Digging Deeper into the Methods of Computational Chemistry

by

Joshua A. Kammeraad

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry and Scientific Computing)
in the University of Michigan
2020

Doctoral Committee:

Associate Professor Paul M Zimmerman, Chair
Professor Charles L. Brooks III
Professor Robert Krasny
Professor Roseanne Sension

Joshua A. Kammeraad

joshkamm@umich.edu

ORCID iD: 0000-0003-0386-7198

Dedication

To God be all the glory.

Acknowledgements

First, I would like to thank my graduate mentor Paul Zimmerman for his mentorship. Paul learned my quirkiness and how to individually mentor me. His focus and direction kept me on track and his patience in challenging seasons is appreciated. I would also like to thank the rest of my graduate committee for their advice and feedback throughout this PhD program.

I would like to acknowledge the rest of the Zimmerman lab, particularly Ian Pendleton, Alan Chien, and Andrew Molina for welcoming me into the lab and engaging in deep conversations of an academic and personal nature. I would like to thank Cody Aldaz for letting me be a part of his cool projects. Thank you Ambuj Tewari and everyone involved in our machine learning collaborations: Jack, Mina, Eric, Exequiel, Ziping, Tarun, and Eunjae. Ryan Hayes and Allison Roessler, thank you for all the times together in prayer for each other and our labs. I am also grateful for all of the mentors and teachers who prepared me for graduate school, particularly professors DeJongh, Pennings, Polik, and Cinzori for their support throughout my undergraduate education. Thank you to my cross country teammates including my roommates Tim Lewis and David Dolfin as well as Hania Szymczak, Michelle Kerr, and Jess Gaines for all of our nerdy, serious, and silly running conversations that kept me sane and the long nights and weekends of studying together that prepared me so well for this program.

Thank you Dave Brzezinski, Jack Geddes, David Taylor, and Paul Webb from Grace Bible Church for your mentorship throughout my graduate studies. I am deeply grateful to my spiritual brothers and sisters in Impact Graduate Christian Fellowship without whom I could not have survived this program, including my accountability partners for their encouragement and support of my continued growth: Sinsar Hsie, Mark Dong, Calvin Montana, Matt Cui, Alex Wang, and Joseph Tu. A special thanks to each of the brothers and sisters I've had the privilege of serving closely with: Nancy Wu, Josh Cheng, Sara Timberlake, and Anita Luong. Our partnerships have sculpted and sharpened my character and your patience and support in my weakness is appreciated. Thank you Jasmine Jones for being so relentlessly selfless including spontaneously helping prepare me for my oral defense while I recovered from illness.

James Tan and Matt Hughes, I have a deep appreciation for your loyal brotherhood in Christ and fighting with me through some of the hardest times of the past 5 years. To my family, thank you for being so loving and supportive of all stages of my education. Finally, I am eternally grateful to Jesus Christ my Savior and Lord for this beautiful, elegant universe and the capacity to explore it.

Table of Contents

List of Figures

List of Schemes

List of Tables

Abstract

This dissertation applies a skeptical but hopeful analytical paradigm and the tools of linear algebra, numerical methods, and machine learning to a diversity of problems in computational chemistry. When the foundation underlying a project is undermined, the primary purpose of the project becomes digging into the nature and structure of the problem. A common theme emerges in which assumptions in an area are challenged and a deeper understanding of the problem structure leads to new insights.

In chapter 2, this approach is exploited to approximate derivative coupling vectors, which together with the difference gradient span the branching planes of conical intersections between electronic states. While gradients are commonly available in many electronic structure methods, the derivative coupling vectors are not always implemented and ready for use in characterizing conical intersections. An approach is introduced which computes the derivative coupling vector with high accuracy (direction and magnitude) using energy and gradient information. The new method is based on the combination of a linear-coupling two-state Hamiltonian and a finite-difference Davidson approach for computing the branching plane. Benchmark cases are provided showing these vectors can be efficiently computed near conical intersections.

In chapter 3, this approach yields a countercultural explanation for what machine learning algorithms have learned in modeling a chemical reactivity dataset. Data-driven models of chemical reactions, a departure from conventional chemical approaches, have recently been shown to be statistically successful using machine learning. These models, however, are largely black box in

character and have not provided the kind of chemical insights that historically advanced the field of chemistry. The chapter examines the knowledgebase of machine learning models—what does the machine learn?—by deconstructing black box machine learning models of a diverse chemical reaction dataset. Through experimentation with chemical representations and modeling techniques, the analysis provides insights into the nature of how statistical accuracy can arise, even when the model lacks informative physical principles. By peeling back the layers of these complicated models we arrive at a minimal, chemically intuitive model (and no machine learning involved). This model is based on systematic reaction type classification and Evans-Polanyi relationships within reaction types which are easily visualized and interpreted. Through exploring this simple model, we gain deeper understanding of the dataset and uncover a means for expert interactions to improve the model's reliability.

In chapter 4, human - algorithm interaction is explored as a paradigm for generating representative ensembles of conformers for organic compounds, a challenging problem in computational chemistry with implications on the ability to understand and predict reactivity. The approach utilizes the molecular editor IQmol as an interface between chemists and reinforcement learning algorithms with the cheminformatics package RDKit as a backbone. Conformer ensembles are evaluated by uniqueness and the approximation they yield of the partition function. Prototype results are presented for a standard reinforcement learning algorithm tested on linear alkanes and chemist manipulation of a fragment of the biomolecule lignin. Future aims and directions for this young project are discussed.

The concluding chapter reflects on the broader lessons learned from conducting the dissertation. It discusses open questions and potential paradigms for pursuing them.

Chapter 1. Introduction

The methods developed in this dissertation draw from a breadth of mathematical and computational disciplines ranging from linear algebra, numerical methods, and machine learning.[1–3] The targets for application also span a breadth of chemical disciplines including photochemistry,[4–7] catalysis,[8,9] and biochemistry. However, within this diversity is a unity in the attitude and approach taken, and what it suggests about the underlying nature of reality and scientific progress. The attitude is one of exploration followed by skepticism coupled with hope for an underlying, deeper conceptual foundation.

Chapter Overviews

Chapter 2

Photochemistry is challenging to understand and model experimentally because key reactivity often occurs quickly, which makes mechanisms and intermediates difficult to observe and identify. Computer simulations can contribute to this understanding, but simulating photochemistry also poses interesting challenges. Photochemistry involves transitions between electronic states through conical intersections which are difficult to model numerically. A key quantity in modeling transitions through conical intersections is the derivative coupling vector which numerically describes the nuclear-electronic coupling underlying state transitions.

Prior to this work, the primary means of obtaining this quantity was through electronic structure calculations, and the scientific community spends significant effort keeping implementations of the derivative coupling vector up to date with an ever evolving framework of

electronic structure theories. What is missing is an accurate general approach to approximating the direction and magnitude of the derivative coupling vector that is immediately accessible as new electronic structure theory improvements are developed.

In chapter 2, we develop a general method for approximating the derivative coupling vector within the framework of any given electronic structure implementation that has the energy and nuclear gradient available. This makes it possible for future theoretical developments in electronic structure to more rapidly be accessible for photochemical simulations.

Chapter 3

Machine learning has recently gained popularity as an approach for modeling complex data relationships in a plethora of domains. This hype has been in part driven by machine learning's quantitative success in highly nonlinear problems previously believed to necessitate human-like heuristics and intuition. The hype has carried machine learning into growing popularity in chemistry, and quantitative results are once again promising.

However, metrics for success in chemistry are nuanced. In chapter 3, we suggest that the machine learning community in chemistry should consider how to evaluate results in a robust and chemically meaningful manner. While machine learning can obtain impressive quantitative success on seemingly reasonable benchmarks of predicting chemical reactivity, this chapter will show that such learning provides little of scientific value. We provide substantial evidence to justify what we believe to be an underrepresented perspective in the field.

Chapter 4

Generation of conformers, different configurations of molecules accessible without breaking of forming chemical bonds, is a challenging chemical problem. An adequate sample of low energy conformers is necessary for accurate thermodynamic modeling, but determining which

of the combinatorially large configurations are significant in affecting the thermodynamics is difficult. Accurate solutions are available in the specific cases of small molecular systems (by enumerating over all possibilities) or when the potential energy surface is well-behaved enough for molecular dynamics simulations to tractably sample ergodically (though at high cost). Some progress has been made in more general cases, but improving accuracy across the breadth of chemically interesting systems remains an active area of research.

In chapter 4, we begin to apply similar paradigms as in chapters 2 and 3 to this problem of conformer generation. We take what we learned about the nature of integrating learning from humans and algorithms and develop a prototype for utilizing the strengths of each.

Chapter 5

Chapter 5 provides a conclusion to the dissertation. It attempts to summarize and tie together the projects undertaken while exploring potential future directions. A discussion follows on principles which were valuable in the conducting of the dissertation work and should continue to prove valuable in future explorations.

Themes

From these chapters, a common theme emerges through the projects undertaken in this dissertation. First, a method is proposed that leverages algorithmic understanding and insights in an attempt to improve the tractability of a recurring challenge encountered in computational studies of chemical problems. Initially, the details are shaped through experience, but the method is developed essentially as expected. The method performs roughly as anticipated and would be considered acceptable within the relevant subdisciplines. However, subsequent adversarial self-critique of the method questions the underlying assumptions and conceptual foundation upon

which the method was built. This leads to further analysis and investigation of the principles and concepts underlying the method, which in turn become the main focus of the project.

In chapter 2, the manifestation of this theme centered around the linear algebra of the derivative coupling vector.[7] In much of the scientific literature, two orthogonal branching plane vectors are defined or determined and referred to as the difference gradient and the derivative coupling vector. Thus, our project's initial idea produced the vector in the branching plane orthogonal to the difference gradient. Initial testing suggested that our method approximated the derivative coupling vector fairly well as calculated directly through an electronic structure package. However, while investigating the deviation I noticed that the electronic structure package derivative coupling vector was not orthogonal to its own difference gradient. Upon further digging into literature we uncovered deeper complexity in the derivative coupling vector and how it is used that was obfuscated by inconsistent usage of terminology in the literature and lack of a consensus way of thinking about and talking about the branching plane.[10] Looking into how different scientists understood the structure of the branching plane led to a new, simple method to parameterize the branching plane.[11] This led to a significant improvement over the original method because the parameterized model not only provided a nearly perfect approximation to the derivative coupling vector, but also allowed for the method to do more than originally intended. The model additionally yielded an approximation for the location of a nearby conical intersection, a key feature of the potential energy surface in photochemistry. This approximation is better than standard optimizations following the difference gradient and using the derivative coupling vector merely to define the branching plane subspace.

In chapter 3, the overarching theme manifested in a project in which we attempted to optimize digital representations of chemical concepts for effectiveness in machine learning of

chemical reactivity. The key structure on which the project is built is the interplay between data, representation, and optimization strategy. It is well known that each of these concepts is important to effective algorithmic learning of chemistry and that they interact in complex and subtle ways, but to jointly optimize these factors is a subject of current interest in the research community.[12–14] We took a computational dataset of simulated reaction data generated within our research group and began exploring representations and testing them with common machine learning algorithms. We quickly obtained quantitatively respectable results given the complexity of our dataset using a chemically meaningful representation, suggesting that common machine learning algorithms were able to learn something chemically meaningful from the reaction data. However, we determined through adversarial self-critique and observation that what we thought would be the most chemically important concepts for an algorithm to "learn" were not actually essential to machine learning algorithms' predictive power.[15,16] We explored this further and demonstrated how machine learning algorithms were actually learning in a way that is counterintuitive to chemists and unlikely to generalize. We showed this point through generating a minimally chemically meaningful representation of reactivity that could still successfully train a quantitative model. This finding is significant because what we learned is applicable to many current projects in the community attempting to apply machine learning to various chemical problems.[16,17]

While the project described in chapter 4 is still in the early stages, we hope that application of a similar approach as in the other projects will prove fruitful and uncover deeper understanding. We feel that the field of conformer generation is still searching for a coherent framework within which to explore solutions and deeper understanding would be valuable towards this effort.

Chapter 2. Estimating the Derivative Coupling Vector using Gradients

This chapter is largely based upon published work:
Kammeraad, J. A.; Zimmerman, P. M. Estimating the Derivative Coupling Vector Using Gradients. J. Phys. Chem. Lett. 2016, 7 (24), 5074–5079.
https://doi.org/10.1021/acs.jpclett.6b02501.

Main Text

Strong coupling between nuclear and electronic degrees of freedom leads to highly interesting chemical phenomena, such as photo-induced reactions.[4,18,19] To enable meaningful descriptions of these ultrafast processes, atomistic simulations can be insightful, but are challenging to perform because accurate electronic structure and dynamics tools must seamlessly work together in tandem. Therefore, new methods to make these simulations simpler and broadly applicable are in demand.

The most difficult to describe aspects of ultrafast vibronic processes are the dynamics near conical intersections, which are regions of the potential energy surface where the Born-Oppenheimer approximation breaks down. Conical intersections (CI) describe the nuclear configurations where two electronic states intersect and are spanned by two vectors, the difference gradient and the derivative coupling. While the difference gradient can be formed from the two states' gradients, making them easily computable, the derivative coupling vector is often less readily available. While derivative coupling vectors can be computed in a separate step after the gradients, these are not implemented for many levels of electronic structure theory.[20] Only in the last few years, for instance, did these vectors become available for the widely used time-dependent density functional theory,[21,22] and novel wave function methods such as Restricted Active Space

Spin-Flip[23–25] have no derivative coupling algorithm available. Even modern implementations of CASCI and CASSCF, for example efficient algorithms for GPUs[26–28] require nontrivial effort for constructing derivative coupling code. In special cases such as multiconfigurational, multistate perturbation theory, the derivative coupling vector can implemented using similar terms as the gradient.[29,30] Such methods, however, cannot be applied in general to many electronic structure theories.

A number of useful methods have been proposed to approximate the information that derivative coupling vectors would otherwise provide, were they available.[31] For conical intersection geometry optimization, difference gradients along the optimization path can provide a sufficient approximation to the branching plane to reach convergence.[32] In surface hopping trajectories,[6,33,34] a molecular system passing through a "trivial crossing" with low coupling can be made to remain on the same diabatic state, which is physically correct.[35] For nontrivial crossings, interpolation of the wave function can be used to quantify the time-derivative coupling,[36] which is the dot product of the derivative coupling with the nuclear velocity.[37–39] While these methods are useful to enhance the study of CI's and nonadiabatic dynamics, none provide a means to determine coupling vectors without a standard derivative coupling calculator. Without this vector, for instance, the velocity rescaling of surface hopping trajectories[33,40] cannot be performed, and conical intersection regions remain incompletely characterized.

In this chapter, branching planes and derivative coupling vectors are shown to be computable using only energy and gradient information. To do this at low cost, the branching plane[41] is constructed from the squared-energy-gap (ΔE^2) Hessian using an iterative diagonalization method that avoids building the full Hessian. Subsequently, the difference gradient and derivative coupling are determined using a model Hamiltonian, allowing the branching plane

to yield accurate estimates of the true derivative coupling vector and its magnitude. Since only energies and gradients are required, the proposed method has the potential for wide application in enabling derivative coupling vectors to be computed for many electronic structure theories.

To compute the derivative coupling using potential energy surface (PES) information, two components are used: 1. A representation of the branching plane, and 2. A model Hamiltonian with a two-state intersection. To construct the first piece, note that the adiabatic PESs involved in a conical intersection are not differentiable at a point of conical intersection,[20] so they are not directly useful for this purpose. The energy gap squared (ΔE^2), however, contains valuable information about the conical intersection and is more well behaved in its vicinity[42] (Figure 2-1). Similar to the regular PESs, where two vectors modulate the energy gap between the states, near the CI the ΔE^2 surface is constant except when moving in the branching plane. This 2-dimensional subspace is therefore spanned by 2 eigenvectors of the ΔE^2 Hessian corresponding to nonzero eigenvalues. Assuming there are only two adiabatic states in the nearby vicinity, these eigenvectors will correspond to $\vec{x}$ and $\vec{y}$, which we denote as two orthogonal vectors spanning the branching plane.[10] All other eigenvalues of this Hessian will be approximately 0 because they do not modulate the energy gap between states.

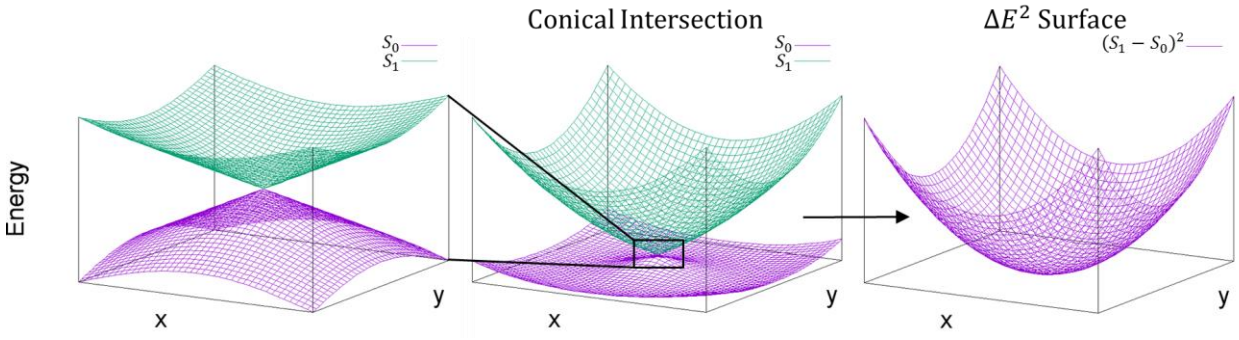
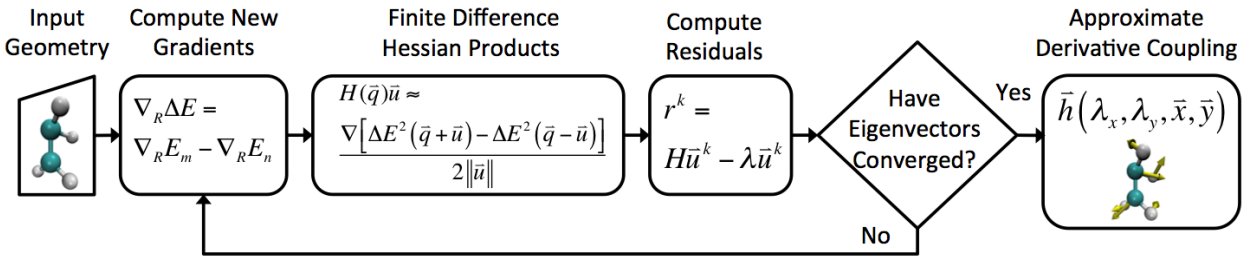


Figure 2-1. Conical intersection showing close up view (left), wide view (middle), and representation in terms of ΔE^2 (right).

Scheme 2-1. Overall process for constructing derivative coupling vectors.



To compute the eigenvectors corresponding to the upwards-curving directions of the ΔE^2 surface, the finite difference Davidson[43–45] method is used. Similar to the strategy of Sharada et al, which diagonalizes a single-surface Hessian without ever constructing that Hessian,[46] the ΔE^2 Hessian's lowest eigenvalues and corresponding eigenvectors can be found using only energies and difference gradient information. This works because the product of the Hessian with a unit vector $H_f(\bar{q})\vec{\mathrm{u}}$ is the rate of change of $\nabla f(\bar{\mathrm{q}})$ when moving in the direction $\vec{\mathrm{u}}$. Therefore, as shown in Scheme 2-1, only energy and gradient computations are required because the Davidson algorithm requires Hessian-vector products. The full finite-difference ΔE^2 Hessian is therefore not required in this procedure. As shown below, the Davidson iterative diagonalization procedure requires a smaller number of gradients to converge to the true branching plane than would be required to construct the full finite difference Hessian.

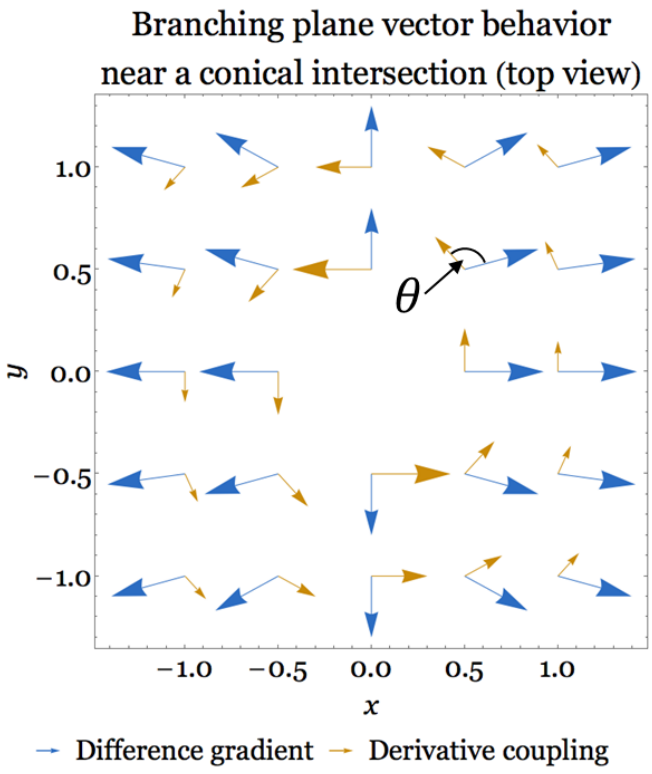


Figure 2-2. Derivative coupling and difference gradient in the region near a conical intersection. Arrow sizes are proportional to vector magnitude to the 0.5 power to improve visualization.

While constructing the branching plane is an important first step in determining the derivative coupling vector, the branching plane alone *does not* uniquely determine the derivative coupling direction. In general, the derivative coupling is not orthogonal to the difference gradient (see Figure 2-2), as was recently affirmed in Lindh et al.[10] To overcome this challenge and provide accurate derivative coupling vectors from the branching plane, a model Hamiltonian,

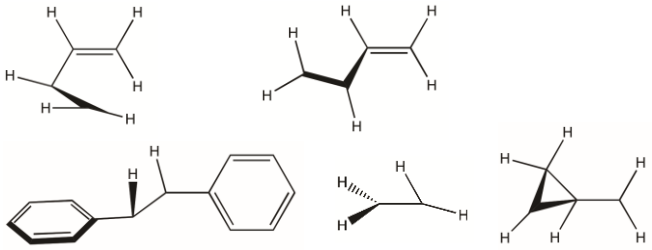
$$H^e = \left(s_x(x) + s_y(y) + s_{\vec{z}}(\vec{z})\right) I + \begin{pmatrix} gx & hy \\ hy & -gx \end{pmatrix} \quad (1)$$

can be utilized. This expansion resembles the diabatic models of Köppel,[11] which are designed to provide a meaningful representation of the electronic wave function in regions near a conical intersection. For this 2-state model, the eigenvalues of the ΔE^2 Hessian are $8g^2$ and $8h^2$. The eigenvectors corresponding to the nonzero eigenvalues are the $\vec{x}$ and $\vec{y}$ directions of the branching plane. Using this model, the Davidson procedure gives the branching plane vectors and the g and

h values, so x and y can be easily computed. Directly evaluating the model gives the angle between the difference gradient and derivative coupling vectors (θ in Figure 2-2) as well as the derivative coupling direction and magnitude. See the Methods section for further details. Extensions to the model to allow 3-state intersections may be possible by extending the Hamiltonian and considering a higher dimensional branching plane.[47,48]

The computational cost of this strategy is dominated by the cost of the electronic structure gradients required to form the Hessian vector products of the Davidson procedure. It will be shown below that only handful of gradients is required to reach convergence. Limitations of this algorithm are that it assumes only two close-lying electronic states, and proximity to a conical intersection where the model Hamiltonian remains meaningful (i.e. linear coupling). The accuracy, therefore, is expected to decrease with distance from the two-state conical intersection. This estimate, however, should be accurate in the high coupling regions where most population transfer occurs between the states.

Scheme 2-2. Conical intersections investigated in this work.



Conical intersection geometries from previous studies[49,50] were used as test cases for the new algorithm (see Scheme 2-2). These therefore represent geometries where the chosen model Hamiltonian (Eqn. 1) should be reasonably accurate and the derivative coupling vector has a large magnitude. While the derivative coupling vector has infinite magnitude at any point of conical intersection, only inside an extremely small radius around the CI seam will this become a problem.

For these geometries, gaps between electronic states are between 0.04 and 0.8 mHa, representing accurate but not exact conical intersections (Table 2-2).

The above test cases were used in choosing appropriate convergence criteria for the Davidson method. For ethene, convergence was reached after 2 iterations using a total of 14 gradients, compared to diagonalization of the full finite difference Hessian. Computing the full finite difference Hessian would require 72 gradients, demonstrating that the finite difference Davidson method is relatively efficient, even for this small molecule with relatively few degrees of freedom. The residual magnitude at convergence was $1.85 \cdot 10^{-4}$, and therefore we chose 10^{-3} as the residual threshold for subsequent computations. The angle between the CASSCF derivative coupling vector and the one computed using the proposed algorithm was 0.075 degrees, corresponding to near-perfect overlap. The magnitude of this vector was 1140.9 a.u., compared to the exact value of 1162.4 a.u. directly from CASSCF. In sum, using relatively few gradient computations, a nearly exact derivative coupling vector was found.

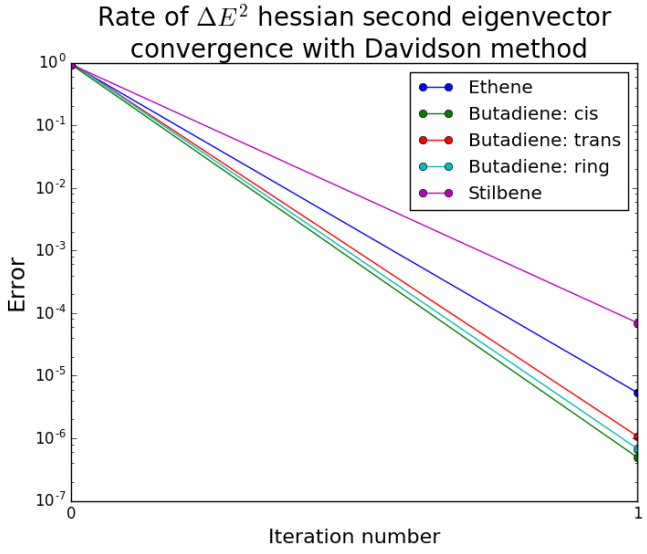


Figure 2-3. Convergence of the Davidson method at various conical intersections. Error $= 1 - |\langle$*Davidson eigenvector*$|$*true eigenvector*$\rangle|$.

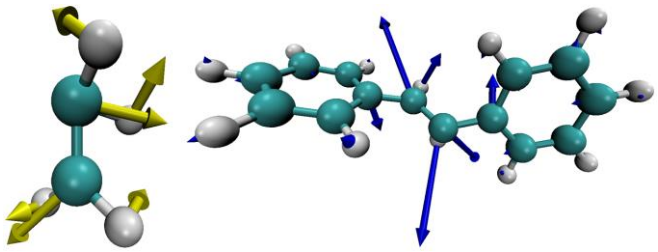


Figure 2-4. Approximate derivative coupling vectors from the Davidson procedure. Exact vectors are visually indistinguishable.

Similar convergence was seen in the other 4 test cases, all of which contain significantly more degrees of freedom than ethene. To reach the chosen 10^{-3} residual threshold, 3 matrix products (14 gradients) were required in each case. This suggests that the Davidson method converges using a small fraction of the dimensionality of the full space (the full finite difference Hessian requires 312 gradients for stilbene). The lack of increase in iterations from ethene to stilbene despite the significant increase in molecular size is encouraging. The coupling vector overlap, angle between difference gradient and derivative coupling vectors, and derivative coupling magnitudes are shown in Table 2-1 for the range of conical intersections of Scheme 2-2. Across the board, the proposed method estimates derivative coupling vectors to high accuracy, and even captures their magnitude to significant precision.

Table 2-1. Benchmark results compared to exact derivative coupling computations at conical intersections.

	Derivative coupling overlap	θ g vs h (deg)	Exact θ g vs h (deg)	Magnitude h	Magnitude h (exact)	Energy gap (au)	Magnitude of error: ‖(h calc) - (h exact)‖
Ethene	1.000	63.13	63.14	1140.9	1162.4	8.98E-05	21.50
Butadiene: cis	1.000	98.25	98.25	1993.6	1994.5	6.78E-05	1.65
Butadiene: trans	1.000	74.01	74.03	1848.9	1848.5	4.71E-05	2.10
Butadiene: ring	1.000	112.87	112.89	1120.9	1121.1	6.41E-05	1.00
Stilbene	1.000	110.10	109.94	101.8	101.6	7.17E-04	1.10

To ensure that these results hold in the vicinity of CI's, not just very close to the intersection, a set of 30 geometries near the ethene minimum energy CI were generated and examined. These structures are random displacements of magnitude 0.1, 0.2, and 0.3 Å from the CI, with ten structures at each distance. The geometries therefore represent the region where nonadiabatic trajectory simulations would most likely cross from one electronic surface to the next. The residual error, which is the magnitude of the difference between the exact and computed derivative coupling vectors, remains small in this region, with an average error of 2.64% (Table 2-2). Similar results were found for 30 geometries near butadiene's trans CI, where errors were even lower, 0.79% on average. The most serious errors in these two cases occur when the derivative coupling is small: in ethene, one geometry with an exact derivative coupling magnitude of 2.7 a.u. results in a residual of 0.29 a.u., which is 11%. Similarly, the worst-case butadiene geometry has coupling magnitude 1.4 a.u. and an error of 0.35 a.u. As expected, the model holds up quite well near CI's, and becomes less accurate in regions where the derivative coupling is small. Those errors are bearable, however, because small derivative couplings have less influence on the resulting dynamics than the larger couplings close to the intersection.

Table 2-2. Benchmark results in the vicinity of conical intersections. Units of a.u.

	Residual Error (avg)	Residual Error (max)	Avg. Error (%)
Ethene	0.16	0.35	2.64
Butadiene: trans	0.13	0.29	0.79

In summary, this work presents a novel means to compute derivative coupling vectors using only energy and gradient information. The method opens up new avenues for using a wider variety of electronic structure methods to analyze conical intersections and perform dynamics simulations. Specifically, any multistate method with an available gradient can now be used to

form accurate derivative coupling vectors, at least in regions near conical intersections where the linear model of Eqn. 1 remains accurate. We anticipate this strategy will be highly useful in contexts where the derivative coupling is unavailable through other means.

Methods

Diagonalizing the model Hamiltonian of Eqn. 1 to find the eigenvalues (adiabatic energies) as a function of the coordinates and evaluating their difference squared yields

$$\Delta E^2(x, y, \vec{z}) = 4((\mathrm{gx})^2 + (h\mathrm{y})^2) \tag{2}$$

Evaluating the Hessian of this surface therefore gives eigenvectors along the $\vec{x}$ and $\vec{y}$ axes, with corresponding eigenvalues $8g^2$ and $8h^2$. The eigenpairs of this Hamiltonian are then assumed to correspond to the Davidson method's eigenpairs. To make this correspondence, the appropriate values of x and y in the real chemical system must be determined. Since determining x and y is equivalent to finding the distance to the CI along the $\vec{x}$ and $\vec{y}$ directions, the vector from the current geometry to the minimum on the ΔE^2 surface contains this information. Finding x and y is thus analogous to computing a Newton step on the ΔE^2 surface.

$$x = \frac{\vec{\mathrm{u}} \cdot \vec{\mathrm{x}}}{\lambda_x} \qquad y = \frac{\vec{\mathrm{u}} \cdot \vec{y}}{\lambda_y} \qquad \vec{\mathrm{u}} = \nabla \Delta E^2 \tag{3}$$

where λ_x and λ_y are the eigenvalues for the real system. Once x and y are available, the derivative coupling can be computed from Eqn. 1. This is done using the coefficients of the normalized eigenvectors of the Hamiltonian, $c_{\mathrm{m}}^{\mathrm{n}}(x, y)$, which are independent of all coordinates except x and y,

$$\phi_1 = c_1^1(x, y)\langle\psi_1| + c_2^1(x, y)\langle\psi_2| \tag{4}$$

$$\phi_2 = c_1^2(x, y)\langle\psi_1| + c_2^2(x, y)\langle\psi_2|$$

The derivative coupling $\left\langle \phi_1 \middle| \frac{\partial}{\partial R} \phi_2 \right\rangle$ can be expressed in terms of derivatives of these coefficients with respect to nuclear coordinates $\vec{x}$, $\vec{y}$, and $\vec{z}_i$.

$$\left\langle \phi_1 \middle| \frac{\partial}{\partial x} \phi_2 \right\rangle = c_1^1(x,y) \frac{\partial}{\partial x} c_1^2(x,y) + c_2^1(x,y) \frac{\partial}{\partial x} c_2^2(x,y) = \frac{-ghy}{2g^2x^2 + 2h^2y^2} \tag{5}$$

$$\left\langle \phi_1 \middle| \frac{\partial}{\partial y} \phi_2 \right\rangle = \frac{ghx}{2g^2x^2 + 2h^2y^2} \tag{6}$$

and for all non-branching coordinates,

$$\left\langle \phi_1 \middle| \frac{\partial}{\partial z_i} \phi_2 \right\rangle = 0 \tag{7}$$

Finally, a change of basis creates the derivative coupling vector in the Cartesian coordinate system of the real system.

The computational complexity of computing energies and gradients needed to find Hessian products varies by electronic structure method but is often $\Theta(\mathrm{N}^{\mathrm{p}})$, $4 \leq \mathrm{p} \leq 8$. Diagonalization of the Davidson subspace matrix formally scales with n^3 where n is the dimension of the matrix. The total computational cost associated with diagonalizing the subspace matrix at all Davidson iterations is thus $\sum_{m=0}^{num\ iterations} m^3 = \Theta((num\ iterations)^4)$. Overall, the cost of the Davidson procedure is dominated by the electronic structure computations. In our tests, the time spent in a single Davidson iteration is less than 2 seconds for 1000-dimension random matrices.

All energies, gradients, and benchmark derivative coupling vectors were computed using the Molpro[51] implementation of Complete Active Space Self-Consistent Field (CASSCF)[52] with an active space of 2 electrons in 2 orbitals for ethene and stilbene and an active space of 4 electrons in 4 orbitals for butadiene. The S0 and S1 intersections were specifically studied. Geometries for the conical intersections were computed in Molpro (ethene) or found from Sicilia et. al.

(butadiene)[50] or Quenneville et. al. (stilbene)[49]. The 6-31G* basis set was used for butadiene and the 6-31G** basis set was used for ethene and stilbene.

A finite difference step size of 10^{-3} Å was used for the Hessian vector products. Initial Davidson vectors consist of the exact difference gradient and a random orthogonal vector. Davidson iterations proceed until the residuals have a magnitude less than 10^{-3}. Correction vectors are not added if the component of the normalized expansion vector orthogonal to the current Davidson subspace has a weight below 0.05.

Acknowledgements

This work was partially supported by the National Science Foundation Grant CHE-1551994. The authors would like to thank Cody Aldaz for help with CASSCF computations and Garrett Meek for helpful discussions.

Chapter 3. What Does the Machine Learn? Knowledge Representations of Chemical Reactivity

This chapter is largely based upon work currently under peer review.
Authors: Joshua A. Kammeraad, Jack Goetz, Eric Walker, Ambuj Tewari, Paul M. Zimmerman

Introduction

A great deal of excitement has been growing among physical scientists and engineers about machine learning. This excitement stems from a host of interesting examples from the data science field, including widely reported advances in image recognition, artificial intelligence in games, and natural language processing that have demonstrated extremely high levels of performance and even abilities beyond expert human capabilities. Substantial efforts have therefore been made to bring the tools of machine learning to bear upon the physical sciences,[53–57] with some of the most interesting chemical applications being in the areas of reactions and synthesis.[17,58–61] Chemistry, however, is traditionally driven by a combination of concepts and data, with its own heuristics, models, and hypothesis-making approach to research. It is our view that the contrast in approach between purely data-driven research and concept-driven research begs questions such as: What is the machine's representation of knowledge? What does the machine learn? It is these questions that will lead to more effective synergies between machine learning and the chemical sciences, as useful answers will involve explainable and interpretable concepts, not merely machine abstraction and black-box decision making. The intent of this chapter is to provide some preliminary indications of how current generation machine learning tools operate on chemical data, in partial

answer to these two questions. Our emphasis will be on application to computer prediction of chemical reactions, a key target for recent generations of machine learning methods.

The potential for computers to assist in synthesis has a long history, dating back to original proposals by E. J. Corey in the 60s.[62–64] These ideas were focused on the possibility for expert systems to encode known chemical principles into a systematic framework for predicting synthetic routes. Expert systems, however, fell out of favor due to the tedious encoding of rules and the rule exceptions required to maintain usability and accuracy across a diversity of reaction types. While recent efforts have challenged this conclusion,[65] the manual efforts needed to construct quality expert systems have by no means decreased. Alternatively, machine learning methodologies give the appearance of being particularly fit for encoding chemical reaction data without substantial human intervention and tinkering. To date, millions of reactions have been reported and are available in online databases, motivating recent efforts to use methods such as neural networks to build predictive tools for synthesis planning.[66–73]

Nonlinear regressions—which include deep neural networks[74–78]—form the basis for machine learning to represent complex relationships between input and output variables.[79] These methods can represent arbitrarily complex maps between any number of input variables and output results,[80] and can simply be applied to data, often with excellent statistical results. Since expert understanding of the meaning behind the data is not needed, the application of nonlinear regressions to encode chemical reaction is vastly different than applying expert systems (i.e. where specific rules are manually encoded and easily understood). In the specific case of neural networks, "hidden layers" constitute the intermediate representations that are used to make predictions. While these layers may well encode concepts and heuristics, they are indeed *hidden*, and do not provide transparent or interpretable reasons for decisions made by the network. In other popular

nonlinear techniques, “kernel” functions are used, where similarity between pairs of data points determines the structure of the predictions. Kernels are relatively interpretable compared to the hidden layers of neural networks, as similarity in the feature space is the core concept that can be understood.

To improve interpretability, data scientists might make use of input features that are comprehensible to chemists. Typical machine learning features involve graph-based features[81–84] (e.g. based on covalent attachments in molecules), strings (e.g. SMILES[85]), hashing, or substructure analysis, and these techniques have been widely used in drug design applications. Metrics such as Tanimoto distances,[81] which are measures of similarity between molecules, provide some grounding to chemical concepts, but are otherwise not trivial to interpret. In contrast, atomic charges or orbital energies derived from quantum chemistry, for instance, might be used alongside conventional physical organic descriptors[86,87] to capture chemical principles in quantitative form.[88,89] Progress in this area is useful and ongoing, but more insight is needed into the relationship between the physical content of these features and how machine learning models make use of the features.

Whereas machines have no prior expectations of the meaning of input features, chemists are clearly the opposite.[90] Chemists use explainable, physical features to make predictions, and they have strong expectations about how their models should behave based on these features.[91] In the case of a polar reaction, an atom with a high positive charge might be expected to react with an atom of large negative charge, due to Coulomb interactions. This fundamental physical interaction is described by chemists in terms of electronegativity and bond polarity, which are chemically specific descriptors that are highly useful for predicting reaction outcome. Due to these

relationships, invoking atomic charge as a descriptor brings in a wealth of expectations for an expert chemist, due to their knowledge of firmly established physical laws.

Machine learning models thus face a significant challenge in providing advances in chemical reactions (Figure 3-1), as it is not obvious how they are rooted in physical reality, or whether they use chemical features in a way that in any way resembles chemical thought. In the machine learning world, it is known that neural networks focus on distinctly different regions of images compared to humans when recognizing objects,[92] and yet still reach high accuracy. In the text that follows, this issue is investigated in detail by examining a dataset of chemical reactions with two qualitatively distinct, powerful machine learning methods. In short, we will show deep neural network and support machine (SVM) models to be quantitatively accurate, but missing a basic, qualitative representation of physical principles. Using this knowledge, it will be shown that a well-known, interpretable chemical principle better describes this dataset—and even provides higher quantitative accuracy than machine learning. Based on these results, Figure 3-1 outlines our viewpoint of the relationship between current-generation machine learning methods and chemical methods. This figure will be discussed in more detail in the discussion section after the main results of this study.

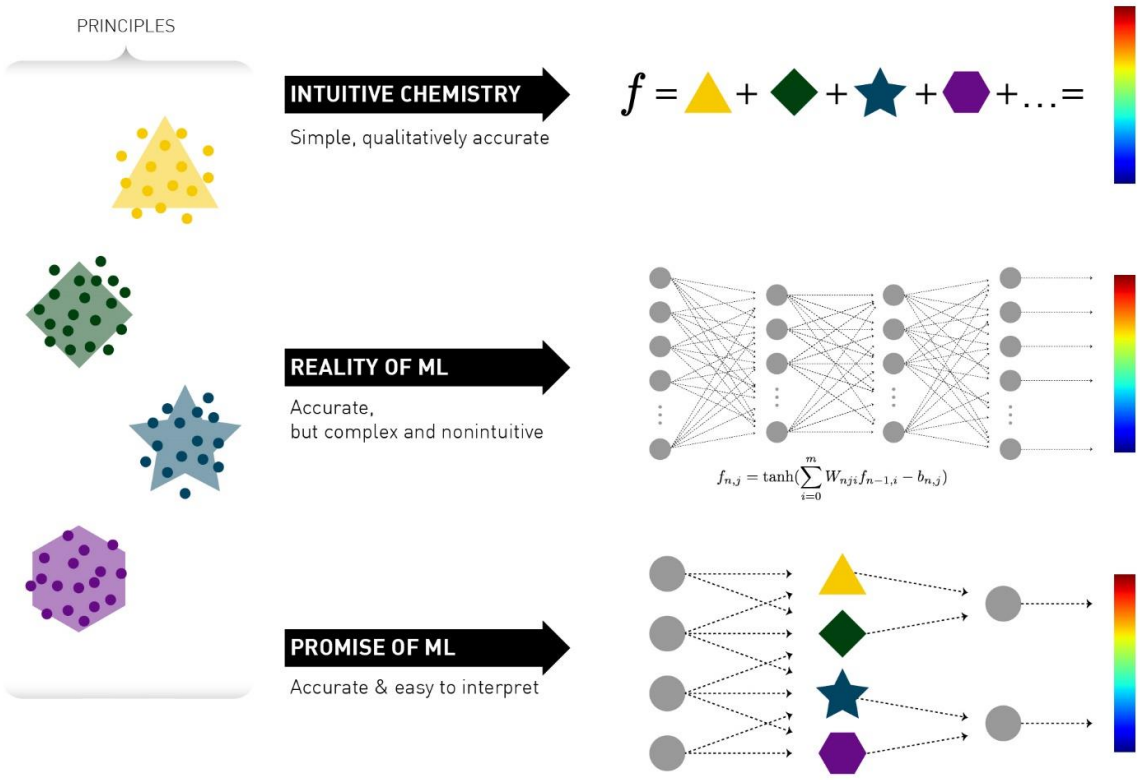


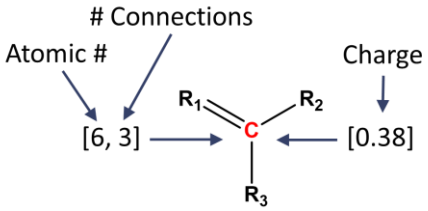
Figure 3-1. Overview of status of machine learning for chemical reactions. The popular deep neural networks are shown in the middle row, where the internal "hidden" representations are hoped to be equivalent to the third row, where the principles behind the predictions are chemically intuitive concepts.

A First Challenge: Representing Chemical Data

For algorithmic techniques to learn relationships between chemical properties and reaction outcomes,[88,89,93–98] the representation of those features is vitally important. A basic principle used here and elsewhere[66,67] is to consider reactions as being composed of bond breaking and bond forming events. This places the features squarely into the chemical domain, and automatically injects accepted chemical principles into the choice of representation: chemical bonding is an *a priori* accepted concept that does not need to be "learned" by the machine. This assumption in turn allows each reaction to be expressed in terms of atom-centered properties (possibly including

neighboring atoms, next neighbors, etc.), such that characteristics of the features are dominated by properties of the reactive atoms. The choice of reactive-atom-centered properties therefore gives a list (a vector) of real numbers that specify a particular reaction. Many choices are conceivable for this feature list.

To represent an atom, one approach is to consider features of the molecular graph centered on the (reactive) atom (Scheme 3-1). Prior efforts in this area have used graphs in a similar way, where in some contexts the assignment of this graph is a key step to classify reactions,[70] and in others, graphs are key frameworks for the ranking of reactions.[66,67,72,73] To form such graphs in the present context, the atomic number, number of covalent bonds, and formal hybridization can be used, where hybridization can usually be inferred from the former two properties. To build a more detailed picture of the atomic environment, these three features can also be added for the atom's neighbors, or next neighbors, as appropriate. While the features themselves are easy to determine, a number of atoms are involved in any particular reaction. The order of these atoms in a feature vector may influence a machine learning algorithm's results, so in this work the ordering of the atoms is standardized according to a prescription given in the computational details section.



Scheme 3-1. Atomic representations based on atomic connectivity and first principles computation. Similar features are available through the neighbors to the central atom, allowing more contextual information to inform the model.

Atomistic simulations can also be used to derive properties of atoms and molecules using procedures that are now considered routine. These techniques can provide a wealth of chemically relevant information, for instance energies and shapes of molecular and atomic orbitals, atomic charges, molecular multipole moments, and excitation energies. While more expensive to calculate

than graphical features, these features are expected to provide more precise, physically meaningful information compared to purely graphical features. In this work, charges and effective hybridization (i.e. a measure of s/p character for an atom) from natural bond order[99] (NBO) calculations are specifically considered as chemically informative atomic features.

In addition to graphical and quantum-chemical features, the energy of reaction is a particularly informative feature for predicting reaction outcome. The energy of reaction (ΔE) is simple to compute with quantum chemistry and provides a basic thermodynamic principle that directly relates to reaction outcome: increasingly positive energies of reaction correspond to reduction in reactivity. ΔE for a single reaction can be found in seconds to minutes on modern computers, and the activation energy—which will be the focus of the predictions herein—costs at least an order of magnitude more computational time, even with advanced algorithms for its evaluation.[100,101]

Relationships Between Representations

To understand how choices of feature representations affect ability for machine learning to predict reaction outcomes, a machine learning model was set up based on two databases of chemical reactions (723 elementary steps, and 3,862 elementary steps). These reactions—described further in the computational details—come from first principles atomistic simulations of reaction pathways.[102,103] The simulations cover two reaction classes: one of interest to atmospheric chemistry,[104–107] and the other to CO_2 reduction chemistry.[108–110] The choice of this dataset allows two significant advantages over other datasets: 1. Activation energies are available for feasible as well as infeasible reactions, and 2. Noise and uncertainties are decreased, as all datapoints were generated with the same simulation method. In sum, the two datasets include a

host of polar and radical reactions, involving unimolecular and bimolecular elementary steps. While we report primarily on the first dataset in this chapter, the Appendix will show that the second dataset behaves similarly to the first, with little differences in statistical errors and interpretation compared to the first dataset.

Two types of regression techniques were chosen as nonlinear machine learning models for further study: neural networks (NNs) and SVM. Both are considered powerful tools with strong theoretical foundations[80,111] in the machine learning community, but the SVM provides simpler, less ambiguous choices of model setup compared to NNs. Vitally, the NN approach is believed to be able to form internal features that represent the core quantities for accurate predictions. To test this hypothesis, a number of network topologies were constructed and tested, with the most generalizable model being presented in the main text (see Appendix for full details). These methods are therefore expected to predict activation energies for chemical reactions to high accuracy, assuming that the input feature representation is meaningful. In addition, the least-squares (LS) variant of SVM—LS-SVM[112]—can provide error bars on all predictions, giving it an internal validation metric to gauge generalizability.

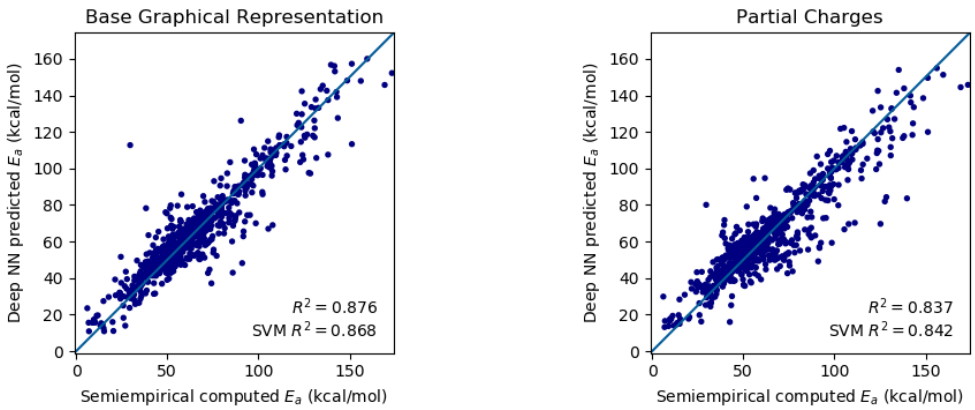


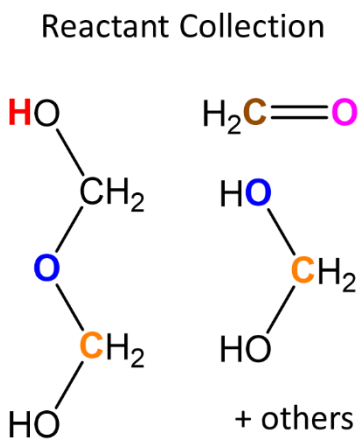
Figure 3-2. Comparison of graphical and quantum chemical feature sets in deep neural network modeling.

For the first round of machine learning modeling, graphical features of reactive atoms, augmented by the energy of reaction, were utilized as features for the NN and the SVM. Upon cross-validation and testing on data points outside of the training set, a good correlation (NN: R^2=0.88 SVM: R^2=0.87) is found between quantum chemical activation energies (E_a) and machine learning estimates of the same quantities (Figure 3-2, left). While higher R^2 values have been found for larger datasets with millions of data points (e.g. potential energies from quantum chemistry),[113,114] these R^2 values are more typical of machine learning studies of chemical reactions.[115] The Appendix shows the error distribution for SVM matches the expected error distribution over the entire dataset (Figure 3-10), indicating that these error estimates are reliable. Similar models without graphical features or energy of reaction showed much lower R^2 values (Figure 3-10). In short, NN and LS-SVM using the chemically relevant graphical and reaction energy features provided quantitative estimates for activation energies that it was not trained on, and reasonable estimates of uncertainties in the LS-SVM case. By these statistical metrics, NN and SVM are each successful at learning activation barriers from first principles simulations.

Next, the quantum chemically derived atomic charges were used as features in place of the graphical features (Figure 3-2, right). Being sensitive to electronic structure of the reactive

molecules and atoms, these charges should in principle be more detailed descriptors than graphical features. The quantum chemical features performed similarly to purely graphical features in terms of test set R^2 (SVM: 0.84 vs. 0.87 NN: 0.84 vs. 0.88). Correlations between predicted and actual error (Figure 3-10) further show that LS-SVM can predict activation energies just as well using either graphical or quantum chemical features, with consistent uncertainties. While the NN provided a slight advantage using graphical features compared to the atomic charges, the difference was not dramatic.

The similar utility of graphical and electronic features suggests that the two sets contain similar information. We hypothesized that one feature set implies the other: the atomic connectivity around each reactive atom dictates the physical charge. To test this hypothesis, all molecules in the benchmark set were collected, and specific atom types extracted based on the graphical features. For example, a trivalent, sp^2 carbon would be one atom type, distinct from a tetravalent, sp^3 carbon. Atomic charges across this set were averaged on an atom-type by atom-type basis, yielding a lookup table that maps atom type to a characteristic charge. The mean change in charge associated with this averaging is small (0.05 a.u. vs. the original charges), suggesting that the charge assignments are reasonable.



Atomic #	Valency	Examples	Partial Charge	Mean Charge
8	2	O	-0.57	-0.56
		⋮		
		O	-0.55	
8	1	O	-0.42	-0.42
6	4	C	0.25	0.24
		⋮		
		C	0.23	
6	3	C	0.23	0.23
1	1	H	0.15	0.14
		⋮		

Figure 3-3. Method for generating the average charge features. First, the reactant molecules are collected and charges are computed for all atoms. For each atom in all of these reactants, atoms with equivalent connectivity are aggregated, and their partial charges averaged. The mean charges are used for all atoms of each respective type in machine learning.

The NN and SVM models trained on the graphically derived electronic properties of atoms (Figure 3-4, top left) show similar prediction accuracy for SVM (R^2=0.83) and slightly worse for NN (R^2=0.80). This similarity suggests that the graph implicitly contains sufficient information to reproduce meaningful electronic features, which in turn work well in building effective NN and SVM models. For the purposes of predicting activation energy in the benchmark set of reactions, these qualitatively different feature sets appear to be equally successful. Up until this point, the NN and SVM modeling of elementary chemical reactions of main group elements is performing well, and has no obvious deficiencies.

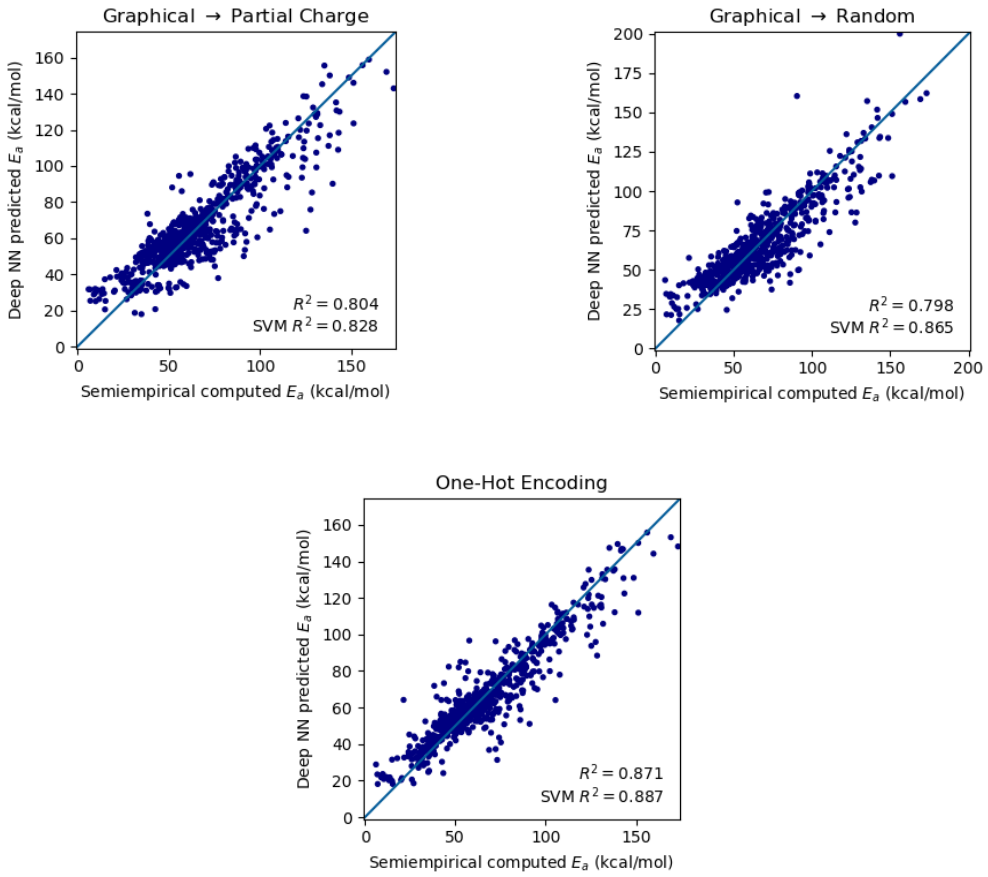


Figure 3-4. Top left: NN results using electronic features derived from graphical features. Top right: NN results based on random values of atomic charges. There is no physical meaning to these charges in the sense that they have no value in representing Coulomb interactions. Bottom: One-hot encoding of reaction types using graphical atomic features.

Deconstruction of Machine Model-Making

At this point in our study, an important insight has been gained with respect to representing chemical information. When expert chemists look at a 2D chemical structure (e.g. a ChemDraw), deep properties are inferred based on their knowledge, intuition, and experiences. Chemists can identify reactive centers, hypothesize the most likely transformations to occur, and propose experiments to reduce uncertainty in challenging cases.[116,117] This expert skill is the concept-

centered approach mentioned in the introduction, which relies on physical properties inferred from the 2D structure (for example, atomic charge).

Since a 2D chemical structure is equivalent to its graph, one might suppose that the machine is inferring principles and properties in a way similar to the expert. The graph implies electronic features, which are the same physical properties that dictate chemical reactivity. While this is easy to imagine and is the hoped-for goal of machine learning, such principles are by no means necessary for nonlinear machine learning tools to provide quantitative accuracy. Not only could the machine develop an entirely alternative viewpoint not held by chemists, it could also be making predictions using properties an expert would consider physically incorrect.

The second possibility appears to be closer to the truth. As the next numerical experiment, the machine learning models were built using *random* values of atomic charge. Instead of using (physically meaningful) average values of charge from graphically derived atom types, each atom type was assigned to a random number from a standard Gaussian distribution. Using the randomized "charges", the two machine learning models performed similarly to the previous models, with R^2=0.86 for SVM and R^2=0.80 for NN, showing approximately equal quantitative accuracy (Figure 3-4). The atomic charge used by SVM therefore must be a *label*, not a physical measure; increasing or decreasing this number does not reflect a varying chemical environment, but simply a renaming of the label. Adjacency or proximity between two of these charges holds no particular meaning, as the random charges have no particular relationship with physical charge.

Reestablishing Chemical Concepts

If electronic or graphical features of atoms are simply labels, it is likely that using "good" labels would yield a somewhat better procedure. An improvement in accuracy should result

because the charges might be mistakenly seen by the NN or SVM to be "ordered" (…-0.2 < -0.1 < 0.0 < 0.1…), which is unrealistic given that the actual ordering is random. A good labeling procedure would not entail any artificial ordering, and this can be done with one-hot encoding. This encoding entails constructing a set of features with values of 0 or 1, where each feature is treated independently of the others. A single one-hot feature corresponds to a particular assignment of atom type based on the graph, just like in the feature averaging strategy discussed above (but with no charge assignment).

A small increase in machine learning predictive performance is observed when using one-hot encoded atom types, giving a test set R^2 of 0.87 (NN) and 0.89 (SVM) (Figure 3-4). This R^2 is slightly higher than that of the random features, and close to or better than the best-case models with the other feature types (0.88 NN and 0.87 SVM). This result suggests that the machine learning models using labels of atomic type appear fully sufficient to reach quantitative accuracy. The implications of this simplified feature representation are important to understanding nonlinear regressions in machine learning, and will thus be further discussed.

The high accuracy achieved using one-hot labels challenges whether machine learning requires quantitative physical principles as underlying features for making accurate predictions. Recall that the reaction feature vector is simply a composite of the atomic features of reactive atoms, augmented by the energy of reaction. Where graphical features and properties derived from quantum chemistry remain close to basic principles such as periodic trends, covalency, and electronic structure, atom labels contain no such properties. A one-hot encoding of a 3-valent carbon is equally different from a 2-valent carbon or a hydrogen in an O-H bond. In other words, all one-hots are unique labels with no special relationships to each other, much less physical relationships. This uniqueness means that (in the feature set) a pair of atom types of the same

element are just as different from each other as a pair of atom types with different elements! Periodic trends, bonding patterns, and electronic properties are lost to such atom labels that do not contain this information.

To push this hypothesis even further, a *k*-nearest neighbors model was applied to the dataset using the base graphical features. With $k = 2$, predictions are made by assuming that the average of the two most closely related data points gives the unknown data point. In this case, an R^2 of 0.86 on the test sets was achieved with the one-hot encoding feature set (Figure 3-11). This surprising result suggests that machine learning is doing little more than memorizing,[118] as predictions are made to reasonably high accuracy by mere similarity with training data points. No believable trends in physical properties are possible using only pairs of data points.

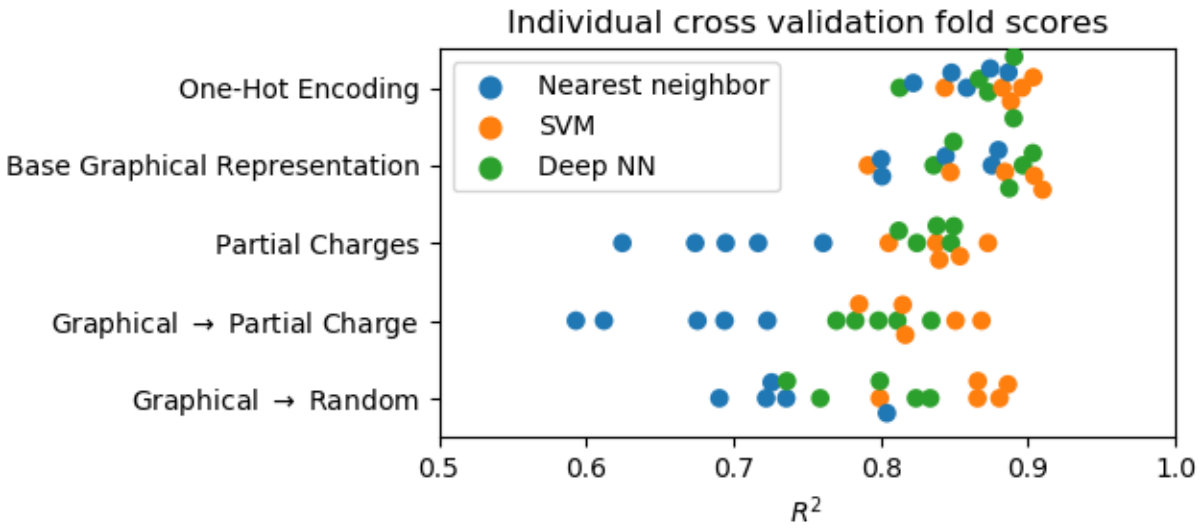


Figure 3-5. Comparison of three machine learning approaches using various representations of the underlying features. Each filled circle line is an R^2 on a cross-validated test set, so there are 5 R^2 values per method/feature combination.

The analysis so far (Figure 3-5, and statistically summarized in Table 3-3) suggests that the nonlinear regressions of this work are largely agnostic to the underlying feature representations (with the exception of the energy of reaction, which is important and we will focus upon shortly). The Appendix shows analysis of a larger dataset, with one order of

magnitude additional data points (3862); no qualitative change in outcome was observed, and only minor differences in quantitative accuracy were found. We therefore ask whether a highly simplified representation of chemical information may be just as effective as the machine learning. When atomic features are represented by simple labels, reaction types therefore are just composites of these labels. Incidentally, chemists have worked with labeled reaction types for centuries: they are called *named reactions*. For each reaction type, simple relationships have been developed to relate molecular properties to reaction rate. This approach will provide a much more transparent picture of reactions than nonlinear regression.

Evans-Polanyi Relationships

At this point, it is clear that the machine learning views reactions categorically, rather than by any deeper physical relationship. The well-known Evans-Polanyi relationship can also do the same, where a linear trend between activation energy and energy of reaction is constructed. The statistical errors on the top-10 most prevalent reaction types are shown in Table 3-1. Comparison of statistical accuracy of Evans Polanyi compared to SVM and NN for common reaction types (RMSE, kcal/mol). Evans Polanyi errors are based on leave-one-out cross validation with RMSE reported for the hold-out points.. In this data set certain reaction types appear repeatedly, and the trends in reactivity fit well to the linear relationship (first row). The SVM model is able to perform almost as well as Evans-Polanyi for the same reactions, with an overall RMSE about 6% higher. The NN model is similar, at 5% higher overall error than Evans Polanyi. This trend remains when analyzing the full data set, shown in *Figure 3-6*, which affirms that the Evans Polanyi is slightly numerically improved over the SVM and NN models. See the Appendix, Figure 3-16, showing that the same picture holds when analyzing the second data set, which was generated using Density Functional Theory.

Table 3-1. Comparison of statistical accuracy of Evans Polanyi compared to SVM and NN for common reaction types (RMSE, kcal/mol). Evans Polanyi errors are based on leave-one-out cross validation with RMSE reported for the hold-out points.

	1	2	3	4	5	6	7	8	9	10		**Total**
Evans Polanyi	5.00	4.98	4.69	4.86	5.12	4.13	6.63	9.09	6.12	1.99		**5.35**
One-Hot SVM	5.69	6.41	4.45	5.70	4.64	4.67	6.12	7.34	7.24	2.56		**5.68**
One-hot DNN	5.71	5.93	5.84	4.73	4.13	5.01	5.54	8.42	5.90	3.07		**5.62**
# Data Points	44	39	26	21	18	15	15	15	15	15		**223**

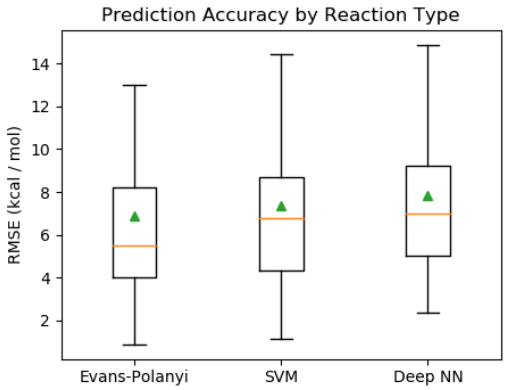


Figure 3-6. Error distributions for all Data Set 1 reaction types with at least 3 data points.

Figure 3-7 shows a hydrolysis reaction as an interesting example (reaction type 1 of Table 3-1). The Evans-Polanyi relationship on these 44 data points gives an R^2 of 0.74, and provides a simple interpretation: water-assisted elimination of ROH at an sp^3 carbon has barriers that trend with energy of reaction. While this statement is not particularly profound, it is easily constructed and can be performed for any reaction type represented by at least two points in the dataset. Further analysis of the data in Figure 3-7 (top), however, shows this reaction is somewhat more nuanced. While in the original feature set rings were not identified, these were found to be important. The

data points of Figure 3-7 therefore divide themselves into two sets: **A**. reactions without 4-membered rings, and **B**. reactions involving 4-membered ring breakup. The **B** reactions break the 4-membered ring and release significant strain, and sit to the left of the other data points in Figure 3-7 (lower ΔE). In region **B**, the Evans-Polanyi relationship has a nearly flat slope. Removing these data points increases the R^2 of the **A** region to 0.81, indicating an improved linear fit. Predicting **A** and **B** data regions separately gives an overall RMSE of 3.37 kcal/mol compared to 4.40 kcal/mol for the original, single Evans-Polanyi relationship.

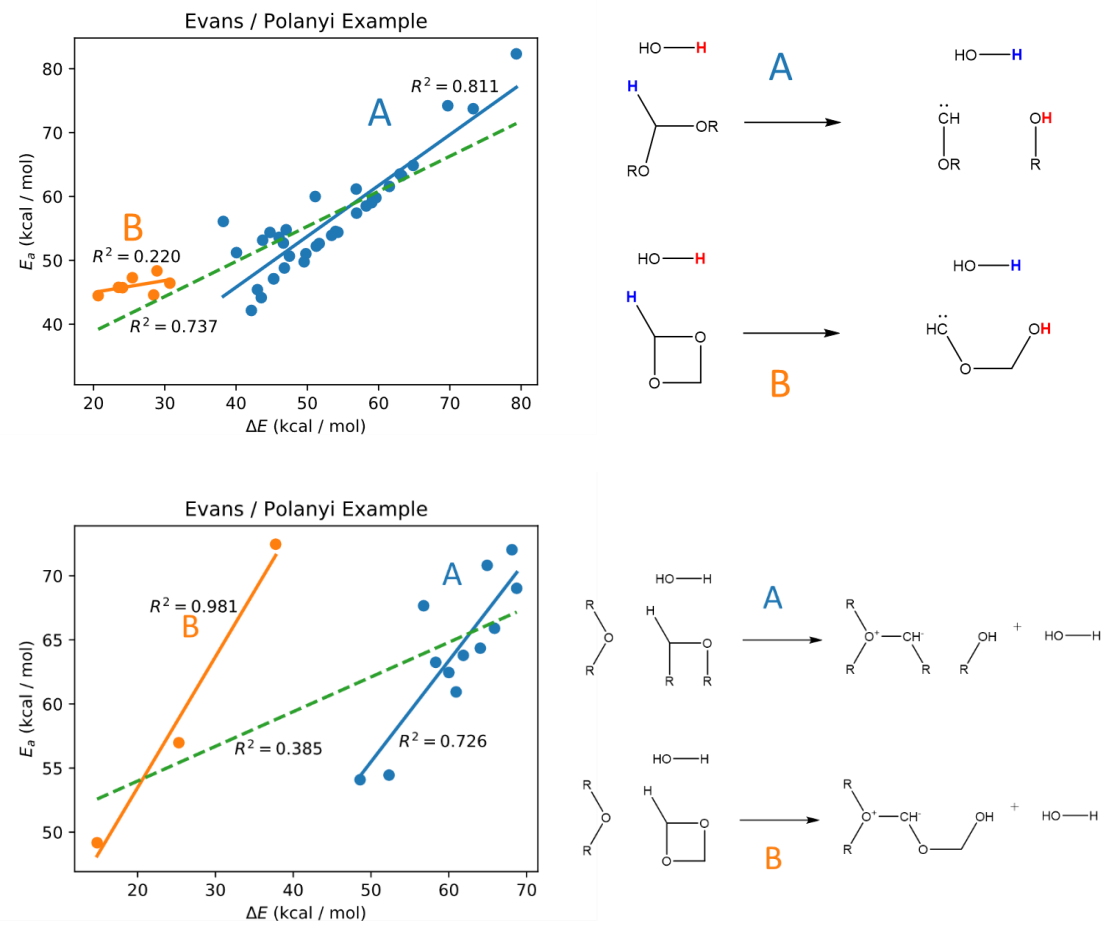


Figure 3-7. Top: An example Evans / Polanyi from a reaction type with many examples in the dataset. Bottom: Bimodal Evans / Polanyi for a second reaction type. The dashed green lines represent the (poor) linear fits when including all data points.

The Evans-Polanyi relationship can break down within specific sets of reactions, giving an indication that the chemistry is more complex than originally envisioned.[119] For example, an Evans-Polanyi plot with a multimodal structure suggests that there are significant mechanistic differences within the reaction type.[120] One such "bad" Evans-Polanyi relationship was easily identified within the dataset.

The reaction type of Figure 3-7, bottom illustrates this point well (reaction type 9 of Table 3-1). The single-line relationship is poor (R^2=0.39), and 3 points on the left appear to be well-separated from the points on the right. While this is insufficient data for statistical significance, mechanistic differences are responsible for the bimodal structure in this example. Examining the individual reactions revealed that the 3 data points differed qualitatively from the others, and involved release of strain from a 4-membered ring. This shifted the reaction energies (ΔE) significantly downward for elementary steps that otherwise had the same reaction classification. Dividing the two cases based on the ring-release criterion provides two Evans-Polanyi relationships with R^2 of 0.98 and 0.73, indicating good fits to the linear relationships.

Discussion

The above results and analysis of a chemical reaction data set highlights a certain tension between machine learning and chemical approaches. Whereas chemistry usually seeks explanations based on physical properties—and inherently cares whether those physical properties are real—machine learning approaches can reach their criteria for success (test-set statistical accuracy) without achieving a convincing relationship to chemical principles.[16,118] While the machine approach could in theory provide physical relationships, there is no reason to believe this will come automatically with currently available algorithms, which are agnostic to expert

knowledge. In the cases examined above, it is reasonable to conclude the machine learning models do slightly more than memorizing values from clusters of data points, where those clusters happened to be similar reaction types.

This limitation applies just as well to similarity-based SVM models as to deep NN machine learning tools. In the latter case, NNs provide no obvious correspondence between their hidden representations and chemical concepts, though in principle these hidden representations could be valuable. Such a valuable hidden representation, however, is clearly not present when formed in the two datasets of this study, as the NN was unable to generalize its predictions beyond the specific reaction types that appeared in the input vector.

The two questions posed in the introduction (What is the machine's representation of knowledge? What does the machine learn?) can be succinctly answered, at least in the case of the NN and SVM models used herein. Since NN and SVM recognize similarity between data points, it does not appear to greatly matter what form the input data comes in. Since the features can take many forms and still discriminate between reaction classes, these features need not be physically grounded. SVM therefore learns to recognize reaction types based on similarity within an abstract feature space. The NN performs similarly, does not provide any additional generalizability, and does so in a less transparent manner. While it is possible that machine learning through NNs can provide improved representations of chemistry with larger datasets, no improvement in statistical accuracy was found on a second dataset with 3,862 reactions (see Appendix, especially Figure 3-16).

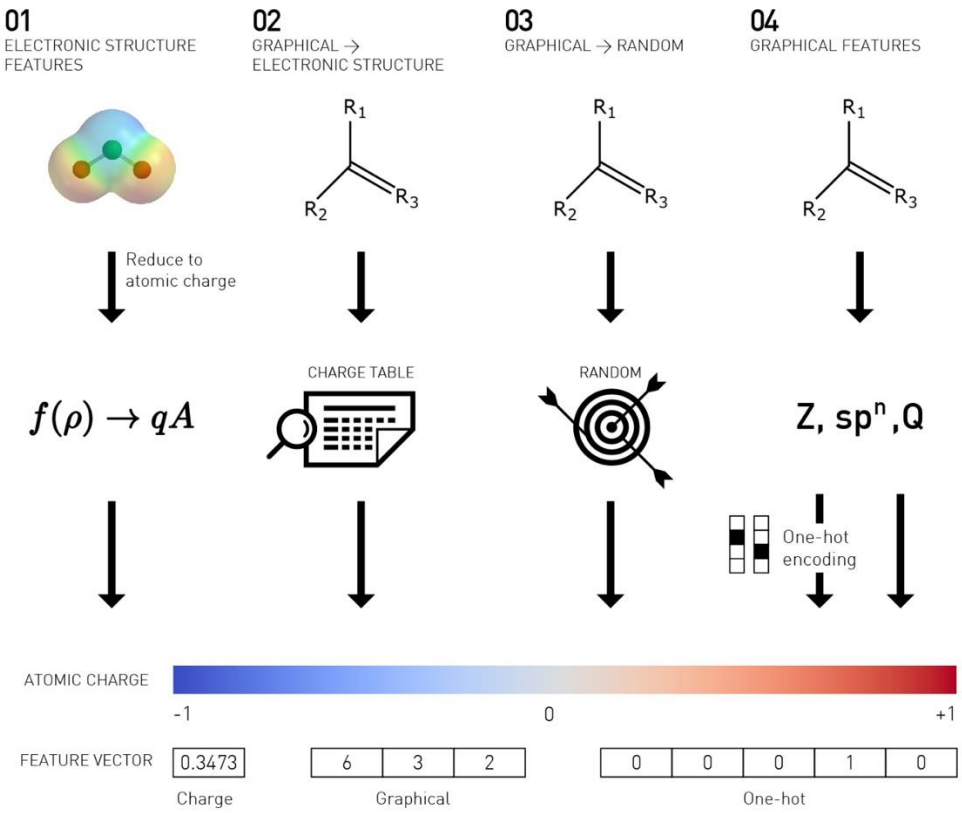


Figure 3-8. Summary of feature experimentation steps. All feature types produce similar results in deep neural network or SVM regression, including random atomic charge assignments and one-hot labels. The machine learning algorithms treat all atom types as completely unique, and essentially unrelated to one another.

Despite these concerns, however, machine learning still has strong abilities. It can operate directly on data and quickly give quantitative accuracy, in contrast to the chemical approach which relies on existing knowledge and highly developed insight. Certain questions of value therefore deserve further consideration:

1. Does the method solve an unsolved chemical problem? Or does it simply reproduce what is known?
2. Does the method offer clear advantages in time to solution compared to existing approaches?
3. Does the method provide transferable chemical insight, where transferable refers to ability to work well outside of the current dataset?

In our opinion, contemporary approaches used by expert chemists address points (1) and (3). New approaches for handling chemical problems are being developed by domain scientists for (2). In the area of chemical reactions, some progress has been made using machine learning to achieve (2) as well, but not necessarily (1), and a few examples of (3) within specific domains.[14,55]

While there remains a lot of room for new machine learning approaches for chemical problems that may perform at a much higher level, one fundamental difficulty remains.

Figure 3-1 compared three types of models for relating data to predicted outcomes. The first most closely resembles expert procedures, where knowledge is represented in precise, explainable concepts developed over years of experience. These concepts are clearly understood, and chemists know the contexts in which each concept may be applied. In many cases, simple mathematical expressions can be written down that show the relationship between the physical properties and the outcome of interest (i.e. Table 3-1 and *Figure 3-6*). In the second case (in the middle of Figure 3-1), machine learning performs a complicated transformation of raw features into a hidden representation, which in turns leads to quantitative predictions. The second case provides no clear interpretation of how it obtains its high accuracy, and this is essentially what is expected of current-generation machine learning methods. In the third case shown at the bottom of Figure 3-1, an idealized machine learning setup takes raw chemical features (e.g. graphs), and relates them to concepts that are recognizable to chemists. This represents an automatic reduction in dimensionality of the feature set into more concise features that are primarily predictive of outcome. While this is a beautiful procedure, more work will be needed to achieve such a goal.

While these three procedures may seem like three equivalent means to the same end, in practice this is far from the truth. The two procedures using interpretable features employ a *low-dimensionality, transferable representation* of the chemical information, which is an incredibly important advantage (Figure 3-7). With a low-dimensionality representation, predictive accuracy can be obtained with exponentially fewer data points compared to a high-dimensionality representation.[121] Consider for instance the (linear) Evans-Polanyi relationship: given perhaps 3 data points, the data can be fit and predictions made. An SVM or neural network with an input

feature vector of dimension 10 can do little to nothing with 3 data points. In addition, chemical principles are backed up by physical considerations, making them much more likely to be transferable outside of the current training/test set. For example, in polar reactions the Coulomb relationship states that positive and negative charges attract, leading to faster reactions (and physical charges are required to capture this relationship in full). Physical models built directly from physical features will therefore be the most generalizable predictive tools.

The low dimensionality representation of knowledge expressly used by expert chemists allows them to operate in uncertain domains and make considerable progress in developing new chemical reactions. Machine learning in high dimensional spaces is, on the other hand, unlikely to provide any value for new chemistries where the number of data points is low. The concern raised in question (3) seems to require low dimensionality and an underlying physicality in models and feature space, which deviates substantially from contemporary machine learning methods.

Conclusions

The present investigation started with an analysis of feature representations for machine learning of chemical reaction barrier heights. Atomic labels that lacked physical trends were found to be the basis for which the model made its predictions, and recognition of reaction types was the full basis for this model. This analysis showed that the machine learning method was simply recalling reaction types, and we therefore give a tentative, weak answer to "What does the machine learn?" The machine learns to recognize the reaction types that were already encoded directly in the input features.

The machine learning model was subsequently replaced by a simple, well-known chemical principle called the Evans-Polanyi relationship. Statistically, the linear Evans-Polanyi model slightly outperformed the nonlinear machine learning models (by about 5% RMSE), and provided

a simple interpretation of the results. This low-dimensionality model (2 parameters per reaction type) is algorithmically and conceptually easier to apply, and can be evaluated using chemical principles, making it transferable to new reactions within the same class. While Evans-Polanyi relationships are not expected to be universal,[119,120] they provide a metric for reactivity that can be easily applied and tested, and give a starting point for more complex models to be proposed.

The interpretable superiority—alongside reasonable statistical accuracy—of a simple chemical relationship compared to nonlinear machine regression suggests that deeper analysis is needed of machine learning methods for chemical sciences.[16] The approaches should not be used as black boxes, and careful investigations are required to reveal whether simpler, more easily interpreted methods could replace the complex workings of these machines. It should be recalled that machine learning tools have seen their greatest benefits when working with giant datasets that are not well-understood. Chemical research is not necessarily in this limit: chemists understand their data and do not necessarily have available millions of poorly understood data points that are ripe for machine learning models.

Computational Details

Reaction Representations

To represent a reaction, which involves bond forming and/or breaking events, the representations of the two atoms involved in the bond were concatenated. Consistency in ordering is important to ensuring that driving coordinates involving the same atoms are treated appropriately when algorithmically learning. Therefore the atoms' representations were sorted in descending order, which provides a unique representation. Due to this ordering, however, if two driving coordinates share an atom in common, it is possible that the two driving coordinates will appear to have no atoms in common.

Representing a reaction using a collection of bond changes is somewhat complex, however, due to the two types of driving coordinates (formed and broken bonds) and a variable number of driving coordinates of each type. Therefore separate representations for the sets of formed and broken bonds were created and concatenated. For each type's representation we utilized pooling to generate a fixed length representation from a variable number of driving coordinates (Scheme 3-2). Min, mean, and max pooling were tested as each of these seems plausibly important in conveying chemical meaning, with mean pooling not utilized in the final feature representation. Our representation also tested a few reaction level features in addition to the aggregate atomic representations. These were the number of bonds formed, number of bonds broken, and ΔE of the reaction (the former two were not used in the final machine learning strategy). While obtaining ΔE requires geometry optimizations, this step is much lower in computational cost than optimizing a reaction path including its associated transition state.[101] The various atomic feature sets examined in the main text are denoted in Table 3-2.

Scheme 3-2. Graphical feature vector for machine learning applications. While more complicated feature vectors were examined (e.g. including nearest neighbor atom descriptors), none showed substantial improvement over this simple choice. See the Appendix for additional test cases.

Feature vector (graphical feature sets, for results reported in main text)

ΔE	Max(add)	Min(add)	Max(break)	Min(break)

Representation of additions or breaks to covalent connections graph, second line is an example

Higher atomic #	*Coordination #*	*Lower atomic #*	*Coordination #*
8	*1*	*6*	*3*

Table 3-2. Feature sets for atomic representations.

Feature set	Description	Size of atom representation	Overall feature set size (8n+1)
One-hot	One-hot encoded atom type (atom type determined by base graphical representation)	5 (# of atom types in PM6 dataset)	41
Base graphical	Atomic # and coordination #	2	17
Partial charge	Effective atomic charge	1	9
Graphical → partial charge	Average partial charge of all atoms of an atom's type	1	9
Graphical → random	A random real number is drawn from a normal distribution for each atom type. This number is used to represent all atoms of this type.	1	9

Dataset

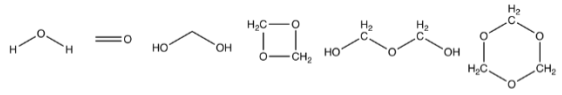
The Z-Struct reaction discovery method[122–124] was used to combinatorically propose intramolecular and intermolecular reactions between small-molecule reactants which include carbon, hydrogen, and oxygen (Scheme 3-3, Dataset 1). Even with these relatively simple reactants, the full extent of elementary reactions that may appear when the species are combined is unknown, due to the significant number of plausible changes in chemical bonding. Based on their relevance to atmospheric chemistries[104–107] and the difficulty in studying the host of possibilities using experiment, details of these reactions are best provided via first principles simulation. For this study, a systematic simulation approach was used to generate this set of possibilities. Specifically, the Z-Struct technique used the Growing String Method (GSM)[101] to search for reaction paths with optimized transition states for each proposed reaction (thousands of

possibilities). Postprocessing scripts then attempted to include only reactions that were unique and well converged single elementary steps. Machine learning tests exposed a few (<10) outliers that passed the automated filters but were clearly incorrect and were manually removed. The PM6 method as implemented in MOPAC[125–127] was used as the underlying potential energy surface. The resulting dataset contained 723 unique reactions from 6 original reactants.

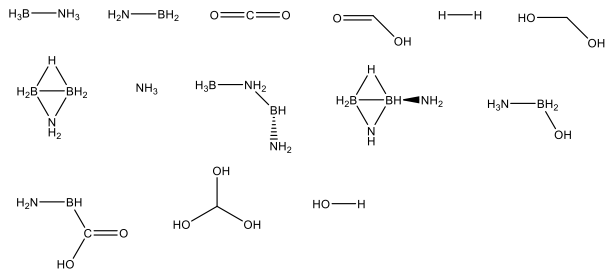
To confirm scalability of the methodology to a larger, higher quality dataset, a second set of reactant molecules was examined (Scheme 3-3, Dataset 2). This larger, more chemically complicated set of reactants was examined at the density functional theory (B3LYP/6-31G**) level, using the same ZStruct/GSM strategy to generate a second dataset of reactions. Dataset 2 includes nitrogen and boron in addition to carbon, oxygen, and hydrogen, so many types of reactions were possible and nearly half of the reactions were the only reaction of their type. These single-instance reactions were removed, leaving 3,862 reactions in the Dataset 2. For analysis on this dataset, see the Appendix. No qualitatively significant changes were observed compared to Dataset 1.

Scheme 3-3. Reactants involved in Dataset 1 and Dataset 2. Results in this chapter from Dataset 1, with Dataset 2 analyzed in the Appendix.

Dataset 1:



Dataset 2:



Machine Learning Pipeline

For the machine learning pipeline, each feature set was extracted from the dataset to give the aggregate reaction representation including the relevant atomic representation of reactive atoms and reaction level features. The features were standardized to zero mean and unitary standard deviation except in the case of one-hot encoding, in which the atomic representation was one-hot encoded and the energy of reaction was scaled to standard deviation of 3 to balance its influence. This reaction representation was provided as input into an LS-SVM[112] with radial basis function kernel that can compute confidence intervals. Since the dataset size is relatively small by machine learning standards, cross-validation was used to tune hyperparameters and generate generalization predictions on all data points. For final predictions, 5-fold cross validation was used for all models. For nearest neighbors, no hyperparameters were trained by cross validation. For SVM, within each split of outer cross validation, hyperparameters for the test set were chosen using 3-fold cross validation within the training folds. Deep NN training was more resource intensive so hyperparameters were chosen globally by 3-fold cross validation on the entire dataset. In the final 5-fold cross validation weights and biases were trained only on training folds but the globally chosen hyperparameters were used for all folds. Data was leaked into the models through comparisons between classes of algorithms and feature sets. Examining extreme outliers in early predictions uncovered a few clearly invalid data points (e.g. reaction profile lacking a single, defined transition state) that evaded automated filters for validating the data generation process, so these data points were removed manually. Additionally, since R^2 is sensitive to outliers and can be dominated by a single extreme outlier, when generating the plots and metrics above all predictions were clipped into the interval [0, 200] kcal / mol.

For the charge averaging in Figure 3-3, the charges for all reactive atoms in all driving coordinates in all reactions in the dataset were grouped into atom types by element and coordination number. Within each atom type, the mean of all charges of all atoms of each type was computed and the charge of each atom within the type was set to this mean charge. This counting strategy implies that, for example, if there are more methanediol reactions involving the hydroxyl hydrogen than the alkyl hydrogen, then the charge on the hydroxyl hydrogen will be effectively weighted heavier in the charge averaging.

Acknowledgements

The authors thank the NSF (1551994) and the NIH (R35GM128830) for support of this work.

Appendix to Chapter 3

Note About Neural Network Topologies

A hyperparameter grid search was conducted using cross validation to determine the neural network architecture for each feature set. Available parameters to the search were: {'depth': [1, 2], 'width': [50, 100, 150, 200], 'epochs': [1000, 5000, 10000], 'beta': [1., 10., 100., 1000.]} where epochs is the number of training cycles and beta is the L2 regularization weight. Dropout with a rate of 0.5 was used throughout. The chosen parameters were then applied to train and predict neural networks with a separate cross validation shuffling to generate the cross validation predictions reported in the main text.

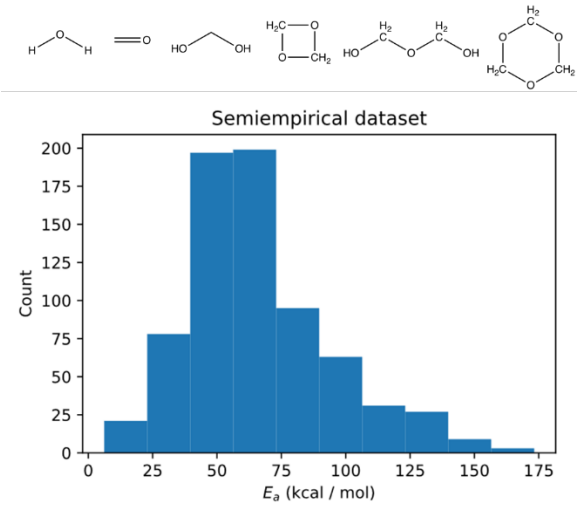
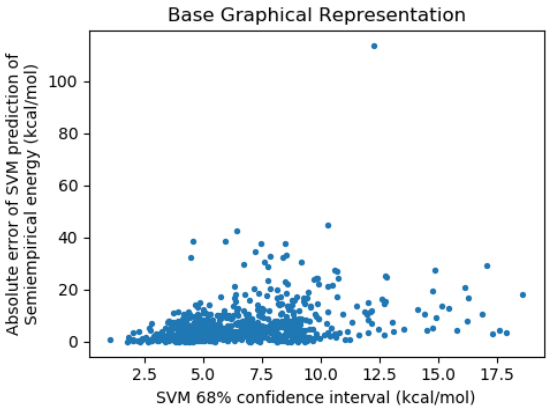
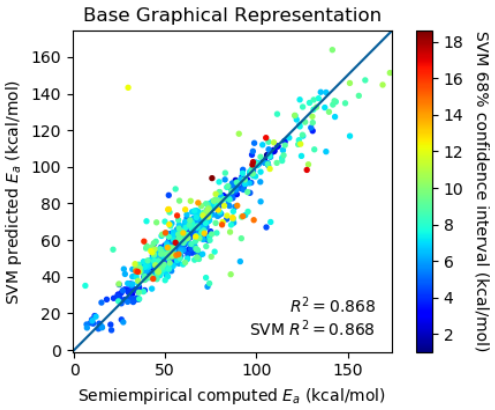
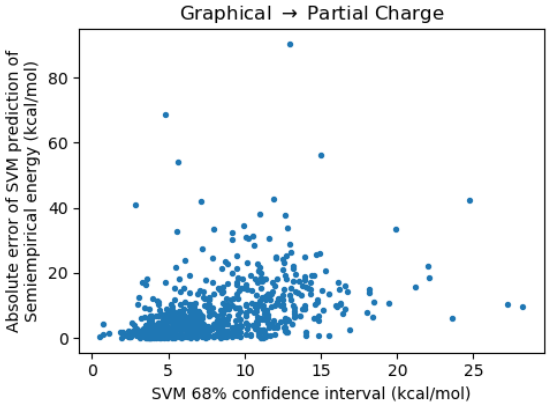
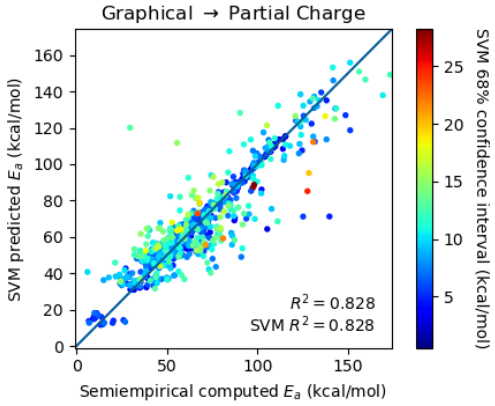


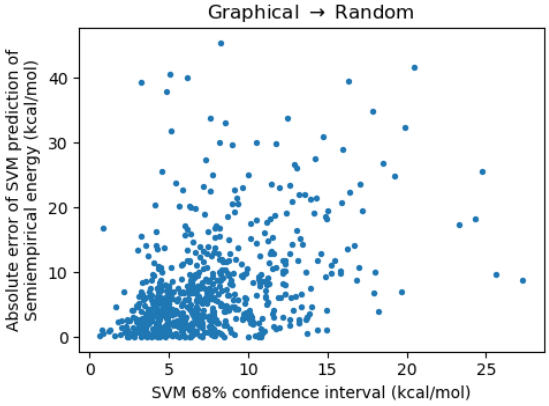
Figure 3-9. Top*: reactants used in PM6 dataset (Data Set 1).* Bottom*: distribution of activation barriers for* PM6 *dataset (energies via MOPAC).*

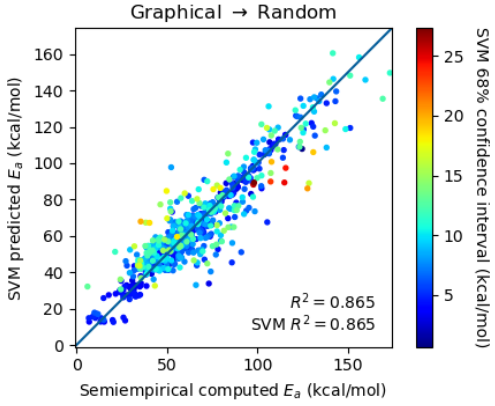

Base Graphical Representation
Absolute error of SVM prediction of Semiempirical energy (kcal/mol)
SVM 68% confidence interval (kcal/mol)

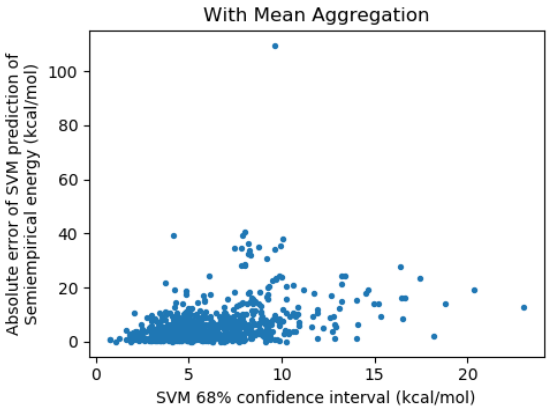

Base Graphical Representation
SVM predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
SVM 68% confidence interval (kcal/mol)
$R^2 = 0.868$
SVM $R^2 = 0.868$

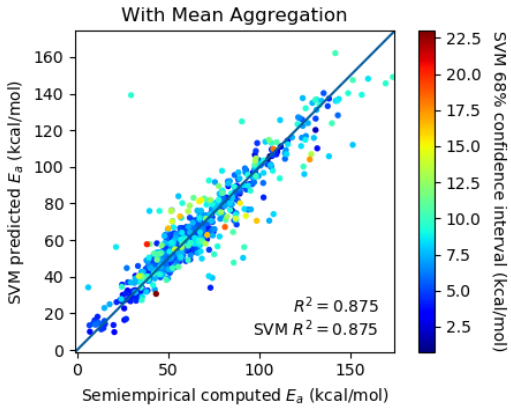

Graphical → Partial Charge
Absolute error of SVM prediction of Semiempirical energy (kcal/mol)
SVM 68% confidence interval (kcal/mol)

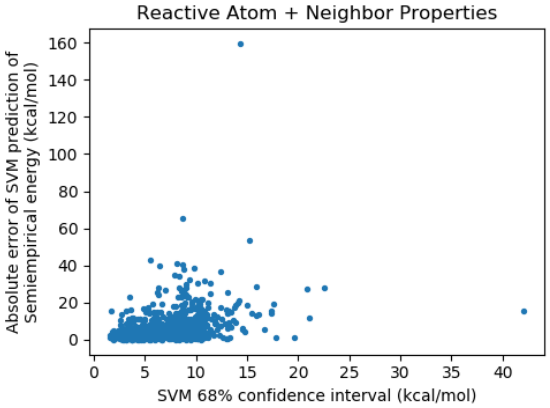

Graphical → Partial Charge
SVM predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
SVM 68% confidence interval (kcal/mol)
$R^2 = 0.828$
SVM $R^2 = 0.828$

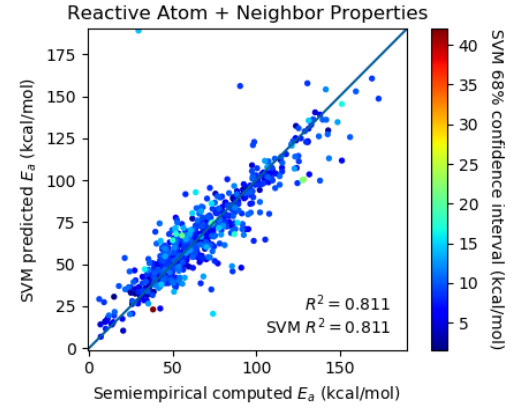

Graphical → Random
Absolute error of SVM prediction of Semiempirical energy (kcal/mol)
SVM 68% confidence interval (kcal/mol)

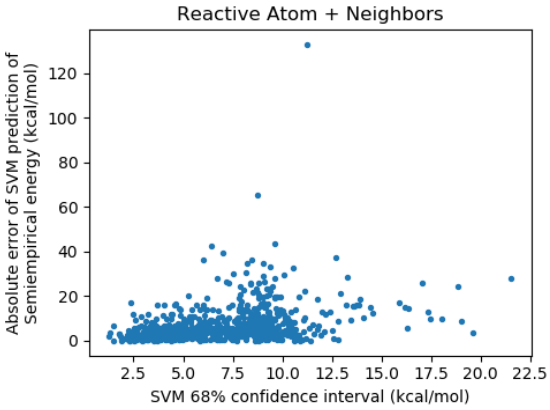

Graphical → Random
SVM predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
SVM 68% confidence interval (kcal/mol)
$R^2 = 0.865$
SVM $R^2 = 0.865$

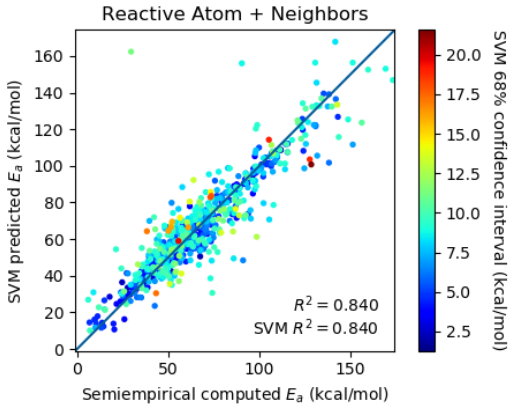

With Mean Aggregation
Absolute error of SVM prediction of Semiempirical energy (kcal/mol)
SVM 68% confidence interval (kcal/mol)

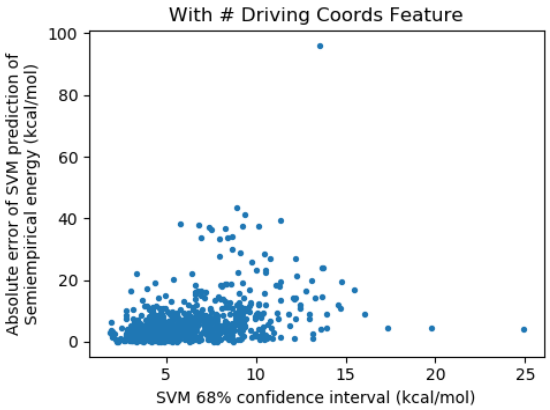

With Mean Aggregation
SVM predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
$R^2 = 0.875$
SVM $R^2 = 0.875$
SVM 68% confidence interval (kcal/mol)

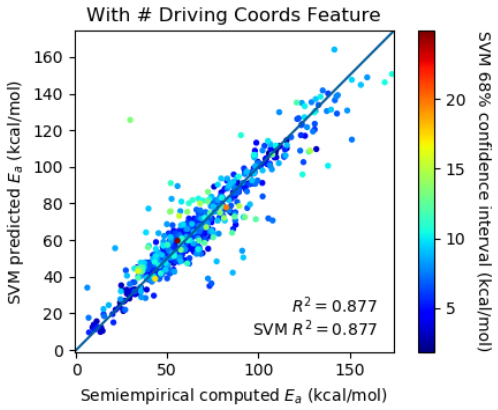

Reactive Atom + Neighbor Properties
Absolute error of SVM prediction of Semiempirical energy (kcal/mol)
SVM 68% confidence interval (kcal/mol)

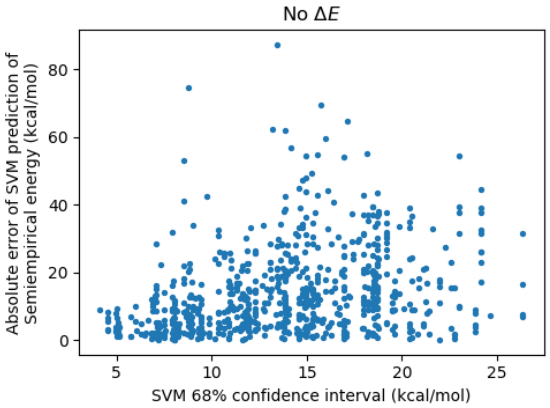

Reactive Atom + Neighbor Properties
SVM predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
$R^2 = 0.811$
SVM $R^2 = 0.811$
SVM 68% confidence interval (kcal/mol)

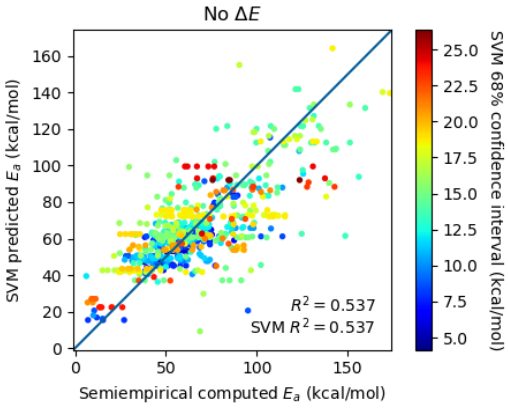

Reactive Atom + Neighbors
Absolute error of SVM prediction of Semiempirical energy (kcal/mol)
SVM 68% confidence interval (kcal/mol)

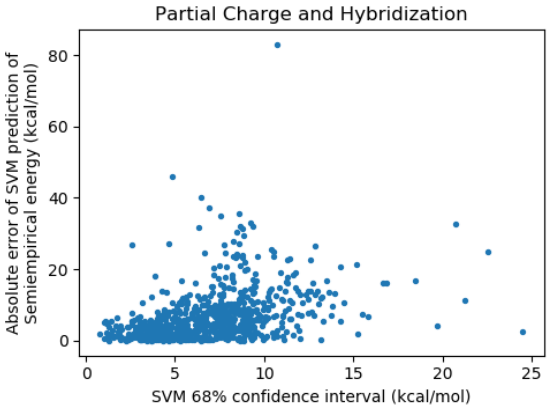

Reactive Atom + Neighbors
SVM predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
$R^2 = 0.840$
SVM $R^2 = 0.840$
SVM 68% confidence interval (kcal/mol)

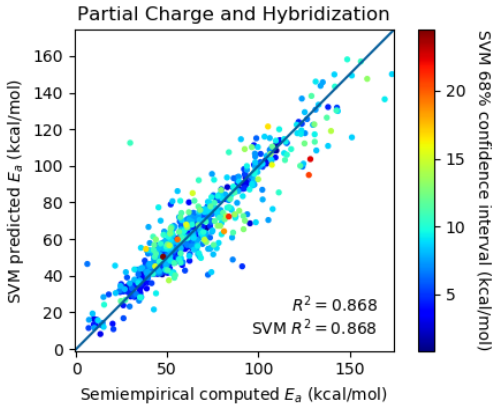

With # Driving Coords Feature
Absolute error of SVM prediction of Semiempirical energy (kcal/mol)
SVM 68% confidence interval (kcal/mol)


With # Driving Coords Feature
SVM predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
$R^2 = 0.877$
SVM $R^2 = 0.877$
SVM 68% confidence interval (kcal/mol)


No ΔE
Absolute error of SVM prediction of Semiempirical energy (kcal/mol)
SVM 68% confidence interval (kcal/mol)


No ΔE
SVM predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
$R^2 = 0.537$
SVM $R^2 = 0.537$
SVM 68% confidence interval (kcal/mol)


Partial Charge and Hybridization
Absolute error of SVM prediction of Semiempirical energy (kcal/mol)
SVM 68% confidence interval (kcal/mol)


Partial Charge and Hybridization
SVM predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
$R^2 = 0.868$
SVM $R^2 = 0.868$
SVM 68% confidence interval (kcal/mol)

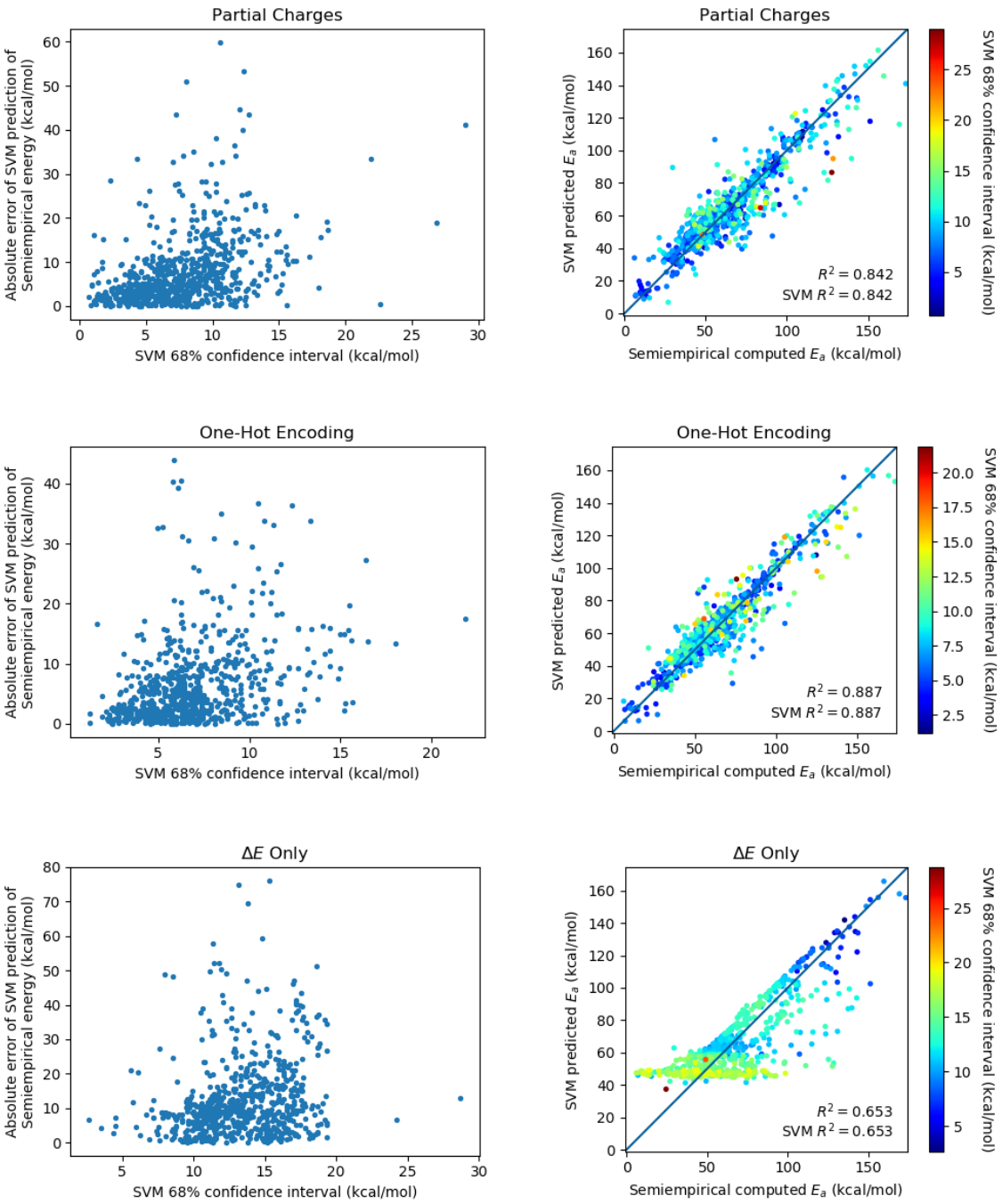
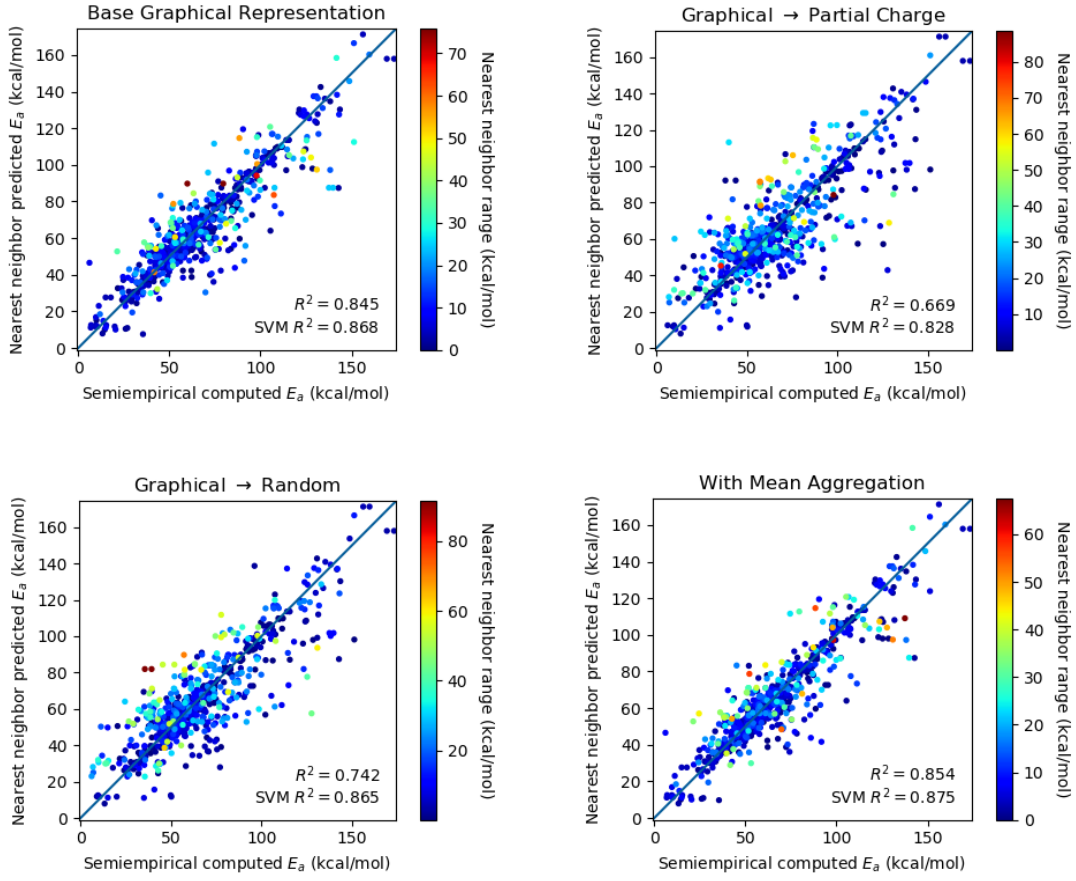
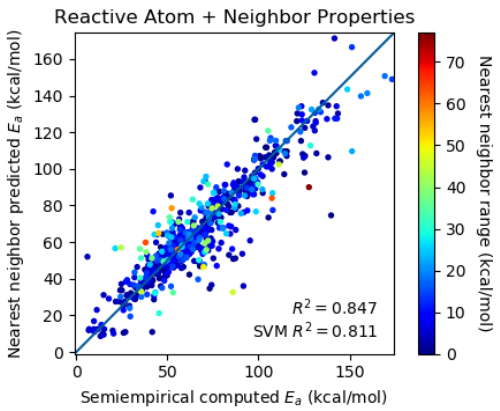


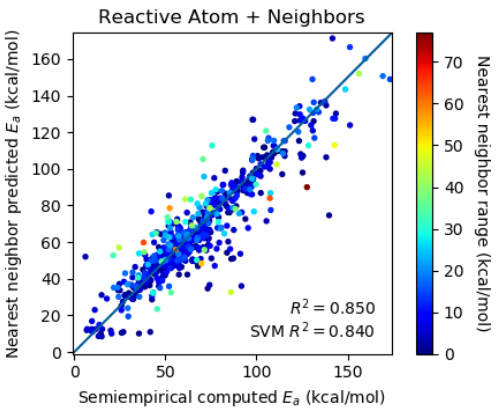
Figure 3-10. Comparison of additional feature sets for the PM6 dataset (Data Set 1). "No ΔE" is the original graphical representation without energy of reaction. "Reactive Atom + Neighbors" is the original graphical extended to include atomic numbers of neighbors. "Reactive Atom + Neighbor Properties" is the same feature set as the previous but including coordination number of neighbors.

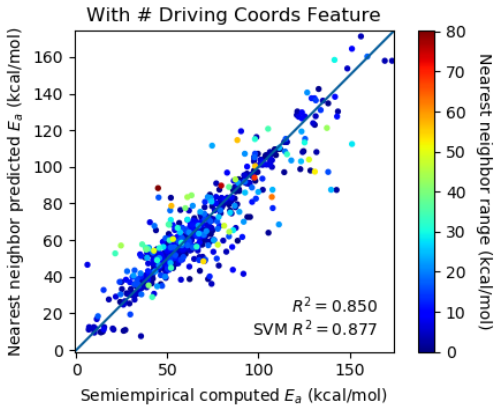
Note About Representing Atoms

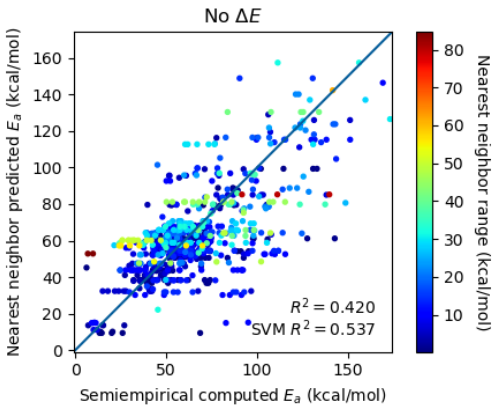
A challenge for representing atoms is aligning atoms' neighbors' representations such that an algorithmic learning technique can appropriately determine similarity between atoms with different numbers and permutations of neighbors. We ordered adjacent atoms in descending order using their representation as a key so that highest atomic number neighbors are first. Zeroes were padded up to 4 neighbors since only our data is restricted to main group elements. This creates invariance to permutations of neighbors when scanning a molecular structure but does not guarantee alignment of similar neighbors across atoms with different neighborhoods.




Reactive Atom + Neighbor Properties
Nearest neighbor predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
Nearest neighbor range (kcal/mol)
$R^2 = 0.847$
SVM $R^2 = 0.811$


Reactive Atom + Neighbors
Nearest neighbor predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
Nearest neighbor range (kcal/mol)
$R^2 = 0.850$
SVM $R^2 = 0.840$


With # Driving Coords Feature
Nearest neighbor predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
Nearest neighbor range (kcal/mol)
$R^2 = 0.850$
SVM $R^2 = 0.877$


No ΔE
Nearest neighbor predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
Nearest neighbor range (kcal/mol)
$R^2 = 0.420$
SVM $R^2 = 0.537$

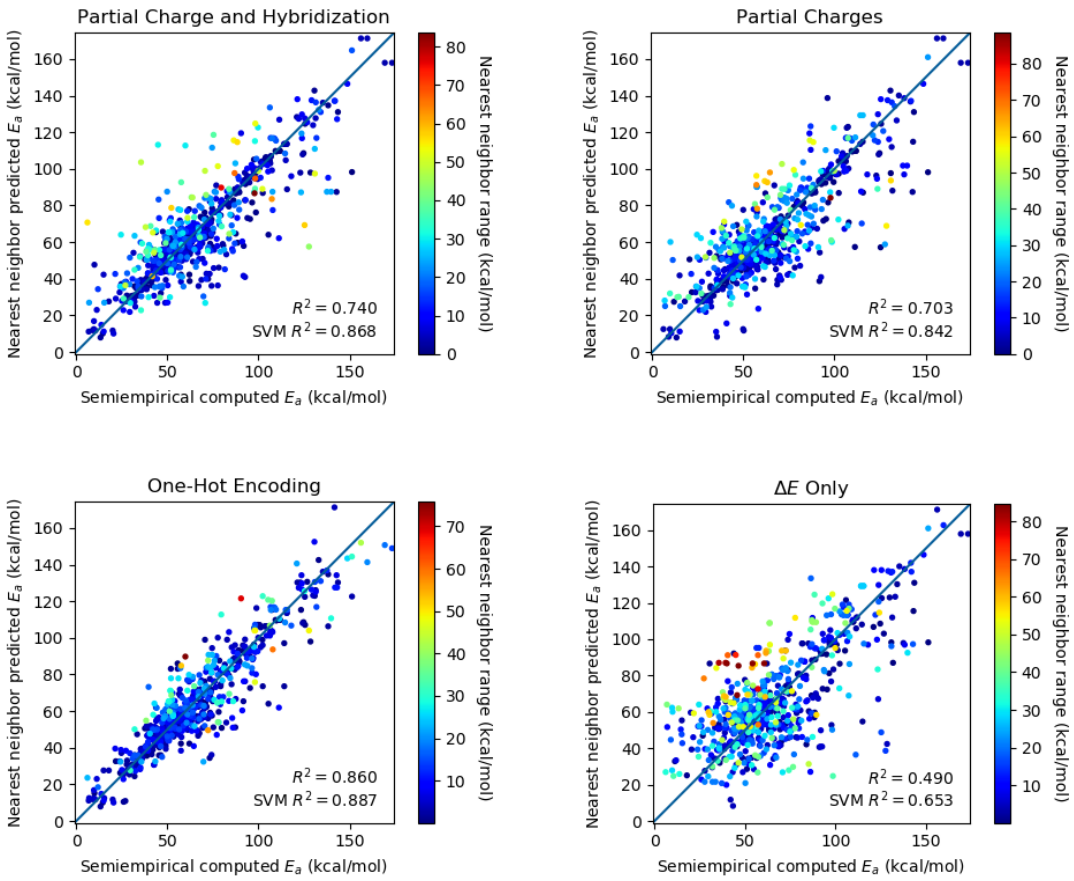


Figure 3-11. 2-nearest neighbor using L1 norm on the PM6 dataset (Data Set 1).

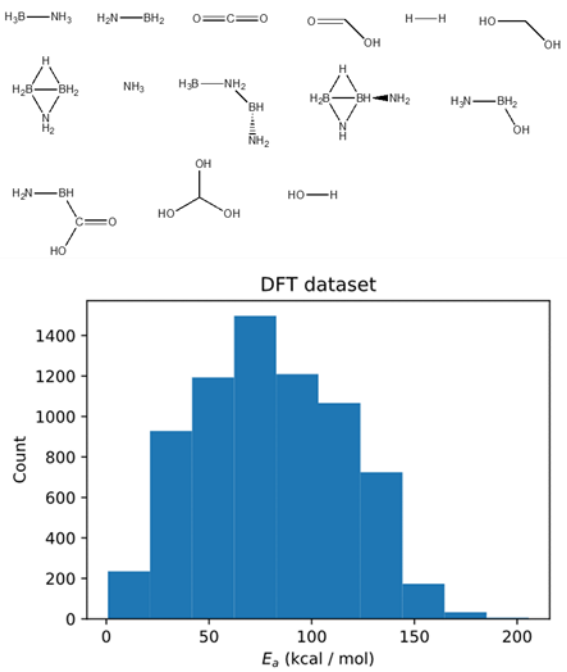


Figure 3-12. Top*: reactants used in DFT dataset (Data Set 2).* Bottom*: distribution of activation barriers for* DFT *dataset.*

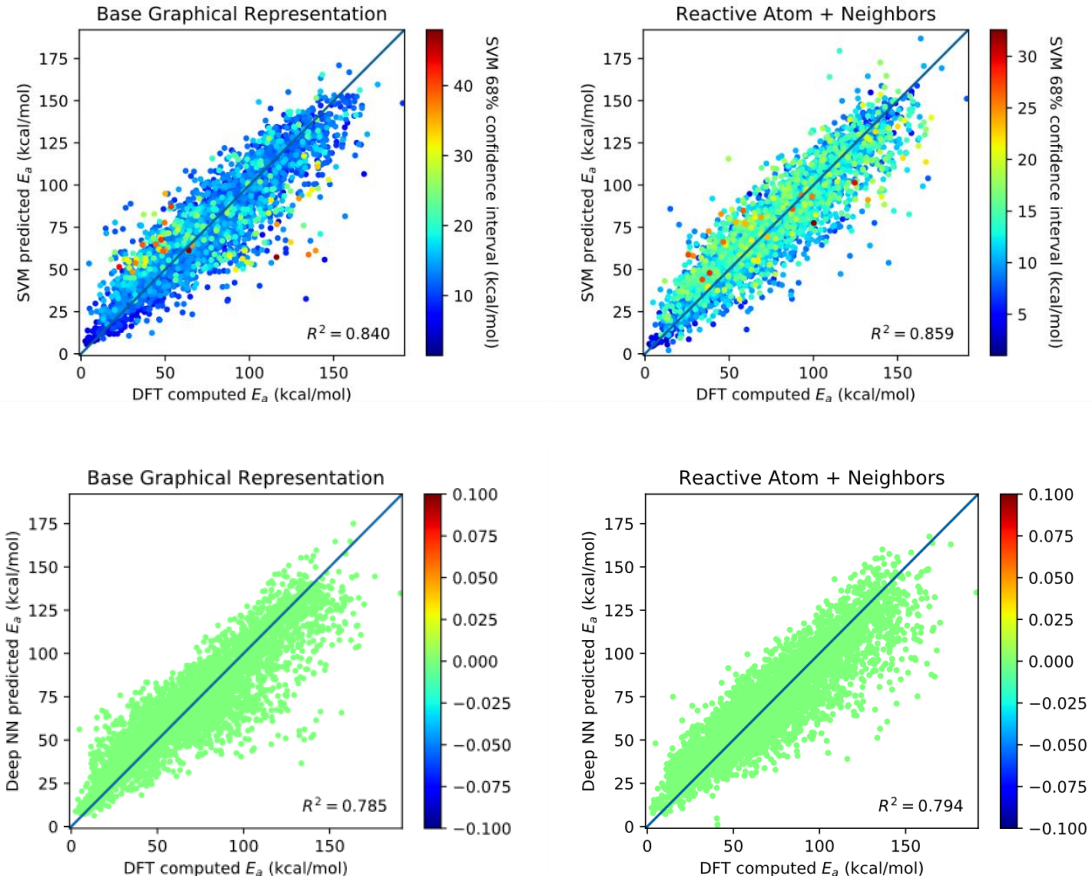


Figure 3-13. Cross validation SVM and NN predictions using graphical feature sets for a larger, DFT generated dataset (Data Set 2). Left: without reactive atom neighbor information. Right: with reactive atom neighbor information. Due to longer NN training time a narrower hyperparameter grid search was used for this larger dataset.

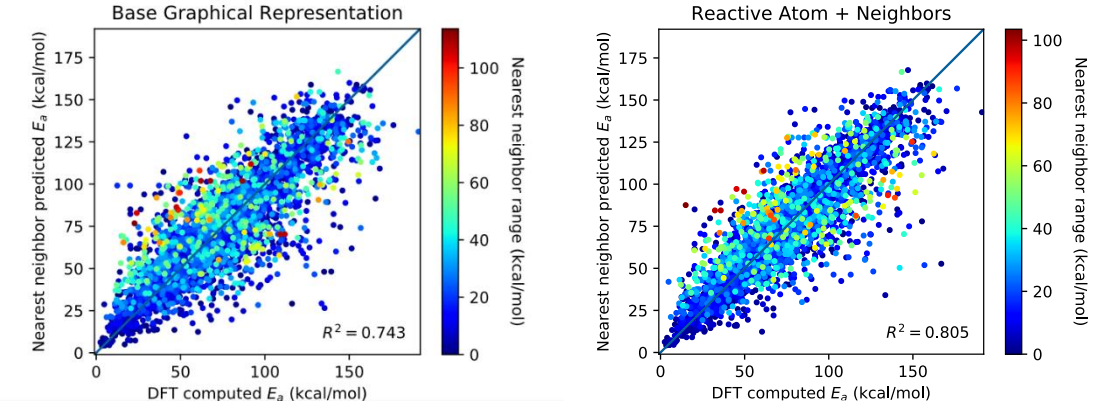
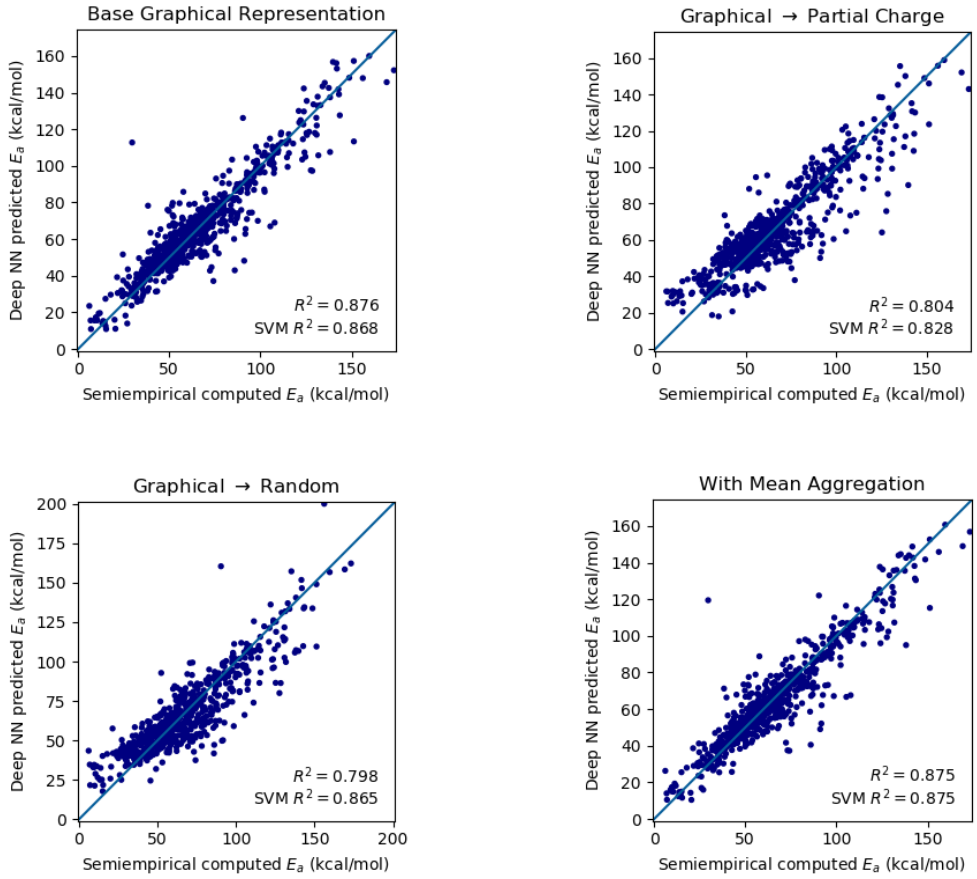
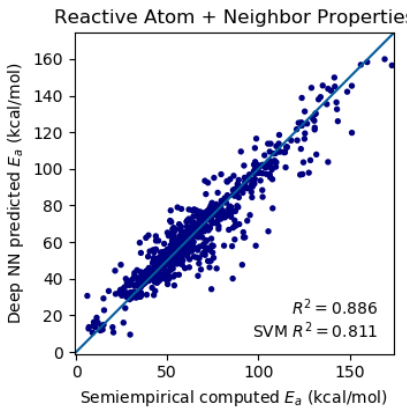
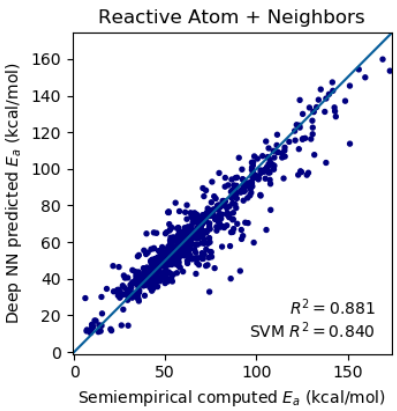
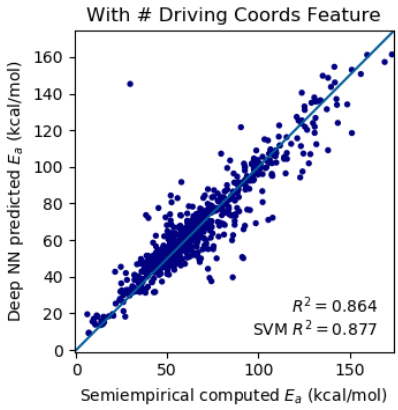


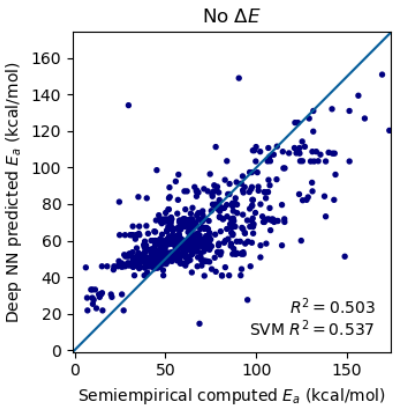
Figure 3-14. Cross validation nearest neighbor predictions using graphical feature sets for a larger, DFT generated dataset (Data Set 2). Left: without reactive atom neighbor information. Right: with reactive atom neighbor information.




Reactive Atom + Neighbor Properties
Deep NN predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
$R^2 = 0.886$
SVM $R^2 = 0.811$


Reactive Atom + Neighbors
Deep NN predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
$R^2 = 0.881$
SVM $R^2 = 0.840$


With # Driving Coords Feature
Deep NN predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
$R^2 = 0.864$
SVM $R^2 = 0.877$


No ΔE
Deep NN predicted E_a (kcal/mol)
Semiempirical computed E_a (kcal/mol)
$R^2 = 0.503$
SVM $R^2 = 0.537$

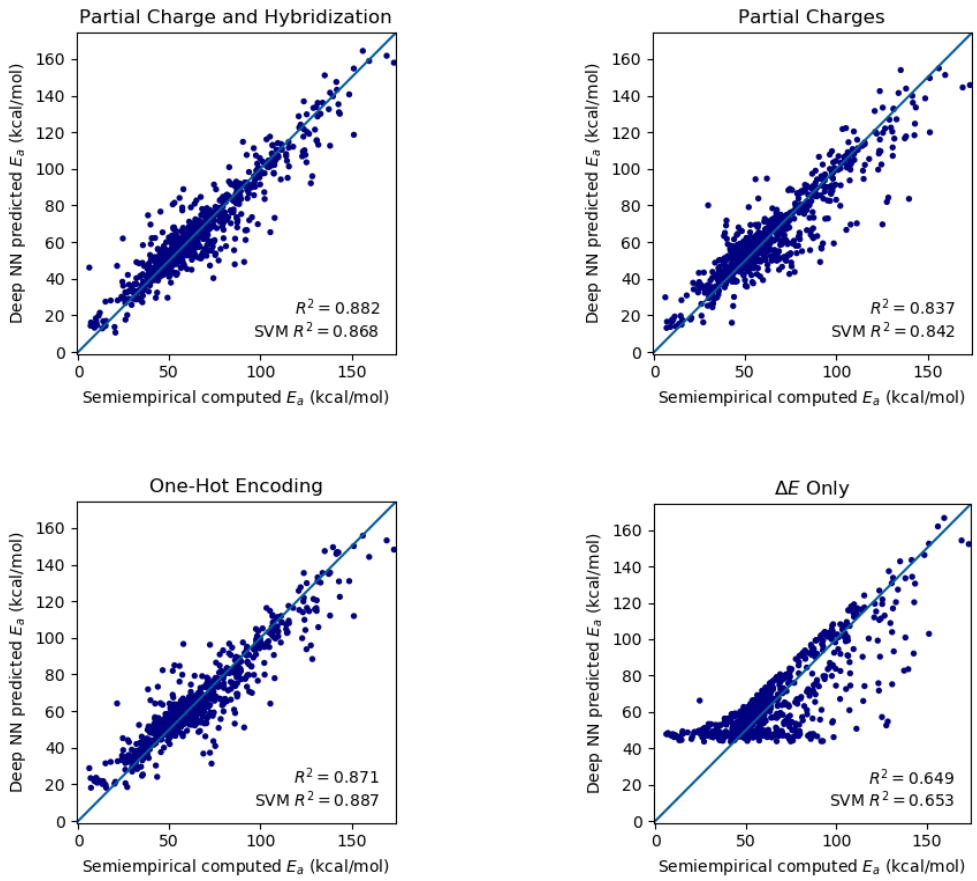


Figure 3-15. Cross validation neural network predictions for the PM6 dataset (Data Set 1).

Table 3-3. Cross validation accuracy metrics for various feature sets for the PM6 dataset (Data Set 1), using cross-validated SVM, prior to clipping of predictions (see computational details). All features sets include ΔE unless mentioned otherwise.

Feature Set	Deep NN RMSE	Deep NN R^2	SVM RMSE	SVM R^2	Nearest neighbor RMSE	Nearest neighbor R^2
One-Hot Encoding	10.05	0.87	9.42	0.89	10.46	0.86
ΔE Only	16.58	0.65	16.48	0.65	19.97	0.49
No ΔE	19.73	0.50	19.04	0.54	21.31	0.42
Base Graphical Representation	9.85	0.88	10.16	0.87	11.02	0.84
With Mean Aggregation	9.88	0.88	9.88	0.88	10.69	0.85
With # Driving Coords Feature	10.30	0.86	9.81	0.88	10.82	0.85
Reactive Atom + Neighbors	9.65	0.88	11.20	0.84	10.82	0.85
Reactive Atom + Neighbor Properties	9.45	0.89	12.16	0.81	10.95	0.85
Partial Charges	11.30	0.84	11.10	0.84	15.24	0.70
Partial Charge and Hybridization	9.59	0.88	10.17	0.87	14.27	0.74
Graphical → Partial Charge	12.39	0.80	11.60	0.83	16.09	0.67
Graphical → Random	12.72	0.79	10.30	0.86	14.22	0.74

Table 3-4. Cross validation individual fold R^2scores for various feature sets and machine learning methods for the PM6 dataset, prior to clipping of predictions (see computational details). All features sets include ΔE unless mentioned otherwise.

SVM

One-Hot Encoding	0.843577	0.888484	0.896051	0.88267	0.90372
ΔE Only	0.551851	0.621795	0.734003	0.640476	0.665406
No ΔE	0.581949	0.525919	0.566465	0.442783	0.542775
Base Graphical Representation	0.847456	0.904343	0.791368	0.884489	0.910004
With Mean Aggregation	0.860574	0.908262	0.795156	0.889134	0.921684
With # Driving Coords Feature	0.865336	0.89489	0.815525	0.881316	0.923949
Reactive Atom + Neighbors	0.847665	0.887473	0.707273	0.857064	0.906946
Reactive Atom + Neighbor Properties	0.839273	0.864569	0.649155	0.863037	0.863271
Partial Charges	0.839876	0.837863	0.805523	0.853894	0.872897
Partial Charge and Hybridization	0.852133	0.887105	0.833266	0.868516	0.89117
Graphical → Partial Charge	0.81507	0.816751	0.785411	0.850913	0.868538
Graphical → Random	0.799479	0.8808	0.86601	0.865813	0.886464

Neural Network

One-Hot Encoding	0.812949	0.89067	0.866857	0.873036	0.890363
ΔE Only	0.552556	0.604397	0.734359	0.637334	0.66509
No ΔE	0.554199	0.493017	0.50853	0.407946	0.525863
Base Graphical Representation	0.849437	0.896672	0.836042	0.887387	0.90343
With Mean Aggregation	0.860651	0.894645	0.83429	0.867791	0.910598
With # Driving Coords Feature	0.867318	0.878123	0.790274	0.870174	0.917089
Reactive Atom + Neighbors	0.851819	0.877388	0.893912	0.883857	0.884794
Reactive Atom + Neighbor Properties	0.870189	0.880546	0.901302	0.892616	0.876934
Partial Charges	0.838143	0.824697	0.8478	0.812297	0.849779
Partial Charge and Hybridization	0.854699	0.886603	0.901381	0.867473	0.885452
Graphical → Partial Charge	0.782832	0.798538	0.834385	0.770037	0.81138
Graphical → Random	0.736377	0.799496	0.83369	0.731495	0.823997

Nearest Neighbor

One-Hot Encoding	0.822227	0.848206	0.85855	0.886844	0.874478
ΔE Only	0.305229	0.446834	0.603531	0.484038	0.528397
No ΔE	0.264301	0.446632	0.578325	0.291091	0.416233
Base Graphical Representation	0.80037	0.800987	0.875568	0.844182	0.87999
With Mean Aggregation	0.82549	0.807954	0.888901	0.840618	0.886454
With # Driving Coords Feature	0.820709	0.783299	0.893746	0.849815	0.885022
Reactive Atom + Neighbors	0.840162	0.783615	0.871371	0.884616	0.868071
Reactive Atom + Neighbor Properties	0.829246	0.789065	0.87171	0.870896	0.8647
Partial Charges	0.624606	0.694794	0.76103	0.674093	0.716838
Partial Charge and Hybridization	0.59516	0.689128	0.815201	0.751419	0.790728
Graphical → Partial Charge	0.612087	0.675633	0.723065	0.593043	0.694023
Graphical → Random	0.69051	0.725975	0.804281	0.735925	0.722267

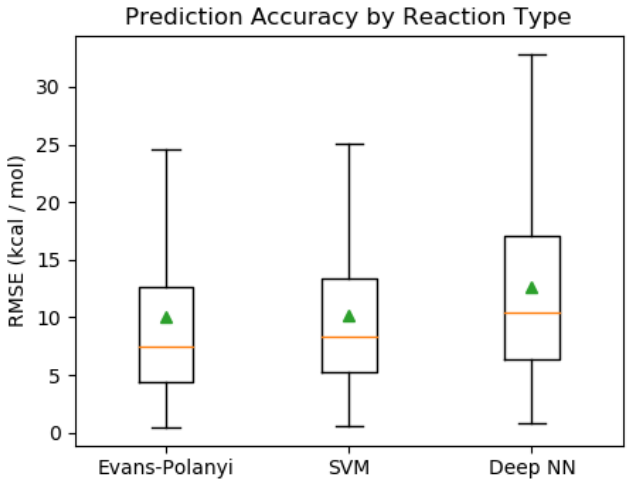


Figure 3-16. DFT dataset (Data Set 2), box and whisker plot of the RMSEs on each of the reaction types with at least 3 data points.

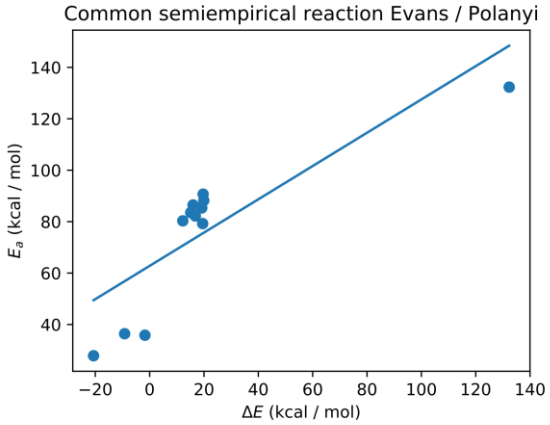


Figure 3-17. This mean for the Evans / Polanyi RMSE in Figure 3-16 is noticeably higher relative to the median because the data point to the far right in this Evans / Polanyi relationship performs especially poorly under leave one out cross validation even under regularization.

Note on Data Postprocessing

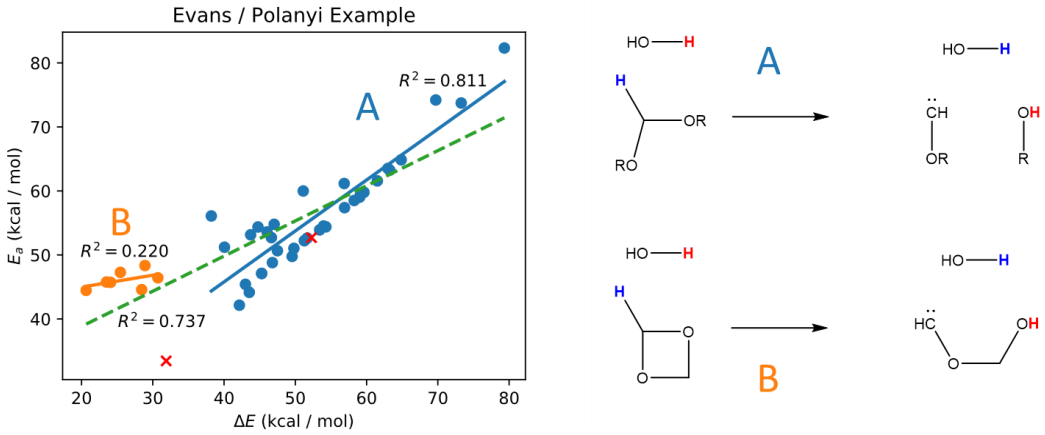


Figure 3-18. Figure 3-7 from the main text showing 2 outliers (Data Set 1).

Figure 3-7 originally contained 2 orange data points labeled as red 'x's in Figure 3-18. Upon further investigation, the lower left red 'x' was found to be a 2 elementary step reaction that was not identified as such by the automatic reaction profile filtering, which otherwise eliminated most of such data points from the dataset. The upper right red 'x' was found to have a product geometry with a C-O ring bond slightly above the bond distance cutoff such that data processing scripts counted the ring as broken but physically, ring strain was still clearly present. Thus both of these data points represent breakdowns in the automated general rules built into the data processing pipeline. Such points weaken the physicality of the data representation and thus should hinder a model's ability to learn on the data. These data points were removed due to being physically incorrect, but since only outlier data points were manually examined this could create a statistical bias, particularly if there are other data points that do not correspond to well defined single elementary steps but coincidentally fit the reaction trends.

It should be noted that both of these data points represent edge cases that are likely not the norm but still prevalent enough to affect many other reaction types in the dataset. The simplicity and interpretability of the reaction categorization and Evans / Polanyi approach facilitates a

human-in-the-loop strategy for examining individual reaction types, gaining chemical insights, and applying what was learned generally. For example, the data points above could motivate improvements in the automated approach to identifying reaction paths that do not correspond to a single elementary step. This would improve data processing for all reaction types, not just the one highlighted in this example. This therefore represents a transfer of knowledge in that the machine facilitates a human's learning on a single reaction type and the human then facilitates the machine's application of what was learned to a multitude of reaction types. In the process, the human gains a deeper understanding of the dataset and the machine's ability to model the dataset improves.

Reactions Appearing in Data Set 1

Due to the numerous variations of molecules involved in these reaction types, a simplified synopsis of the major reaction types is given below. Since each reaction type corresponds to multiple individual reaction steps, the following abbreviations are used: R = CH_2, $CHCH_2$ and so forth, and RH = CHR, etc, when the RH bond is reacting. R• appears after H abstraction from RH. Reacting molecules are delineated in Figure 3-9.

A few comments on the overall quality of this data are needed. First, Data Set 1 contains activation barriers that will appear too low compared to more accurate levels of theory. This is an inherent limitation in semiempirical methods (i.e. PM6 in MOPAC) in general, but the results from said method are considered "correct" by the machine learning tools. Data Set 1 quality is thus limited from the perspective of chemical accuracy. To address this issue, Data Set 2 was created using a higher level of theory—density functional theory—to ensure the analysis does not depend strongly on level of theory. The specific density functional level of theory was B3LYP/6-31G**, and the Data Set 2 reaction classes include (and largely resemble) those reported in Li et al, *J. Phys. Chem. A* **2016**, *120*, 1135-1144. All conclusions in the main text (discussing Data Set 1)

were affirmed by analysis of the second data set (*vide supra*). In addition to this discussion of accuracy, both data sets are unique in that they contain high as well as low barrier reactions, which is a missing feature of most reaction data sets available in the literature. The inclusion of "negative" results allows a more careful examination of the quality of machine learning predictions of reactivity, by forcing the machine to rate reactions on an activation energy scale, rather than simply ranking the most likely reactions.

Top reaction types:

Example reaction ID-2253:
H shuttle through H_2O for H transfer to O
RHOR + H_2O → R• + HOR + H_2O
E_a > 34 kcal/mol

Example reaction ID-4328:
H transfer to O
RH + ROR → R• + HOR + R
E_a > 45 kcal/mol

Example reaction ID-4971:
H shuttle through H_2O for R-H addition to O
ROR + RH + H_2O → R_2OHR + H_2O
E_a > 43 kcal/mol

Example reaction ID-2680:
Insertion of R-H into ROR
RH + R_2O → ROHR + R
E_a > 50 kcal/mol

Example reaction ID-6498:
Dihydrogen elimination
RH + RH → R• + R• + H_2
E_a > 46 kcal/mol

Example reaction ID-4983:
H transfer to O
RH + H_2O → R• + H_3O*
*this species 3 H are stabilized by bridges across O atoms
E_a > 37 kcal/mol

Example reaction ID-1178:
Insertion of water into cyclic dimer of formaldehyde
$c(RO)_2 + H_2O \rightarrow$ HOROROH
E_a > 31 kcal/mol

Example reaction ID-1640:
Ring-opening H transfer
$c(RO)_2 + c(RO)_2 \rightarrow c(R\bullet O)(RO) + CH_3OCH_2O\bullet$
E_a > 66 kcal/mol

Example reaction ID-2203:
Ring-opening H shuttle to form C-O bond
$c(RO)_2 + c(RO)_2 + H_2O \rightarrow$ c(RO)2CHOCH2OH + H_2O
E_a > 49 kcal/mol

Example reaction ID-11178:
ROH shuttle of H from C to O
HOROR + ROH→ H2OR•OR + ROH
E_a > 42 kcal/mol

Two additional reaction types
(selected due to appearance early in data set):

Reaction ID-27:
2+2 dimerization of aldehyde
R=O + R=O → $c(RO)_2$
E_a = 30 kcal/mol

Reaction ID-521:
Insertion of aldehyde into ring with R-H activation
$c(RO)_2$ + R=O → HORORORH
E_a = 32 kcal/mol

Chapter 4. Human – Algorithm Interactive Approach to Conformer Generation

This chapter is based on a highly collaborative project with the research group of Ambuj Tewari. Student contributors: Tarun Gogineni, Exequiel Punzalan, and Ziping Xu.

Introduction

The dynamic between humans and computers has been a focus of significant research.[13] Computers can perform many types of quantitative calculations orders of magnitude faster than the most skilled human and can even outcompete humans in complex abstract games such as chess and go. Likewise, certain tasks that require minimal effort for a human are extremely difficult for artificial intelligence as illustrated by the concept of Captcha. Since humans and computers have different strengths, much study has been devoted to combining or integrating the strengths of humans and computers to perform tasks more proficiently than either individually.[128] While this shows much promise, effectively combining these strengths and facilitating effective and exchange of meaning between humans and computers is challenging and currently must be understood in a domain specific context.

In chemistry, human driven experiments and computational modeling[129] are both active areas of research. Recently, machine learning approaches[130–132] have become popular as tools to model and approximate other computational methods, effectively allowing them to be employed at higher volume. In conformer generation, machine learning has been used to for various purposes including predicting molecular geometry from the chemical graph[133] and sampling from the Boltzmann distribution of equilibrium states.[12] Combined human-computer approaches have been employed such as in protein folding and design.[134]

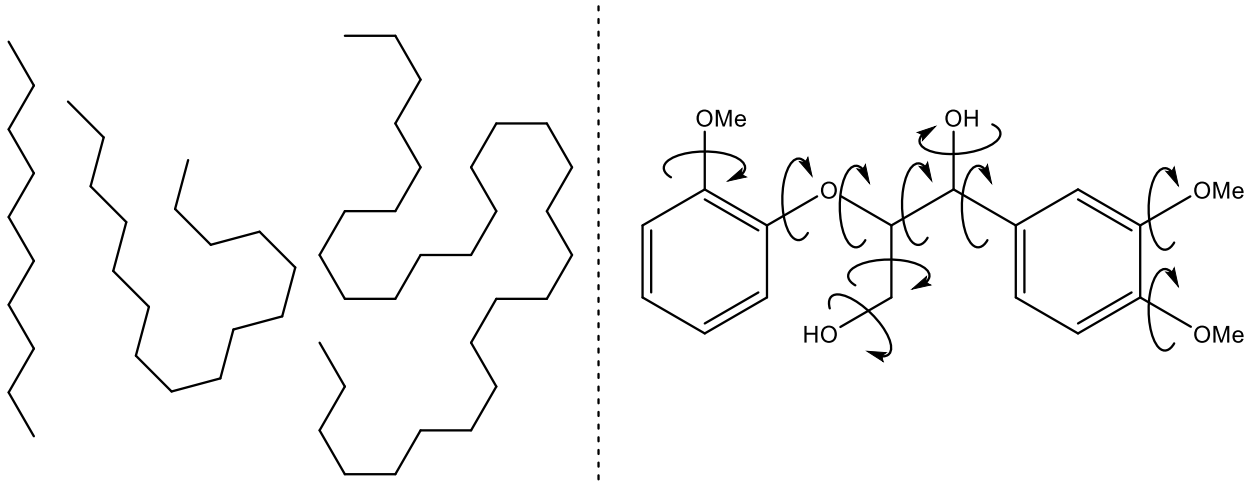


Figure 4-1. Left: even molecules as simple as alkanes can adopt vastly many conformations. Right: rotatable bonds of a small lignin fragment we used to test our methodology (see results).

We sought to engage with this challenge in an actively studied area of computational chemistry called conformer generation. Conformer generation[135] involves determining the relevant conformers of a chemical system for a particular purpose, such as calculating the relative likelihood of structures via free energy or investigating the feasibility of a particular chemical reaction. Conformers are different organizations and arrangements of atoms in a molecule that can interconvert without breaking or forming any chemical bonds. Conformational change involves restructuring of a molecule by rotating parts of the molecule about rotatable single bonds.

Conformers are important for numerous reasons.[136–138] Thermodynamic properties of molecules depend on the probability distribution over the entire ensemble of conformers accessible at the relevant temperature. Thus, a representative ensemble of conformers with accurate energies is necessary to accurately predict thermodynamic quantities.[139]

While some torsions are functionally codependent in order for the molecule to avoid self-collision (see Figure 4-1), they are sufficiently independent to yield an effective exponential scaling of plausible conformers in the range of molecular size generally applicable in biological chemistry and other areas of chemical interest where conformers are important. This means that

exhaustive evaluation of possible conformers quickly becomes computationally prohibitive as molecule size increases. Millions of conformers are already possible with 15 rotatable bonds, which is often exceeded by chemically relevant molecules and polymers.

Only some of the plausible conformers exist in solution and noticeably contribute to macroscopic thermodynamic properties. In common cases this number is hundreds or thousands of conformers but still a microscopic fraction of the number of plausible conformers. A combination of physical and statistical principles elucidates that at equilibrium, conformer populations follow a Boltzmann distribution in which the relative populations of conformers decay exponentially as a function of the conformer energy.

Methods of conformer generation are generally grouped into two primary categories. Systematic methods[140–142] deterministically enumerate conformers by combinatorially exploring the torsional landscape. For molecules with few rotatable bonds these methods can effectively capture all possible conformers. However, for molecules with many rotatable bonds the number of possible conformers grows exponentially, and quickly the challenge becomes choosing which conformers to generate without *a priori* knowledge of which will be low in energy. Stochastic methods[143–145] of conformer generation make no attempt to capture all conformers but use a random sampling approach to attempt to capture an ensemble of conformers that is statistically representative of the overall set. Molecular dynamics is commonly employed for this purpose because it samples conformations with realistic bias towards low energy conformations. However, accurate molecular dynamics simulations cannot currently be performed for the time scales necessary to sample all relevant degrees of freedom, particularly for hindered conformational change that requires an energetic barrier to be crossed.

Successful conformer generation strategies must maximize accuracy of the geometries and energies of conformers generated as well as the number of conformers while minimizing costs of resources. There is still much room for growth in achieving this balance because there are many chemical systems of industrial interest for which the conformers cannot be accurately modeled with the current best practices and state of the art methods in the field.

Reinforcement learning[3,146,147] is the interaction of an algorithmic agent with an environment. The reinforcement learning paradigm is an iterative process consisting of a representation of the state of the environment at each iteration. The algorithmic agent chooses from a set of possible actions from each state which results in a change of state for the next iteration and a numerical reward is given. The goal of the agent is to optimize the total numerical reward over the long run. Having different approaches for approximating the function of interest has been shown to be valuable, even for just a less accurate but cheaper approximation function.[148]

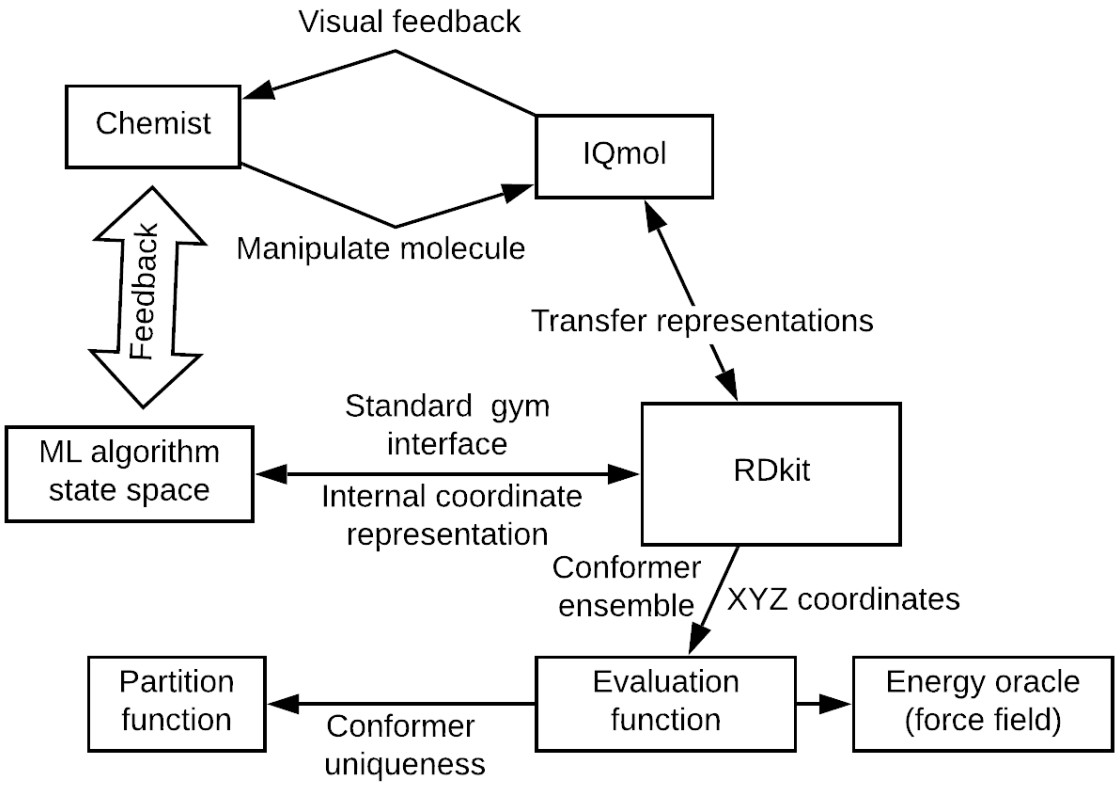


Figure 4-2. Flowchart of our framework built to encourage human – algorithm interaction.

Effective interaction between an expert chemists and algorithmic tools developed by data scientists includes the technical challenge of developing an interface that facilitates meaningful interaction. Towards this end we augmented the IQmol molecular editor to record chemist interaction with a molecule. IQmol provides functionality for selecting specific atoms and bonds, even within rings, and rotating or translating them relative to the rest of the molecule while preserving bonds. This means that a chemist can observe a structure and instantly perform geometric manipulations that affect the conformation but not the graphical bonding connectivity of a molecule. Local force field geometry optimization can also be performed instantly. Our

augmented IQmol saves the current geometry to a file each time the chemist signals the completion of a molecular modification by locally optimizing a new conformer.

We employ the suite of cheminformatics tools available in python in the central algorithmic management of facilitating interaction with humans, interaction with algorithmic tools, and analysis and evaluation tools. We used RDKit and DeepChem to manage ensembles of conformers and convert between the differing representations needed by various aspects of the workflow. The human interaction and evaluation methods both employ a 3D cartesian coordinate representation. For algorithmic learning, we employed a representation which encodes only the rotatable torsions. Since our methodology is focused on torsion angles rather than overall translation and rotation or bond lengths and angles, representing a conformer by its torsional angles reduces the effective dimensionality explored by algorithmic tools. This reduces the difficulty of the task imposed on an algorithmic learner by embedding basic human understanding of the problem into the representation. In some sense this is a basic instance of human-computer interaction.

A comprehensive method of evaluating an ensemble of conformers and individual conformers within an ensemble is important in training, validating, and comparing conformer generation processes. The partition function is the key quantity which needs to be approximated effectively. A method can do this by generating a sufficiently large and representative ensemble of unique conformers and accurately approximating their energies.

The partition function quantifies the relationship between the energy of an individual conformation and its frequency of existence at thermal equilibrium which is central to the various applications of conformer generation. With an accurate partition function, entropies and relative free energies can be accurately approximated. The partition function has the form $Z = \sum_i e^{-\beta E_i}$ where the sum index is over all possible conformations. The lower the energy of a conformer, the

more it contributes to the partition function. To estimate this function effectively, we need an ensemble of conformers that represents the entirety of the low energy regions of conformational space. Symmetries and degeneracies need to be accounted for such that the multiplicity of conformers matches that of the underlying statistics and physics. The energies of the conformers also need to be computed accurately in order for the partition function to model reality.

Conformers exist in the continuous space of cartesian geometry, but for the purposes of thermodynamics each local energetic minimum on the conformational potential energy surface should be counted exactly once. Handling this appropriately when generating conformers is nontrivial because geometric optimization is numerical, and the potential energy surface is complex so generated conformers may be nearly but not exactly identical. We experimented with multiple methods for evaluating conformer uniqueness.

The root mean squared distance (RMSD) method of conformer uniqueness computes a distance metric on conformer space by rotating and translating the conformers so as to minimize the RMSD of the atomic coordinates in 3D space and setting a threshold on this minimal RMSD which defines unique conformers. This method is simple and straightforward to implement but it scales nonlinearly with molecule size, so a single threshold is inadequate to determine unique conformers across a range of molecule sizes.

Torsion fingerprints[149] is a method designed to mediate challenges of using RMSD as a distance metric. It constructs a weighting of torsions based on their centrality in a molecule in a way that more closely matches with a chemist's intuition about molecular similarity. It also normalizes differences to 1 which makes the metric able to handle molecules of different sizes without having to determine an appropriate nonlinear scaling for the metric.

Even with a good ensemble of conformers, accurate energies for each conformer are necessary to accurately approximate the partition function. There is a plethora of methods for computing energies of molecules all along the computational cost vs accuracy tradeoff. Accuracy is also dependent on the type of molecular system so in principle different functions from conformer to energy could be employed in different applications. For initial testing, the forcefield MMFF94[150] was used due to its low cost and availability which allows for rapid prototyping and experimentation with workflow and other variables.

The Gym interface, a standard in reinforcement learning, was used to express conformer generation as a reinforcement learning problem. Use of the Gym interface facilitates facile application of the array of reinforcement learning tools expressed in the interface to the conformer generation problem. To express conformer generation as a reinforcement learning problem the notions of state and action need to be contextualized to conformers. We defined a state as a discretized set of torsion angles for all rotatable bonds in a chemical system. For linear alkanes, this was reduced to a sequence of the torsions along the bond and discretized into trans ('t'), and both gauche ('g+', 'g-') configurations. Actions were defined to be any possible rotation of any combination of rotatable bonds. These actions can either be expressed as rotations relative to the current state or as absolute actions synonymous with the states they lead to. Rewards are given based on the contribution of the generated conformer that results from taking a given action to the partition function. Conformers within a single trajectory that have already been generated receive a reward of 0 because duplicate conformers do not improve the estimated partition function.

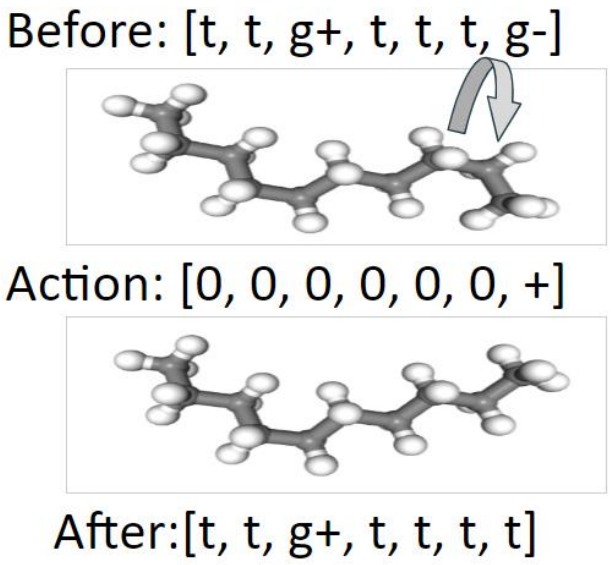


Figure 4-3. Illustration of reinforcement learning state and action representations for linear alkanes.

There are multiple important characteristics of this embedding from the perspective of reinforcement learning. First, allowing any combination of rotations of torsions means that any state is accessible via a single action from any other state. Second, all actions of rotating bonds lead to a predetermined end state so all actions are deterministic. Third, since there is a deterministic action from any state to any other state, there is theoretically no incentive to explore a temporarily undesirable state for the purpose of optimizing for delayed reward.

For modeling a function from conformer to next conformer, neural networks were used. In initial tests with linear alkanes, this was a long short-term memory (LSTM) network. LSTM is effective for linear alkanes because the data is naturally sequential. Information about the molecular environment can be passed in through the memory. The LSTM receives a sequence of discretized torsions of a linear alkane and outputs a sequence of discretized rotations to perform on the alkane which results in a new conformation. The actions can either be coded as an absolute position (e.g. trans) for the current torsion or as a discretized rotation relative to the current torsion angle (e.g. +120 degrees).

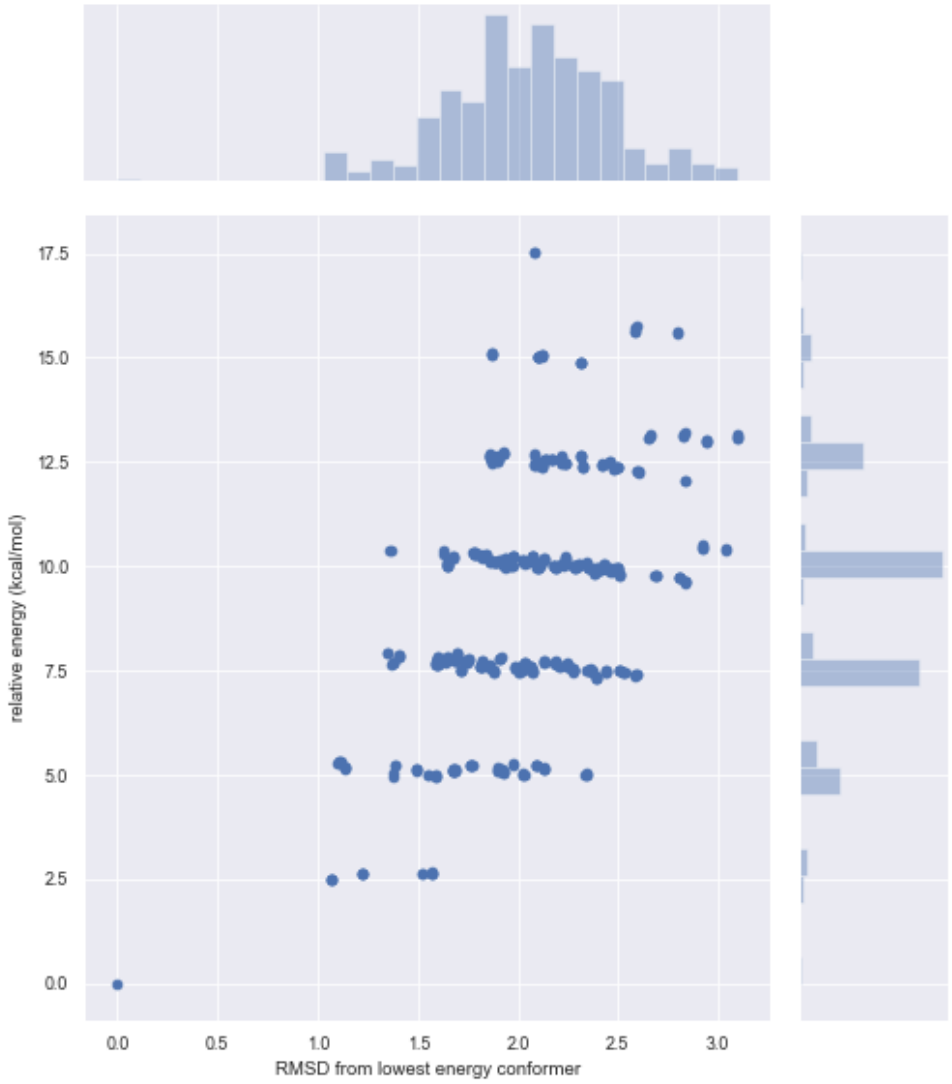


Figure 4-4. Conformers of decane fall into energy bands based on the number of torsions that are not in a trans orientation. Figure courtesy of Exequiel Punzalan.

Initial testing of reinforcement learning algorithms was performed on linear alkanes. For small linear alkanes, the lowest energy conformer is a straight chain in which all torsions are in a trans configuration. Other low energy conformers generally have mostly trans torsions but one or more gauche torsions at various places along the chain (see Figure 4-4).

We used the A2C reinforcement learning approach to train an LSTM on alkanes of lengths 4-8 and predict using the LSTM on nonane. We started in a random configuration which will likely

contain significant steric clashes resulting in a high energy and low initial score. The A2C method reaches a ceiling of training accuracy within 30,000 conformations. However, the test accuracy has not leveled off after the model has been training on 50,000 conformations. This suggests that the model is continuing to be driven towards a more generalizable representation of the relationship between conformation and energy even within the space of representations that perform essentially optimally on the training molecules.

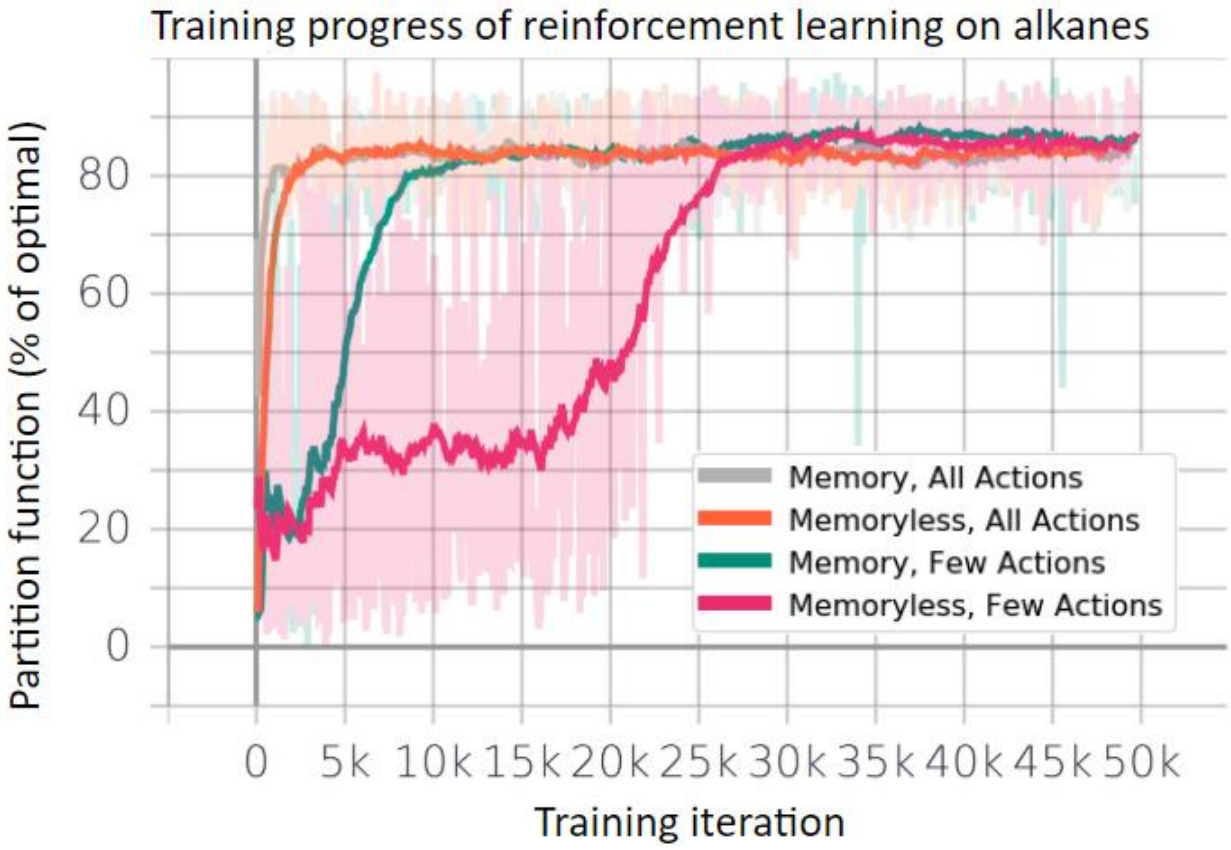


Figure 4-5. Training performance throughout training of advantage actor critic reinforcement learning algorithm training an LSTM to generate ensembles of n-alkane conformers. A short alkane (length 4-8) was repeatedly randomly selected followed by generation and evaluation of 200 conformers. "Memory" refers to allowing the LSTM that takes the last torsion of a conformer as input to transfer its memory to the first torsion of the next conformer within the ensemble. "Few actions" refers to defining the action space as relative rotations from the previous conformer. "All actions" refers to providing direct, absolute actions to all torsion angles independent of the current conformer. Training iteration is the total number of conformers generated. Dark lines are smoothed for visual clarity. Courtesy of Tarun Gogineni.

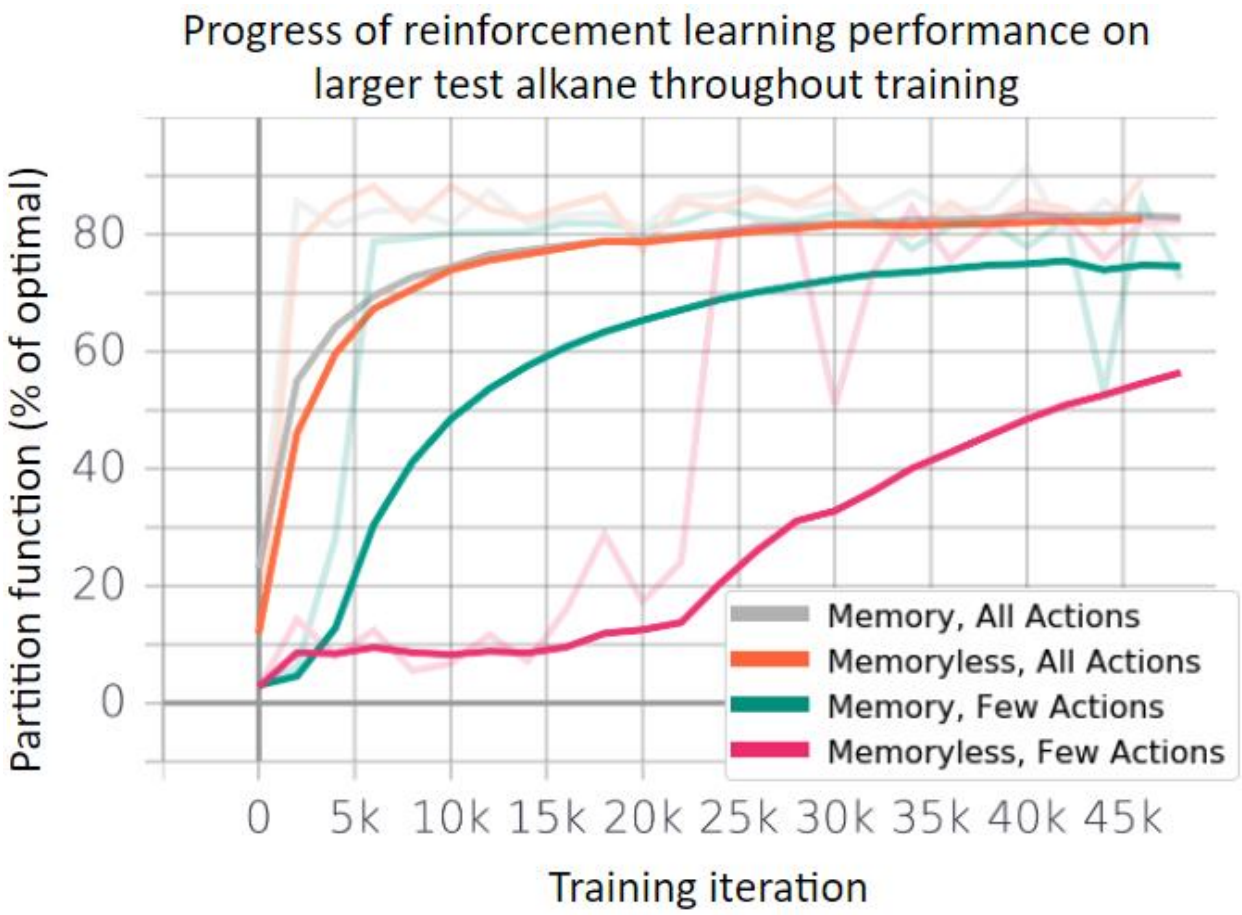


Figure 4-6. Test performance of LSTM throughout training of advantage actor critic reinforcement learning algorithm. Performance tested by generating ensembles of nonane conformers which is larger than the alkanes used in training. Training iteration is the total number of conformers generated in training before the test. Dark lines are smoothed for visual clarity. Courtesy of Tarun Gogineni.

The reward depends on whether a conformer is unique, so treating the most recent conformer as the current state is non-markovian. This is because whether or not a conformer has been generated previously impacts the reward but would not be a part of the state. Representing the state as the ensemble of all conformers generated could overwhelm an algorithm with too much diffuse information. A middle approach is to allow the LSTM memory to carry forward from the last torsion of one conformer to the first torsion of the next conformer. This at least allows for the possibility of maintaining some representation of the conformer space that has already been explored. Whether this has actually occurred is unclear but when applying this memory technique training does progress noticeably faster.

For alkanes, the conformers which will have the largest contribution to the partition function are those with the most trans torsions. However, how to arrive at a trans torsion is dependent on the starting configuration. This means that for alkanes there is an effective additional layer of obscurity in simply learning what actions to take to arrive at a trans configuration from all possible starting configurations. As a benchmark, we substituted the action space of specifying rotations with an action space in which the absolute discretized angles are specified. This should lead to a simpler rule for generating low energy conformers for alkanes, and this is reflected in significantly faster learning. For more complicated molecules, this could be a useful benchmark for evaluating the difficulty of conformational problems and the generalizability of models. For a sufficiently challenging conformational problem, intuitively having a reasonable starting conformation and making minimal rotations from it should be more efficient than learning absolute angles over the entire conformational space. The distinction between thinking about generating conformers "from scratch" or by rotating other conformers has paradigmatic implications as well as implications on how such models could be expected to generalize.

Chemists have already learned the basic principles underlying conformers and their energetics, so after using alkanes to learn the IQmol interface we needed a system with more complexity to gather interesting data. A small lignin fragment was used as an initial benchmark (see Figure 4-1, right). Lignin[151] is chemically interesting because it is highly abundant naturally and could have potential to be converted into biofuels but its relative stability, polymeric irregularity, and conformational complexity render utilizing it as a challenging exercise. and Figure 4-7 shows the author's results from generating conformers in IQmol. Most of the effort went into attempting to line up hydrogen bonds in the most favorable configurations paired with rotating other bonds not expected to significantly impact the energy. In addition to potentially

generating useful conformers, humans think creatively and immersing them in a problem can create opportunities for ideas that would otherwise be unlikely to occur. For example, after running out of ideas the author determined that taking a section of the molecule and translating it to the opposite side of the molecule would encourage the force field optimization to minimize other rotations in the long path back to stability. Since force fields maintain bonding information regardless of atomic locations large, stochastic changes may be beneficial algorithmically as well.

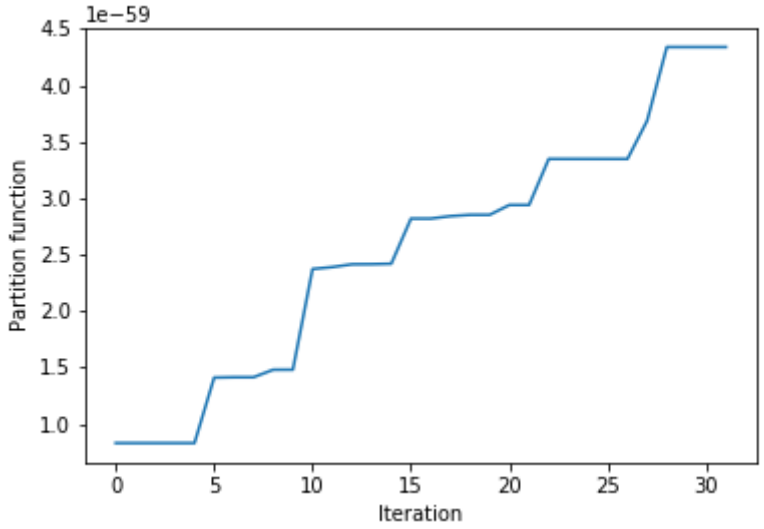


Figure 4-7. Partition function progress from recording the author's manipulation of a small lignin fragment in IQmol. Figure courtesy of Exequiel Punzalan.

Future Directions

This collaborative project towards the goal of generating conformers is young and efforts thus far have focused on building a platform and framework for human-computer interaction. Combining algorithmic and human strengths has been a valuable concept in the previous projects so it is likely that such a framework has the potential to continue to be fruitful. A few additional steps are necessary to enable effective conformer generation on chemically interesting molecules presently unreachable with current methods. Four of the main steps are graphical instead of sequence input, ring actions, machine to human interface, and improved energy function.

The reinforcement learning state space can be effectively encoded as a sequence of torsions for linear molecules such as those used in initial testing. This simplification allows for simpler neural network architectures to be applied in the reinforcement learning. While torsions could be serialized for arbitrary molecules with branching and rings, neural networks that take sequence data rely on a meaningful connection between location in the sequence and meaning relative to other elements of the sequence. However, serialization of torsions is highly arbitrary and forces an algorithm to effectively relearn valuable information that is inherent in the bonding structure of a molecule. Thus, a graphical representation of the state space and neural network methods that operate on graph data will be preferable. Some methods already exist in this area, and investigation is currently underway into how to appropriately apply them.

The action space thus far has been a set of torsion modifications which makes sense for linear and branched structures. However, for rings the space is more subtle because most combinations of ring torsions will result in inconsistency with the ring constraint. A few strategies are possible. First, actions that break the ring constraint could be technically allowed but penalized so heavily in the reward function that a reinforcement learning algorithm must learn to avoid them. Second, a palate of multi-torsion actions within rings that span the torsional space could be developed and used instead of an arbitrary combination of single torsional changes.

The current interface allows for passing of information from a chemist to an algorithm. However, valuable information could also be passed from an algorithm to a chemist such as a current set of the most diverse low energy conformers as starting points for chemist exploration. This will require an interface to be developed which will likely involve using IQmol to generate editable visualizations of structures produced algorithmically.

Finally, the MMFF94 force field currently used in testing is sufficient for method development but will likely lack the quantitative accuracy necessary to generate an accurate partition function for molecules of chemical interest. A plethora of methods exist for generating approximate electronic energies for conformers and the best choice is molecule dependent.

Chapter 5. Final Thoughts

While the conclusions to the individual projects are found within those chapters, this conclusion will take a broader perspective and examine the principles learned while carrying out these projects and the outlook going forward. First, a highlight of the key takeaways from previous chapters.

In chapter 2, an approximation scheme for computing derivative coupling vectors was developed that allows for derivative coupling vectors to be approximated via any electronic structure method which has an implementation for the energy and force (i.e. nuclear gradient).[152] This creates potential for modelers of light induced chemistry[5,18,48] to more quickly adopt state of the art electronic structure theories as improved methods continue to be developed. This was made possible by a deeper understanding of conical intersections and employing a simple, intuitive model.

In chapter 3, inadequacies in current metrics for evaluating machine learning performance for chemical reactivity were exposed, and potential methodological solutions were explored to circumvent the lack of generalizability in many machine learning models. This opens the door for other scientists to avoid common pitfalls and devote effort towards more productive modeling paradigms. This was made possible by a deeper understanding of the interaction between how chemistry is represented for machine learning and how machine learning models "learn" from data. A simple, intuitive model was again critical to gaining understanding.

The project described in chapter 4 is in too preliminary of a stage to have definitive conclusions, but the paradigms gleaned from the previous projects have begun to be applied to another intriguing challenge in computational chemistry, that of conformer generation.[135] A similar approach will hopefully once again prove fruitful in deepening understanding and reframing the problem in a way that promotes efficient progress by other scientists.

Additionally, much was learned about the process of research in the midst of conducting the research in this dissertation. The most interesting insights came through skepticism and digging deeper beyond the original goals of a project. Uncovering the underlying structure in an area requires stubborn care and confidence that uncovering the true structure of the methodology will drive the field forward. Assumptions were frequently invalid, and things are not always as they seem. It was often easy to be convinced that results happen for the seemingly natural explanation when this is not the case.[15]

Open Questions and Research Opportunities

Looking forward, many questions and open pathways for research remain. From chapter 2, the diabatic model Hamiltonian used to calculate the derivative coupling vector has potential to provide much more information, such as an estimate for the location of a nearby minimum energy conical intersection. An approximation for the derivative coupling vector at nearby points in atomic coordinates could also be extracted, or information from a previous nearby geometry could be used to improve accuracy and convergence at another geometry. The ability to utilize a model as a starting point for a model at a nearby geometry is appealing because many methods that utilize derivative coupling vectors require large numbers of them at points within a localized region of the potential energy surface.[34,153] For example, molecular dynamics simulations involve trajectories in which adjacent time steps will be very close. Near a conical intersection, the

adiabatic potential energy surfaces are highly non-smooth, making utilizing information from a nearby geometry difficult within standard Taylor Series based frameworks. Within a diabatic model, however, utilizing information from nearby geometries is much more plausible.

In chapter 3, the potential of human-computer interaction over the simple linear Evans / Polanyi modeling procedure was only mildly exploited. Even basic interactions led to deeper human understanding of the dataset, but the intuition learned could have been applied to all types of reactions in the dataset and further exploited to improve the modeling procedure. Iteratively implementing model improvements based on human intuition learned from observing breakdowns of an evolving representation and intuitive model seems promising. Could such an iteratively grown model prove generalizable under extrapolation rather than merely interpolation? Could the model effectively generalize to new types of reactions with less new data?[1,154]

In chapter 4 the potential of human-computer interaction in chemistry has much more to be explored. What other types of chemical problems could benefit from human-computer interaction? How can we generalize to broader or bigger chemical problems? What are the ultimate efficiency and potential from these interactions?

How to Approach Future Challenges

The means of reaching these broader goals will be through developing multidisciplinary scientific leaders, building a strong research community with a multidisciplinary mindset, and active engagement with other scientists.[155] This will require care for the overall health of the scientific community, commitment to the pursuit of truth by following the data even when it leads to unexpected places, and a willingness to challenge the community to think broadly and deeply.

Developing Multidisciplinary Scientific Leaders

Numerous traditional disciplines contributed to this work, suggesting that building up the comprehensive understanding that is required to truly understand methods and apply them robustly takes time and investment. There are no shortcuts. In chapter 3, the key insights were developed through caring for how the machine was actually learning and a willingness to challenge the assumption that quantitative accuracy of machine learning methods implies that the means of obtaining this accuracy is consistent with how chemists perceive the relationship between features and reactivity. Once an anomaly was uncovered, commitment to following these results at the expense of the original goal of the project was necessary to build deeper understanding. The project could have been carried to completion as originally hoped and would have appeared impressive, but it ultimately would have provided little scientific value to the community.

For this reason, digging deeper can lead to a need to challenge the scientific community. Current prevailing models, methods, and interpretations in computational chemistry form a foundation on which computational investigations into the chemical reactivity of specific types of reactions are built. Many such investigations lead to the situation where many different chemists' understanding of the reactions they study is influenced by the theoretical and computational methods that have become common as a means of justifying or grounding mechanistic hypotheses. Challenging current understanding in an area propagates into connected areas, making correcting misunderstandings important.

Building a Strong Research Community with a Multidisciplinary Mindset

No matter how talented, no individual researcher can understand the full picture surrounding a project. Relational and conceptual grounding in multiple fields is necessary in order to be truly multidisciplinary. Building strong relationships between chemists and data scientists is

necessary for people to truly develop methods that will be technically robust, theoretically grounded, and provide what chemists desire. The person developing the method also needs to personally care about and be invested in people like those who make up the intended beneficiaries.[156] Strong community must traverse disciplinary boundaries in order for methodologies developed in one area to make an impact in another. Machine learning techniques have been recently applied in a variety of areas in chemistry, but one of the limitations to their wider adoption is a sentiment among some experimental chemists that these computational techniques will not solve their real problems.

An illustrative example comes from a collaboration related to the project described in chapter 2. We were developing a method to traverse excited state potential energy surfaces to computationally investigate photochemical reactions.[100] Confusion arose from a chemical perspective when the computational technique failed to find all of the products observed experimentally despite attempting to be an exhaustive search, suggesting incompleteness or inaccuracy in the modeling. I suggested that my collaborator search along maxima in the directional search from the point of state transition. This was counterintuitive because usually minimal paths on potential energy surfaces are most important. However, in the case of excited to ground state transitions the minimum on the excited state, which would be a logical direction of momentum for a molecule to carry through the transition, actually corresponds to a maximum on the ground state. Including this insight in the method led to the ability to model additional chemical products observed in experiment. However, this type of insight requires coupling of geometrical intuition in multidimensional space with a chemical understanding of reaction mechanism in the context of potential energy surfaces.

Active Engagement with Other Scientists

Active engagement with scientists outside of one's field also promotes creativity and appropriate incorporation of the value of these fields into one's own. An illustrative example comes from a collaboration meeting in which I was facilitating collaboration between chemists and statisticians. One of the chemists was reporting results of attempting an approach suggested by the statisticians and reported that they were actually getting a negative R^2 which was confusing to both the chemists and the statisticians. I asked a simple question which shed light on the issue: "What is the distribution of catalyst frequency like?". It turned out that the majority of the reactions in the dataset had used the same type of catalyst. To the chemist thinking in the language of chemistry, such details can seem so obvious, insignificant, or mundane that they do not naturally enter the discussion. Of course most of the time people would use the standard catalyst for that type of reaction. To the statistician thinking in the language of statistics, such details are clearly important and worthy of mention. Of course you would want to know if the distribution of your categorical feature is extremely far from uniform and is dominated by a single value! However, due to the language and cultural barrier, the important information is not naturally communicated from the chemists to the statisticians. Someone who understands both cultures and speaks both languages is needed to translate. The ability of a collaborative team to communicate across cultural and disciplinary language barriers has a transformative impact on the team.

Conclusion

I hope that these principles will enable real progress in the ability of humans to create and interact with machines in ways that facilitate deeper understanding of truth, effective stewardship of our environment, personal growth,[157] and meaningful loving relationships. Specific questions I hope will be answered towards this end are how an algorithm can learn scientific meaning and

effectively teach what it has learned to humans, as well as how human and computer interaction can be better utilized to facilitate joint learning.

Bibliography

(1) Weiss, K.; Khoshgoftaar, T. M.; Wang, D. A Survey of Transfer Learning. *J. Big Data* **2016**, *3* (1), 9. https://doi.org/10.1186/s40537-016-0043-6.

(2) Bishop, C. M. *Pattern Recognition and Machine Learning*; 2006; Vol. 4. https://doi.org/10.1117/1.2819119.

(3) Sutton, R.; Barto, A. *Reinforcment Learning*; 2018.

(4) Albini, A. *Photochemistry*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2016. https://doi.org/10.1007/978-3-662-47977-3.

(5) Matsika, S.; Krause, P. Nonadiabatic Events and Conical Intersections. *Annu. Rev. Phys. Chem.* **2011**, *62*, 621–643. https://doi.org/10.1146/annurev-physchem-032210-103450.

(6) Levine, B. G.; Martínez, T. J. Isomerization through Conical Intersections. *Annu. Rev. Phys. Chem.* **2007**, *58*, 613–634. https://doi.org/10.1146/annurev.physchem.57.032905.104612.

(7) Worth, G. A.; Cederbaum, L. S. Beyond Born-Oppenheimer: Molecular Dynamics through a Conical Intersection. *Annu. Rev. Phys. Chem.* **2004**, *55*, 127–158. https://doi.org/10.1146/annurev.physchem.55.091602.094335.

(8) Wolczanski, P. T.; Chirik, P. J. A Career in Catalysis: John E. Bercaw. *ACS Catal.* **2015**, 150209095624001. https://doi.org/10.1021/acscatal.5b00076.

(9) Campbell, C. T. The Degree of Rate Control: A Powerful Tool for Catalysis Research. *ACS Catal.* **2017**, *7* (4), 2770–2779. https://doi.org/10.1021/acscatal.7b00115.

(10) Fdez. Galván, I.; Delcey, M. G.; Pedersen, T. B.; Aquilante, F.; Lindh, R. Analytical State-Average Complete-Active-Space Self-Consistent Field Nonadiabatic Coupling Vectors: Implementation with Density-Fitted Two-Electron Integrals and Application to Conical Intersections. *J. Chem. Theory Comput.* **2016**, *12* (8), 3636–3653. https://doi.org/10.1021/acs.jctc.6b00384.

(11) Köppel, H. Regularized Diabatic States and Quantum Dynamics on Intersecting Potential Energy Surfaces. *Faraday Discuss.* **2004**, *127* (0), 35–47. https://doi.org/10.1039/B314471B.

(12) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann Generators: Sampling Equilibrium States of Many-Body Systems with Deep Learning. *Science (80-.).* **2019**, *365* (6457), eaaw1147. https://doi.org/10.1126/science.aaw1147.

(13) Duros, V.; Grizou, J.; Xuan, W.; Hosni, Z.; Long, D. L.; Miras, H. N.; Cronin, L. Human versus Robots in the Discovery and Crystallization of Gigantic Polyoxometalates. *Angew. Chemie - Int. Ed.* **2017**, *56* (36), 10815–10820. https://doi.org/10.1002/anie.201705721.

(14) Häse, F.; Fdez. Galván, I.; Aspuru-Guzik, A.; Lindh, R.; Vacher, M. How Machine Learning Can Assist the Interpretation of Ab Initio Molecular Dynamics Simulations and Conceptual Understanding of Chemistry. *Chem. Sci.* **2019**, *10* (8), 2298–2307. https://doi.org/10.1039/C8SC04516J.

(15) Chuang, K. V.; Keiser, M. J. Adversarial Controls for Scientific Machine Learning. *ACS Chem. Biol.* **2018**, *13* (10), 2819–2821. https://doi.org/10.1021/acschembio.8b00881.

(16) Chuang, K. V; Keiser, M. J. Comment on "Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning." *Science (80-.).* **2018**, *362* (6416), eaat8603. https://doi.org/10.1126/science.aat8603.

(17) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science (80-.).* **2018**, *360* (6385), 186–190. https://doi.org/10.1126/science.aar5169.

(18) Bernardi, F.; Olivucci, M.; Robb, M. a. Potential Energy Surface Crossings in Organic Photochemistry. *Chem. Soc. Rev.* **1996**, *25* (5), 321. https://doi.org/10.1039/cs9962500321.

(19) Yarkony, D. R. Conical Intersections: Diabolical and Often Misunderstood. *Acc. Chem. Res.* **1998**, *31* (8), 511–518. https://doi.org/10.1021/ar970113w.

(20) Levine, B. G.; Coe, J. D.; Martínez, T. J. Optimizing Conical Intersections without Derivative Coupling Vectors: Application to Multistate Multireference Second-Order Perturbation Theory (MS-CASPT2). *J. Phys. Chem. B* **2008**, *112* (2), 405–413. https://doi.org/10.1021/jp0761618.

(21) Zhang, X.; Herbert, J. M. Analytic Derivative Couplings for Spin-Flip Configuration Interaction Singles and Spin-Flip Time-Dependent Density Functional Theory. *J. Chem. Phys.* **2014**, *141* (6), 064104. https://doi.org/10.1063/1.4891984.

(22) Ou, Q.; Fatehi, S.; Alguire, E.; Shao, Y.; Subotnik, J. E. Derivative Couplings between TDDFT Excited States Obtained by Direct Differentiation in the Tamm-Dancoff Approximation. *J. Chem. Phys.* **2014**, *141* (2), 024114. https://doi.org/10.1063/1.4887256.

(23) Zimmerman, P. M.; Bell, F.; Goldey, M.; Bell, A. T.; Head-Gordon, M. Restricted Active Space Spin-Flip Configuration Interaction: Theory and Examples for Multiple Spin Flips with Odd Numbers of Electrons. *J. Chem. Phys.* **2012**, *137*, 1–12. https://doi.org/10.1063/1.4759076.

(24) Mayhall, N. J.; Head-Gordon, M. Increasing Spin-Flips and Decreasing Cost: Perturbative Corrections for External Singles to the Complete Active Space Spin Flip Model for Low-Lying Excited States and Strong Correlation. *J. Chem. Phys.* **2014**, *141* (4), 044112.

https://doi.org/10.1063/1.4889918.

(25) Mayhall, N. J.; Goldey, M.; Head-Gordon, M. A Quasidegenerate 2nd-Order Perturbation Theory Approximation to RAS-NSF for Excited States and Strong Correlations. *J. Chem. Theory Comput.* **2014**, *10*, 589–599. https://doi.org/10.1021/ct400898p.

(26) Fales, B. S.; Levine, B. G. Nanoscale Multireference Quantum Chemistry: Full Configuration Interaction on Graphical Processing Units. *J. Chem. Theory Comput.* **2015**, *11* (10), 4708–4716. https://doi.org/10.1021/acs.jctc.5b00634.

(27) Hohenstein, E. G.; Luehr, N.; Ufimtsev, I. S.; Martínez, T. J. An Atomic Orbital-Based Formulation of the Complete Active Space Self-Consistent Field Method on Graphical Processing Units. *J. Chem. Phys.* **2015**, *142* (22), 224103. https://doi.org/10.1063/1.4921956.

(28) Snyder, J. W.; Hohenstein, E. G.; Luehr, N.; Martínez, T. J. An Atomic Orbital-Based Formulation of Analytical Gradients and Nonadiabatic Coupling Vector Elements for the State-Averaged Complete Active Space Self-Consistent Field Method on Graphical Processing Units. *J. Chem. Phys.* **2015**, *143* (15), 154107. https://doi.org/10.1063/1.4932613.

(29) Mori, T.; Glover, W. J.; Schuurman, M. S.; Martinez, T. J. Role of Rydberg States in the Photochemical Dynamics of Ethylene. *J. Phys. Chem. A* **2012**, *116* (11), 2808–2818. https://doi.org/10.1021/jp2097185.

(30) Mori, T.; Kato, S. Dynamic Electron Correlation Effect on Conical Intersections in Photochemical Ring-Opening Reaction of Cyclohexadiene: MS-CASPT2 Study. *Chem. Phys. Lett.* **2009**, *476* (1–3), 97–100. https://doi.org/10.1016/j.cplett.2009.05.067.

(31) Page, C. S.; Olivucci, M. Ground and Excited State CASPT2 Geometry Optimizations of Small Organic Molecules. *J. Comput. Chem.* **2003**, *24* (3), 298–309. https://doi.org/10.1002/jcc.10145.

(32) Maeda, S.; Ohno, K.; Morokuma, K. Updated Branching Plane for Finding Conical Intersections without Coupling Derivative Vectors. *J. Chem. Theory Comput.* **2010**, *6* (5), 1538–1545. https://doi.org/10.1021/ct1000268.

(33) Tully, J. C. Molecular Dynamics with Electronic Transitions. *J. Chem. Phys.* **1990**, *93* (2), 1061. https://doi.org/10.1063/1.459170.

(34) Ben-Nun, M.; Quenneville, J.; Martínez, T. J. Ab Initio Multiple Spawning: Photochemistry from First Principles Quantum Molecular Dynamics. *J. Phys. Chem. A* **2000**, *104* (22), 5161–5175. https://doi.org/10.1021/jp994174i.

(35) Wang, L.; Prezhdo, O. V. A Simple Solution to the Trivial Crossing Problem in Surface Hopping. *J. Phys. Chem. Lett.* **2014**, *5* (4), 713–719. https://doi.org/10.1021/jz500025c.

(36) Pittner, J.; Lischka, H.; Barbatti, M. Optimization of Mixed Quantum-Classical Dynamics:

Time-Derivative Coupling Terms and Selected Couplings. *Chem. Phys.* **2009**, *356* (1–3), 147–152. https://doi.org/10.1016/j.chemphys.2008.10.013.

(37) Hammes-Schiffer, S.; Tully, J. C. Proton Transfer in Solution: Molecular Dynamics with Quantum Transitions. *J. Chem. Phys.* **1994**, *101* (6), 4657. https://doi.org/10.1063/1.467455.

(38) Meek, G. A.; Levine, B. G. Evaluation of the Time-Derivative Coupling for Accurate Electronic State Transition Probabilities from Numerical Simulations. *J. Phys. Chem. Lett.* **2014**, *5* (13), 2351–2356. https://doi.org/10.1021/jz5009449.

(39) Meek, G. A.; Levine, B. G. Accurate and Efficient Evaluation of Transition Probabilities at Unavoided Crossings in Ab Initio Multiple Spawning. *Chem. Phys.* **2015**, *460*, 117–124. https://doi.org/10.1016/j.chemphys.2015.06.007.

(40) Herman, M. F. Nonadiabatic Semiclassical Scattering. I. Analysis of Generalized Surface Hopping Procedures. *J. Chem. Phys.* **1984**, *81* (1984), 754. https://doi.org/10.1063/1.447708.

(41) Sicilia, F.; Bearpark, M. J.; Blancafort, L.; Robb, M. A. An Analytical Second-Order Description of the S 0 /S 1 Intersection Seam: Fulvene Revisited. *Theor. Chem. Acc.* **2007**, *118* (1), 241–251. https://doi.org/10.1007/s00214-007-0320-8.

(42) Bearpark, M. J.; Robb, M. A.; Bernhard Schlegel, H. A Direct Method for the Location of the Lowest Energy Point on a Potential Surface Crossing. *Chem. Phys. Lett.* **1994**, *223* (3), 269–274. https://doi.org/10.1016/0009-2614(94)00433-1.

(43) Davidson, E. The Iterative Calculation of a Few of the Lowest Eigenvalues and Corresponding Eigenvectors of Large Real-Symmetric Matrices. *J. Comput. Phys.* **1975**, *17*, 87–94. https://doi.org/10.1016/0021-9991(75)90065-0.

(44) Liu, B. The Simultaneous Expansion Method for the Iterative Solution of Several of the Lowest Eigenvalues and Corresponding Eigenvectors of Large Real-Symmetric Matrices. *Numer. Algorithms Chem. Algebr. Methods, Lawrence Berkeley Lab. Univ. Calif.* **1978**, 49–53.

(45) Leininger, M. L.; Sherrill, C. D.; Allen, W. D.; Schaefer, H. F. Systematic Study of Selected Diagonalization Methods for Configuration Interaction Matrices. *J. Comput. Chem.* **2001**, *22* (13), 1574–1589. https://doi.org/10.1002/jcc.1111.

(46) Sharada, S. M.; Bell, A. T.; Head-Gordon, M. A Finite Difference Davidson Procedure to Sidestep Full Ab Initio Hessian Calculation: Application to Characterization of Stationary Points and Transition State Searches. *J. Chem. Phys.* **2014**, *140* (16), 164115. https://doi.org/10.1063/1.4871660.

(47) Matsika, S.; Yarkony, D. R. Accidental Conical Intersections of Three States of the Same Symmetry. I. Location and Relevance. *J. Chem. Phys.* **2002**, *117* (15), 6907. https://doi.org/10.1063/1.1513304.

(48) Coe, J. D.; Martínez, T. J. Competitive Decay at Two- and Three-State Conical Intersections in Excited-State Intramolecular Proton Transfer. *J. Am. Chem. Soc.* **2005**, *127* (13), 4560–4561. https://doi.org/10.1021/ja043093j.

(49) Quenneville, J.; Martínez, T. J. Ab Initio Study of Cis−Trans Photoisomerization in Stilbene and Ethylene. *J. Phys. Chem. A* **2003**, *107* (6), 829–837. https://doi.org/10.1021/jp021210w.

(50) Sicilia, F.; Blancafort, L.; Bearpark, M. J.; Robb, M. A. Quadratic Description of Conical Intersections: Characterization of Critical Points on the Extended Seam. *J. Phys. Chem. A* **2007**, *111* (11), 2182–2192. https://doi.org/10.1021/jp067614w.

(51) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. Molpro: A General-Purpose Quantum Chemistry Program Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (2), 242–253. https://doi.org/10.1002/wcms.82.

(52) Roos, B. O. The Complete Active Space Self-Consistent Field Method and Its Applications in Electronic Structure Calculations. In *Advances in Chemical Physics*; 1987; Vol. 69, pp 399–445. https://doi.org/10.1002/9780470142943.ch7.

(53) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9* (8), 3404–3419. https://doi.org/10.1021/ct400195d.

(54) Pereira, F.; Xiao, K.; Latino, D. A. R. S.; Wu, C.; Zhang, Q.; Aires-de-Sousa, J. Machine Learning Methods to Predict Density Functional Theory B3LYP Energies of HOMO and LUMO Orbitals. *J. Chem. Inf. Model.* **2017**, *57* (1), 11–21. https://doi.org/10.1021/acs.jcim.6b00340.

(55) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533* (7601), 73–76. https://doi.org/10.1038/nature17439.

(56) St. John, P. C.; Kairys, P.; Das, D. D.; McEnally, C. S.; Pfefferle, L. D.; Robichaud, D. J.; Nimlos, M. R.; Zigler, B. T.; McCormick, R. L.; Foust, T. D.; et al. A Quantitative Model for the Prediction of Sooting Tendency from Molecular Structure. *Energy & Fuels* **2017**, *31* (9), 9983–9990. https://doi.org/10.1021/acs.energyfuels.7b00616.

(57) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4* (2), 268–276. https://doi.org/10.1021/acscentsci.7b00572.

(58) Struebing, H.; Ganase, Z.; Karamertzanis, P. G.; Siougkrou, E.; Haycock, P.; Piccione, P. M.; Armstrong, A.; Galindo, A.; Adjiman, C. S. Computer-Aided Molecular Design of

Solvents for Accelerated Reaction Kinetics. *Nat. Chem.* **2013**, *5* (11), 952–957. https://doi.org/10.1038/nchem.1755.

(59) Kayala, M. a; Baldi, P. ReactionPredictor: Prediction of Complex Chemical Reactions at the Mechanistic Level Using Machine Learning. *J. Chem. Inf. Model.* **2012**, *52* (10), 2526–2540.

(60) Ulissi, Z. W.; Medford, A. J.; Bligaard, T.; Nørskov, J. K. To Address Surface Reaction Network Complexity Using Scaling Relations Machine Learning and DFT Calculations. *Nat. Commun.* **2017**, *8*, 14621. https://doi.org/10.1038/ncomms14621.

(61) Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. Controlling an Organic Synthesis Robot with Machine Learning to Search for New Reactivity. *Nature* **2018**, *559* (7714), 377–381. https://doi.org/10.1038/s41586-018-0307-8.

(62) Corey, E. J.; Wipke, W. T. Computer-Assisted Design of Complex Organic Syntheses. *Science (80-.).* **1969**, *166* (3902), 178–192. https://doi.org/10.1126/science.166.3902.178.

(63) PENSAK, D. A.; COREY, E. J. LHASA—Logic and Heuristics Applied to Synthetic Analysis; 1977; pp 1–32. https://doi.org/10.1021/bk-1977-0061.ch001.

(64) Corey, E. J. General Methods for the Construction of Complex Molecules. *Pure Appl. Chem.* **1967**, *14* (1), 19–38. https://doi.org/10.1351/pac196714010019.

(65) Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chemie Int. Ed.* **2016**, *55* (20), 5904–5937. https://doi.org/10.1002/anie.201506101.

(66) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. *Nature* **2018**, *555* (7698), 604–610. https://doi.org/10.1038/nature25978.

(67) Coley, C. W.; Barzilay, R.; Jaakkola, T. S.; Green, W. H.; Jensen, K. F. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **2017**, *3* (5), 434–443. https://doi.org/10.1021/acscentsci.7b00064.

(68) Segler, M. H. S.; Waller, M. P. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chem. - A Eur. J.* **2017**. https://doi.org/10.1002/chem.201605499.

(69) Segler, M. H. S.; Waller, M. P. Modelling Chemical Reasoning to Predict and Invent Reactions. *Chem. - A Eur. J.* **2017**. https://doi.org/10.1002/chem.201604556.

(70) Schneider, N.; Lowe, D. M.; Sayle, R. A.; Landrum, G. A. Development of a Novel Fingerprint for Chemical Reactions and Its Application to Large-Scale Reaction Classification and Similarity. *J. Chem. Inf. Model.* **2015**, *55* (1), 39–53. https://doi.org/10.1021/ci5006614.

(71) Schneider, N.; Stiefl, N.; Landrum, G. A. What's What: The (Nearly) Definitive Guide to Reaction Role Assignment. *J. Chem. Inf. Model.* **2016**, *56* (12), 2336–2346. https://doi.org/10.1021/acs.jcim.6b00564.

(72) Jin, W.; Coley, C. W.; Barzilay, R.; Jaakkola, T. Predicting Organic Reaction Outcomes with Weisfeiler-Lehman Network. **2017**, No. Nips, 1–10.

(73) Schwaller, P.; Gaudin, T.; Lányi, D.; Bekas, C.; Laino, T. "Found in Translation": Predicting Outcomes of Complex Organic Chemistry Reactions Using Neural Sequence-to-Sequence Models. *Chem. Sci.* **2018**, *9* (28), 6091–6098. https://doi.org/10.1039/C8SC02339E.

(74) Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein-Ligand Scoring with Convolutional Neural Networks. **2016**, 1–47. https://doi.org/10.1021/acs.jcim.6b00740.

(75) Duvenaud, D.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Gómez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. *J. Chem. Inf. Model.* **2015**, *56* (2), 399–411. https://doi.org/10.1021/acs.jcim.5b00572.

(76) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8* (4), 3192–3203. https://doi.org/10.1039/C6SC05720A.

(77) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. **2017**.

(78) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8*, 13890. https://doi.org/10.1038/ncomms13890.

(79) Marcus, G. Deep Learning: A Critical Appraisal. *arXiv Prepr. arXiv1801.00631* **2018**, 1–27.

(80) Hornik, K. Approximation Capabilities of Multilayer Feedforward Networks. *Neural Networks* **1991**, *4* (2), 251–257. https://doi.org/10.1016/0893-6080(91)90009-T.

(81) Leach, A. R.; Gillet, V. J. *An Introduction To Chemoinformatics*; Springer Netherlands, 2007. https://doi.org/10.1007/978-1-4020-6291-9.

(82) Fernández-De Gortari, E.; García-Jacas, C. R.; Martinez-Mayorga, K.; Medina-Franco, J. L. Database Fingerprint (DFP): An Approach to Represent Molecular Databases. *J. Cheminform.* **2017**. https://doi.org/10.1186/s13321-017-0195-1.

(83) Rogers, D. J.; Tanimoto, T. T. A Computer Program for Classifying Plants. *Science (80-.).* **1960**, *132* (3434), 1115–1118. https://doi.org/10.1126/science.132.3434.1115.

(84) Bajusz, D.; Rácz, A.; Héberger, K. Why Is Tanimoto Index an Appropriate Choice for

Fingerprint-Based Similarity Calculations? *J. Cheminform.* **2015**. https://doi.org/10.1186/s13321-015-0069-3.

(85) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28* (1), 31–36. https://doi.org/10.1021/ci00057a005.

(86) Libman, A.; Shalit, H.; Vainer, Y.; Narute, S.; Kozuch, S.; Pappo, D. Synthetic and Predictive Approach to Unsymmetrical Biphenols by Iron-Catalyzed Chelated Radical–Anion Oxidative Coupling. *J. Am. Chem. Soc.* **2015**, *137* (35), 11453–11460. https://doi.org/10.1021/jacs.5b06494.

(87) Hammett, L. P. The Effect of Structure Upon the Reactions of Organic Compounds. Temperature and Solvent Influences. *J. Chem. Phys.* **1936**, *4* (9), 613–617. https://doi.org/10.1063/1.1749914.

(88) Christian, A. H.; Niemeyer, Z. L.; Sigman, M. S.; Toste, F. D. Uncovering Subtle Ligand Effects of Phosphines Using Gold(I) Catalysis. *ACS Catal.* **2017**, *7* (6), 3973–3978. https://doi.org/10.1021/acscatal.7b00757.

(89) Orlandi, M.; Coelho, J. A. S.; Hilton, M. J.; Toste, F. D.; Sigman, M. S. Parameterization of Noncovalent Interactions for Transition State In-Terrogation Applied to Asymmetric Catalysis. *J. Am. Chem. Soc.* **2017**, jacs.7b02311. https://doi.org/10.1021/jacs.7b02311.

(90) Seeman, J. I. The Curtin-Hammett Principle and the Winstein-Holness Equation: New Definition and Recent Extensions to Classical Concepts. *J. Chem. Educ.* **1986**, *63* (1), 42. https://doi.org/10.1021/ed063p42.

(91) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; University Science Books, 2006.

(92) Das, A.; Agrawal, H.; Zitnick, L.; Parikh, D.; Batra, D. Human Attention in Visual Question Answering: Do Humans and Deep Networks Look at the Same Regions? *Comput. Vis. Image Underst.* **2017**, *163*, 90–100. https://doi.org/10.1016/j.cviu.2017.10.001.

(93) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114* (10), 105503. https://doi.org/10.1103/PhysRevLett.114.105503.

(94) Janet, J. P.; Kulik, H. J. Resolving Transition Metal Chemical Space: Feature Selection for Machine Learning and Structure–Property Relationships. *J. Phys. Chem. A* **2017**, *121* (46), 8939–8954. https://doi.org/10.1021/acs.jpca.7b08750.

(95) Huo, H.; Rupp, M. Unified Representation of Molecules and Crystals for Machine Learning. **2017**.

(96) Huang, B.; von Lilienfeld, O. A. Understanding Molecular Representations in Machine

Learning: The Role of Uniqueness and Target Similarity. *J. Chem. Phys.* **2016**, *145* (16), 161102. https://doi.org/10.1063/1.4964627.

(97) Jaeger, S.; Fulle, S.; Turk, S. Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition. *J. Chem. Inf. Model.* **2018**, acs.jcim.7b00616. https://doi.org/10.1021/acs.jcim.7b00616.

(98) Pronobis, W.; Tkatchenko, A.; Müller, K.-R. Many-Body Descriptors for Predicting Molecular Properties with Machine Learning: Analysis of Pairwise and Three-Body Interactions in Molecules. *J. Chem. Theory Comput.* **2018**, *14* (6), 2991–3003. https://doi.org/10.1021/acs.jctc.8b00110.

(99) Weinhold, F.; Landis, C. R. Natural Bond Orbitals and Extensions of Localized Bonding Concepts. *Chem. Educ. Res. Pr.* **2001**, *2* (2), 91–104. https://doi.org/10.1039/B1RP90011K.

(100) Aldaz, C.; Kammeraad, J. A.; Zimmerman, P. M. Discovery of Conical Intersection Mediated Photochemistry with Growing String Methods. *Phys. Chem. Chem. Phys.* **2018**, *20* (43), 27394–27405. https://doi.org/10.1039/C8CP04703K.

(101) Zimmerman, P. M. Single-Ended Transition State Finding with the Growing String Method. *J. Comput. Chem.* **2015**, *36* (9), 601–611. https://doi.org/10.1002/jcc.23833.

(102) Dewyer, A. L.; Zimmerman, P. M. Finding Reaction Mechanisms, Intuitive or Otherwise. *Org. Biomol. Chem.* **2017**. https://doi.org/10.1039/c6ob02183b.

(103) Dewyer, A. L.; Argüelles, A. J.; Zimmerman, P. M. Methods for Exploring Reaction Space in Molecular Systems. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2018. https://doi.org/10.1002/wcms.1354.

(104) Feldmann, M. T.; Widicus, S. L.; Blake, G. A.; Kent, D. R.; Goddard, W. A. Aminomethanol Water Elimination: Theoretical Examination. *J. Chem. Phys.* **2005**, *123* (3), 034304. https://doi.org/10.1063/1.1935510.

(105) Toda, K.; Yunoki, S.; Yanaga, A.; Takeuchi, M.; Ohira, S.-I.; Dasgupta, P. K. Formaldehyde Content of Atmospheric Aerosol. *Environ. Sci. Technol.* **2014**, *48* (12), 6636–6643. https://doi.org/10.1021/es500590e.

(106) Behera, S. N.; Sharma, M.; Aneja, V. P.; Balasubramanian, R. Ammonia in the Atmosphere: A Review on Emission Sources, Atmospheric Chemistry and Deposition on Terrestrial Bodies. *Environ. Sci. Pollut. Res.* **2013**, *20* (11), 8092–8131. https://doi.org/10.1007/s11356-013-2051-9.

(107) Ge, X.; Shaw, S. L.; Zhang, Q. Toward Understanding Amines and Their Degradation Products from Postcombustion CO 2 Capture Processes with Aerosol Mass Spectrometry. *Environ. Sci. Technol.* **2014**, *48* (9), 5066–5075. https://doi.org/10.1021/es4056966.

(108) Zimmerman, P. M.; Zhang, Z.; Musgrave, C. B. Simultaneous Two-Hydrogen Transfer as

a Mechanism for Efficient CO 2 Reduction. *Inorg. Chem.* **2010**, *49* (19), 8724–8728. https://doi.org/10.1021/ic100454z.

(109) Li, M. W.; Pendleton, I. M.; Nett, A. J.; Zimmerman, P. M. Mechanism for Forming B,C,N,O Rings from NH 3 BH 3 and CO 2 via Reaction Discovery Computations. *J. Phys. Chem. A* **2016**, *120* (8), 1135–1144. https://doi.org/10.1021/acs.jpca.5b11156.

(110) Zhang, J.; Zhao, Y.; Akins, D. L.; Lee, J. W. CO 2 -Enhanced Thermolytic H 2 Release from Ammonia Borane. *J. Phys. Chem. C* **2011**, *115* (16), 8386–8392. https://doi.org/10.1021/jp200049y.

(111) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20* (3), 273–297. https://doi.org/10.1007/BF00994018.

(112) Suykens, J. A. K.; Van Gestel, T.; De Brabanter, J.; De Moor, B.; Vandewalle, J. *Least Squares Support Vector Machines*; World Scientific Pub. Co.: Singapore, 2002.

(113) Wei, J. N.; Duvenaud, D.; Aspuru-Guzik, A. Neural Networks for the Prediction of Organic Chemistry Reactions. *ACS Cent. Sci.* **2016**, *2* (10), 725–732. https://doi.org/10.1021/acscentsci.6b00219.

(114) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *J. Chem. Theory Comput.* **2017**, *13* (11), 5255–5264. https://doi.org/10.1021/acs.jctc.7b00577.

(115) Fooshee, D.; Mood, A.; Gutman, E.; Tavakoli, M.; Urban, G.; Liu, F.; Huynh, N.; Van Vranken, D.; Baldi, P. Deep Learning for Chemical Reaction Prediction. *Mol. Syst. Des. Eng.* **2018**, *3* (3), 442–452. https://doi.org/10.1039/C7ME00107J.

(116) Grossman, R. B. *The Art of Writing Reasonable Organic Reaction Mechanisms*, 2nd ed.; Springer, 2000.

(117) Carey, F. A.; Sundberg, R. J. *Advanced Organic Chemistry: Part B: Reaction and Synthesis*, 5th ed.; Springer, 2010.

(118) Wallach, I.; Heifets, A. Most Ligand-Based Classification Benchmarks Reward Memorization Rather than Generalization. *J. Chem. Inf. Model.* **2018**, *58* (5), 916–932. https://doi.org/10.1021/acs.jcim.7b00403.

(119) Ess, D. H.; Houk, K. N. Theory of 1,3-Dipolar Cycloadditions: Distortion/Interaction and Frontier Molecular Orbital Models. *J. Am. Chem. Soc.* **2008**, *130* (31), 10187–10198. https://doi.org/10.1021/ja800009z.

(120) Liu, F.; Yang, Z.; Yu, Y.; Mei, Y.; Houk, K. N. Bimodal Evans–Polanyi Relationships in Dioxirane Oxidations of Sp 3 C–H: Non-Perfect Synchronization in Generation of Delocalized Radical Intermediates. *J. Am. Chem. Soc.* **2017**, *139* (46), 16650–16656. https://doi.org/10.1021/jacs.7b07988.

(121) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press, 2016.

(122) Zimmerman, P. M. Automated Discovery of Chemically Reasonable Elementary Reaction Steps. *J. Comput. Chem.* **2013**, *34* (16), 1385–1392. https://doi.org/10.1002/jcc.23271.

(123) Pendleton, I. M.; Pérez-Temprano, M. H.; Sanford, M. S.; Zimmerman, P. M. Experimental and Computational Assessment of Reactivity and Mechanism in C(Sp 3)–N Bond-Forming Reductive Elimination from Palladium(IV). *J. Am. Chem. Soc.* **2016**, *138* (18), 6049–6060. https://doi.org/10.1021/jacs.6b02714.

(124) Jafari, M.; Zimmerman, P. M. Uncovering Reaction Sequences on Surfaces through Graphical Methods. *Phys. Chem. Chem. Phys.* **2018**. https://doi.org/10.1039/c8cp00044a.

(125) Maia, J. D. C.; Urquiza Carvalho, G. A.; Mangueira, C. P.; Santana, S. R.; Cabral, L. A. F.; Rocha, G. B. GPU Linear Algebra Libraries and GPGPU Programming for Accelerating MOPAC Semiempirical Quantum Chemistry Calculations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3072–3081. https://doi.org/10.1021/ct3004645.

(126) Stewart, J. J. P. Stewart Computational Chemistry. *MOPAC2012*. 2012.

(127) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13* (12), 1173–1213. https://doi.org/10.1007/s00894-007-0233-4.

(128) Kartoun, U.; Stern, H.; Edan, Y. A Human-Robot Collaborative Reinforcement Learning Algorithm. *J. Intell. Robot. Syst.* **2010**, *60* (2), 217–239. https://doi.org/10.1007/s10846-010-9422-y.

(129) Wang, L. P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering Chemistry with an Ab Initio Nanoreactor. *Nat. Chem.* **2014**, *6* (12), 1044–1048. https://doi.org/10.1038/nchem.2099.

(130) Ulissi, Z. W.; Tang, M. T.; Xiao, J.; Liu, X.; Torelli, D. A.; Karamad, M.; Cummins, K.; Hahn, C.; Lewis, N. S.; Jaramillo, T. F.; et al. Machine-Learning Methods Enable Exhaustive Searches for Active Bimetallic Facets and Reveal Active Site Motifs for CO 2 Reduction. *ACS Catal.* **2017**, *7* (10), 6600–6608. https://doi.org/10.1021/acscatal.7b01648.

(131) John, I.; Adam, R.; Chris, S.; Debora, M. Learning Protein Structure with a Differentiable Simulator. *Iclr* **2019**, 1–21.

(132) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science (80-.).* **2019**, *363* (6424), eaau5631. https://doi.org/10.1126/science.aau5631.

(133) Mansimov, E.; Mahmood, O.; Kang, S.; Cho, K. Molecular Geometry Prediction Using a Deep Generative Graph Neural Network. **2019**, 1–29.

(134) Koepnick, B.; Flatten, J.; Husain, T.; Ford, A.; Silva, D.-A.; Bick, M. J.; Bauer, A.; Liu, G.; Ishida, Y.; Boykov, A.; et al. De Novo Protein Design by Citizen Scientists. *Nature* **2019**, *570* (7761), 390–394. https://doi.org/10.1038/s41586-019-1274-4.

(135) Hawkins, P. C. D. Conformation Generation: The State of the Art. *J. Chem. Inf. Model.* **2017**, *57* (8), 1747–1756. https://doi.org/10.1021/acs.jcim.7b00221.

(136) Gilbert, M. M.; Demars, M. D.; Yang, S.; Grandner, J. M.; Wang, S.; Wang, H.; Narayan, A. R. H.; Sherman, D. H.; Houk, K. N.; Montgomery, J. Synthesis of Diverse 11- and 12-Membered Macrolactones from a Common Linear Substrate Using a Single Biocatalyst. *ACS Cent. Sci.* **2017**, *3* (12), 1304–1310. https://doi.org/10.1021/acscentsci.7b00450.

(137) Bender, T. A.; Morimoto, M.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Supramolecular Host-Selective Activation of Iodoarenes by Encapsulated Organometallics. *J. Am. Chem. Soc.* **2019**, *141* (4), 1701–1706. https://doi.org/10.1021/jacs.8b11842.

(138) Zhang, K. Da; Ajami, D.; Gavette, J. V.; Rebek, J. Complexation of Alkyl Groups and Ghrelin in a Deep, Water-Soluble Cavitand. *Chem. Commun.* **2014**, *50* (38), 4895–4897. https://doi.org/10.1039/c4cc01643b.

(139) Štejfa, V.; Fulem, M.; Růžička, K. First-Principles Calculation of Ideal-Gas Thermodynamic Properties of Long-Chain Molecules by R1SM Approach—Application to n -Alkanes. *J. Chem. Phys.* **2019**, *150* (22), 224101. https://doi.org/10.1063/1.5093767.

(140) Vansteenkiste, P.; Pauwels, E.; Van Speybroeck, V.; Waroquier, M. Rules for Generating Conformers and Their Relative Energies in N-Alkanes with a Heteroelement O or S: Ethers and Alcohols, or Sulfides and Thiols. *J. Phys. Chem. A* **2005**, *109* (42), 9617–9626. https://doi.org/10.1021/jp051910b.

(141) Tasi, G.; Mizukami, F.; Pálinkó, I.; Csontos, J.; Gyorffy, W.; Nair, P.; Maeda, K.; Toba, M.; Niwa, S. I.; Kiyozumi, Y.; et al. Enumeration of the Conformers of Unbranched Aliphatic Alkanes. *J. Phys. Chem. A* **1998**, *102* (39), 7698–7703. https://doi.org/10.1021/jp981866i.

(142) Tasi, G.; Mizukami, F.; Csontos, J.; Gyorffy, W.; Pálinkó, I. Quantum Algebraic - Combinatoric Study of the Conformational Properties of n-Alkanes. II. *J. Math. Chem.* **2000**, *27* (3), 191–199. https://doi.org/10.1023/A:1026472102742.

(143) Tsujishita, H.; Hirono, S. CAMDAS: An Automated Conformational Analysis System Using Molecular Dynamics. *J. Comput. Aided. Mol. Des.* **1997**, *11* (3), 305–315. https://doi.org/10.1023/A:1007964913898.

(144) Chan, L.; Hutchison, G.; Morris, G. Bayesian Optimization for Conformer Generation. **2018**, 1–19. https://doi.org/10.26434/chemrxiv.7228940.v3.

(145) Vainio, M. J.; Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47* (6), 2462–2474.

https://doi.org/10.1021/Ci6005646.

(146) Sutton, R. S.; Precup, D.; Singh, S. Between MDPs and Semi-MDPs: A Framework for Temporal Abstraction in Reinforcement Learning. *Artif. Intell.* **1999**, *112* (1–2), 181–211. https://doi.org/10.1016/S0004-3702(99)00052-1.

(147) Vinyals, O.; Babuschkin, I.; Czarnecki, W. M.; Mathieu, M.; Dudzik, A.; Chung, J.; Choi, D. H.; Powell, R.; Ewalds, T.; Georgiev, P.; et al. Grandmaster Level in StarCraft II Using Multi-Agent Reinforcement Learning. *Nature* **2019**, No. August. https://doi.org/10.1038/s41586-019-1724-z.

(148) Kandasamy, K.; Dasarathy, G.; Oliva, J. B.; Schneider, J.; Poczos, B. Multi-Fidelity Gaussian Process Bandit Optimisation. **2016**, 1–45.

(149) Schulz-Gasch, T.; Schärfer, C.; Guba, W.; Rarey, M. TFD: Torsion Fingerprints As a New Measure To Compare Small Molecule Conformations. *J. Chem. Inf. Model.* **2012**, *52* (6), 1499–1512. https://doi.org/10.1021/ci2002318.

(150) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17* (5–6), 490–519. https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P.

(151) Kleine, T.; Buendia, J.; Bolm, C. Mechanochemical Degradation of Lignin and Wood by Solvent-Free Grinding in a Reactive Medium. *Green Chem.* **2013**, *15* (1), 160–166. https://doi.org/10.1039/C2GC36456E.

(152) Kammeraad, J. A.; Zimmerman, P. M. Estimating the Derivative Coupling Vector Using Gradients. *J. Phys. Chem. Lett.* **2016**, *7* (24), 5074–5079. https://doi.org/10.1021/acs.jpclett.6b02501.

(153) Malhado, J. P.; Bearpark, M. J.; Hynes, J. T. Non-Adiabatic Dynamics Close to Conical Intersections and the Surface Hopping Perspective. *Front. Chem.* **2014**, *2* (November), 97. https://doi.org/10.3389/fchem.2014.00097.

(154) Pan, S. J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22* (10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191.

(155) Tan, J. Impact 2019 Proposal. *Personal Communication* **2019**, 1–4.

(156) Han, S.; Mao, L.; Gu, X.; Zhu, Y.; Ge, J.; Ma, Y. Neural Consequences of Religious Belief on Self-Referential Processing. *Soc. Neurosci.* **2008**. https://doi.org/10.1080/17470910701469681.

(157) Schwartz, J. M. Neuroplasticity and Spiritual Formation. *The Table* **2019**.