test set degrades for four successive steps. This criteria would stop the optimization before the drop in RMS error that occurs near iteration 325 in Figure 4.1. This may indicate the need for a global optimization procedure [110]. However, this drop does

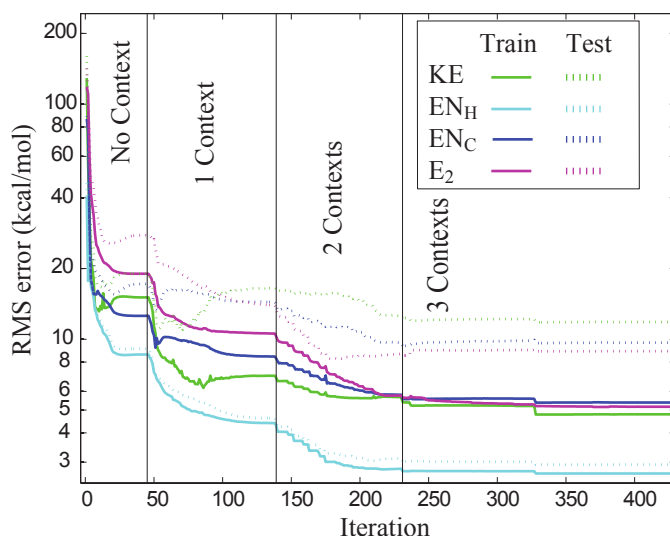


Figure 4.2: Error in the individual operators, excluding E_{tot} , for the optimization of Figure 4.1. The RMS error is obtained using Eq. 4.2, with summations limited to a particular operator type.

not lead to improved performance on the total energy and so does not suggest a need to alter our stopping criteria. The performance on the test set roughly tracks that of the training set, with a final performance for the total energy only 30% higher than the training set, 4 kcal/mol average for the training set as opposed to 3 kcal/mol for the test set.

Figure 4.2 shows the average error for each operator type. The error for these operators is substantially higher than that of the total energy in Figure 4.1, indicating that there is considerable cancellation of error when the operators are summed to give the total energy. The cancellation of errors is likely related to the variational nature of the Hartree-Fock method, which minimizes the total energy under the constraint of a single Slater determinantal wavefunction. Since only the total energy is minimized, the better performance observed for E_{tot} versus other operators may not be surprising.

The performance of the model of Figure 4.1 on molecule types not included in the training is shown in Figure 4.3 through Figure 4.5. Figure 4.3 shows the error in the absolute energy, Eq. 4.2, in which case the performance is quite poor for molecules not included in the fit. However, the absolute energy of a quantum chemical model has little meaning since it is only energy differences that can be measured experimentally. Figure 4.4, shows the error of Eq. 4.4, which

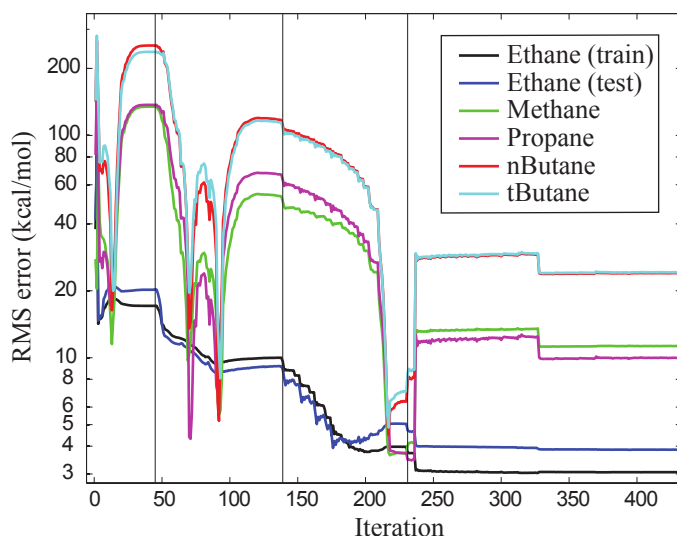


Figure 4.3: RMS error, Eq. 4.2, in the total energy, E_{tot} , for a variety of molecules, using the pLL model of Figure 4.1.

reflects the performance expected for such energy differences. The performance on molecules not included in the fit is comparable to that obtained on the test set of ethane molecules. This verifies that a model trained on one type of molecule can be transferred to other similar molecules. Figure 4.5 shows the error of Eq. 4.4, summed over all operators except that correspond to E_{tot} . The results show that performance on molecules types not included in the training is comparable to the performance of the ethane test set.

The better performance obtained for E_{tot} than individual operators indicates a substantial cancellation of errors upon addition of operators to obtain the total energy. This suggests that improved performance for E_{tot} may be obtained by weighting E_{tot} more strongly in the objective function of Eq. 4.2. The performance on energy differences, computed via Eq. 4.4, is shown as a function of such a weighting parameter in Figure 4.6 and Figure 4.7. Figure 4.6 shows that the performance for the total energy tends to improve with increased weighting of the total energy. This includes performance on molecule types not included in the training of the model, although the improvement on the training set is substantially larger than that on the test sets. Figure 4.7 shows that the improved performance on the total energy comes at a cost to performance on the individual operators, with the error becoming 1000s of kcal/mol when individual operators are

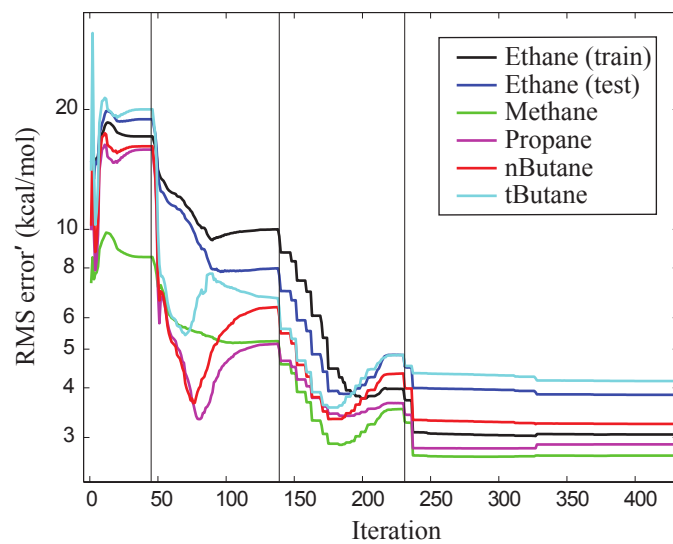


Figure 4.4: RMS error, Eq. 4.4, in the total energy, E_{tot} , for a variety of molecules, using the pLL model of Figure 4.1. This error reflects performance on the change in total energy arising from changes in geometry or environment.

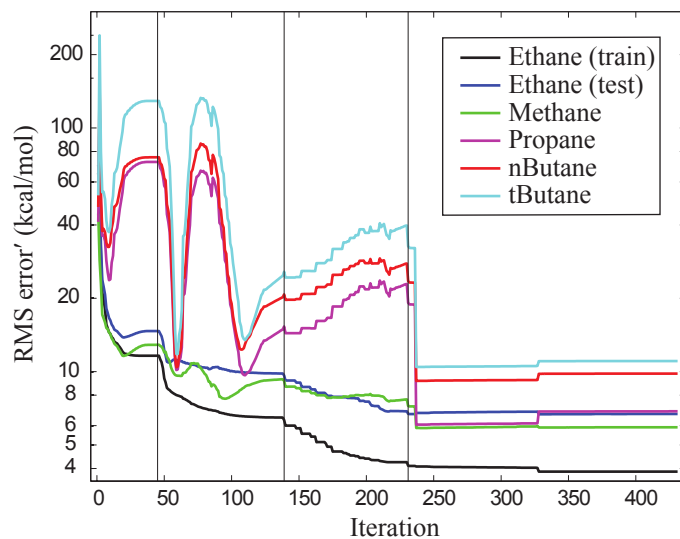


Figure 4.5: RMS error associated with individual operators for the fits of Figure 4.4, computed using Eq. 4.4 with summation over all operators except E_{tot} .

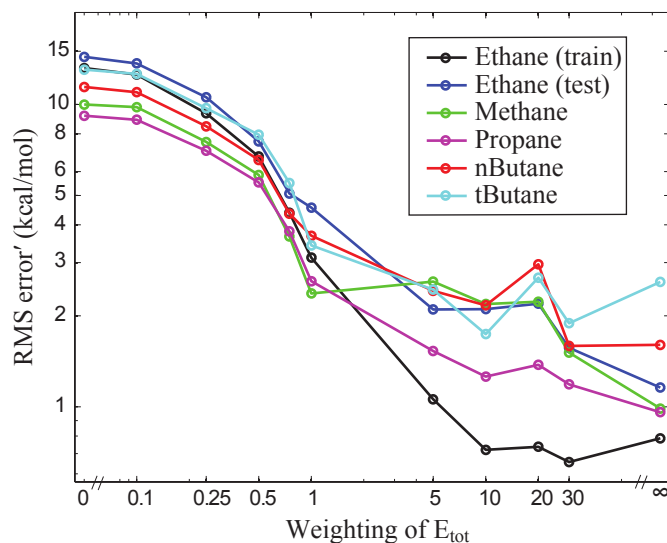


Figure 4.6: RMS error, Eq. 4.4, in total energy of various molecules, obtained from fitting the pLL model to ethane, with the total energy multiplied by a weighted factor.

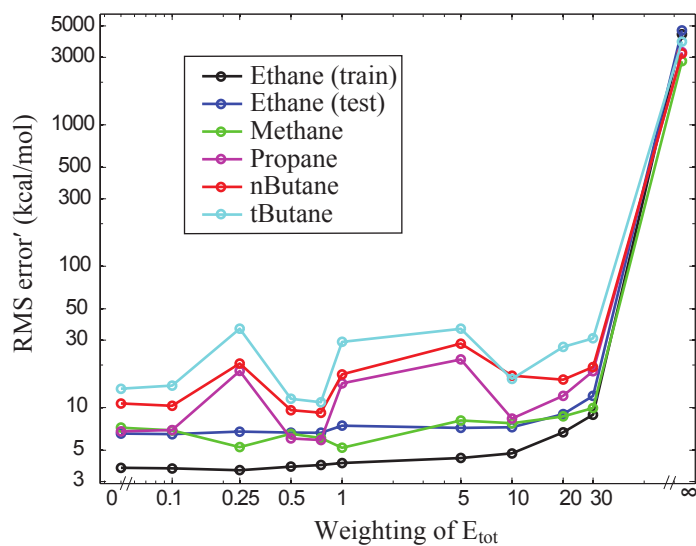


Figure 4.7: RMS error associated with individual operators for the fits of Figure 4.6, computed using Eq. 4.4 with summation over all operators except E_{tot} .

	Error in Energy Eq. 4.2					Error in Energy Differences Eq. 4.4				
	KE	EN_H	EN_C	E_2	E_{tot}	KE	EN_H	EN_C	E_2	E_{tot}
Ethane (train)										
Initial	938	45	716	311	545	126	21	86	112	17
No Context	25	27	25	29	7.6	24	26	25	28	7.6
Context	6.6	3.6	5.4	7.1	0.7	6.6	3.6	5.3	7	0.7
Ethane (test)										
Initial	904	51	693	352	542	142	23	93	111	15
No Context	33	28	31	42	7.2	32	28	31	42	7.2
Context	11	4.2	9.2	11	2.1	11	4.2	9.2	11	2.1
Methane										
Initial	424	47	717	161	279	63	22	67	51	8.8
No Context	197	28	482	213	60	13	28	44	18	6.2
Context	38	5.2	144	110	2.2	8	5.1	13	8.5	2.2
Propane										
Initial	1456	39	737	372	831	81	17	56	66	6.1
No Context	182	21	252	248	55	18	20	206	51	5.2
Context	49	4.3	97	250	5.9	10	4.3	12	14	1.3
nButane										
Initial	1959	40	734	503	1104	82	18	65	80	7.7
No Context	379	22	328	608	101	19	22	200	64	4.7
Context	136	6.3	196	657	19	14	6.3	27	32	2.2
tButane										
Initial	1893	45	718	554	1109	119	18	66	90	8
No Context	327	22	416	535	91	34	22	343	47	4.9
Context	107	5.7	195	680	19	16	5.7	23	37	1.7

Table 4.3: Detailed errors (in kcal/mol) from the fit in Figure 4.6 corresponding to $w=10$

not included in the fit (a weight of infinity). Fits to just the total energy therefore obtain good performance on the total energy, but very poor performance for individual operators. This poor performance on individual operators does not appear to harm transfer of model parameters between molecular systems, as reasonable performance is obtained for the total energy of molecule types not included in the training of the model, even when the molecule is trained on just total energy (see results for a weight of infinity in Figure 4.6).

A weighting of 10 in Figure 4.6 and Figure 4.7 reduces the error in the total energy on the test molecules, while having little impact on the errors for the individual operators. Table 4.3 summarizes the results obtained from this fit. The left portion of the table shows errors in the

absolute energies, while the right shows errors in energy differences. The rows labeled initial refer to differences between STO-3G (LL) and 6-31G (HL). The initial disagreement regarding energy differences, Eq. 4.4, is large for individual operators and considerably smaller for the total energy, E_{tot} . This indicates substantial cancellation of errors between the operators of STO-3G and 6-31G. The use of constant parameters, labeled no context in the table, reduces the error in the total energy by about a factor of two for both the train and test data sets, while substantially improving the agreement on individual operators. The decrease of only a factor of two in the error associated with the total energy speaks to the complex nature of the data set being explored here. The environments are perturbing the electron density in ways that cannot be captured by simply scaling the matrix elements of the minimal basis Hamiltonian. When the model parameters are made functions of context, the model improves substantially. With context included, the error for E_{tot} of test molecules is reduced by a factor of between 3.5 and 7 relative to the initial error. Performance on individual operators is also improved substantially, although these remain considerably larger than the error in E_{tot} .

4.4 Conclusions

This work explores a form of semiempirical model in which parameters are embedded into a LL *ab initio* theory and adjusted to obtain agreement with a HL *ab initio* theory. The results suggest that the approach provides a well-controlled approximation based on molecular similarity.

The approximation introduces a predictable level of accuracy. Estimates of the error introduced by the approximation are obtained from the performance of the model on test data. Here, we trained a model to data on ethane and monitored convergence of the training process on a test set containing also data on just ethane. The performance on this ethane test set correlated well with the performance seen on molecules (methane, propane, and butane) not included in the training process. This suggests that the performance on test data provides a reliable measure of the error introduced by the approximation.

The model also has a means for systematically improving the accuracy. The general approach explored here may be improved by increasing either the sophistication of the LL *ab initio* theory into which the parameters are embedded, or by improving the means through which the parameters are embedded. The current implementation considered only the latter, by making the parameters functions of the molecular context such as bond lengths, atomic charges and bond orders. Incremental addition of the context variables to the model led to steady improvements in the accuracy of the model, as judged by performance on test data.

The results presented here suggest this approach holds promise for creating models of organic systems. The next chapter builds on this work by considering organic molecules containing N, O and F. A more systematic comparison of various means for embedding the parameters in the Hamiltonian is also carried out.

Chapter 5

An Embedded Parameters Model of CNOFH Containing Molecules

5.1 Introduction

This chapter extends the model of Chapter 4 to a broader range of molecules, consisting of hydrocarbons substituted with amine, hydroxyl, and fluoro groups. A broader range of model forms is also explored, by comparing a variety of approaches through which parameters may be embedded into the low-level (LL) *ab initio* Hamiltonian, to form a parameterized low-level (pLL) models.

A number of means are explored to improve training of the model. That the parameters of the pLL model are associated with molecular fragments has the advantage of allowing model parameters to be transferred between different molecules. However, challenges in training the model result from the parameters being associated with molecular fragments while model performance is measured on the molecule as a whole. The training data includes total energy and frontier orbital energies as target properties the model is intended to predict. In addition, the extent to which inclusion of additional properties can enrich the training data and so lead to better model performance is explored. These additional properties include decomposition of the total energy

by operator into kinetic energy, electron-nuclear and electron-electron components. Inclusion of data regarding the molecular charge distribution is also explored, to include information related to longer-range interactions, Eq. 1.3.

Section 5.2 describes the data used to train and test the pLL models. Section 5.3 describes the various approaches used to apply scaling factors to the operator matrix elements of the LL Hamiltonian. The manner in which these scaling factors are made functions of the molecular context is described in Section 5.4. Training of pLL models is described in Section 5.5 and the results are discussed in Section 5.6. Section 5.7 reflects on the outcomes and discusses possible next steps.

5.2 Chemical Data

The data used to train and test the model is constructed by selecting a set of molecules across which molecular similarity may be expected to hold, generating geometries of these molecules that are distorted from the minimum-energy geometry in some controlled manner, and finally placing these molecules in external electrostatic environments. The combination of a molecular geometry and an environment will be referred to as a molecular “instance”.

The geometries are distorted in a manner that is meant to provide uniform sampling over some prespecified range of perturbations to both bond lengths and bond angles. The approach utilizes a function, $\mathbf{R}(\xi)$, that maps the internal coordinates of the z-matrix, ξ , to all bond lengths and bond angles, $\mathbf{R}$, of the molecule. A set of target bond lengths and angles is generated as

$$\mathbf{R}^{\text{target}} = \mathbf{R}(\xi^{\text{opt}}) + \mathbf{X} \quad (5.1)$$

where ξ^{opt} are the internal coordinates of the optimized structure and $\mathbf{X}$ is a list of random numbers that specify distortions in bond lengths and angles. A uniform random distribution is used for $\mathbf{X}$ to ensure that the geometries well sample the specified range of bond lengths and

angles. Because there are typically more bond lengths and angles than internal coordinates, $\mathbf{R}(\xi)$ is not directly invertible. The internal coordinates are therefore adjusted to give the best match to $\mathbf{R}^{\text{target}}$ by minimizing the following root mean square (RMS) deviation,

$$\sum_{i=1}^N [w_i (R_i(\xi) - R_i^{\text{target}})]^2, \quad (5.2)$$

where N is the total number of bond lengths and angles and w_i is a weight that is inversely proportional to the width of the uniform random number distribution used in generating $\mathbf{R}^{\text{target}}$.

In cases where a dihedral angle between groups is sampled, this is done by creating a z-matrix where a single internal coordinate controls the relative rotation of the two groups. This coordinate is assigned to a randomly generated value and not altered during the minimization of Eq. 5.2. For propane, propylene and butane, the range of dihedral angles is restricted to prevent overlapping groups.

To include information on the manner in which the systems respond to external electrostatic environments, each molecule is placed in static electric fields applied along the X, Y and Z directions. A magnitude of 0.00333 a.u. was chosen, which is sufficient to induce a dipole of 0.1 Debye in water. Each molecular geometry thus leads to four instances, one instance in zero field and three instances for fields along the X, Y and Z directions.

Section 5.1 argued that the degree to which long-range interactions are well described by the pLL model is related to the degree to which the density of Eq. 1.2 agrees with that of the HL model. To obtain a measure of electron density that is comparable across basis sets, the interaction of the electron density with a set of point charges surrounding the molecule, E_{chg} , is computed. One hundred locations for the point charges are first constructed using the Chelpg method. [111] Values are then assigned to these charges from a uniform random distribution of -0.08 e to +0.08 e . Twenty such external point charge environments are created for each molecular geometry.

For each instance in the dataset, the following properties are generated.

E_{tot} The total energy of the molecule, without inclusion of the interaction with an external field. The external field thereby serves to perturb the electron density and E_{tot} measures the electronic energy associated with this perturbation.

E_{orb} The energy of the HOMO and LUMO orbitals. Because the calculations are being done at the SCF level, these are Koopman’s theory estimates of the ionization potential and electron affinity.

KE, EN_A , E_2 Expectation values of the operators that make up the Hamiltonian. KE is the total kinetic energy. EN_A is the interaction of the electron density with the A^{th} nucleus, and E_2 is the total electron-electron repulsion energy. Unlike decomposition of the energy into atoms or molecular fragments, the decomposition of the energy by operator is uniquely defined and directly comparable across basis set.

E_{chg} The interaction energy between the electron density and the randomized point charges described above.

The model parameters are adjusted to obtain agreement between the pLL and HL models for the above quantities, as described in Section 5.5.

The following data sets were generated. Models were trained on the first three datasets, with the remaining used for model testing.

ethane Includes all 16 unique ways to place between zero and two substituents, selected from -NH₂, -OH, and -F, on ethane. The optimized geometry, at the 6-31G level, is included along with distorted geometries created as described above. The bond lengths are uniformly distorted by ± 0.2 Å, bond angles are distorted by $\pm 10^\circ$, and rotation about the central C-C bond is randomized. For each such geometry, four instances are created corresponding to no external field, and external fields along the X, Y and Z axes of 0.0033 atomic units. A training and validation set are created, both of which contain the optimized geometry along with 10 distorted geometries in the training set and 9 in the validation set, leading to 640 instances in the training and 576 instances in the validation set.

ethylene This dataset is generated in a manner identical to the *ethane* dataset, except that the substituents are attached to ethylene and the bond lengths are uniformly distorted by ± 0.15 Å. Just as in ethane, the geometries include full sampling of rotation about the central carbon-carbon double bond. The bond length distortions were reduced to make the disagreements between the initial LL and the HL model for E_{tot} comparable.

combined A combination of the *ethane* and *ethylene* datasets.

propane Includes all unique combinations of zero to two of the above substituents on propane, but with at most one substituent attached to each carbon. The bond lengths are distorted by ± 0.2 Å, bond angles are distorted by $\pm 10^\circ$, and rotation about one carbon-carbon bond is randomized subject to avoiding overlapping groups. A total of 576 instances are included in this dataset.

propylene This dataset is generated in a manner similar to the propane dataset, except that the substituents are attached to propylene, the bond lengths are uniformly distorted by ± 0.15 Å, and rotation about the double bond is randomized. The substituents are either on each end of the molecule or adjacent across the double bond.

t-butane This dataset is identical to the *ethane* dataset but has two methyl groups attached to one of the carbons. This tests the ability of a model trained on molecules in which carbon has at most three heavy atom substituents to transfer to molecules in which carbon has four heavy atom substituents. This dataset consists of 576 instances.

The magnitude of geometric distortions used above was chosen to provide data that is challenging to model yet is on molecules of sufficiently small size that models can be trained in an efficient manner. The average change in energy associated with the distortions in the *combined* dataset is 65 kcal mol^{-1} . Table 5.1 shows disagreements between the HL model and traditional SEQC models for the *combined* dataset. The relatively large values point towards the level of difficulty associated with these datasets, although this is only qualitative because the differences are quoted with respect to the HL model while the SEQC models are parametrized to experimental data.

Method	E_{tot} kcal mol ⁻¹	E_{HOMO} eV	E_{LUMO} eV
CNDO/2	58.97	2.98	1.62
INDO	59.94	2.43	1.56
MNDO	34.49	1.99	3.55
AM1	32.95	1.84	3.37
PM3	34.81	2.09	3.78

Table 5.1: RMS differences between the HL model and SEQC methods for the *combined* dataset. E_{tot} refers to the change in energy associated with geometric distortions and applied fields (see Section 5.5).

The amount of training data for the *ethane* and *combined* datasets is shown in Table 5.2. The values that most influence the training are E_{tot} and E_{orb} . With the default weighting of E_{chg} in the objective function of Section 5.5, E_{chg} does not play a strong role in the fitting. The effects of increasing the weighting of E_{chg} are explored in Section 5.6. Comparisons are also made between models trained with and without inclusion of the decomposed energies (KE, EN_A and E_2).

	<i>ethane</i>	<i>combined</i>
E_{tot}	688	1376
E_{orb}	1408	2816
E_{chg}	13376	26752
KE	688	1376
EN_A	6665	11954
E_2	688	1376

Table 5.2: Amount of training data in the datasets used to develop pLL models.

5.3 Embedding Parameters in the Low-Level Hamiltonian

Flexible model forms are created by embedding parameters into a low-level (LL) *ab initio* model, to form a parametrized low-level (pLL) model. The parameters of the pLL model are then adjusted to obtain agreement with a higher-level (HL) *ab initio* model. We have chosen to use SCF solutions for both the pLL and HL model, such that the empirical parameters are used

only to compensate for differences related to the basis set. For the HL model, a split-valence basis set with polarization functions (6-31G**) is used. For the valence electrons of the pLL model, the STO-3G minimal basis is used. For the core electrons of the pLL model, the core orbitals of the 6-31G basis are used, in order to make the description of the core electrons in the pLL model equivalent to that in the HL model. This combination of 6-31G core functions and STO-3G valence functions will be referred to as the modified STO-3G basis (mSTO-3G). The unparametrized low-level (LL) model therefore corresponds to SCF solutions within the mSTO-3G basis.

To initialize the pLL model, the matrix elements of each of the operators appearing in the electronic Hamiltonian are evaluated in the mSTO-3G basis of the LL model

$$KE_{i,j} \quad EN_{i,j}^A \quad (ij|kl) \quad H_{i,j}^{env} \quad (5.3)$$

where KE is the kinetic energy operator, EN^A is the interaction of the electrons with the A^{th} nucleus, $(ij|kl)$ are the two-electron integrals in chemists notation, and H^{env} is the interaction with the external electrostatic environment. The pLL model embeds parameters into the mSTO-3G Hamiltonian by multiplying subblocks of these matrices by scaling factors, S . Different scaling factors are used for different operator types (KE , EN_A , E_2) and for differing elements (H, C, N, O, F) involved in the sub-block being modified. The scaling factors, S , depend on context through the form

$$S = 1 + p_0 + \sum_{i=1}^{N_c} p_i c_i \quad (5.4)$$

where c_i are context variables that describe the local environment of the atom or bond associated with the sub-block being modified (Section 5.4). The total number of parameters, p_i , is then related to both the number of scaling factors and the degree to which these are made context sensitive, N_c . A number of different approaches, or policies, are explored for embedding the scaling factors. These policies differ along two dimensions. One dimension is whether different

scaling factors are used for σ versus π interactions between p orbitals, indicated as $\sigma\pi$ policies. The other dimension is whether scaling factors are applied directly to the two-electron matrix elements, $2elec$, or to the matrix elements of the J and K operators in the Fock matrix, JK . Note that matrix elements that are not scaled retain their LL values, and scaling is not applied to matrix elements involving core orbitals.

For diagonal blocks of one-electron operators, $O_{i,j}$ where i and j are orbitals on the same atom, one scaling factor is used for the s orbitals and a different scaling factor is used for the subblock corresponding to the p orbitals. For off-diagonal blocks of one-electron operators, $O_{i,j}$ where i and j are orbitals on different atoms, scaling is applied only when i and j are on bonded atoms. Two different policies are used for these off-diagonal blocks. In the default policy, different scaling factors are used for matrix elements connecting s with s , p with p , and s with p orbitals, leading to three parameters for each pair of elements. The $\sigma\pi$ policy adds an additional scaling factor per pair of heavy elements by using different parameters for σ versus π orientations of the p orbitals. (This is done by temporarily rotating the p orbitals such that one p orbital of each atom lies along the internuclear axis.) For the attraction between the electron density and the A^{th} nucleus, $EN_{i,j}^A$ is modified according to the above rules only when i , j , or both i and j reside on atom A . H^{env} is not modified.

For electron-electron interactions, two different policies are explored. The $2elec$ policy modifies the two-electron integrals, $(ij|kl)$, directly. Scaling factors are applied only to those integrals that are retained in the NDDO approximation [35] but, unlike NDDO, the remaining integrals retain their LL values instead of being set to zero. For on-atom integrals, where i , j , k , and l are on the same atom, the three-parameter form developed by Slater [31, 80, 112] is used, which leads to parameters F_0 , G_1 , and F_2 per heavy element. The unscaled values of F_0 , G_1 and F_2 are first determined from the two-electron integrals of the LL basis, and these values are then multiplied by scaling factors, leading to one scaling factor for hydrogen and three scaling factors for each heavy element. Between bonded atoms, all integrals $(ij|kl)$ where i and j are on one atom while k and l are on the other atom, are multiplied by a single scaling factor, leading to one scaling

factor for each pair of elements. A scaling factor between non-bonded atoms is included only for non-bonded hydrogen atoms closer than 2.8 Å, because this leads to a small improvement in model predictions. This nonbonded scaling uses bond order as its single context variable, c_1 of Eq. 5.4.

The *JK* policy is an alternative to the *2elec* policy based on scaling of the Coulomb, J, and exchange, K, matrices of the Fock operator. The scaling is applied in each iteration of the self-consistent solution of the Roothan equations. The unscaled two-electron integrals of the LL model are first used to evaluate the J and K matrices. The policies discussed above for the one-electron operators are then used to scale the J and K matrices, with different scaling factors used for J versus K.

5.4 Context Sensitive Scaling Factors

A given set of scaling factors is likely to be valid only over some limited range of molecules. To extend this range, the scaling factors are made functions of the current context of the atom or bond by making them linear functions of context variables, c_i of Eq. 5.4. In training of the model parameters (Section 5.5), the context sensitivity is turned on sequentially in the order shown in Table 5.3. The first level of context involves only bond lengths and so only introduces simple geometry dependence into the scaling factors. The second and third level of context go beyond geometry to include aspects of the electronic structure of the molecule through atomic charges and bond orders. As charge is pushed onto (or pulled from) an atom, we expect the charge density to expand (or contract). This effect is not present in the minimal basis of the LL model. Making the scaling factors functions of atomic charge and bond orders is an attempt to compensate for this. For scaling factors applied to the diagonal, on-atom, matrix elements, the atomic charge seems most relevant and so this is included at the second level of context sensitivity. For scaling factors applied to off-diagonal blocks, the bond order seems more relevant and so this is included at the second level of context. The third level crosses these, using bond order for diagonal blocks

and atomic charges for off-diagonal blocks. The context variables are calculated using the LL model.

level	on-atom	between-atom
c1 ^a	r: Average bond length to bonded atoms	r: Bond length
c2 ^b	q: Mulliken charge on the atom	bo: Bond order
c3	bo: Average bond order to bonded atoms	q: charge difference between atoms

Table 5.3: Context variables, c_i of Eq.5.4, for the three levels of context added sequentially to the model during training.

^a r refers to deviations of the bond length from 1.1Å for C-H bonds and 1.5Å for all other bonds.

^b q refers to deviations from average values of 0.112 for H, -0.335 for C, -0.713 for N, -0.639 for O, and -0.417 for F.

5.5 Training the pLL Model

The parameters of the pLL model are trained by minimizing

$$Obj(\mathbf{p}) = \sum_i^{N_{data}} w_i^2 \left(X_i^{HL} - X_i^{pLL}(\mathbf{p}) \right)^2 + R(\mathbf{p}) \quad (5.5)$$

where Obj is the objective to be minimized, $\mathbf{p}$ is a vector containing all parameters in the scaling factors of Eq. 5.4, X_i^{HL} and $X_i^{pLL}(\mathbf{p})$ are predictions of the HL and pLL models respectively, and w_i sets the relative weight of the computed properties in the objective. The sum includes all molecular instances and all properties of these instances, leading to a total of N_{data} values to be fit. $R(\mathbf{p})$ is the regularization function described below and is included only for the training dataset. The properties that are optionally included in the objective are those listed in Section 5.2.

Because only energy differences are relevant to making predictions regarding molecules, properties related to absolute energies (E_{tot} , KE , EN_A , E_2) are included in the objective of Eq. 5.5 in only a relative manner. This is done by taking, as a reference state for each molecule, the optimized geometry of that molecule in no external environment. The quantity X_i of Eq. 5.5 is then the difference between the computed energy and that computed for the reference state.

The objective therefore includes only the change in energy associated with distorting the molecule and placing it in an external electrostatic environment. For the remaining properties of Section 5.2 (E_{orb} , and E_{chg}), X_i of Eq. 5.5 is the value itself, not the difference relative to the reference state.

The weights, w_i , are based on the relative accuracy desired for the predicted quantities and here are set to $(2 \text{ kcal mol}^{-1})^{-1}$ for E_{tot} and $(0.1 \text{ eV})^{-1}$ for E_{orb} . For E_{chg} , the default value of $0.016 \text{ (kcal mol}^{-1})^{-1}$ is sufficiently low that inclusion does not degrade performance on E_{tot} or E_{orb} . In Section 5.6, the impacts of increasing the weight of E_{chg} is examined. The weighting of decomposed energies is $1/30^{th}$ that of E_{tot} , as discussed further in Section 5.6.

The parameters are optimized using the trust-region reflective algorithm [108] as implemented within Matlab. [109] To help prevent negative and overly large scaling factors from being explored, p_0 of Eq. 5.4 is constrained to lie between -1 and 2. The inclusion of context sensitive parameters is done in stages, such that all non-context sensitive parameters, p_0 of Eq. 5.4, are first optimized, with all other parameters set to zero. As parameters associated with additional levels of context are added, all of the currently active parameters are optimized, using the values from the previous level of context as initial values.

The regularization function, $R(\mathbf{p})$ of Eq. 5.5, has the form

$$R(\mathbf{p}) = N_{data}C\|\mathbf{p} - \mathbf{p}_{init}\| \quad (5.6)$$

where $\mathbf{p}_{init}$ is the initial value of the parameters for that level of context. For the initial fit of non-context sensitive parameters, $\mathbf{p}_{init} = \mathbf{0}$ and thus the regularization adds a penalty for scaling factors that deviate strongly from 1. As context is added, the regularization penalizes large deviations from the scaling factors that were obtained at the previous context level. Initial studies on smaller datasets led to a value of 6 for C .

To help prevent overtraining, decisions on when to terminate the optimization are based on performance of the model on the validation dataset. The training dataset is used to calculate the

objective and its Jacobian, which the optimization algorithm then uses to generate a new trial set of parameters. If performance on the validation dataset degrades for four sequential optimization steps, this is taken as a sign of overtraining and the optimization is terminated. This process is applied at each level of context discussed above.

The linear dependence of the scaling factors on model parameters, Eq. 5.4, enables computational optimizations. For each molecular instance, the derivative of each operator with respect to each of the parameters is first computed. This data is then distributed such that different processor cores handle different molecular instances. This allows the operators to be constructed quickly for arbitrary parameters. The Jacobian for the objective of Eq. 5.5 is obtained from finite differences, using a step size of 0.01. Analytical evaluation of the Jacobian, as obtained from the Hellmann-Feynman theorem, does not lead to as high quality in the final fits. Optimization of the parameters on the *ethane* training datasets required about 100 iterations and took about 18 hours on 12 processor cores.

5.6 Results

Figure 5.1 shows iterative improvement of the objective function of Eq. 5.5 during model training, with vertical lines indicating addition of context dependence to the model. The training objective includes the regularization of Eq. 5.6 and the strong decrease in the training objective on addition of context reflects the redefinition of the regularization in terms of deviations of the parameters from those obtained at the previous context level (Section 5.5). The greatest reduction in error occurs in the first stage, where the scaling parameters are constants, referred to as context level zero. Including the first level of context, Table 5.3, adds bond-length dependence to the parameters and also leads to a significant reduction in error. The degree to which higher levels of context lead to additional improvements is smaller and varies depending on dataset and policy.

Models were trained on the *ethane*, *ethylene* and *combined* datasets using the four policies

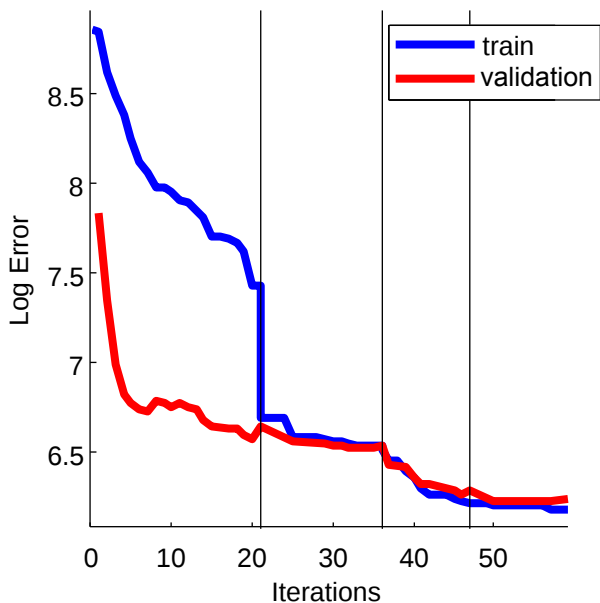


Figure 5.1: The objective of Eq. 5.5 versus iteration during training of a model on the *combined* dataset. The pLL model uses the *JK* policy and decomposed energies are included in the objective. Vertical lines indicate addition of the contexts of Table 5.3.

discussed in Section 5.3. In addition, fits were done both with and without inclusion of the decomposition of the energy by operator (Section 5.2). Here, the focus is on models trained on the *ethane* (Figures 5.2 and 5.3) and on the *combined* datasets (Figures 5.4 and 5.5 and Tables 5.5 and 5.4). Models trained on ethane were not tested on ethylene or propylene since the training set did not include molecules with carbon-carbon double bonds.

For some sets of parameters in a pLL model, the SCF iterations may fail to lead to a converged density matrix. In such cases, predictions obtained with the unconverged density matrix are retained, and the RMS errors in E_{tot} rise to hundreds of kcal mol⁻¹. During model training, inclusion of unconverged values in evaluation of the objective of Eq. 5.5 provides a sufficiently large penalty that the optimization algorithm rejects sets of parameters that lead to such instabilities. In reporting performance on model testing, inclusion of such unconverged values provides an indication of the degree to which a model leads to poor or unstable performance.

Three general trends emerge. First is that using separate parameters for σ and π interactions

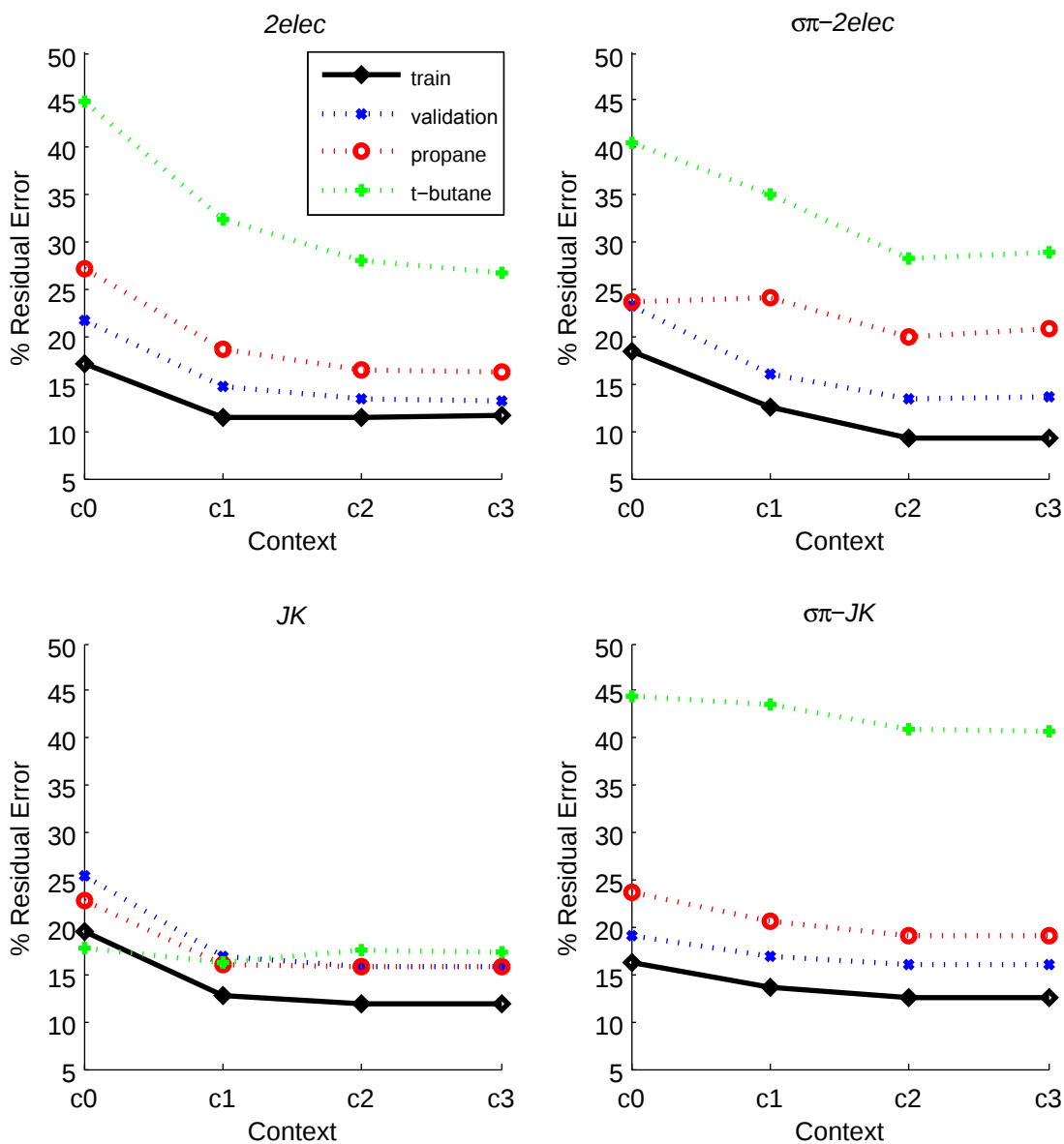


Figure 5.2: Residual error in E_{tot} as a function of context level for pLL models trained on the *ethane* dataset, without inclusion of decomposed energies in the objective of Eq. 5.5. Residual error is RMS disagreement between the pLL and HL models, quoted relative to initial disagreement between the LL and HL models. Panels refer to the policies of Section 5.3.

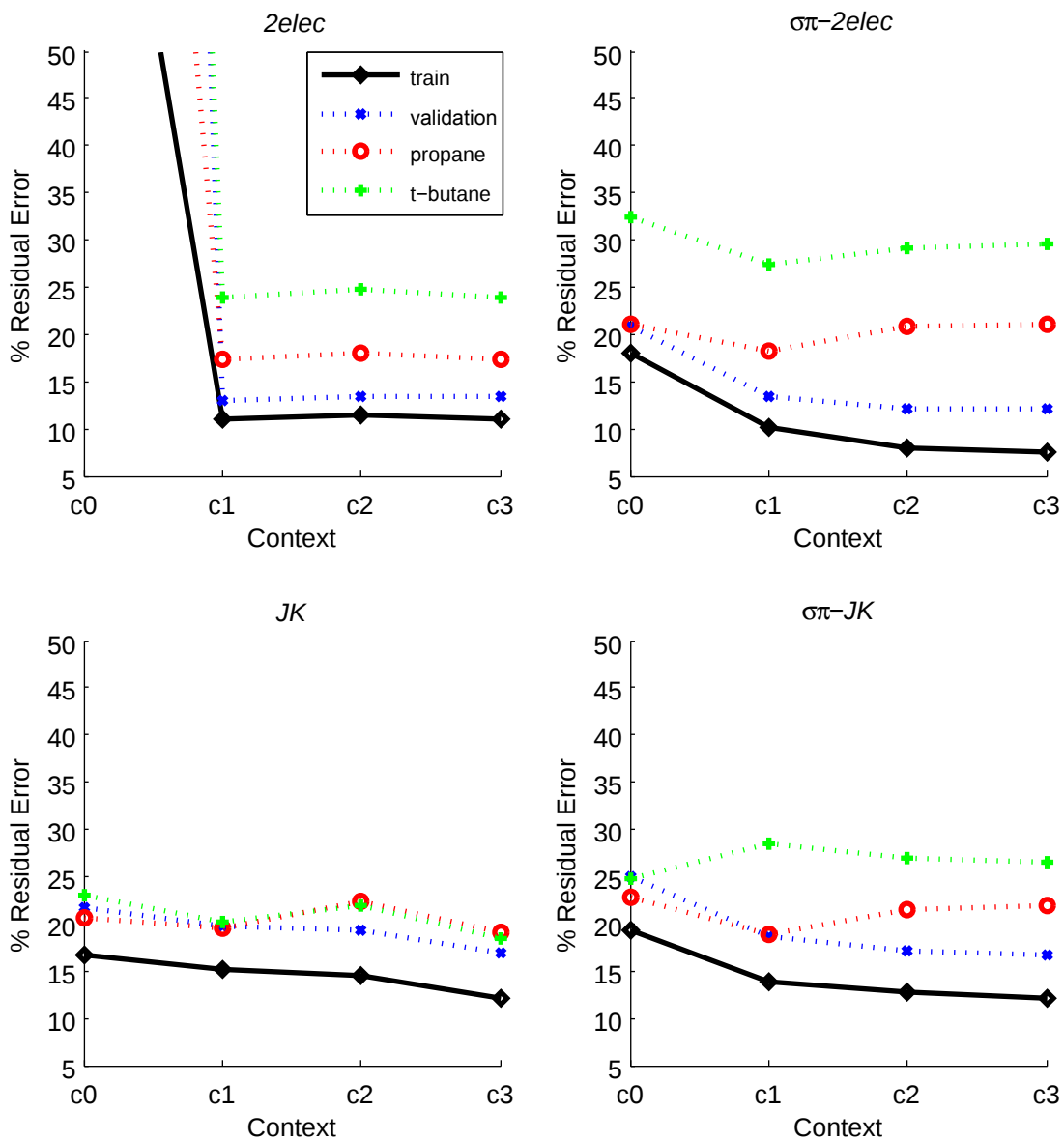


Figure 5.3: Residual error in E_{tot} as a function of context level for pLL models trained on the *ethane* dataset, with inclusion of decomposed energies in the objective of Eq. 5.5. Conventions are as in Fig. 5.2.

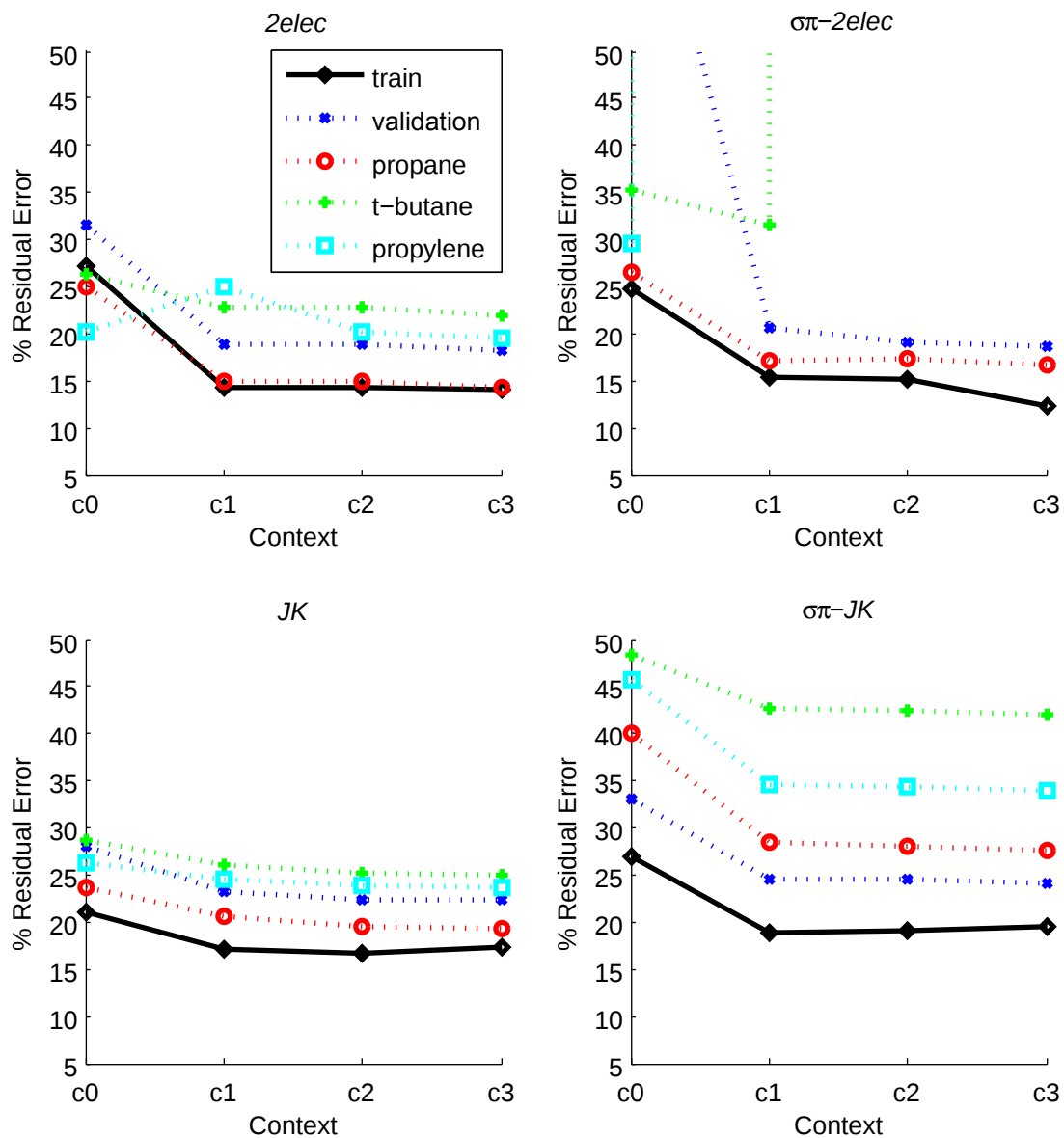


Figure 5.4: Residual error in E_{tot} as a function of context level for pLL models trained on the *combined* dataset, without inclusion of decomposed energies in the objective of Eq. 5.5. Conventions are as in Fig. 5.2.

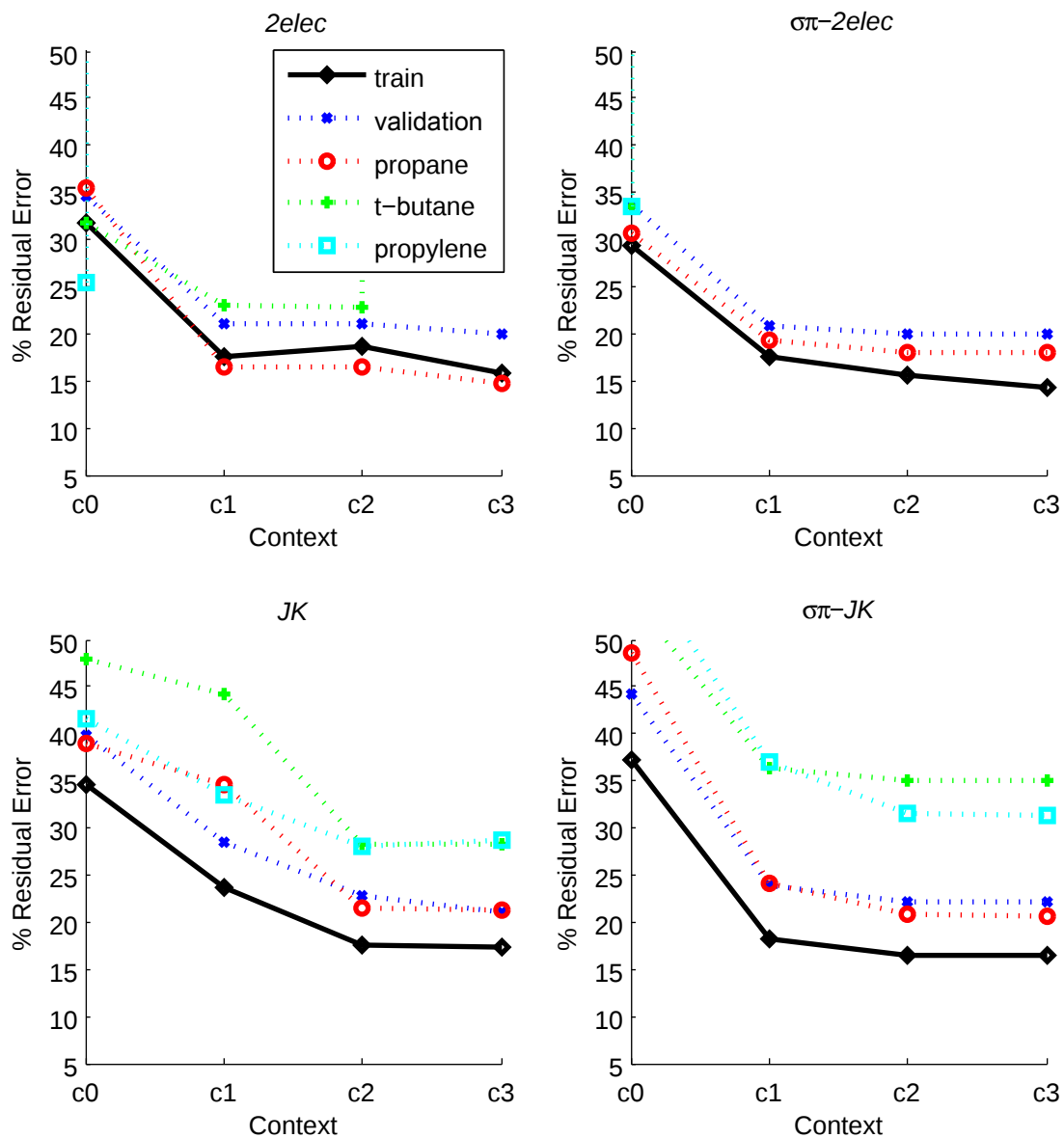


Figure 5.5: Residual error in E_{tot} as a function of context level for pLL models trained on the *combined* dataset, with inclusion of decomposed energies in the objective of Eq. 5.5. Conventions are as in Fig. 5.2.

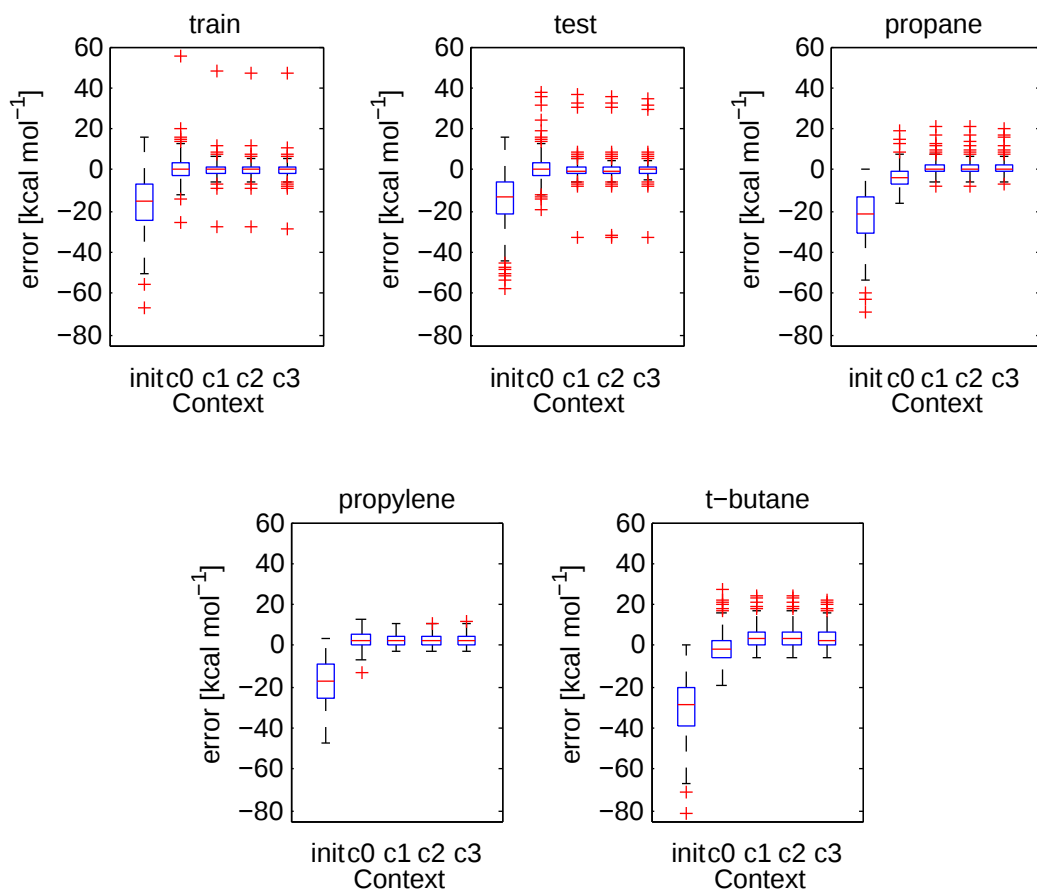


Figure 5.6: Boxplots of disagreements in E_{tot} between HL and the best performing pLL model. The pLL model uses the *2elec* policy and is trained on the *combined* dataset without inclusion of decomposed energies. Initial refers to the unparametrized LL model.

	Context	Train		Validation		Propane		t-Butane		Propylene		
		n_{par}	E_{tot}	E_{orb}	E_{tot}	E_{orb}	E_{tot}	E_{orb}	E_{tot}	E_{orb}	E_{tot}	E_{orb}
Initial			21.47	4.44	19.21	4.43	26.68	5.38	32.69	5.31	20.03	2.78
2elec	c0	75	5.86	0.25	6.07	0.26	6.64	0.39	8.58	0.44	4.04	0.26
	c1	150	3.05	0.15	3.63	0.17	3.96	0.34	7.44	0.63	4.99	0.17
	c2	224	3.08	0.15	3.62	0.17	3.97	0.34	7.48	0.65	4.05	0.26
	c3	298	3.00	0.14	3.52	0.17	3.84	0.35	7.18	0.66	3.90	0.26
$\sigma\pi$ -2elec	c0	83	5.33	0.26	15.09	0.31	7.10	0.38	11.50	0.45	5.93	0.27
	c1	166	3.28	0.15	3.94	0.18	4.60	0.31	10.32	0.55	311.65	1.17
	c2	248	3.27	0.14	3.66	0.17	4.65	0.44	1514.45	2.28	740.76	1.74
	c3	330	2.66	0.13	3.59	0.16	4.44	0.50	2046.02	2.24	255.79	1.20
JK	c0	108	4.49	0.23	5.35	0.24	6.31	0.41	9.34	0.58	5.23	0.36
	c1	216	3.68	0.20	4.45	0.21	5.47	0.33	8.49	0.48	4.91	0.27
	c2	324	3.56	0.20	4.28	0.20	5.19	0.32	8.21	0.48	4.76	0.26
	c3	432	3.70	0.20	4.29	0.20	5.12	0.33	8.13	0.48	4.72	0.26
$\sigma\pi$ -JK	c0	125	5.77	0.24	6.33	0.26	10.64	0.40	15.75	0.44	9.13	0.20
	c1	250	4.03	0.20	4.68	0.21	7.55	0.36	13.89	0.40	6.93	0.19
	c2	374	4.09	0.20	4.72	0.21	7.47	0.36	13.82	0.40	6.85	0.19
	c3	498	4.20	0.21	4.62	0.21	7.36	0.36	13.70	0.40	6.79	0.19

Table 5.4: RMS errors for models trained on the *combined* dataset without inclusion of decomposed energies. n_{par} is the number of parameters in the model. Large RMS errors reflect failures of SCF iterations to converge (Section 5.6). Units are eV for E_{orb} and kcal mol⁻¹ for all other quantities.

between p orbitals does not improve model performance. Although the $\sigma\pi$ -2elec and $\sigma\pi$ -JK models do, in some cases, outperform their 2elec and JK counterparts on the training dataset, performance on the test datasets is not substantially improved and in many cases degrades.

A second general trend is that models trained on the *combined* dataset typically perform better than those trained on *ethane* alone. The inclusion of ethylene in the *combined* dataset has the advantage of extending the training data to include a broader class of molecules. However, inclusion of ethylene also requires a single model to describe a broader class of molecules, including large amplitude rotation about a double bond. The improvement in performance seen for the *combined* dataset suggests that the benefits accruing from training on more diverse data overrides the challenges associated with using a single model to describe a more diverse class of molecules. This is a promising result with regards to the ability of the approach explored here to develop models that are applicable to a diverse range of molecules.

	Context	Train		Validation		Propane		t-Butane		Propylene		
		n_{par}	E_{tot}	E_{orb}	E_{tot}	E_{orb}	E_{tot}	E_{orb}	E_{tot}	E_{orb}	E_{tot}	E_{orb}
Initial			21.47	4.44	19.21	4.43	26.68	5.38	32.69	5.31	20.03	2.78
<i>2elec</i>	c0	75	6.83	0.29	6.65	0.30	9.46	0.57	10.36	0.55	5.08	0.24
	c1	150	3.78	0.18	4.05	0.18	4.42	0.32	7.50	0.31	524.89	1.78
	c2	224	3.98	0.18	4.03	0.18	4.38	0.32	7.44	0.31	580.27	1.94
	c3	298	3.39	0.18	3.83	0.18	3.95	0.30	751.20	1.82	627.40	2.03
$\sigma\pi$ - <i>2elec</i>	c0	83	6.32	0.29	6.49	0.29	8.18	0.53	10.93	0.49	6.69	0.27
	c1	166	3.78	0.18	4.01	0.18	5.14	0.31	1105.67	1.41	851.01	1.60
	c3	248	3.36	0.17	3.82	0.19	4.82	0.29	4234.52	4.28	1743.58	4.06
	c3	330	3.08	0.17	3.83	0.18	4.82	0.31	4531.66	4.79	2109.05	4.36
<i>JK</i>	c0	108	7.39	0.37	7.64	0.33	10.38	0.39	15.66	0.50	8.30	0.27
	c1	216	5.05	0.31	5.45	0.31	9.18	0.38	14.39	0.47	6.71	0.22
	c3	324	3.78	0.25	4.37	0.27	5.71	0.32	9.18	0.44	5.60	0.28
	c3	432	3.70	0.24	4.05	0.25	5.67	0.31	9.19	0.42	5.72	0.28
$\sigma\pi$ - <i>JK</i>	c0	125	7.99	0.27	8.45	0.28	12.92	0.36	17.86	0.48	11.85	0.22
	c1	250	3.90	0.24	4.56	0.25	6.42	0.37	11.84	0.42	7.39	0.20
	c3	374	3.52	0.23	4.26	0.24	5.54	0.36	11.39	0.40	6.28	0.22
	c3	498	3.51	0.22	4.23	0.24	5.49	0.36	11.40	0.40	6.25	0.21

Table 5.5: RMS errors for models trained on the *combined* dataset with inclusion of decomposed energies. Conventions are as in Table 5.4.

A third general trend is that addition of decomposed energies does little to improve model performance and often substantially degrades performance. The initial disagreement between the LL and HL models is 5 to 20 times larger for the individual energy components than for the total energy. This indicates a large cancellation of errors occurs as the expectation values of the individual operators are summed to give the total energy. The weight of decomposed energies in Eq. 5.5 is set to $1/30^{th}$ that of E_{tot} . With this value, the model performance on E_{tot} of the training data is degraded by less than 20%. The question is then the degree to which inclusion of this substantial amount of additional data (Table 5.2) impacts transfer to other molecules. For the *ethane* dataset, comparison of Figs. 5.2 and 5.3 shows little benefit from including decomposed energies. In some cases, such as for the *2elec* policy without context, inclusion of decomposed energies leads to instabilities in the resulting model. The cases where including decomposed energies improves transfer are for the $\sigma\pi$ policies. Similar results are seen for the *combined* dataset in Figs. 5.4 and 5.5 where again, inclusion of decomposed energies improves transfer primarily for the $\sigma\pi$ policies. However, the best performing models are those with the *2elec* and *JK* poli-

cies, for which inclusion of decomposed energies either does little to improve or substantially lowers transfer. This can be rationalized in terms of the larger number of parameters present in the $\sigma\pi$ policies (Table 5.4). Inclusion of decomposed energies is helpful for specifying these parameters, however, better performance is obtained by using a policy with fewer parameters.

The best performing model is that obtained from the *2elec* policy trained on the *combined* dataset without inclusion of decomposed energies. Fig. 5.6 shows the distribution of errors in E_{tot} for the training and test datasets. The distributions show that the pLL model has low systematic error such that the RMS errors reflect primarily the width of the distribution. Outliers are also present for both the initial LL model and the trained pLL models. Removal of these outliers would lower the reported RMS errors by approximately 0.3 to 0.5 kcal mol⁻¹, for the well-performing models.

Dataset	Weight (kcal mol ⁻¹) ⁻¹	Train			Validation			Propane			tButane			propylene		
		E_{tot}	E_{chg}	E_{orb}	E_{tot}	E_{chg}	E_{orb}	E_{tot}	E_{chg}	E_{orb}	E_{tot}	E_{chg}	E_{orb}	E_{tot}	E_{chg}	E_{orb}
<i>ethane</i>	0.016	2.55	4.15	0.18	2.57	4.14	0.23	4.31	5.90	0.27	8.72	7.32	0.37	-	-	-
	0.16	2.42	3.98	0.18	2.48	3.98	0.23	4.32	5.74	0.26	8.58	7.13	0.35	-	-	-
	0.20	2.19	3.50	0.16	2.24	3.49	0.20	3.99	5.06	0.22	7.58	6.33	0.30	-	-	-
	0.27	2.53	3.31	0.16	2.51	3.31	0.20	4.05	4.74	0.22	7.45	5.98	0.30	-	-	-
	0.40	2.71	3.09	0.17	2.93	3.10	0.19	4.18	4.32	0.25	7.22	5.51	0.33	-	-	-
	0.80	2.81	2.98	0.20	3.40	3.01	0.22	4.86	4.07	0.29	7.02	5.17	0.36	-	-	-
<i>combined</i>	0.016	3.00	4.27	0.14	3.52	4.51	0.17	3.84	4.86	0.35	7.18	6.11	0.66	3.91	5.46	0.26
	0.16	3.09	4.25	0.15	3.65	4.46	0.17	3.92	4.80	0.44	7.36	6.03	0.76	4.15	5.40	0.31
	0.20	3.78	3.53	0.15	4.14	3.72	0.15	5.15	4.57	0.45	10.20	5.83	0.88	5.43	4.93	0.34
	0.27	4.00	3.38	0.17	4.45	3.56	0.18	5.49	4.57	0.51	11.59	5.89	1.17	22.60	5.11	0.56
	0.40	3.99	3.36	0.16	4.47	3.54	0.18	5.59	4.57	0.52	11.88	5.91	1.19	24.84	5.11	0.58
	0.80	4.00	3.28	0.17	4.53	3.46	0.19	6.09	4.59	0.57	13.99	6.02	1.39	89.12	5.51	1.04

Table 5.6: Effects of increasing the weighting of E_{chg} in the objective of Eq. 5.5, for pLL models using the *2elec* policy with full context trained without inclusion of decomposed energies.

The effects of increasing the weighting of E_{chg} in the objective of Eq. 5.5 is examined in Table 5.6, for the best performing model. For training on the *combined* dataset, increasing the weight by an order of magnitude from the default value of 0.016 (kcal mol⁻¹)⁻¹ leads to an improved description of the charge distribution, as measured by E_{chg} , at the expense of E_{tot} . According to the arguments accompanying Eq. 1.3, an improved description of the charge distribution should lead to improvements in the longer-range interactions. However, the performance of the model on the larger test molecules degrades with increased weighting of E_{chg} . This suggests

the longer-range interactions are not the primary source of error in those systems. A somewhat different behavior is observed for the smaller *ethane* dataset. Here, as the the weighting of E_{chg} is increased, the training error in E_{tot} first decreases and then increases. This suggests that additional information regarding interaction of the charge density with the external charges aids the training process at intermediate weights and, only at large weights, does improved performance on E_{chg} come at the expense of degraded performance on E_{tot} . Similar behavior is seen for the validation dataset, which includes different geometries of the training molecules. For the larger test molecules, increasing the weight of E_{chg} improves the performance on E_{tot} . These observations can be rationalized by viewing the interaction with external charges as simply additional data. For the smaller *ethane* dataset, this additional data aids the training. For the larger *combined* dataset, placing increased weight on this additional data degrades performance.

5.7 Discussion

The goal of this work is to develop flexible and systematically-improvable means to take advantage of molecular similarity in quantum chemical computations. The approach explored here embeds parameters in a LL *ab initio* Hamiltonian and adjusts these to obtain agreement with predictions of a HL *ab initio* Hamiltonian. This approach bridges between the flexible models of machine learning, such as neural nets, and the model Hamiltonians of SEQC. Model forms based on quantum chemical Hamiltonians may have advantages that stem from being more closely related to chemical phenomena than generic forms such as neural nets. For example, models trained on small molecules may incorporate sufficient information regarding molecular fragments as to be applicable to larger systems. Here, models trained on ethane and ethylene transfer reasonably well to propane, propylene and butane. In addition, models trained on ethane and ethylene perform better than models trained on ethane alone. This indicates that the benefits gained from including the additional ethylene data in the training override the challenges associated with using a single model to describe a more diverse class of molecules.

The general approach of embedding parameters in a LL *ab initio* Hamiltonian provides a flexible approach to model construction. Considerable flexibility may stem from using different LL Hamiltonians. However, the current work considers only different schemes, or policies, for embedding scaling factors in a minimal basis *ab initio* Hamiltonian. The policies differ along two dimensions. One dimension is whether different scaling factors are used for σ versus π interactions between p orbitals, indicated as $\sigma\pi$ policies. For the datasets considered here, better performance is obtained when σ versus π interactions were not treated separately. The other dimension along which the policies differ is whether scaling factors were applied directly to the two-electron matrix elements, $2elec$, or to the matrix elements of the J and K operators in the Fock matrix, JK . The $2elec$ policy led to somewhat better performance than the JK policy. That the best performing policy, $2elec$, is the policy with the fewest parameters may indicate that the parameters in the other policies are underdetermined (Table 5.4). If this is the case, the other policies may perform better on datasets involving larger classes of molecules.

Another means through which the model is made flexible and improvable is by making the scaling factors functions of context variables that capture the environment of the molecular fragment. Here, the model is first trained with no context dependence. A first level of context is added by making the scaling factors linear functions of bond lengths. This leads to substantial performance enhancements. Two additional levels of context are added that relate to the electronic structure, via the charge and bond order predicted by the unparametrized LL model. Depending on the policy and training dataset, these additional levels can lead to additional small improvements. Context variables are a means to add considerable flexibility to the model, but taking better advantage of this flexibility may require discovery of variables that better describe the molecular context.

Extending the types of data used to train the model is a potential means to improve model performance. One source of additional data is the interaction with external charges. This provides information on the electron density that is comparable across LL and HL models. Inclusion of this additional data leads to small improvements for the *ethane* dataset but not for the *combined*

dataset. Another source of additional data is decomposition of the energy by operator. Inclusion of such data improves performance for $\sigma\pi$ policies in some cases, but tends to degrade performance for the other policies. This is consistent with the above conclusion that the larger number of parameters in the $\sigma\pi$ policies leads to an underdetermined model. The improvement in the $\sigma\pi$ policies resulting from inclusion of decomposed energies is, however, not sufficient to make them competitive with the other policies. Decomposition of the energy by operator is therefore not found to lead to benefits in model training.

Decomposition of the energy by operator has the advantage of being uniquely defined, such that it can be applied unambiguously in both the LL and HL models. Decomposition of the energy by molecular fragment may provide information more relevant to the training, since the embedded parameters are associated with local interactions on and between bonded atoms. However, such decompositions are not unique and require division of the electron density in a manner that is compatible across the LL and HL models. [81, 113]

A means through which future work may increase the flexibility of the pLL models is by relaxing some constraints present in the current approach to applying scaling factors. In particular, the p orbitals on an atom are currently treated as equivalent such that, for example, a single scaling factor is used for the p sub-block of the kinetic energy operator on a heavy atom. When the Quambo method is used to transform the high-level electron densities to a minimal basis form, the resulting p orbitals do not retain this equivalency. [81] The quambo orbitals are well localized on individual sites, this may allow the energies to be decomposed into atom and bond contributions. This would substantially increase the amount of data used to train the model, and also bring in more detailed information on the nature of the fragments

Addition of parametrized core-core potentials is an additional avenue through which the model can be made more flexible. The current pLL models do not include such terms because the core electrons are included explicitly using basis functions that are identical to those in the HL model. In SEQC and DFTB, core-core terms substantially enhance the accuracy. A possible approach is to follow the above training of the pLL model with an additional stage that

trains only core-core terms to reproduce HL molecular forces. This would be analogous to the DFTB approach, which first derives the electronic Hamiltonian and then fits the core-core potentials. [43, 45, 46, 47] The distinction being that here both the electronic Hamiltonian and the core-core potentials would be obtained empirically.

In the current work, the parameters in the pLL model were used only to compensate for errors arising from the use of a minimal basis set in the LL model. The ability of this approach to compensate for errors arising from the absence of electron correlation in a LL model is yet to be explored. The ability of parametrized LL models to reduce errors in the original LL model by over 75%, for molecules larger than those included in the training data, suggests this approach has promise for using molecular similarity to reduce the computational cost of quantum chemical computations.

The results in the previous chapter suggested that charge and bond order are useful context variables. This may be because the data set studied in that chapter included molecules placed in strongly perturbing electrostatic environments. Here, where weaker external environments are employed, those context variables did not significantly improve performance. In future work, it may be useful to explore alternative context variables, such as those related to local geometry, e.g. bond angles.

For future work, it is also critical to test the resulting models on much larger molecules. This is necessary to fully evaluate the potential of this approach to address challenging computational problems. The results presented in this chapter lay the ground work for such studies and provide initial results.

Chapter 6

Accelerating Hartree Fock convergence

6.1 Introduction

Many computational methods in science and engineering use fixed-point iteration ($x_{i+1} = f(x_i)$ for $i = 0, 1, 2, \dots$) to attempt to find fixed points of f , i.e., points at which $f(x) = x$. These methods generate a trajectory of iterates, $x_0, x_1, x_2, \dots$, that ideally converge to a fixed point x . This chapter explores the application of imitation learning to fixed-point iteration to accelerate convergence and improve stability. Imitation learning, also called learning from demonstration, is designed to learn a policy that imitates an expert’s method of performing a target task: we gather training data by posing task instances to the expert, and train a function approximator to mimic the expert’s mapping from situations to actions.

The most straightforward way to apply imitation learning to fixed-point iteration would be to treat our original update function as the expert: train a cheaper function approximator to imitate the original, more expensive update. (There are two reasonable variants of this approach: in the first we need only a trajectory from our original fixed-point iteration, and we train our learner to produce the i th point on this trajectory at step i , no matter what the current point x_i is. In the second variant we need black-box access to the original update function f , and we train our learner to produce $f(x_i)$ when the current point is x_i .) [114] demonstrate that this type of

imitation learning can be successful in calculating fixed points, specifically in the context of belief propagation for computing marginals of a posterior distribution.

However, scientific and engineering applications often have two important differences from typical uses of imitation learning, making the straightforward approach problematic. First, to achieve a useful speedup, we do not need to completely eliminate calls to the original update function; it is enough to reduce their frequency. Second, considerable effort has often gone into developing and fine-tuning the original update function; simply replacing it with a naive function approximator would throw away much of this effort. In particular, the original update function often encodes physical constraints and intuitions that would be difficult to enforce with a naive function approximator; dropping these constraints and intuitions would lead to solutions that are not acceptable to domain experts, even if we were able to maintain or improve overall approximation error.

So instead of completely replacing the original update function, we propose to train our imitation learner to skip some of the steps from our original sequence of iterates: given the *output* of the update function at steps $1 \dots t$, the learner tries to predict a good *input* for the update function at step $t + \Delta$. By doing so, we replace two or more calls to the expensive update function with a single call, saving computation; but we still only task our learner with the easier job of producing a good input to the update function instead of mimicking its output. The difference is that, unlike the output, the input need not respect physical constraints or intuitions: we assume that the original update function itself is responsible for enforcing these, and is capable of repairing small violations.

As a test case for our methodology, we accelerate the Hartree-Fock method from computational quantum chemistry. Hartree-Fock is a fixed-point algorithm that approximates the electronic structure and energy of molecules. It is a specific instance of a broader class of mean-field-theory approaches to many-body problems. In such mean-field theories, the effect of all individuals on any given individual is approximated by a single averaged effect, thus reducing a many-body problem to a one-body problem. In quantum chemistry, the mean field arises from

the averaged charge distribution of all electrons, as described by the electron density, $\rho(\mathbf{r})$. The Hartree-Fock iterations thereby generate a sequence of electron densities that ideally converge to a fixed point. Quantum chemical methods that go beyond mean-field theory typically begin with the results of a Hartree-Fock calculation, making the Hartree-Fock algorithm a pervasive component of quantum chemistry [115]. Portions of this chapter has been published. [116]

6.2 Background

6.2.1 Hartree-Fock

The core problem of quantum chemistry is to compute the distribution of electrons in a molecule given the positions of its nuclei. The Hartree-Fock method efficiently approximates this distribution. From a computational standpoint, its relevant properties are:

- It is a fixed-point algorithm, whose iterates are single-electron density functions $\rho(\mathbf{r})$, represented as symmetric matrices, ρ , using a fixed set of basis functions $\{\chi_i(\mathbf{r})\}_{i=1}^{N_{\text{basis}}}$.
- The basis set size N_{basis} is typically at least linear in the number of nuclei in the molecule we are considering.
- Each iteration of Hartree-Fock is expensive: among other operations, we must iterate over all 4-tuples of basis functions to construct our mean-field approximation (nominally taking $O(N_{\text{basis}}^4)$ time, although scalings closer to $O(N_{\text{basis}}^3)$ are routinely achieved by taking advantage of sparsity [115]).

Hartree-Fock is shown as Algorithm 1. The lines inside the **while** loop constitute the fixed point update, mapping ρ_i to ρ_{i+1} . A number of extensions to the base Hartree-Fock method have been proposed; most relevant to the current study is an approach called Direct Inversion in the Iterative Subspace, or DIIS [82]. In DIIS, we no longer take the output of iteration t directly as the input to iteration $t + 1$. Instead, at the beginning of iteration t , we build an input density matrix ρ'_t as a linear combination of the past several output density matrices $\rho_t, \rho_{t-1}, \rho_{t-2}, \dots$,

and use ρ'_t to calculate the Fock operator $\hat{F}$ in the first line of the **while** loop. The coefficients of the linear combination are chosen by minimizing some measure of the error of the resulting iterates. Various methods have been used to define the error measure and to find the weights that minimize it [83, 117, 118]. However, identifying weights that lead to the fastest and most stable convergence remains an open question [85, 119]. The relevance of DIIS is that its overall

Data: Coordinates of atomic nuclei; basis set χ_i ; number of electrons N ; initial density matrix ρ_0 ; termination criterion $\delta > 0$
Result: Density matrix with approximately-minimum energy
 Calculate overlap matrix: $S_{ij} \leftarrow \langle \chi_i, \chi_j \rangle$
 Initialize $t \leftarrow 0$
while $t = 0$ or $|E_t - E_{t-1}| > \delta$ **do**
 Calculate the Fock matrix $F \leftarrow F(\rho_t)$
 Solve generalized eigenproblem $Fc = \epsilon Sc$ for eigenpairs c_a, ϵ_a
 Build matrix C : stack eigenvectors c_a side by side, two copies of each, starting from lowest energy ϵ_a , until we have added N columns
 Update density matrix $\rho_{t+1} = CC^\top$
 Update energy $E_{t+1} \leftarrow$ sum of ϵ_a for columns of C
 $t \leftarrow t + 1$
end

Algorithm 1: Hartree-Fock

form is the same as that of our proposed method: we train a function approximator (in this case linear regression) to produce a good input to our original fixed-point update step (in this case the Hartree-Fock update). In contrast to existing DIIS methods, however, we view the problem of training our function approximator as one of imitation learning: we explicitly take into account the effect of our learned weights, not just on the immediate error, but on the overall degree to which the fixed-point iteration tracks the behavior of our desired fixed-point iteration. By so doing, we hope to achieve faster and more stable convergence than existing methods.

6.2.2 Imitation learning

In imitation learning, we wish to discover how to solve a sequential decision problem by using expert demonstrations. We are given access to the expert’s policy π^* , which is a (possibly ran-

domized) mapping from states of the world s to actions $a = \pi^*(s)$. And, we are given access to a world model M : for any state s and action a , M is a (possibly randomized) mapping to the next state $s' = M(s, a)$. Using the policy and the world model we can sample training trajectories: for each trajectory, we start in a designated start state $s_0 = \text{start}$, execute action $a_0 = \pi^*(s_0)$, transition to state $s_1 = M(s_0, a_0)$, and repeat for T steps. We can then use these trajectories to train a predictor that, given the current state, predicts the expert's action. The key difficulty in imitation learning is that prediction errors early in a trajectory can cause us to deviate from the expert's distribution over states; so, we can easily drift into areas of the state space where we have little or no training data, causing a growing cascade of errors.

In our experiments, we use the DAgger algorithm for imitation learning [86]. DAgger (Algorithm 2) avoids the above difficulty by using its current hypothesized policy at each iteration to gather additional training trajectories, so that it gains experience on how to correct its own errors.

Data: Policy class Π , expert's policy π^* , horizon T , world model M

Result: Best $\hat{\pi}_j$ on validation

Initialize $D \leftarrow \emptyset$

Initialize $\hat{\pi}_1$ to any policy in Π

for $j=1$ **to** N **do**

 Let $\pi_j = \beta_j \pi^* + (1 - \beta_j) \hat{\pi}_j$

 Sample a trajectory $s_0, s_1, \dots, s_T$ using π_j and M

 Construct dataset $D_j = \{(s_t, \pi^*(s_t))\}$ of visited states and their expert actions

 Aggregate datasets: $D \leftarrow D \cup D_j$

 Train policy $\hat{\pi}_{j+1}$ by supervised learning on D

end

Algorithm 2: DAgger

6.3 Learning the Policy

To apply imitation learning methods like DAgger to speed up fixed-point iterations like Hartree-Fock, we need to specify three things: the world model M , the policy class Π , and the expert policy π^* . For our target problem of accelerating fixed-point updates, the world model is simple:

Table 6.1: Example: applying DAgger on expert demonstrations with step size = 2

			Hartree-Fock iterations (with step size = 2)				
DAgger iterations	iter		iter 1	iter 2	iter 3	...	iter $x = \frac{n}{2}$
	1	Objective	$(\rho_0) \rightarrow \rho_2$	$(\rho_0, \rho_2) \rightarrow \rho_4$	$(\rho_0, \rho_2, \rho_4) \rightarrow \rho_6$	...	$(\rho_{2i})_{i=0}^{x-1} \rightarrow \rho_n$
		Result	$(\rho_0) \rightarrow \rho'_2$	$(\rho_0, \rho_2) \rightarrow \rho'_4$	$(\rho_0, \rho_2, \rho_4) \rightarrow \rho'_6$	...	$(\rho_{2i})_{i=0}^{x-1} \rightarrow \rho'_n$
	2	New objective		$(\rho_0, \rho'_2) \rightarrow \rho_4$	$(\rho_0, \rho_2, \rho'_4) \rightarrow \rho_6$	...	$((\rho_{2i})_{i=0}^{x-2}, \rho'_{2(x-1)}) \rightarrow \rho_n$
		Result		$(\rho_0, \rho'_2) \rightarrow \rho''_4$	$(\rho_0, \rho_2, \rho'_4) \rightarrow \rho''_6$	...	$((\rho_{2i})_{i=0}^{x-2}, \rho'_{2(x-1)}) \rightarrow \rho''_n$
	3	New objective			$(\rho_0, \rho'_2, \rho''_4) \rightarrow \rho_6$	...	$((\rho_{2i})_{i=0}^{x-3}, (\rho_{2i}^{[i-(x-3)]})_{i=x-2}^{x-1}) \rightarrow \rho_n$
		Result			$(\rho_0, \rho'_2, \rho''_4) \rightarrow \rho'''_6$	...	$((\rho_{2i})_{i=0}^{n-3}, (\rho_{2i}^{[i-(x-3)]})_{i=x-2}^{x-1}) \rightarrow \rho'''_n$
	$\vdots$	$\vdots$				$\ddots$	$\vdots$
	$x = \frac{n}{2}$	New objective					$(\rho_{2i}^{[i]})_{i=0}^{x-1} \rightarrow \rho_n$
		Result					$(\rho_{2i}^{[i]})_{i=0}^{x-1} \rightarrow \rho_n^{[x]}$

our state is the history of iterates (e.g., density matrices) we have visited so far, and our action is to append another iterate to this history.

We have already mentioned our policy class: at each iteration i of our fixed-point calculation, we apply a function approximator with parameter vector $c^{(i)}$ to the history of past iterates, and then we pass the output of the function approximator into the original fixed-point update (e.g., the base Hartree-Fock update). In our experiments, the function approximator is a linear combination: $c^{(i)}$ is a vector of weights on previous density matrices.

Finally, there are two reasonable choices for an expert policy. In both cases we pick a step size Δ ; recall that the goal is to replace Δ calls to the original update function with just a single call.

The simpler choice of expert policy is to work from one or more expert demonstrations—in our case, each demonstration is a sequence of density matrices $\rho_0, \rho_1, \rho_2, \dots$. At step i we train our function approximator to predict $\rho_{i+\Delta}$ of the expert, no matter what the current iterate is. In addition to simplicity, this choice of expert has the advantage that we can gather expert demonstrations however we like: we don't have to be able to make calls to the expert at training time. We use this choice of expert policy in our experiments below.

The second choice of expert policy is to use Δ calls to our original update function f : e.g.,

if the current iterate is ρ and $\Delta = 2$, then the expert predicts that we should move to $f(f(\rho))$. This choice of expert has the possible advantage that it could adapt better to mistakes by the learner. However, in our application it is substantially more expensive, since it requires calling the update function repeatedly during training; so, we defer experiments with this expert policy to future work.

We introduce the notation $(\rho_i, \rho_j, \dots) \rightarrow \rho_k$ to represent: (i) creating a linear combination of the density matrices $\rho_i, \rho_j, \dots$ and (ii) carrying out one iteration of the Hartree-Fock algorithm to generate the next iterate ρ_k . We use a different set of coefficients for each step of the trajectory; write $\hat{c}_i$ for the coefficients at step i . The training data is a set of molecular examples, and we train the coefficients to minimize a regularized sum of squared errors over these examples, subject to some constraints (details in Section 6.4).

The training process is visualized in Table 6.1, in which Hartree-Fock iterations are shown as columns and Dagger iterations are shown as rows. We begin by training a policy for the first iteration of Hartree-Fock. In this case DAgger has only one iteration, in which a policy is trained on the objective $(\rho_0) \rightarrow \rho_2$. The learned policy uses the coefficients $\hat{c}_1^{[1]}$, where the superscript indicates DAgger iteration and the subscript indicates Hartree-Fock iteration. The density matrices generated from this learned policy are referred to as ρ'_2 , where the number of primes indicates the DAgger iteration.

The training process then moves onto the second Hartree-Fock iteration, which has two DAgger iterations. In the first DAgger iteration, a policy is trained on $(\rho_0, \rho_2) \rightarrow \rho_4$. The learned policy uses coefficients $\hat{c}_2^{[1]}$ and generates induced states ρ'_4 . The second DAgger iteration uses an objective, symbolized $(\rho_0, \rho'_2) \rightarrow \rho_4$, that includes states, ρ'_2 , induced from the learned policy of the previous Hartree-Fock iteration. For each molecular example, the objective selects between the expert density, ρ_2 , and the induced one, ρ'_2 , with probability β_2 . (β_j of Algorithm 2 is set to 0.5^{j-1} : sampling of expert states decreases with DAgger iteration.) The resulting learned policy uses $\hat{c}_2^{[2]}$ and generates induced states, ρ''_4 . At this point, there are no additional induced states to include in the training and so the DAgger iteration terminates, leading to the upper triangular

structure of Table 6.1.

In the third Hartree-Fock iteration, the DAgger iterations train objectives $(\rho_0, \rho_2, \rho_4) \rightarrow \rho_6$, $(\rho_0, \rho_2, \rho'_4) \rightarrow \rho_6$, and $(\rho_0, \rho'_2, \rho''_4) \rightarrow \rho_6$, respectively, such that each DAgger iteration samples induced states earlier in the trajectories.

6.4 Experimental Design

The computational experiments use the approach of Section 6.3 to train a policy for accelerating Hartree-Fock and compare the results to some baseline approaches. Of particular interest is the degree to which a policy trained on one class of molecules can transfer to a different class of molecules.

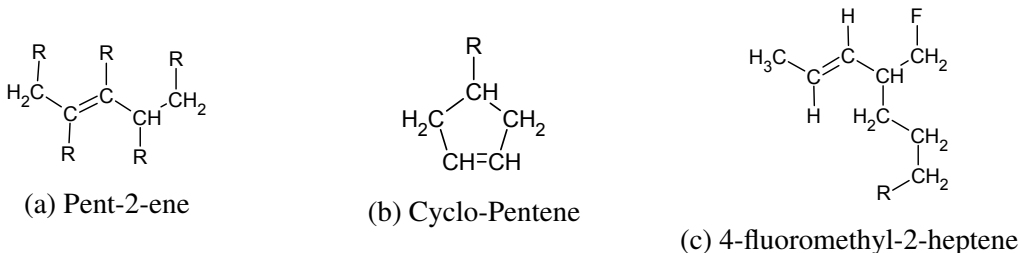


Figure 6.1: The chemical structure of molecules in the datasets, R = H, F, OH or NH₂

We generated the following three data sets, with the first being used for training. Each data set consists of a set of molecules, as defined by the bonding pattern between the atoms, and a set of distinct geometries of these molecules, as defined by distortions of the structure away from the equilibrium geometry. The geometric distortions are generated using a random uniform distribution of $\pm 0.5 \text{ \AA}$ for bond lengths and $\pm 10^\circ$ for bond angles. This approach of uniform sampling generates highly distorted structures. To prevent inclusion of structures that are of little interest in chemical applications, structures are rejected if there are contacts closer than $3 \text{ \AA}$ between non-bonded atoms. In addition, each molecular configuration is placed in 4 different electrical field environments (1 with no field + 3 different fields for X, Y and Z directions).

pent2ene 15 unique molecules corresponding to single substitution of pent-2-ene: we place a

single substituent at one of the positions indicated by R in Figure 6.1a, with all other R’s being hydrogen (-H). The substituents are fluorine (-F), hydroxyl (-OH) and amine (-NH₂). The training set includes, for each molecule, the equilibrium geometry and three distorted geometries. Distortion includes a random free rotation about the leftmost carbon-carbon single bond of Figure 6.1a.

cycloPentene Includes cyclo-pentene and its three singly-substituted analogues (Figure 6.1b), each in the equilibrium geometry and 4 distorted geometries.

fluoroMethylHeptene Includes the three singly-substituted species of 4-fluoromethyl-2-heptene (Figure 6.1c), each in its equilibrium geometry and 7 distorted geometries. Distortion includes a random free rotation about the leftmost carbon-carbon single bond of Figure 6.1c.

For each molecular instance, we use a heuristic approach to generate a high-quality expert demonstration. We start from a steady-state density matrix, ρ_n , obtained from a standard fixed-point algorithm [82]. We first train a policy that takes the initial density matrix to the final density matrix directly, $(\rho_0) \rightarrow \rho_n$, and use this policy to generate ρ_1 . We next train a policy on $(\rho_0, \rho_1) \rightarrow \rho_n$ and use it to generate ρ_2 , and so on. Expert demonstration is constructed from using the densities from DIIS [82]. To further expand the training data, expert demonstrations are constructed starting from two starting points, $\rho_0 = 0$ and $\rho_0 = \mathbf{I}$. In our experiments we compare the learned DAgger policy both to this “IL-final” policy and to the “DIIS” policy.

In training a policy to the objective, $(\rho_a, \rho_b, \dots) \rightarrow \rho_c$, the error is defined as the sum of the distance of the density matrix, $\|\rho - \rho_c\|$, and the molecular energy, $|E(\rho) - E(\rho_c)|$ in Hartrees, from the target. We optimize the policy coefficients using the trust-region reflective algorithm [108] as implemented within Matlab [109]. Since we are assuming only black-box access to the expert (and since the Hartree-Fock update step contains a number of operations, such as eigenvalue calculations, that are difficult to differentiate analytically), we computed gradients for the optimizer using finite differences.

To help ensure the stability of our learned policy, we constrain the coefficients c_i to lie in

$[0, 1]$. As in DIIS methods, it is likely desirable to have the coefficients sum to one [84], which we promote by adding regularizers to the objective, one for each step i of the Hartree-Fock iteration:

$$R^{(i)} = w(\mathbf{1}^\top \hat{c}^{(i)} - 1)^2 \quad (6.1)$$

where the weight w was empirically adjusted to a value of 30.

6.5 Results

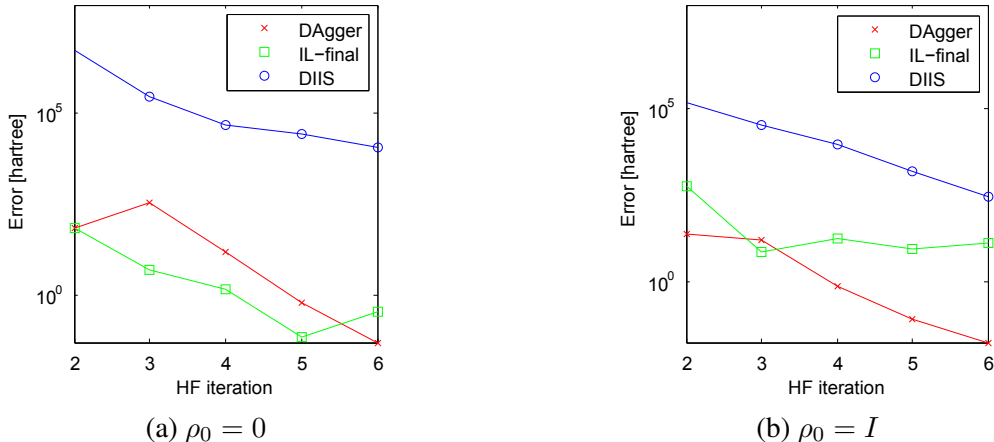


Figure 6.2: Average error in energy as a function of iteration tested on *pent2ene* test dataset

The error from “DAgger” and “IL-final” decreases more rapidly than the “DIIS” policy for the initial 6 iterations in all cases. Looking at the *pent2ene* test dataset and *cycloPentene* dataset (figure 6.2 and 6.3) “DAgger” performs better than “IL-final” by the 6th iteration. Testing on the *fluoroMethylHeptene* dataset containing larger molecules (figure 6.4) “IL-final” has a lower error throughout the iterations, however, “DAgger” is not far off.

These results demonstrate that imitation learning can accelerate fixed-point iteration in a manner that generalizes to situations not included in the training data.

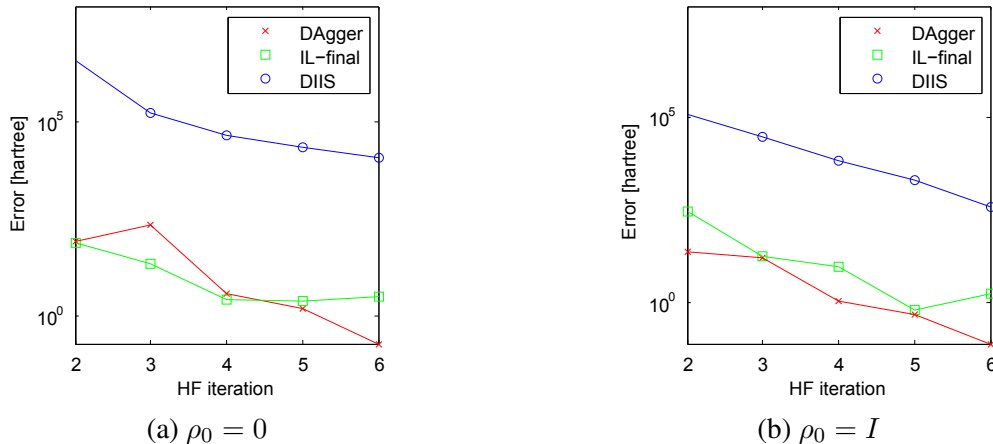


Figure 6.3: Average error in energy as a function of iteration tested on *cycloPentene* dataset

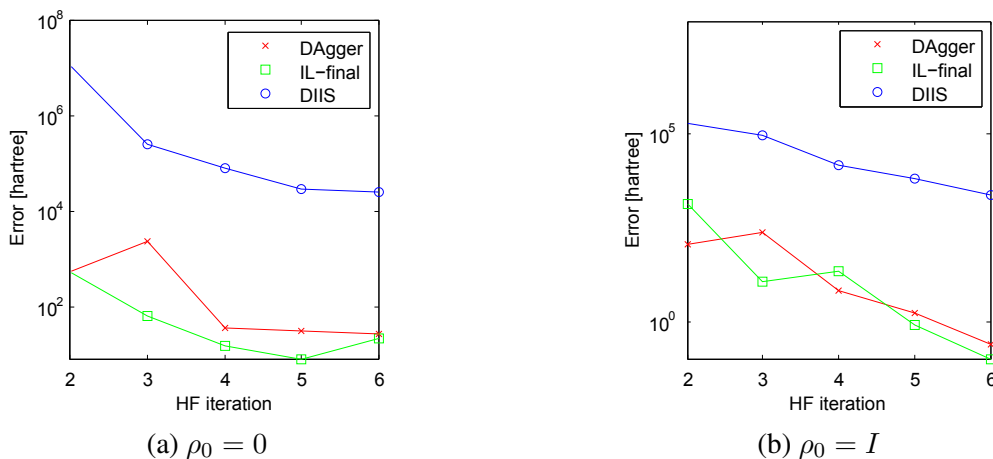


Figure 6.4: Average error in energy as a function of iteration tested on *fluoroMethylHeptene* dataset

6.6 Conclusion

The work explores a new application of imitation learning: accelerating the fixed-point iterations that are common in science and engineering. The mapping to imitation learning is quite general in that it requires only black-box access to the update function from an existing fixed-point algorithm. The intuition is simple: we accelerate convergence by attempting to skip $\Delta - 1$ out of every Δ calls to the original update function, replacing the skipped calls by a learned mapping. The specific case considered here, the Hartree-Fock algorithm from quantum chemistry, allowed

us to test the ability of policies trained on one situation (chemical system) to transfer to different situations (chemical systems). Our results indicate successful transfer: imitation learning leads to policies that transfer better between systems than competing approaches.

In future work, it may be useful to explore fragment based methods. In such an approach, blocks of the density matrix corresponding to different molecular fragments could be updated independently. This could improve transfer between molecules, because models trained on molecules composed of a certain set of fragments may transfer well to other molecules composed of those same fragments.

Another direction for future work is use of this approach in quantum chemical calculation of the minimum-energy geometry of a molecule. In such geometry optimizations, SCF calculations are run on the same molecule in a set of geometries that gradually converge to the minimum-energy geometry. A model trained on the results from the initial SCF calculations may transfer well to later SCF calculations.

Bibliography

- [1] F. Hund. History of quantum theory. 1980. URL http://inis.iaea.org/Search/search.aspx?orig_q=RN:13654573. 1.1
- [2] M. Karplus and R. N. Porter. *Atoms and Molecules. An introduction for students of physical chemistry*. Benjamin Cummings Publishing Company, 1971. 1.1
- [3] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications, 1996. 1.1
- [4] P.W. Atkins and R. Friedman. *Molecular Quantum Mechanics (4th Edit)*. 2005. 1.1
- [5] Kostas Gavroglu and Ana Simões. *Neither Physics nor Chemistry: A History of Quantum Chemistry*. 2011. 1.1
- [6] John A. Pople, Martin Head-Gordon, Douglas J. Fox, Krishnan Raghavachari, and Larry A. Curtiss. Gaussian-1 theory: A general procedure for prediction of molecular energies. *The Journal of Chemical Physics*, 90(10):5622, 1989. ISSN 00219606. doi: 10.1063/1.456415. URL <http://adsabs.harvard.edu/abs/1989JChPh..90.5622P>. 1.1
- [7] Larry A. Curtiss, Christopher Jones, Gary W. Trucks, Krishnan Raghavachari, and John A. Pople. Gaussian-1 theory of molecular energies for second-row compounds. *The Journal of Chemical Physics*, 93(4):2537, August 1990. ISSN 00219606. doi: 10.1063/1.458892. URL <http://scitation.aip.org/content/aip/journal/jcp/93/4/10.1063/1.458892>. 1.1

- [8] L A Curtiss, K Raghavachari, G W Trucks, and J A Pople. Gaussian-2 Theory for Molecular-Energies of 1st-Row and 2nd-Row Compounds. *Journal of Chemical Physics*, 94(11):7221–7230, 1991. URL <GotoISI>://A1991FL83700029. 1.1
- [9] LA Curtiss and K Raghavachari. Gaussian-2 theory for molecular energies of first-and second-row compounds. *The Journal of chemical physics*, 94:7221–7230, 1991. URL <http://link.aip.org/link/?JCPSA6/94/7221/1>. 1.1
- [10] Larry A. Curtiss, Krishnan Raghavachari, Paul C. Redfern, Vitaly Rassolov, and John A. Pople. Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *The Journal of Chemical Physics*, 109(18):7764, November 1998. ISSN 00219606. doi: 10.1063/1.477422. URL <http://scitation.aip.org/content/aip/journal/jcp/109/18/10.1063/1.477422>. 1.1
- [11] James B. Foresman, Martin Head-Gordon, John A. Pople, and Michael J. Frisch. Toward a systematic molecular orbital theory for excited states. *The Journal of Physical Chemistry*, 96(1):135–149, January 1992. ISSN 0022-3654. doi: 10.1021/j100180a030. URL <http://dx.doi.org/10.1021/j100180a030>. 1.1
- [12] Filipp Furche and Reinhart Ahlrichs. Adiabatic time-dependent density functional methods for excited state properties. *The Journal of Chemical Physics*, 117(16):7433, October 2002. ISSN 00219606. doi: 10.1063/1.1508368. URL <http://scitation.aip.org/content/aip/journal/jcp/117/16/10.1063/1.1508368>. 1.1
- [13] Roberto Improta, Vincenzo Barone, Giovanni Scalmani, and Michael J. Frisch. A state-specific polarizable continuum model time dependent density functional theory method for excited state calculations in solution. *The Journal of Chemical Physics*, 125(5):054103, August 2006. ISSN 00219606. doi: 10.1063/1.2222364. URL <http://scitation.aip.org/content/aip/journal/jcp/125/5/10.1063/1.2222364>. 1.1
- [14] Benjamin J. Lynch and Donald G. Truhlar. How Well Can Hybrid Density Functional Methods Predict Transition State Geometries and Barrier Heights? *The Jour-*

- nal of Physical Chemistry A*, 105(13):2936–2941, April 2001. ISSN 1089-5639. doi: 10.1021/jp004262z. URL <http://dx.doi.org/10.1021/jp004262z>. 1.1
- [15] Vildan Guner, Kelli S. Khuong, Andrew G. Leach, Patrick S. Lee, Michael D. Bartberger, and K. N. Houk. A Standard Set of Pericyclic Reactions of Hydrocarbons for the Benchmarking of Computational Methods: The Performance of ab Initio, Density Functional, CASSCF, CASPT2, and CBS-QB3 Methods for the Prediction of Activation Barriers, Reaction Energetics, an. *The Journal of Physical Chemistry A*, 107(51): 11445–11459, December 2003. ISSN 1089-5639. doi: 10.1021/jp035501w. URL <http://dx.doi.org/10.1021/jp035501w>. 1.1
- [16] Yan Zhao and Donald G. Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theoretical Chemistry Accounts*, 120(1-3):215–241, July 2007. ISSN 1432-881X. doi: 10.1007/s00214-007-0310-x. URL <http://link.springer.com/10.1007/s00214-007-0310-x>. 1.1
- [17] Wolfram Koch and Max C. Holthausen. *A Chemist’s Guide to Density Functional Theory*. Wiley-VCH Verlag GmbH, 2001. 1.1
- [18] Hossein Pasha Ebrahimi and Mohsen Tafazzoli. Probing theoretical level effect on fluorine chemical shielding calculations. *Concepts in Magnetic Resonance Part A*, 42(4):140–153, 2013. ISSN 1552-5023. doi: 10.1002/cmr.a.21271. URL <http://dx.doi.org/10.1002/cmr.a.21271>. (document), 1.1
- [19] E Prodan and W Kohn. Nearsightedness of electronic matter. *Proceedings of the National Academy of Sciences of the United States of America*, 102:11635–11638, 2005. ISSN 0027-8424. doi: DOI10.1073/pnas.0505436102. URL <GotoISI>: //000231317000014. 1.1
- [20] Christopher A. White, Benny G. Johnson, Peter M.W. Gill, and Martin Head-Gordon. The

- continuous fast multipole method. *Chemical Physics Letters*, 230(12):8 – 16, 1994. ISSN 0009-2614. doi: [http://dx.doi.org/10.1016/0009-2614\(94\)01128-1](http://dx.doi.org/10.1016/0009-2614(94)01128-1). URL <http://www.sciencedirect.com/science/article/pii/0009261494011281>. 1.1
- [21] C A White, B G Johnson, P M W Gill, and M HeadGordon. Linear scaling density functional calculations via the continuous fast multipole method. *Chem. Phys. Lett.*, 253: 268–278, 1996. ISSN 0009-2614. URL <GotoISI>://A1996UL21900012. 1.1
- [22] Robert Polly, Hans-Joachim Werner *, Frederick R. Manby, and Peter J. Knowles. Fast hartreefock theory using local density fitting approximations. *Molecular Physics*, 102(21-22):2311–2321, 2004. doi: 10.1080/0026897042000274801. URL <http://dx.doi.org/10.1080/0026897042000274801>. 1.1
- [23] A Van der Vaart, V Gogonea, S L Dixon, and K M Merz. Linear scaling molecular orbital calculations of biological systems using the semiempirical divide and conquer method. *J. Comput. Chem.*, 21:1494–1504, 2000. ISSN 0192-8651. URL <GotoISI>://000165132900007. 1.1
- [24] Honghui Shang, Hongjun Xiang, Zhenyu Li, and Jinlong Yang. Linear scaling electronic structure calculations with numerical atomic basis set. *Int. Rev. Phys. Chem.*, 29(4):665–691, October 2010. ISSN 0144-235X. URL <http://www.tandfonline.com/doi/abs/10.1080/0144235X.2010.520454>. 1.1
- [25] Mark S. Gordon, Dmitri G. Fedorov, Spencer R. Pruitt, and Lyudmila V. Slipchenko. Fragmentation methods: A route to accurate calculations on large systems. *Chemical Reviews*, 112(1):632–672, 2012. doi: 10.1021/cr200093j. URL <http://dx.doi.org/10.1021/cr200093j>. PMID: 21866983. 1.1
- [26] Ryan M. Richard and John M. Herbert. A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory. *The Journal of Chemical Physics*, 137(6):064113, 2012. doi: <http://dx.doi.org/10.1063/1.4742816>. URL <http://scitation.aip.org/content/aip/journal/jcp/>

137/6/10.1063/1.4742816. 1.1

- [27] Rudolph Pariser and Robert G. Parr. A semiempirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. i. *The Journal of Chemical Physics*, 21(3), 1953. 1.1
- [28] J. A. Pople. Electron interaction in unsaturated hydrocarbons. *Trans. Faraday Soc.*, 49: 1375–1385, 1953. doi: 10.1039/TF9534901375. URL <http://dx.doi.org/10.1039/TF9534901375>. 1.1
- [29] J. A. Pople, D. P. Santry, and G. A. Segal. Approximate selfconsistent molecular orbital theory. i. invariant procedures. *The Journal of Chemical Physics*, 43(10), 1965. 1.1
- [30] J. A. Pople and G. A. Segal. Approximate selfconsistent molecular orbital theory. ii. calculations with complete neglect of differential overlap. *The Journal of Chemical Physics*, 43(10), 1965. 1.1
- [31] J Ridley and M Zerner. An intermediate neglect of differential overlap technique for spectroscopy: pyrrole and the azines. *Theor. Chim. Acta*, 32:111, 1973. 1.1, 1.3, 4.1, 5.3
- [32] Iffat H. Nayyar and Artm E. Masunov. Two-photon absorption spectra predicted by semiempirical methods. *Journal of Computational and Theoretical Nanoscience*, 11(10):2208–2220, 2014-10-01T00:00:00. doi: doi:10.1166/jctn.2014.3629. URL <http://www.ingentaconnect.com/content/asp/jctn/2014/00000011/00000010/art00022>. 1.1, 1.3
- [33] Alexander A. Voityuk. Indo/x: A new semiempirical method for excited states of organic and biological molecules. *Journal of Chemical Theory and Computation*, 10(11): 4950–4958, 2014. doi: 10.1021/ct500717u. URL <http://dx.doi.org/10.1021/ct500717u>. 1.1, 1.3
- [34] M J S Dewar and W Thiel. Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.*, 99:4899, 1977. 1.1

- [35] M J S Dewar, E G Zoebisch, E F Healy, and J J P Stewart. The development and use of quantum-mechanical molecular-models .76. Am1 - a new general-purpose quantum-mechanical molecular-model. *J. Am. Chem. Soc.*, 107:3902–3909, 1985. ISSN 0002-7863. URL <GoToISI>://A1985ALC3500024. 1.1, 5.3
- [36] M J S Dewar, C X Jie, and J G Yu. Sam1 - the 1st of a New Series of General-Purpose Quantum-Mechanical Molecular-Models. *Tetrahedron*, 49:5003–5038, 1993. ISSN 0040-4020. URL <GoToISI>://A1993LF27500003. 1.1, 1.3, 2.4.6
- [37] James JP Steward. Optimization of parameters for semiempirical methods. *Journal of Computational Chemistry*, 10:209–212, 1989. 1.1
- [38] Matthew P Repasky, Jayaraman Chandrasekhar, and William L Jorgensen. PDDG/PM3 and PDDG/MNDO: improved semiempirical methods. *J. Comput. Chem.*, 23(16):1601–22, December 2002. ISSN 0192-8651. URL <http://www.ncbi.nlm.nih.gov/pubmed/12395428>. 1.1
- [39] 1.1
- [40] Gerd B. Rocha, Ricardo O. Freire, Alfredo M. Simas, and James J. P. Stewart. Rm1: A reparameterization of am1 for h, c, n, o, p, s, f, cl, br, and i. *Journal of Computational Chemistry*, 27(10):1101–1111, 2006. ISSN 1096-987X. doi: 10.1002/jcc.20425. URL <http://dx.doi.org/10.1002/jcc.20425>. 1.1
- [41] Wolfgang Weber and Walter Thiel. Orthogonalization corrections for semiempirical methods. *Theoretical Chemistry Accounts*, 103(6):495–506, 2000. ISSN 1432-881X. doi: 10.1007/s002149900083. URL <http://dx.doi.org/10.1007/s002149900083>. 1.1
- [42] Walter Thiel. Semiempirical quantumchemical methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(2):145–157, 2014. ISSN 1759-0884. doi: 10.1002/wcms.1161. URL <http://dx.doi.org/10.1002/wcms.1161>. 1.1

- [43] M Elstner, D Porezag, G Jungnickel, J Elsner, M Haugk, T Frauenheim, S Suhai, and G Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58:7260–7268, 1998. ISSN 1098-0121. URL <GotoISI>://000076058800081. 1.1, 1.3, 1.3, 4.1, 5.7
- [44] M Elstner, T Frauenheim, E Kaxiras, G Seifert, and S Suhai. A self-consistent charge density-functional based tight-binding scheme for large biomolecules. *Phys. Status Solidi B*, 217:357–376, 2000. ISSN 0370-1972. URL <GotoISI>://000085329200018. 1.1, 1.3, 1.3
- [45] Qiang Cui, Marcus Elstner, Efthimios Kaxiras, Thomas Frauenheim, and Martin Karplus. A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method. *J. Phys. Chem. B*, 105(2):569–585, 2001. ISSN 15206106. doi: 10.1021/jp0029109. URL <http://pubs.acs.org/doi/abs/10.1021/jp0029109>. 1.1, 1.3, 1.3, 5.7
- [46] M Elstner. The SCC-DFTB method and its application to biological systems. *Theor. Chem. Acc.*, 116:316–325, 2006. ISSN 1432-881X. doi: DOI10.1007/s00214-005-0066-0. URL <GotoISI>://000239958600028. 1.1, 1.3, 1.3, 4.1, 5.7
- [47] Michael Gaus, Qiang Cui, and Marcus Elstner. DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.*, 7(4):931–948, April 2012. ISSN 1549-9618. URL <http://dx.doi.org/10.1021/ct100684s>. 1.1, 1.3, 1.3, 5.7
- [48] M S Gordon, J M Mullin, S R Pruitt, L B Roskop, L V Slipchenko, and J A Boatz. Accurate Methods for Large Molecular Systems. *J. Phys. Chem. B*, 113:9646–9663, 2009. ISSN 1520-6106. doi: Doi10.1021/Jp811519x. URL <GotoISI>://000268139600006. 1.1
- [49] LiHong Hu, XiuJun Wang, LaiHo Wong, and GuanHua Chen. Combined first-principles

- calculation and neural-network correction approach for heat of formation. *J. Chem. Phys.*, 119(22):11501, November 2003. ISSN 00219606. URL <http://scitation.aip.org/content/aip/journal/jcp/119/22/10.1063/1.1630951>. 1.1
- [50] Xue-Mei Duan, Guo-Liang Song, Zhen-Hua Li, Xiu-Jun Wang, Guan-Hua Chen, and Kang-Nian Fan. Accurate prediction of heat of formation by combining Hartree-Fock/density functional theory calculation with linear regression correction approach. *J. Chem. Phys.*, 121(15):7086–95, October 2004. ISSN 0021-9606. URL <http://scitation.aip.org/content/aip/journal/jcp/121/15/10.1063/1.1786582>. 1.1
- [51] Hui Li, LiLi Shi, Min Zhang, Zhongmin Su, XiuJun Wang, LiHong Hu, and Guan-Hua Chen. Improving the accuracy of density-functional theory calculation: the genetic algorithm and neural network approach. *J. Chem. Phys.*, 126(14):144101, April 2007. ISSN 0021-9606. URL <http://scitation.aip.org/content/aip/journal/jcp/126/14/10.1063/1.2715579>. 1.1
- [52] Jian Sun, Jiang Wu, Tao Song, LiHong Hu, KaiLu Shan, and GuanHua Chen. Alternative Approach to Chemical Accuracy: A Neural Networks-Based First-Principles Method for Heat of Formation of Molecules Made of H, C, N, O, F, S, and Cl. *J. Phys. Chem. A*, 118(39):9120–9131, October 2014. ISSN 1089-5639. URL <http://dx.doi.org/10.1021/jp502096y>. 1.1
- [53] Ting Gao, Li-Li Shi, Hai-Bin Li, Shan-Shan Zhao, Hui Li, Shi-Ling Sun, Zhong-Min Su, and Ying-Hua Lu. Improving the accuracy of low level quantum chemical calculation for absorption energies: the genetic algorithm and neural network approach. *Phys. Chem. Chem. Phys.*, 11(25):5124–9, July 2009. ISSN 1463-9076. URL <http://pubs.rsc.org/en/content/articlehtml/2009/cp/b812492b>. 1.1
- [54] Ting Gao, Shi-Ling Sun, Li-Li Shi, Hui Li, Hong-Zhi Li, Zhong-Min Su, and Ying-Hua Lu. An accurate density functional theory calculation for electronic excitation en-

- ergies: the least-squares support vector machine. *J. Chem. Phys.*, 130(18):184104, May 2009. ISSN 1089-7690. URL <http://scitation.aip.org/content/aip/journal/jcp/130/18/10.1063/1.3126773>. 1.1
- [55] Jianming Wu and Xin Xu. Improving the B3LYP bond energies by using the X1 method. *J. Chem. Phys.*, 129(16):164103, October 2008. ISSN 1089-7690. URL <http://scitation.aip.org/content/aip/journal/jcp/129/16/10.1063/1.2998231>. 1.1
- [56] Jianming Wu, Igor Ying Zhang, and Xin Xu. The X1s method for accurate bond dissociation energies. *Chemphyschem*, 11(12):2561–7, August 2010. ISSN 1439-7641. URL <http://www.ncbi.nlm.nih.gov/pubmed/20669214>. 1.1
- [57] Roman M Balabin and Ekaterina I Lomakina. Neural network approach to quantum-chemistry data: accurate prediction of density functional theory energies. *J. Chem. Phys.*, 131(7):074104, August 2009. ISSN 1089-7690. URL <http://scitation.aip.org/content/aip/journal/jcp/131/7/10.1063/1.3206326>. 1.1
- [58] Volkan Ediz, Anthony C. Monda, Robert P. Brown, and David J. Yaron. Using Molecular Similarity to Develop Reliable Models of Chemical Reactions in Complex Environments. *J. Chem. Theory Comput.*, 5(12):3175–3184, December 2009. ISSN 1549-9618. URL <http://dx.doi.org/10.1021/ct9004195>. 1.1, 4.1
- [59] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.*, 108(5):058301, January 2012. ISSN 0031-9007. URL <http://link.aps.org/doi/10.1103/PhysRevLett.108.058301>. 1.1
- [60] Katja Hansen, Grégoire Montavon, Franziska Biegler, Siamac Fazli, Matthias Rupp, Matthias Scheffler, O. Anatole von Lilienfeld, Alexandre Tkatchenko, and Klaus-Robert Müller. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.*, 9(8):3404–3419, August 2013. ISSN

1549-9618. URL <http://dx.doi.org/10.1021/ct400195d>. 1.1

- [61] Grégoire Montavon, Matthias Rupp, Vivekanand Gobre, Alvaro Vazquez-Mayagoitia, Katja Hansen, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole von Lilienfeld. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.*, 15(9):095003, September 2013. ISSN 1367-2630. URL <http://stacks.iop.org/1367-2630/15/i=9/a=095003>. 1.1
- [62] Edward O. Pyzer-Knapp, Kewei Li, and Alan Aspuru-Guzik. Learning from the harvard clean energy project: The use of neural networks to accelerate materials discovery. *Advanced Functional Materials*, pages n/a–n/a, 2015. ISSN 1616-3028. doi: 10.1002/adfm.201501919. URL <http://dx.doi.org/10.1002/adfm.201501919>. 1.1
- [63] T M Mitchell. *Machine learning*. Burr Ridge, IL: McGraw Hill, 1997. 1.1
- [64] Vasilis Ntziachristos. Fluorescence Molecular Imaging. *Annu. Rev. Biomed. Eng.*, 8(1): 1–33, 2006. URL <http://www.ncbi.nlm.nih.gov/pubmed/16834550>. 1.2, 2.1
- [65] Nina Wache, Alexander Scholten, Thorsten Klüner, Karl-Wilhelm Koch, and Jens Christoffers. Turning On Fluorescence with Thiols - Synthetic and Computational Studies on Diaminoterephthalates and Monitoring the Switch of the Ca^{2+} Sensor Recoverin. *Eur. J. Org. Chem.*, 2012(29):5712–5722, October 2012. ISSN 1434193X. URL <http://doi.wiley.com/10.1002/ejoc.201200879>. 1.2, 2.1
- [66] Seth Olsen and Sean C Smith. Radiationless Decay of Red Fluorescent Protein Chromophore Models via Twisted intramolecular charge-transfer states. *J. Am. Chem. Soc.*, 129(7):2054–65, February 2007. ISSN 0002-7863. URL <http://dx.doi.org/10.1021/ja066430s>. 1.2, 2.1
- [67] A. Prasanna de Silva, Thorfinnur Gunnlaugsson, and Terence E. Rice. Recent Evolution of Luminescent Photoinduced Electron Transfer Sensors. A Review. *Analyst*, 121

- (12):1759, 1996. ISSN 0003-2654. URL <http://pubs.rsc.org/en/content/articlehtml/1996/an/an9962101759>. 1.2, 2.1
- [68] A Prasanna de Silva, Thomas S Moody, and Glenn D Wright. Fluorescent PET (Photoinduced Electron Transfer) Sensors as Potent Analytical Tools. *Analyst*, 134(12):2385–93, December 2009. ISSN 1364-5528. URL <http://pubs.rsc.org/en/content/articlehtml/2009/an/b912527m>. 1.2, 2.1
- [69] Richard A. Bissell, A. Prasanna de Silva, H.Q. Nimal Gunaratne, P.L. Mark Lynch, Glenn E.M. Maguire, Colin P. McCoy, and K.R.A. Samankumara Sandanayake. Fluorescent pet (photoinduced electron transfer) sensors. *Curr. Top. Med. Chem.*, 168:223–264, 1993. doi: 10.1007/3-540-56746-1_12. URL http://dx.doi.org/10.1007/3-540-56746-1_12. 1.2, 2.1
- [70] Luke D Lavis and Ronald T Raines. Bright ideas for chemical biology. *ACS chemical biology*, 3(3):142–155, 2008. 1.2
- [71] André Nadler and Carsten Schultz. The power of fluorogenic probes. *Angewandte Chemie International Edition*, 52(9):2408–2410, 2013. 1.2
- [72] Maroti G Pawar and Seergazhi G Srivatsan. Environment-responsive fluorescent nucleoside analogue probe for studying oligonucleotide dynamics in a model cell-like compartment. *The Journal of Physical Chemistry B*, 117(46):14273–14282, 2013. 1.2
- [73] Eul Hyun Suh, Yu Liu, Stephen Connelly, Joseph C Genereux, Ian A Wilson, and Jeffery W Kelly. Stilbene vinyl sulfonamides as fluorogenic sensors of and traceless covalent kinetic stabilizers of transthyretin that prevent amyloidogenesis. *Journal of the American Chemical Society*, 135(47):17869–17880, 2013. 1.2
- [74] Jeremy S Paige, Karen Y Wu, and Samie R Jaffrey. Rna mimics of green fluorescent protein. *Science*, 333(6042):642–646, 2011. 1.2
- [75] Erik W Debler, Gunnar F Kaufmann, Michael M Meijler, Andreas Heine, Jenny M Mee,

- Goran Pljevaljčić, Angel J Di Bilio, Peter G Schultz, David P Millar, Kim D Janda, et al. Deeply inverted electron-hole recombination in a luminescent antibody-stilbene complex. *Science*, 319(5867):1232–1235, 2008. 1.2
- [76] Xiaohua Li, Xinghui Gao, Wen Shi, and Huimin Ma. Design strategies for water-soluble small molecular chromogenic and fluorogenic probes. *Chemical reviews*, 114(1):590–659, 2013. 1.2
- [77] Jianfang Cao, Tong Wu, Chong Hu, Tao Liu, Wen Sun, Jiangli Fan, and Xiaojun Peng. The nature of the different environmental sensitivity of symmetrical and unsymmetrical cyanine dyes: an experimental and theoretical study. *Physical Chemistry Chemical Physics*, 14(39):13702–13708, 2012. 1.2, 3.1
- [78] Alessandro Biancardi, Tarita Biver, Alberto Marini, Benedetta Mennucci, and Fernando Secco. Thiazole orange (to) as a light-switch probe: a combined quantum-mechanical and spectroscopic study. *Physical Chemistry Chemical Physics*, 13(27):12595–12602, 2011. 1.2, 3.1
- [79] Gloria L Silva, Volkan Ediz, David Yaron, and Bruce A Armitage. Experimental and Computational Investigation of Unsymmetrical Cyanine Dyes: Understanding Torsionally Responsive Fluorogenic Dyes. *J. Am. Chem. Soc.*, 129:5710, 2007. 1.2, 3.1
- [80] J. A. Pople. Approximate Self-Consistent Molecular-Orbital Theory. V. Intermediate Neglect of Differential Overlap. *J. Chem. Phys.*, 47(6):2026, September 1967. ISSN 00219606. URL <http://link.aip.org/link/?JCPSA6/47/2026/1>. 1.3, 4.2.2, 5.3
- [81] WC Lu, CZ Wang, MW Schmidt, L Bytautas, KM Ho, and K Ruedenberg. Molecule intrinsic minimal basis sets. i. exact resolution of ab initio optimized molecular orbitals in terms of deformed atomic minimal-basis orbitals. *J. Chem. Phys.*, 120(6):2629–2637, 2004. 1.3, 5.7

- [82] Péter Pulay. Convergence acceleration of iterative sequences. The case of scf iteration. *Chem. Phys. Lett.*, 73(2):393–398, July 1980. ISSN 00092614. URL <http://www.sciencedirect.com/science/article/pii/0009261480803964>. 1.4, 6.2.1, 6.4
- [83] Xiangqian Hu and Weitao Yang. Accelerating self-consistent field convergence with the augmented roothaanhall energy function. *J. Chem. Phys.*, 132(5):054109, 2010. doi: <http://dx.doi.org/10.1063/1.3304922>. URL <http://scitation.aip.org/content/aip/journal/jcp/132/5/10.1063/1.3304922>. 1.4, 6.2.1
- [84] Konstantin N. Kudin, Gustavo E. Scuseria, and Eric Cancs. A black-box self-consistent field convergence algorithm: One step closer. *J. Chem. Phys.*, 116(19), 2002. 1.4, 6.4
- [85] Thorsten Rohwedder and Reinhold Schneider. An analysis for the diis acceleration method used in quantum chemistry calculations. *J. Math. Chem.*, 49(9):1889, 2011. 1.4, 6.2.1
- [86] Stephane Ross, Geoffrey J. Gordon, and J. Andrew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. *arXiv*, 1011.0686, 2011. 1.4, 6.2.2
- [87] Matteus Tanha, Shiva Kaul, Alex Cappiello, Geoffrey J. Gordon, and David J. Yaron. Embedding parameters in ab initio theory to develop well-controlled approximations based on molecular similarity. <http://arxiv.org/abs/1311.3440>, November 2013. URL <http://arxiv.org/abs/1311.3440>. 2.1
- [88] G Salama and A Scarpa. Enhanced Ca^{2+} Uptake and ATPase Activity of Sarcoplasmic Reticulum in the Presence of Diethyl Ether. *J. Biol. Chem.*, 255(14):6525–6528, 1980. 2.2
- [89] Takeshi Yanai, David P Tew, and Nicholas C Handy. A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.*, 393(1-3):51–57, July 2004. ISSN 00092614. URL <http://www.sciencedirect.com>.

com/science/article/pii/S0009261404008620. 2.3, 3.2

- [90] Denis Jacquemin, Eric A. Perpète, Gustavo E. Scuseria, Ilaria Ciofini, and Carlo Adamo. TD-DFT Performance for the Visible Absorption Spectra of Organic Dyes: Conventional versus Long-Range Hybrids. *J. Chem. Theory Comput.*, 4(1):123–135, January 2008. ISSN 1549-9618. URL <http://dx.doi.org/10.1021/ct700187z>. 2.3, 3.2
- [91] J. Tomasi, B. Mennucci, and E. Cancès. The IEF Version of the PCM Solvation Method: an Overview of a New Method Addressed to Study Molecular Solutes at the QM Ab Initio Level. *J. Mol. Struct.*, 464(1-3):211–226, May 1999. ISSN 01661280. URL <http://linkinghub.elsevier.com/retrieve/pii/S0166128098005533>. 2.3, 3.2
- [92] Gaussian 09, Revision A.1, Frisch, M. J. et al, Wallingford CT. 2009. 2.3
- [93] Eric D. Glendening, David Feller, and Mark A. Thompson. An Ab Initio Investigation of the Structure and Alkali Metal Cation Selectivity of 18-Crown-6. *Journal of the American Chemical Society*, 116(23):10657–10669, November 1994. ISSN 0002-7863. URL <http://dx.doi.org/10.1021/ja00102a035>. 2.3, 2.4.6
- [94] Mark A. Thompson, Eric D. Glendening, and David Feller. The Nature of K⁺/Crown Ether Interactions: A Hybrid Quantum Mechanical-Molecular Mechanical Study. *The Journal of Physical Chemistry*, 98(41):10465–10476, October 1994. ISSN 0022-3654. URL <http://dx.doi.org/10.1021/j100092a015>. 2.3, 2.4.6
- [95] Jonathan W Steed. First- and second-sphere coordination chemistry of alkali metal crown ether complexes. *Coordination Chemistry Reviews*, 215(1):171–221, May 2001. ISSN 00108545. URL [http://dx.doi.org/10.1016/S0010-8545\(01\)00317-4](http://dx.doi.org/10.1016/S0010-8545(01)00317-4). 2.3, 2.4.6
- [96] Michael Kasha. Characterization of Electronic Transitions in Complex Molecules. *Discuss. Faraday Soc.*, 9(c):14–19, 1950. ISSN 03669033. doi: 10.1039/df9500900014.

URL <http://xlink.rsc.org/?DOI=df9500900014>. 2.4.2

- [97] Aimée Tomlinson and David Yaron. Direct INDO/SCI method for excited state calculations. *Journal of Computational Chemistry*, 24(14):1782–1788, 2003. 2.4.6, 3.2
- [98] Rudolph A Marcus. Electron Transfer Reactions in Chemistry. Theory and Experiment. *Rev. Mod. Phys.*, 65(3):599–610, 1993. ISSN 00346861. doi: 10.1103/RevModPhys.65.599. URL <http://link.aps.org/doi/10.1103/RevModPhys.65.599>. 2.4.8
- [99] Elizabeth Rastede, Matteus Tanha, David J. Yaron, Alan S. Waggoner, and Bruce A. Armitage. Spectral Fine Tuning of Cyanine Dyes: Electron Donor-Acceptor Substituted Analogues of Thiazole Orange. *Photochemical and Photobiological Sciences*, 2015. 3.1
- [100] Axel D. Becke. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.*, 98(7):5648, April 1993. ISSN 00219606. doi: 10.1063/1.464913. URL <http://link.aip.org/link/?JCP/98/5648/1>. 3.2
- [101] Denis Jacquemin, Yan Zhao, Rosendo Valero, Carlo Adamo, Ilaria Ciofini, and Donald G. Truhlar. Verdict: Time-Dependent Density Functional Theory Not Guilty of Large Errors for Cyanines. *Journal of Chemical Theory and Computation*, 8(4):1255–1259, April 2012. ISSN 1549-9618. doi: 10.1021/ct200721d. URL <http://dx.doi.org/10.1021/ct200721d>. 3.2
- [102] Denis Jacquemin, Eric A. Perpète, Ilaria Ciofini, and Carlo Adamo. Assessment of the ω B97 family for excited-state calculations. *Theoretical Chemistry Accounts*, 128(1):127–136, July 2010. ISSN 1432-881X. doi: 10.1007/s00214-010-0783-x. URL <http://link.springer.com/10.1007/s00214-010-0783-x>. 3.2
- [103] AE Frisch, JB Foresman, et al. Exploring chemistry with electronic structure methods. *Pittsburgh PA: Gaussian Inc*, 302, 1996. 3.2
- [104] Iliana I Timcheva, Vera A Maximova, Todor G Deligeorgiev, Nikolai I Gadjev, Ram W

Sabnis, and Ivan G Ivanov. Fluorescence spectral characteristics of novel asymmetric monomethine cyanine dyes in nucleic acid solutions. *FEBS letters*, 405(2):141–144, 1997. 3.4

- [105] M J S Dewar, E F Healy, A J Holder, and Y C Yuan. Comments on a Comparison of Am1 with the Recently Developed Pm3 Method. *J. Comput. Chem.*, 11:541–542, 1990. ISSN 0192-8651. URL <GotoISI>://A1990CZ22100012. 4.1
- [106] Benjamin G Janesko and David Yaron. Using molecular similarity to construct accurate semiempirical electronic structure theories. *J. Chem. Phys.*, 121(12):5635–45, September 2004. ISSN 0021-9606. URL <http://www.ncbi.nlm.nih.gov/pubmed/15366987>. 4.1
- [107] Thomas F. Coleman and Yuying Li. On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds. *Math. Program.*, 67(1-3):189–224, October 1994. ISSN 0025-5610. URL <http://www.springerlink.com/index/10.1007/BF01582221>. 4.3
- [108] T F Coleman and Y Li. An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIAM J. Optim.*, 6(2):418–445, 1996. ISSN 10526234. doi: 10.1137/0806023. URL <http://link.aip.org/link/SJOPE8/v6/i2/p418/s1&Agg=doi>. 4.3, 5.5, 6.4
- [109] Matlab version 7.14.0.739. Natick, Massachusetts, The MathWorks Inc., 2012. 4.3, 5.5, 6.4
- [110] Kumara Sastry, D. D. Johnson, Alexis L. Thompson, David E. Goldberg, Todd J. Martinez, Jeff Leiding, and Jane Owens. Optimization of Semiempirical Quantum Chemistry Methods via Multiobjective Genetic Algorithms: Accurate Photodynamics for Larger Molecules and Longer Time Scales. *Mater. Manuf. Processes*, 22(5):553–561, June 2007. ISSN 1042-6914. URL <http://dx.doi.org/10.1080/10426910701319506>. 4.3

- [111] Curt M. Breneman and Kenneth B. Wiberg. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.*, 11(3):361–373, April 1990. ISSN 0192-8651. URL <http://doi.wiley.com/10.1002/jcc.540110311>. 5.2
- [112] John C. Slater. *Quantum Theory of Atomic Structure*. McGraw Hill, 1960. ISBN 0070580421. URL <http://www.amazon.com/Quantum-Theory-Atomic-Structure-SLATER/dp/0070580421>. 5.3
- [113] Richard F. W. Bader. A quantum theory of molecular structure and its applications. *Chem. Rev.*, 91(5):893–928, July 1991. ISSN 0009-2665. URL <http://dx.doi.org/10.1021/cr00005a013>. 5.7
- [114] S. Ross, D. Munoz, M. Hebert, and J. A. Bagnell. Learning message-passing inference machines for structured prediction. 2011. 6.1
- [115] D.B. Cook. *Handbook of Computational Quantum Chemistry*. Dover Books on Chemistry. Dover Publications, 2005. ISBN 9780486443072. URL <http://books.google.com/books?id=Aa0qAwAAQBAJ>. 6.1, 6.2.1
- [116] Matheus Tanha, Tse-Han Huang, Geoffrey J. Gordon, and David J. Yaron. Imitation Learning for Accelerating Iterative Computation of Fixed Points in Quantum Chemistry. *ewrl2015*. 6.1
- [117] Alejandro J. Garza and Gustavo E. Scuseria. Comparison of self-consistent field convergence acceleration techniques. *J. Chem. Phys.*, 137(5):054110, 2012. doi: <http://dx.doi.org/10.1063/1.4740249>. URL <http://scitation.aip.org/content/aip/journal/jcp/137/5/10.1063/1.4740249>. 6.2.1
- [118] Alejandro J. Garza and Gustavo E. Scuseria. Comparison of self-consistent field convergence acceleration techniques. *J. Chem. Phys.*, 137:054110, 2012. 6.2.1
- [119] Konstantin N. Kudin, Gustavo E. Scuseria, and Eric Cancés. A black-box self-consistent

field convergence algorithm: One step closer. *J. Chem. Phys.*, 116:8255, 2010. 6.2.1