Designing and Implementing Computational Methods to Study Catalysis

by

Mina Jafari

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
2019

Doctoral Committee:

Associate Professor Paul Zimmerman, Chair
Assistant Professor Charles McCrory
Associate Professor Ambuj Tewari
Associate Professor Dominika Zgid

Mina Jafari

mjafari@umich.edu

ORCID iD: 0000-0003-1964-5469

To my father, who is a true intellectual, and my mother, for her unwavering support.

ACKNOWLEDGMENTS

My time at the University of Michigan gave me the opportunity to meet great people and make wonderful friends, all of whom helped me grow professionally and personally. First and foremost, I would like to express my sincere gratitude to my advisor, Professor Paul Zimmerman. This work would not have been completed without his support, encouragement, and patience. I am grateful to him for pushing me to think critically about my work, for emboldening me to connect with experts in the field, for his support during personal difficulties, and for supporting my professional development activities outside the lab. His continued encouragement and optimism motivated me during times when research progress was slow. I will also miss his famous carrot cakes and piano performances. I would also like to thank my committee members, Professors Charles McCrory, Ambuj Tewary, and Dominika Zgid for their flexibility and kindness.

Of course, a research group would not be perfect without kind and knowledgeable labmates. Special thanks to all Zimmerman group members, past and present, for their technical help and companionship. I always enjoyed spending time outside the lab with them in our social gatherings while also benefiting from their insights and ideas in my research. Alan Chien and Andrew Molina helped me get started in the lab when I first joined the group, Alan's sense of humor and affability makes him a unique person and made our lab welcoming. Andy Vitek's most memorable characteristic was the way he lounged at his desk! His knowledge about the world always impressed me. Cody Aldaz, Josh Kammeraad, and Michael Robo were my desk-mates whose advice was invaluable to my research, and whose company was a great break from work. I should also thank Josh for helping me through EECS545 class and being so patient and generous with his time, without him I did not stand a chance. Amanda Dewyer's commitment is astonishing, we went through many milestones of our Ph.D. life together and she always kept me motivated and gave me great advice. I would like to thank our post-docs for their help as well. Hyungjun Kim is a smart, hard working, and super nice person who kept helping me with my research even after he left the group. Tanmay Malakar is the Q-Chem expert in our lab who always has the answer to your questions. Matt Hannigan, Alan Rask, Allison Roessler, and Kevin Skinner were other labmates who helped me with various questions and problems.

My collaborators, Professor Alison Narayan and Jessica Yazarians provided the mate-

rial for my final research project and guided me with their insightful comments and ideas. I am sincerely thankful for their time and contributions.

I would like to thank David Braun, our system administrator, for keeping our computational resources up and running and also for his witty comments. Professor Charlie Brooks kept me and the whole computational suite fueled with leftovers from his group meeting lunch and I would like to thank him for that! Professor Eitan Geva gave me the opportunity to work with him on a chemical education project and surprised me with the random facts he knew about Iran. I am grateful that I had the chance to know him and work with him, hope our countries do the same one day. Assistant Dean John Godfrey was my advisor in GRIN and helped me and other Iranian students during very hard times. He always has a great story to tell. Thank you John for all you have done for us and other international students.

Living in Ann Arbor would not have been so enjoyable without my friends! First, I should thank Armin Sarabi for helping me to literally put the chapters of this thesis together! Having a programmer friend pays off. I would also like to thank all my friends who are my moral support and were my family away from home. You made Ann Arbor the greatest city in the world and I am grateful for all the happy memories we created together.

My boyfriend Amirhomayoun played a significant role in my Ph.D. endeavor, starting from my graduate school application until completion of my Ph.D. I should thank him for being my tech guy and answering my numerous programming questions and introducing me to so many great tools that I use on a daily basis. I am grateful to him for listening to me complain about life and research, for creating many great memories, and for providing solutions for my problems. Thank you Homayoun for being a great companion. My sister, Rana, is my best friend for life and partner in crime. I always envied her intelligence and dedication to science. Finally, I would like to thank my parents, for all their sacrifices and for always encouraging me to pursue education and supporting my decisions. Without them, I would have never had the courage to become who I am today. I am forever grateful.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

In recent years, new findings in different branches of chemistry have been highly dependent on computational and theoretical tools. Advancements in computing power and developments in computational chemistry methods has provided a unique opportunity for the field of computational method development to thrive. Investigations using new computational methods are not heavily guided by experiments anymore and need minimal structural and mechanistic insight from experiments for model development. Complex chemical transformations can be studied using these methods with a reasonable computation time. A variety of tools are built by computational chemists to assess the physical properties controlling chemical processes. This dissertation evaluates the viability of two classes of computational methods for examining chemical reactions: 1) reaction discovery tools and 2) multivariate analysis methods. Reaction path and transition state finding methods fall under the first category and provide means to gaining insight into the reaction mechanism and transition state structure at the atomistic level while quantifying kinetic and thermodynamics of the reaction. The ultimate goal of this type of studies is to either explain the experimental observations that do not follow traditional chemical principles or modify reaction conditions or reagents to improve the desired outcome of the reaction. However, in situations where there is not enough information to develop a model for reaction path finding studies, alternative methods such as multivariate analysis could be used to uncover mechanistic details and engineer catalysts at even a cheaper computation cost.

Chapter 1 provides an overview on regression tools and potential energy surfaces and their significance in chemistry. The motivations and limitation of designing and implementing new methods for surface chemistry reactions were also briefly discussed. Linear free energy relationships and their application in mechanistic studies and their relation to quantum mechanical methods were also briefly described. Chapter 2 introduces a new reaction path and transition state finding method called surface growing string method, for exploring surface reactions that is at least two times faster than the conventional nudged elastic band method and finds the structures along the reaction path, the transition state, and the product in one single run. Chapter 3 expands upon the growing string method and

the automation process for orientation sampling, geometry optimization, driving coordinate generation, and finally performing reaction path finding. The new algorithm described in Chapter 3 resulted in the surface-ZStruct program. Using surface-ZStruct, the complete reaction cycle for atomic layer deposition of titanium nitride, a diffusion barrier in microelectronics, is uncovered computationally for the first time. In Chapter 4, the biocatalysis of coumarin cross-coupling reactions is investigated using data analysis tools in an attempt to identify controlling factors of reactivity and selectivity. The crystal structure of the enzyme's active site is not available, and the only information at hand are the structure of the native substrate, structure of other substrates used in the screening process, and the yield of the reactions. Based on the results, we believe the reaction proceeds via a di-radical pathway through double H atom abstractions. The yield is also dependent on the size of the substituent on the coumarin scaffold, however this association varies in different substrate pairs. The results suggest that the enzyme pre-organizes the substrate pair into a complex that facilitates the initial O−H homolytic bond cleavage. Using the size and bond dissociation energies of phenolic OH bonds, the yield of the reaction is predicted within a reasonable accuracy.

CHAPTER 1

Introduction

1.1 Overview

Chemists have long faced the challenge of gaining insight into chemical reactions and reaction mechanisms using available experimental tools. While the evidence gathered through experiments is usually concrete and more trusted than the theoretical proofs, there are cases where it is impossible to explain experimental observations or test potential hypotheses using experiments. In these situations we turn to numerous computational methods that enable us to understand and explain reactions at the atomistic level (Figure 3.1). The most significant advantage of computational methods is their ability to reliably calculate the energy and geometry of the desired states.

Computational chemistry is especially useful for mechanistic understanding of chemical reactions. Atomistic level simulations provide the three-dimensional geometry of reaction intermediates, step-by-step sequence of elementary step reactions, the heat of reaction and activation energy for each elementary step, and the structure of the transition state (TS). Identifying the changes in connectivity happening at the TS structure helps chemists to describe the reaction using physical organic chemistry principles, explain experimental observations, and alter the reaction conditions to achieve a desired product [49, 144]. For example, if a computational mechanistic study unravels a key step to proceed through reductive elimination, changing the identity of the metal center, its oxidation state, or introducing directing or bulky ligands could help in changing the rate of this step [232]. Not only the computational study can help in identifying the key reaction steps, the proposed alternations in reaction conditions or reagent structures could also be tested through simulations before being carried out in experimental laboratories [49].

For chemical processes such as atomic layer deposition, little mechanistic information is known on the gas-phase and surface reactions and experimental techniques are not able to reveal elementary step reactions. The information from sequential reaction steps obtained

from computational studies could be used to optimize the growth conditions to improve the physical properties of the resulting deposited films. Computational studies could also be used to substitute existing precursors with new ones to reduce contamination or reaction temperatures. Similarly, mechanistic details and structural properties on an enzyme's catalytic cycle and active site could be used to design new synthetic routes for natural products and synthesize chemicals with high stereo- and regio-selectivity.

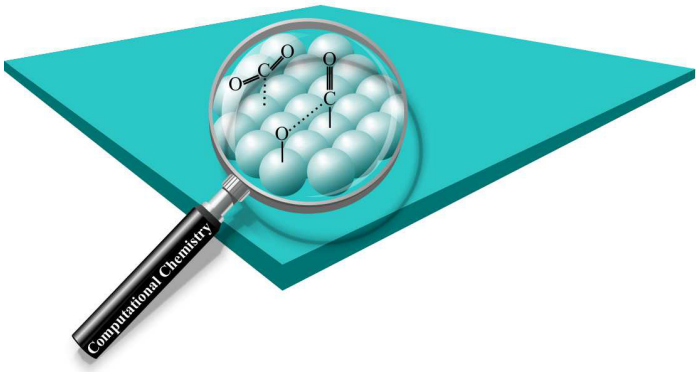


Figure 1.1: Computational methods are a probe for gaining insight into reaction mechanisms and detailing every single step of a reaction.

In addition to powerful computational methods, advances in computing power made it possible to implement and execute more accurate quantum mechanical methods and even compute and visualize the progress of a chemical reaction. Efficient computational methods such as density functional theory (DFT) have been widely used for the resolution of chemical mechanisms because of their low cost-to-accuracy ratio, these methods provide accurate, consistent, and reliable results. Despite the success in computing minimum energy structures, finding unstable TS structures connecting the desired reactant and products is still a challenge. Multiple structures should be generated and optimized in order to find a reaction path (RP) and the TS along that path, or a guess TS structure should be optimized to a saddle point and then to local minima to confirm that the correct TS is identified. Even though a single geometry optimization or energy calculation is trivial, running multiple calculations on a system, especially when the system has more than a few atoms or is periodic, could be time-consuming if efficient methods are not available. So, designing more efficient and reliable computational mechanism discovery methods is an active area of research in computational chemistry and when there is not enough information to set up reaction path simulations, we turn to data analysis methods for explanation. In addition to reaction discovery tools, we have also explored the multivariable linear regression as a tool to understand an enzymatic reaction where very limited information about the enzyme is

available. This work focuses on expanding the capabilities of the growing string method (GSM) and designing and implementing a new method for automated reaction discovery of surface reactions (surface ZStruct).

1.2 Using DFT to Study Heterogeneous Catalysis

Surface reactions, and heterogeneous catalysis in particular, have a profound role in our daily life. Many important industrial processes such as ammonia production, catalytic cracking of gas oil, steam reforming of methane, methanol synthesis, and many other reactions depend on methods and practices of heterogeneous catalysis [86, 157]. Discovering new chemical procedures and catalytic routes depends mainly on experimental methods in chemistry, however these studies in any area of chemistry, including surface reactions, often raise intriguing questions that cannot be easily answered by the experiments themselves. In these cases, computational approaches that mostly rely on DFT methods (due to its accuracy and speed) become handy. Computational results are usually confirmed if they match experimental findings and then are used to explain puzzling experimental observations [76, 193, 197].

Periodic systems such as metal surfaces contain transition metals with hundreds of atoms, each with tens of electrons, making electronic structure calculations expensive and complex. DFT methods use the Kohn-Sham equation to solve the intractable many-electron Schrödinger equation by assuming non-interacting particles and treating each electron separately. Many functionals have been developed to accurately capture the interaction energy through the exchange-correlation term [86, 193]. Additionally, the core electrons of transition metals in surface chemistry calculations are considered frozen or treated using a pseudopotential [86]. The metal surface is usually modeled by a slab with a periodic structure in two or three dimensions. The slab size should be chosen carefully to avoid any undesired interactions between the adsorbates in neighboring slabs. A plane-wave basis set, defined by reciprocal lattice vectors, is used instead of localized functions while the plane-wave expansion must consist of only the plane-waves that satisfy the constraint of the lattice periodicity [86].

Choice of the slab size and the exchange-correlation functional can greatly influence the accuracy of DFT calculations. Here, we are concerned with two types of accuracy, physical accuracy deals with the accuracy of the DFT calculation compared to the "true" value obtained in a perfect experimental setting. Numerical accuracy is the reproducibility and accuracy of the computational results obtained by a researcher compared to the accumulated scientific conclusions [70, 71, 82, 111, 158, 222]. The conventional standard for

measuring both physical and numerical accuracy in heterogeneous catalysis calculations is comparing the lattice constant for experimental and other theoretical results. An extensive discussion on this parameter and its reproducibility can be found here [121]. More details on accuracy of DFT methods are presented in chapter 10 of this book [193].

Multiple benchmark studies have compared the accuracy of plane-wave DFT calculations with experimental results. Calculations on a set of 200 crystal structures including hydrides, borides, oxides, nitrides, carbides, and semiconductors had errors of 2% compared to experimental lattice parameters [6, 148]. Another property relevant to the accuracy of computational heterogeneous catalysis studies is the adsorption energy. This property is calculated for CO and NO molecules on multiple metal surfaces and the comparison suggests that PBE GGA functional of DFT overestimates these values [4, 85], with root mean square deviation of 0.67 and 0.43 eV, respectively. For assessment of ab initio methods' accuracy, an experimental reference is needed to validate the calculation results of different exchange-correlation functionals for thermodynamic studies [141].

1.3 Potential Energy Surfaces

A potential energy surface (PES) is a multidimetional surface that describes the changes in the energy of a system (molecule, cluster, or periodic system) at a given electronic state with respect to changes in its atomic coordinates and provides comprehensive information on all reaction paths. The dimensionality of PES increases linearly with the number of atoms in the system, which makes studying systems with more than a few atoms harder even with efficient computational methods. For a system with N atoms, there are 3N-6 dimensions or degrees of freedom that correspond to vibrational modes of the system (stretching, bending, etc.). There are multiple interestingpoints points on a PES that are important for designing and implementing reaction discovery tools. Minimum energy points on PES represent equilibrium structures (reactants, products, or intermediates) while transition states are first-order saddle points connecting the equilibrium structures [187]. The difference in energy between the starting structure and the TS is called the activation energy and the difference in energy between the starting structure and its first minimum energy intermediate is the heat of reaction. Every chemical reaction proceeds according to a reaction mechanism, which is a step-by-step description of what occurs during a reaction on the molecular level. Each step of the mechanism is known as an elementary process, which describes a single moment during a reaction in which molecules break and/or form new bonds.

A PES with 2 degrees of freedom (DoF) is shown in Figure 1.2 where the x and y axes represent DoF and the z axis is the energy of the system. Starting from the reactant

structure, there are two possible pathways with different TS and heats of reaction. Given the appropriate input parameters, a systematic and efficient reaction discovery tool should be able to find both RPs and the energies for each of the important structures along the path.

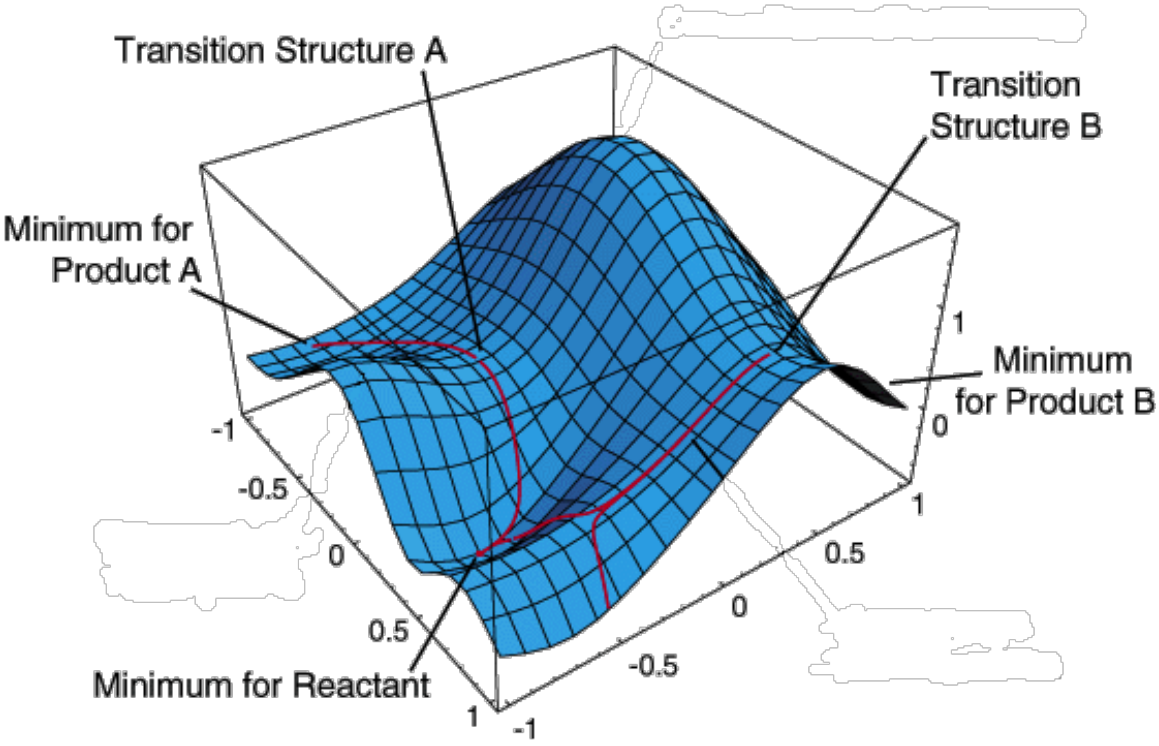


Figure 1.2: A schematic representation of a potential energy surface with 2 degrees of freedom. X and Y axes represent degrees of freedom and the Z axis showd the energy of the system.

1.4 Reaction Discovery Tools

Computational methods for calculating RPs and TSs are an essential part of mechanistic studies because they provide three-dimensional structures of the reagents involved in the reaction along with thermodynamics and activation energies that could be mapped to reaction rates. This information help us to elucidate the chemical process step-by-step, identify the rate limiting step, quantify the parameters that control the reactivity and selectivity of the reaction, and identify thermodynamic or kinetic control. These parameters could be further used to engineer more efficient ligands, catalysts, and other reagents in order to boost the yield of the desired reaction product and eliminate unproductive side reactions.

Despite advancements in computational reaction discovery, calculating RPs and TSs

is still challenging due to the large number of geometry optimization steps required to optimize all the chemical structures along a path. In addition to computing capacity limitations, an efficient RP finding method needs a reliable coordinate system that can handle different types of chemical reactions. It also needs strategies to quickly approach a saddle point, find the direction of the negative curvature to identify an approximate TS geometry, and an efficient optimizer to find the TS from that guess geometry. While many methods might have one or two of these components, an efficient and reliable RP finding method should combine all of them together. Methods developed for RP finding could be categorized into two groups based on the number of required input structures: double-ended [9, 19, 20, 24, 25, 38, 39, 45, 50, 59–61, 67, 74, 78, 79, 81, 90, 92, 94, 105, 106, 109, 112, 126, 138, 147, 165, 166, 168, 172, 176, 179, 180, 191, 192, 204, 208–210, 212, 226, 229, 235] (two input structures) or single-ended [3, 11–17, 22, 23, 26, 30, 37, 42, 44, 48, 55, 64, 83, 91, 95, 96, 100, 108, 119, 123, 124, 134, 136, 137, 142, 153, 160, 161, 170, 173, 174, 181, 185–187, 194, 207, 213, 214, 223, 227, 234, 236, 238] (one input structure). Reaction discovery tools could also be organized into four categories based on method of operation as discussed extensively elsewhere [56].

Reaction discovery tools are developed to automate the RPs and TSs calculation when minimum energy structures are known, but manually exploring the desired region of PES is unsystematic and tedious. Their main advantage is identifying new reaction pathways with minimum input from the user's chemical intuition specially when the system under study is not well characterized. This works describes the development of surface-GSM as RP finding tool for heterogeneous catalysis and surface-ZStruct to explore uncharted areas of surface reactions at ground-state. The top panel of Figure 1.3 shows the reaction network for ammonia borane (NH_3-BH_3) and aminoborane ($NH_2=BH_2$). The computational reaction discovery starts by inputting the optimized geometries of the reactants and designating reactive atoms. ZStruct uses this information as will be described in more detail in Chapter 3 to find elementary step reactions and products starting from the initial input structures. This procedure is carries out for each of the new intermediates until the full reaction network is elucidated. The grayed out reaction steps are not known before hand and the full network is calculated step-by-step using ZStruct. At the end, the calculated activation energies and heats of reaction could be analyzed and summarized in a PES plot as shown in the top right corner of Figure 1.3 and competing reaction pathways, rate limiting steps, and potential thermodynamic traps or energetically inaccessible routes could be identified.

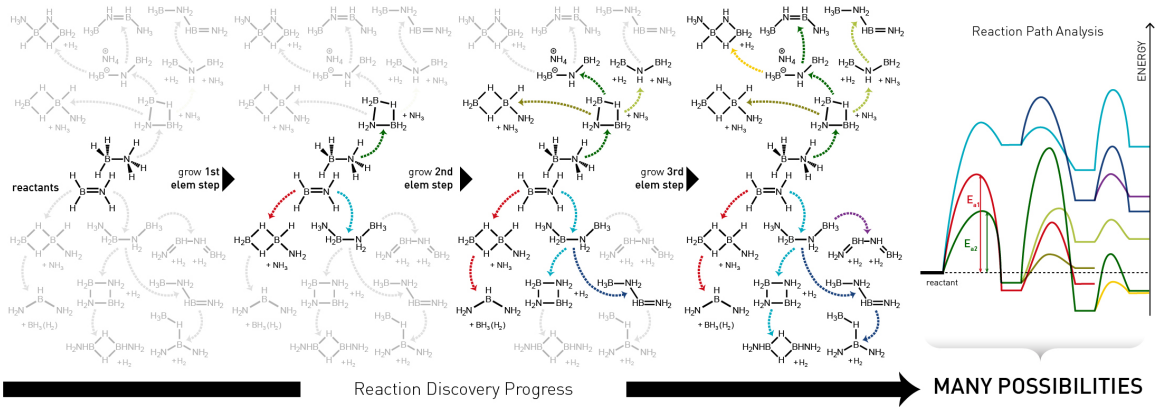


Figure 1.3: Depiction of reaction discovery progress for a bimolecular reaction. The grayed-out parts of the network are not known beforehand and are discovered using computational methods. The thermodynamic and kinetic information of the network with its elementary step reactions could be summarized in a PES plot.

1.4.1 Overview of the Growing String Method

GSM is a computational protocol developed to compute RPs and TSs [101, 168, 233, 235, 238, 239] that operates in single-ended (SE-GSM) and double-ended (DE-GSM) fashions. DE-GSM is used when both end-points are known and the objectives are finding the TS, activation energy, heat of reaction, and gaining insight into the reaction mechanism. SE-GSM is used when the final geometry is unknown or the goal is to explore multiple RPs starting from a given minimum energy structure. GSM has a built-in TS optimizer and computes a RP and exact TS in one single run for both single- and double-ended methods. Another advantage of GSM, making it more efficient compared to other RP finding methods, is its use of internal coordinates, which is a more natural way of representing bonds, angles, and torsions of a chemical system.

DE-GSM operates by inputing both reactant and product structures, converting the coordinates to internals and interpolating a reaction tangent connecting the two structures based on differences in the input coordinates. The RP grows from two ends, with the TS structure connecting two half paths. SE-GSM is more complicated, starting its search from a single input structure (reactant or product) and a set of driving coordinates. Driving coordinates are a combination of bond break and formation moves that would lead from reactant to a desired product, and their reliability also depends on the integration of internal coordinates into the GSM method.

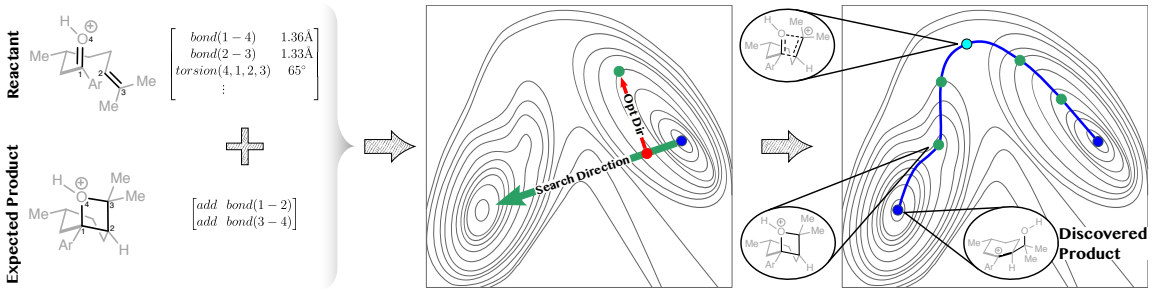


Figure 1.4: A schematic showing the progress of SE-GSM from a single input structure and using driving coordinates.

As shown in Figure 1.4, the reactant is an oxycation that is expected to form an oxetane product by forming a C−C and a C−O bond. In the beginning of the calculation, the geomtry is converted to internal coordinates and the bond lengths, angles, and torsions of the system are measured. Based on the expected product, two bonds between atomic indices 1 and 2, and 3 and 4 is desired to form. The calculated RP reveals that the reaction is asynchronous, based on the newly formed C−C and C−O bond lengths at the TS, and also discovers a more stable product. The expected oxetane product is shown to be an unstable intermediate on the RP that fragments through an energetically favorable ring-opening to result in benzylic carbocation intermediate.

1.4.2 Overview of ZStruct

ZStruct [57, 102, 234, 236] is developed to automate the process of running SE-GSM and systematically form a reaction network by only inputting a reactant structure and identifying the reactive atoms. ZStruct is implemented and validated for surface reactions in ground state for both unimolecular and bimolecular reactions. It aligns input structures automatically and generates multiple sets of driving coordinates based on connectivity information coming from internal coordinates and coordination number rules hard-coded in the program.

The main idea behind ZStruct is using graphs to describe the connection between atoms in a chemical system, where indices represent atoms and edges bonds. Using this model, the connection between atoms could easily be summarized in a matrix form and a connection transformation matrix could delineate the changes in connectivity (Figure 1.5). However, setting up such a matrix automatically requires a reliable coordinate system to accurately describe atomic connectivities.

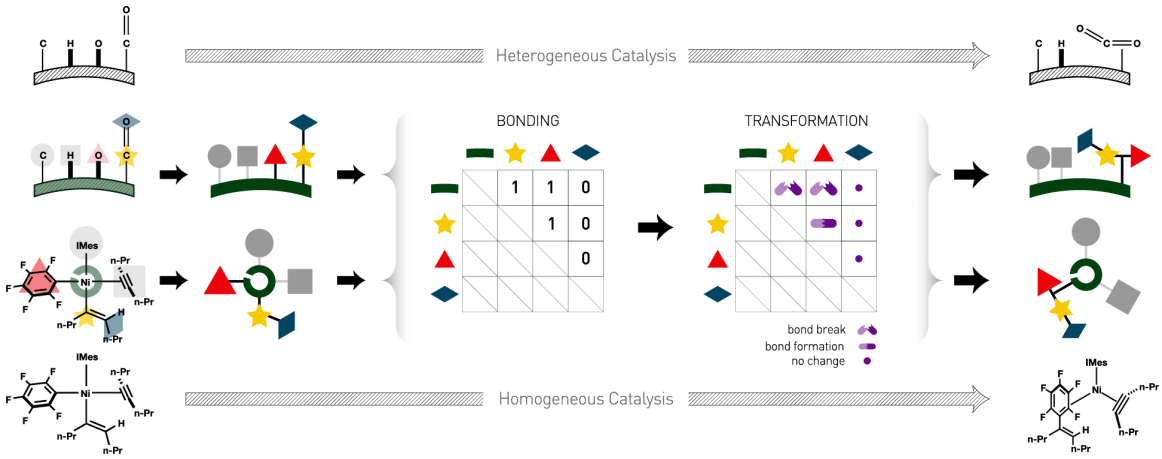


Figure 1.5: Using graphs to represent chemical tranformations of a molecule in a reaction.

Figure 1.5 shows examples from heterogeneous and homogeneous catalysis and how the chemicals in each reaction are transformed to a connectivity table. The top reaction shows a slab with four different chemisorbed species, each represented using a geometrical shape. The reactive atoms (the ones undergoing a bond breaking or formation) are showed in color while the unreactive ones are grayed out. Similarly, the bottom reaction shows the rate limiting step for C−H functionalization reaction of a four-coordinate nickel complex that undergoes reductive elimination to form the styrenyl product. In both cases, the metal center or slab is shown by dark green and active ligands or adsorbates in yellow, navy, and red. Bond formation or breaking is marked by 1 and 0 indicates no change in connectivity.

Despite improvements in the field of small molecule reaction discovery, the study of heterogeneous catalysis is still challenging because it involves surface atoms with high coordination numbers and molecular adsorbates, which make designing a coordinate system that can handle both components at the same time a difficult task. Internal coordinates are a natural way to describe bonds, angles, and torsions in molecular systems and provide a significant benefit for optimization of systems including atoms with low coordination numbers. However, use of internal coordinates becomes impractical when treating periodic systems which include a large number of atoms with high coordination numbers due to the huge number of primitive coordinates that can be present. This difference is expanded upon in Chapter 2. Having multiple binding sites for surface atoms and several potential adsorption sites in adsorbate molecules make the surface RP finding methods more challenging to design. For its validation, surface-ZStruct is used to study propanoic acid dissociation pathway (unimolecular reaction) and atomic layer deposition (ALD) of TiN that is described in Chapter 3, along with partial reaction network of tungsten deposition.

1.5 Linear Regression, The Statistical Chemist's Best Friend

As explained in previous sections, quantum chemistry can be used in RP finding studies to optimize chemical geometries while it can also provide parameters for regression analysis. The list includes bond dissociation energies of various bonds in a structure, Mulliken charges, natural bond orbital analysis that provides natural charge on each atom and orbital occupancies, spin density in radicals, HOMO/LUMO energies, and many other parameters. Some quantum chemistry parameters correspond to real physical properties and can be modified by changing the structure of the chemical. For example the HOMO/LUMO energies are a function of electron-donating/withdrawing property of the groups attached to a chemical scaffold [104].

In addition to quantum mechanical calculations, chemists also rely extensively on linear regression when it comes to establishing a relationship between experimental observations and chemical principles. Famous examples in this area include the Hammett equation or the Evans-Polanyi principle. If a linear relationship between the performance of a reaction and a chemical descriptor of the reagents (catalysts or reactants) is identified, the interpretability of the linear model makes virtual screening of potential reagents trivial. Linear relationships can also be used to gain insight into mechanistic details and identify the rate limiting step when modeling the reaction via RP finding methods is not possible due to limitations such as restricted information on reaction condition or reagent structures.

When a relationship between a parameter obtained from a calculation (like spin density or nucleophilicity) and a measurable reaction outcome (like yield or rate of the reaction) is demonstrated, extra steps should be taken to make sure the correlation is not a result of random fitting, especially when the number of data points is a lot smaller than the number of available features. To further confirm the proposed hypothesis concluded from the regression results, new predictions should be made on the same system and they should be tested by conducting more experiments, or the relationship should be observed in similar chemical systems. Multiple successful studies in this area that combine the information gathered from quantum mechanical calculations with evidence from experimental studies prove the utility of data analysis as a reliable prediction tool in chemistry [88, 183, 201].

1.6 Dissertation Outline

In Chapter 1, a brief overview of RP finding and regression tools was presented. The motivations and limitation of designing and implementing new methods for surface chem-

istry reactions were also briefly explained. ALD of tungsten was explained as an example showing the viability of the surface-ZStruct method. Finally, an alternative approach to RP finding methods that also relies on information obtained from quantum mechanical calculations was introduced.

Chapter 2 presents the details and workflow of the surface growing string method and the test set used for its validation. The performance of double-ended and single-ended growing string methods are benchmarked against the widely used climbing image method (NEB), and GSM is shown to be at least twice as efficient. In addition, GSM has been more reliable than the competing method, converging successfully when NEB fails.

Chapter 3 introduces the surface-ZStruct method for automated orientation sampling and reaction discovery of heterogeneous catalysis reactions. The new method is applied towards a variety of systems, including unimolecular and bimolecular reactions with multiple reaction steps. For the unimolecular case, the propanoic acid dissociation cycle is automatically calculated and the results are compared with other computational studies. The complete reaction cycle of TiN ALD is computationally elucidated for the first time. Surface-ZStruct not only found a reaction network, but it was also shown to be able to handle systems with tens of atoms. This method could be used for any under-studied surface reaction discovery as was demonstrated for ALD W example presented in the introduction.

Chapter 4 four describes studying the biocatalysis of coumarin cross-coupling reactions using data analysis tools to gain insight into a reaction where very limited information is available. The crystal structure of the enzyme's active site is not available, and the only information at hand are the structure of the native substrate, structure of other substrates used in the screening process, and the yield of the reactions. Based on the results, we believe the reaction proceeds via a di-radical pathway through double H atom abstractions. The yield is also dependent on the size of substituent on the coumarin scaffold, however this association varies in different substrate pairs. The results suggest that the enzyme preorganizes the substrate pair into a complex that facilitates the initial O−H homolytic bond cleavage. Using the size and bond dissociation energies of phenolic OH bonds, we are also able to predict the yield of the reaction within a reasonable accuracy.

Chapter 5 includes Final Remarks, in which the findings of the prior Chapters are reviewed, and possible future projects laid out. There have been a great improvement in the efficiency of surface RP finding methods, but novel strategies to handle surface reactions under potential bias will open new avenues for the study of electrochemical surface reactions. Access to large databases of chemical reaction parameters and outcomes coupled with quantum mechanical and TS finding methods will steer the course of mechanistic studies and catalyst design in new directions.

CHAPTER 2

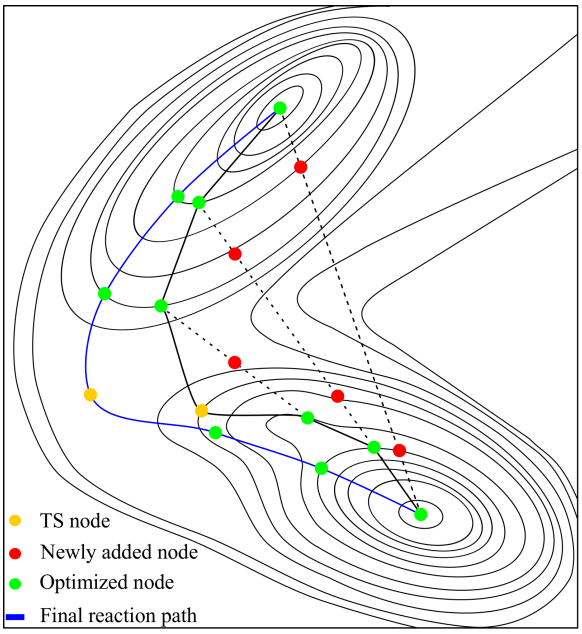
Reliable and Efficient Reaction Path and Transition State Finding for Surface Reactions with the Growing String Method

This Chapter largely based upon published work:
Reproduced with permission from M. Jafari and P.M. Zimmerman, J. Comput. Chem., 2017, 38, 645-658.

2.1 Abstract

The computational challenge of fast and reliable transition state and reaction path optimization requires new methodological strategies to maintain low cost, high accuracy, and systematic searching capabilities. The growing string method using internal coordinates has proven to be highly effective for the study of molecular, gas phase reactions, but difficulties in choosing a suitable coordinate system for periodic systems has prevented its use for surface chemistry. New developments are therefore needed, and presented herein, to handle surface reactions which include atoms with large coordination numbers that cannot be treated using standard internal coordinates. The double-ended and single-ended growing string methods are implemented using a hybrid coordinate system, then benchmarked for a test set of 43 elementary reactions occurring on surfaces. These results show that the growing string method is at least 45% faster than the widely used climbing image-nudged elastic band method, which also fails to converge in several of the test cases. Additionally, the surface growing string method has a unique single-ended search method which can move outward from an initial structure to find the intermediates, transition states, and reaction paths simultaneously. This powerful explorative feature of single ended-growing string method is demonstrated to uncover, for the first time, the mechanism for atomic layer deposition of TiN on Cu(111) surface. This reaction is found to proceed through multiple

hydrogen-transfer and ligand-exchange events, while formation of H-bonds stabilizes intermediates of the reaction. Purging gaseous products out of the reaction environment is the driving force for these reactions.



The growing string method is a powerful tool for locating reaction paths and transition states, using little user input. Herein, we present a new implementation of the growing string method for surface reactions and its application to the atomic layer deposition of titanium nitride.

2.2 Introduction

The information contained in transition state (TS) structures and reaction paths (RP) provides the fundamental atomistic details of reaction mechanisms. From a computational viewpoint, TSs are first-order saddle points on a potential energy surface (PES) representing the connection of two intermediates along a path. The high dimensionality of most PESs, however, makes TS-finding an impossible task unless fast, reliable, and accurate methods are available. Given the great interest in simulation of reactions on surfaces, such as atomic layer deposition (ALD), [73] heterogeneous catalysis, [149, 205] and electrochemical CO_2 reduction, [8] novel tools for TS and RP finding are in demand.

The algorithms designed to locate TSs and RPs are usually classified as single-ended [3, 11–17, 22, 23, 26, 30, 37, 42, 44, 48, 55, 64, 83, 91, 95, 96, 100, 108, 119, 123, 124, 134, 136, 137, 142, 153, 160, 161, 170, 173, 174, 181, 185–187, 194, 207, 213, 214, 223, 227, 234, 236, 238] or double-ended. [9, 19, 20, 24, 25, 38, 39, 45, 50, 59–61, 67, 74, 78, 79, 81, 90, 92, 94, 105, 106,

109, 112, 126, 138, 147, 165, 166, 168, 172, 176, 179, 180, 191, 192, 204, 208–210, 212, 226, 229, 235] Single-ended methods start from a single initial state and refine it systematically to locate a TS. Many single-ended methods require an initial guess-geometry lying close to the desired TS structure, which limits the effectiveness of these approaches. Double-ended methods, on the other hand, connect two structures in a discretized RP and are usually more reliable than single-ended methods due to the endpoints of the path being fixed, so double-ended methods are less likely to diverge to undesired search regions. Most double-ended algorithms do not compute the exact saddle point, so they are usually followed by a local search method (such as the dimer method, [91] mode-tracking, [22] or eigenvector following method [17, 187]) to refine the apparent TS structure to the exact TS. [233]

Given the multitude of methods available for TS and RP finding, these techniques are best summarized by the key components that afford their success. We classify these in four groups: 1.) Strategies for quickly approaching the vicinity of the saddle point, 2.) Estimation of the direction of negative curvature, 3.) Optimizer, and 4.) Coordinate system, which should all operate synergistically to rapidly and reliably locate TSs. In an ideal search algorithm, the combination of these four components should operate with little input from the user. These four areas will now be discussed to set the context for our proposed method.

Starting from an initial state, there are three often-used algorithmic strategies to approach the saddle point region. Minimum-mode following methods find the lowest curvature direction of the Hessian and follow this eigenvector towards the saddle point. [162,227] Alternatively, coordinate driving techniques push the initial structure towards an approximate TS structure along a specified reaction direction. [37, 185, 221] When the initial and final states are known, the highest energy point along an approximate reaction path from a double-ended method can be used as a good estimate of the exact TS. [94, 233]

After obtaining an approximate TS geometry, the direction corresponding to the transition vector must be estimated. While in principle the exact Hessian can be calculated and diagonalized to find the negative curvature direction, the computational cost can be expensive or prohibitive. To reduce this cost, approximate Hessians can be constructed and diagonalized via subspace iteration methods, [54, 120, 190, 227] or alternatively, the reaction tangent at the guess TS from a RP can provide an estimate of this direction. [233]

In addition to the two prerequisites of a good initial TS structure and reaction direction, an efficient optimizer [187] is necessary to direct the TS searches and refine reaction paths. In practice, quasi-Newton [187] methods are widely used because they update approximate Hessians at each optimization step, entirely skipping Hessian computations while still benefiting from PES curvature information. For TS searches, eigenvector following opti-

mizers [233] maximize the energy along the lowest Hessian mode while minimizing in all other directions. These methods tend to converge when the Hessian contains a reasonably accurate eigenvector representing the reaction direction.

The fourth component of interest is the coordinate system which forms the basis in which RPs and TSs are optimized. Cartesian coordinates are often chosen due to their simple implementation, despite internal coordinates (IC) being superior in many respects: chemical bonds are included as intrinsic coordinates, the curvilinear motion of angle bending or torsions are better represented by internals, and ICs have reduced intercoordinate coupling which allows faster optimization. Further advantages include that the interpolation of a RP in ICs avoids the collision of atoms or intersections of bonds, [13, 15, 165, 174, 185, 187, 233, 235, 238] and ICs can accelerate convergence of optimization by a factor of four. [9, 16, 27, 165]

Two common surface-compatible reaction finding methods are the nudged elastic band [105] (NEB) and the dimer [91] method. NEB and its variations [50, 92, 94, 210] interpolate between two structures in Cartesian coordinates to optimize a chain-of-states representation of the reaction path. NEB therefore is frequently used to form the guess for a transition state optimization by the dimer method in a two-step procedure. Multi-step computational procedures are inherently less user-friendly, suggesting new methods for simultaneous RP and TS finding with increased efficiency, reliability, and usability as promising additions to the computational toolkit.

Herein a novel means for systematic TS search and RP finding is implemented in a powerful tool for the study of surface reactions. The method is inspired by GSM [168, 233, 235, 238] and designed as a combined RP optimization and TS search algorithm. When the reactant and product structures are known, the new double-ended GSM (DE-GSM) can be used to calculate a RP and TS at low cost and high fidelity. In cases where the final structure is unknown, single-ended GSM (SE-GSM) can explore a new reaction space based on simple reaction coordinates as input. Detailed comparisons of three investigated methods (DE-, SE-GSM, and CI-NEB) are provided to benchmark their computational cost and reliability. The high usability of SE-GSM for exploring new reactions is demonstrated by showing an atomistic mechanism for the initiation and growth of titanium nitride on Cu(111) surface.

2.3 Method

2.3.1 Growing String Method with Exact TS Search

2.3.1.1 Overview

GSM develops a RP by iteratively adding new nodes and optimizing them until a complete RP with a TS and a stable intermediate on each side of the string are present. The string consists of a discretized set of structures along the RP connecting the reactant and product geometries, and is constructed starting only from the endpoints. By incremental addition of new nodes, GSM rapidly leads to a reasonably well converged RP since it avoids placing nodes at high-energy regions of the PES. [168]

Based on our experience using GSM for molecular systems, [233, 235, 238] we have developed a new method to overcome challenges of RP and TS finding for periodic systems and surface reactions. This method operates through three overall phases: growth, optimization, and exact TS search (Figure A.1), which now will be discussed in detail.

2.3.1.2 Growth Phase

During the growth phase, new nodes are added along the reaction tangent direction and minimized in directions perpendicular to the reaction tangent. The reaction tangent is defined either by interpolation or driving coordinates (see below), and used as constraint to prevent nodes from falling back to local minima. New nodes are added after the gradient at the frontier node drops below a predefined threshold, and the growth phase terminates when either two string fragments are connected (double-ended) or an intermediate on the other side of the string is found (single-ended).

The tangent definition during the growth phase depends on whether the algorithm is double-ended or single-ended. For DE-GSM, reaction tangent for node i pointing to node j is defined as

$$U_C = \alpha_c \sum_k \langle \Delta q | U_k^{(i)} \rangle U_k^{(i)} \tag{2.1}$$

where U_C is the (constrained) tangent direction, Δq is defined to be $\Delta q = q^{p,(j)} - q^{p,(i)}$, q^p are the primitive (hybrid) coordinates, α_c is a normalization factor, and the vectors U_k are the non-redundant (hybrid) coordinates vectors (see Coordinate System for Surfaces). Following the constrained optimization in delocalized IC introduced by Baker et al, [15] Δq is projected onto the non-redundant DOF and then normalized to form a vector space with one extra DOF. The new vector set undergoes Schmidt orthonormalization to form a new coordinate set spanning the constraint vector U_C and the remaining non-redundant DOF.

This procedure allows a reaction path to be represented in any combination of internal and Cartesian coordinates without any problems caused by an over-specified (redundant) set of coordinates.

SE-GSM requires a modification in the tangent definition for the growth phase

$$U_C = \alpha_c \sum_{k=1}^{3N-6} \langle \delta q | U_k^{(i)} \rangle U_k^{(i)} \tag{2.2}$$

where δq is a primitive coordinate vector describing desired changes in connectivity (bond lengths, angles, and torsions). During the growth phase, new nodes are added, one at a time, along the vector U_C and only this frontier node is optimized using U_C as a constraint.

Combining GSM with IC therefore allows an opportunity of using driving coordinates to find TSs starting from a single initial state. The resulting method, SE-GSM, can explore the chemical reaction space without having prior knowledge about the final state. In practice, δq includes reaction coordinates (combination of bonds, angles, and torsions) representing any desired reaction. This includes coordinates not present in the primitive internals of the starting structure, as any reaction coordinates can be trivially added to the coordinate system when needed.

2.3.1.3 Optimization

When the string is fully grown, all the nodes on the string undergo optimization cycles under the constraint U_C of Equation (2.1), which depends on the node's location along the string. During optimization, an approximate Hessian matrix is used to accelerate convergence. This Hessian is formed when a node is created from a diagonal primitive coordinate Hessian, [233] and updated using the BFGS [36, 68, 77, 189] scheme as optimization proceeds. Diagonalization of the Hessian at each node in the non-redundant coordinates provides a set of eigenvectors and eigenvalues which are used in the eigenvector optimizer:

$$\Delta v_i = \frac{-g_i}{H_{ii} + \lambda} \tag{2.3}$$

v_i are the eigenvectors of the Hessian in coordinates U_k, H_{ii} are the corresponding eigenvalues, g_i is the gradient in the eigenvector basis, and λ is a scaling factor.

After the reaction path is converged to a specified threshold, a CI search begins. [94] At the TS node, perpendicular directions are optimized as described by Equation (2.3) while

the U_C direction of the highest energy node is maximized according to:

$$\Delta U_C = \frac{g_c}{\beta} \tag{2.4}$$

U_C is the constraint climbing direction, g_c is the gradient along the U_c, and β is a scaling constant.

The CI search (Equation 2.4) moves the highest energy node towards the vicinity of the saddle point, which is vital to providing an accurate TS guess prior to the exact TS search. At this point, the reaction tangent (U_C) also provides a good approximation to the TS eigenmode.

2.3.1.4 Exact TS search

After the CI search has begun and the RP converges to a predefined gradient threshold, the eigenvector following TS search commences. The eigenvector of the Hessian with highest overlap with the reaction tangent ($\max_i \langle U_C | v_i \rangle$) at the TS node is followed to find the exact TS

$$\Delta v_{RP} = \frac{g_{RP}}{H_{RP} + \lambda} \tag{2.5}$$

where subscript *RP* refers to the vector with maximum overlap. This strategy ensures the correct mode is followed, [233] but requires that the RP be available during the TS search. Therefore GSM with exact TS search has a particular advantage over typical saddle point finding methods which do not simultaneously optimize the RP.

Prior to beginning the TS optimization, the Hessian has no negative eigenvalues because the BFGS scheme enforces a positive definite Hessian. To initiate the exact TS search, the curvature along RP is approximated using the reaction path tangent defined by the nodes neighboring the TS. Projecting this curvature into the Hessian results in a single negative eigenvalue and its corresponding eigenvector, while avoiding the (expensive) computation of the exact Hessian. Details on building this Hessian can be found in the Appendix.

2.3.2 Coordinate System for Surfaces

As many studies have shown, [13–16,124,165,174,184,186,187,235] the motion of molecular systems is best described by IC, which are composed of primitive coordinates such as bonds, angles, and torsions. A specifically useful type of ICs are delocalized ICs [15] which are constructed from a set of primitive internals [174] and fully span the nonredundant coordinate space. These coordinates can be used whenever a set of primitives is available,

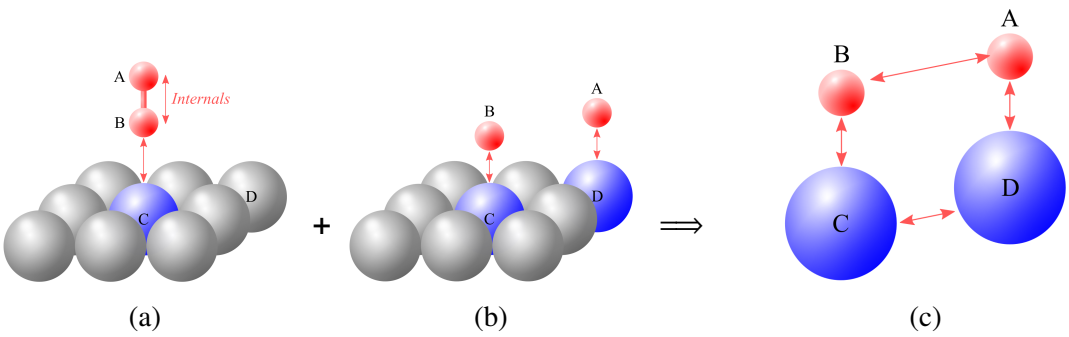


Figure 2.1: Illustration of the hybrid coordinate system for bonds. (a) ,(b), and (c) show bonds in reactant, product, and the union, respectively. Red and blue atoms indicate adsorbate and active surface species, respectively (red: IC only, blue: IC and Cartesians, grey: Cartesians only). Double arrows denote a bond between two atoms.

and provide the significant benefit for optimization of systems including atoms with low coordination numbers.

Use of any type of IC becomes significantly more cumbersome, however, when treating periodic systems which include a large number of atoms with high coordination numbers. For instance with metallic systems, optimizing using ICs is impractical due to the huge number of primitive coordinates that can be present. A mixed coordination system involving Cartesians on high coordination number atoms, and ICs elsewhere, is straightforward. [89, 154, 184]

Such a mixed coordinate system is justified because in a surface reaction only certain atoms require ICs, while others are largely immobile. For instance in a typical reaction, adsorbate atoms move significantly while surface atoms remain relatively immobile and act as binding sites for adsorbate species. Furthermore, only the surface's topmost layer is actively involved in a reaction and bottom layers are stationary supports. As a result, there is no obvious need to include all atoms in the IC set, as Cartesians will easily be able to describe relatively immobile atoms.

Fortunately, a hybrid coordinate system is fully compatible with the delocalized internal coordinates procedure. For each reaction, we define the active surface atoms involved, and assign ICs to these atoms along with all molecular species in the system. All atoms embedded in the surface are assigned Cartesian coordinates, as shown in Figure 2.1. Once the (redundant) set of Cartesians and ICs are available, the delocalized hybrid coordinates are formed with the usual procedure (see Appendix).

2.4 Computational Details

All energy and gradient calculations are performed in a plane wave basis set under periodic boundary conditions as implemented in the Vienna Ab Initio Simulation Package (VASP). [115–118] The PBE functional and projector-augmented wave methods are used to describe the exchange-correlation energy and electron-ion interactions, respectively. An energy cutoff of 300 eV and a smearing parameter of 0.2 eV were employed for the plane waves. The Brillouin zone is sampled using a $1 \times 1 \times 1$ Monkhorst-Pack mesh for all reactions except reactions 5, 7, and 9 where a $2 \times 2 \times 1$ k-point grid is used for the integration and the energy cutoff is set to 400 eV. GSM is implemented in C++ and invokes the Atomistic Simulation Environment (ASE) [10] to provide the quantum mechanical gradients through VASP.

For both single-ended and double-ended GSM the equal spacing of the nodes on each side of the TS node is maintained by a reparameterization step that is performed after each optimization cycle. Reparameterization does not shift the highest energy node after string is fully grown and CI starts.

All CI-NEB calculations used the BFGS optimizer as implemented in ASE and a spring constant of 0.1 $eV/$Å. CI-NEB were considered converged when the RMS gradient on the TS node was below 0.0136 $eV/$Å $\simeq$ 0.0005 Hartree/Å and the total gradient over all the active images was below 2.7 $eV/$Å $\simeq$ 0.1 Hartree/Å to match the GSM's convergence criteria. The calculations that required more than 200 gradient computations per active node (>1,800 gradient calculations for double-ended methods) were terminated and considered unsuccessful.

The chemical structures for this study are created using ASE and the CI-NEB method is employed as implemented in ASE. Reactant and product structures were optimized using the BFGSLineSearch optimizer and were converged when the maximum force on each atom was below 0.05 $eV/$Å.

In the examples that follow, 11 nodes including the two fixed endpoints (9 active nodes) were used to represent the reaction path for double-ended calculations (DE-GSM & CI-NEB) except reaction 1, which has 7 nodes. The input reactant and product structures for the double-ended methods are identical for both methods. The number of nodes for SE-GSM is determined by the method automatically, and typically ranges from 7 to 11 nodes in the reported tests. The three methods under study are compared based on the number of gradient computations required for the convergence of the reaction paths and calculated activation barriers. More details can be found in section A.4 of SI.

2.5 Surface Reaction Validation Test Set

To confirm the efficiency and reliability of the proposed method, an extensive test set of elementary reactions was created. A variety of reaction types, including molecular and dissociative adsorptions, desorptions, and bimolecular and unimolecular reactions, are covered in this set. Most of these reactions have been investigated previously using the NEB method by other researchers. [41, 47, 75, 93, 99, 131, 139, 145, 150, 167, 171, 196, 215–217, 228] In summary, 43 elementary reactions which consists of 9 different metals, one metal oxide, and 7 different surface terminations were studied. Summaries of the reactions are given in Table 2.1.

2.6 Results and Discussion

2.6.1 Overall Performance of Reaction Path Optimization Methods

To evaluate the performance and stability of the three reaction finding methods, their computational cost and success rate will be compared first. Robust methods should converge relatively fast on a wide variety of reactions and successfully calculate a RP and TS in a small number of gradient calculations. The success rate and average number of gradient calls for convergence of DE-GSM, SE-GSM, and CI-NEB are shown in Figure 2.2. DE-GSM was successful in all cases (43 out of 43) while SE-GSM succeeded for 41. CI-NEB converged in 33 out of 43 test cases within 1,800 total gradient calculations, and the reasons for the failures will be discussed in the subsequent section. The average number of gradient calls were 614, 338, and 366 for CI-NEB, DE-GSM, and SE-GSM, respectively, demonstrating that GSM is on average at least 1.8 times faster than CI-NEB.

Taking a closer look at the convergence behavior of the methods provides some insight into the faster convergence of GSM. An example is shown in Figure 2.3 where an addition reaction on Pd(111) (Reaction 8-b) takes place between a hydrogen atom and an ethyl fragment to form ethane. The initially interpolated RP from CI-NEB has a higher RMS gradient compared to DE-GSM's and therefore requires a larger number of optimization iterations to converge. GSM, on the other hand, does not generate all of the nodes at once, which avoids distorted chemical structures with high gradients. This property of maintaining low gradients and small numbers of optimization steps is well-known for GSM, [168] and is fully taken advantage of in DE- and SE-GSM for surfaces.

Additionally, the initial linear path by CI-NEB does not capture the correct asynchronicity in hydrogen and carbon movements. On average, the carbon atom moves 0.03 Å higher

on the surface in CI-NEB's initial path compared to DE-GSM. At the same time, the hydrogen is 0.04 Å closer to the surface in CI-NEB. The root mean square deviations (RMSD) in RPs indicate that DE-GSM's initial reaction path is similar to its final path (RMSD = 0.097), while CI-NEB's deviates more significantly (RMSD = 0.209). In CI-NEB, this difference is seen in the unnecessary half-circular motion of hydrogen adatom on surface before its addition to the ethyl group (Figures A.16, A.15 and Table A.1). Overall, the high quality of GSM's initial RP results in improved performance compared to CI-NEB.

ID	Reaction	ID	Reaction
1	$\mathrm{Au(fcc)} \xrightarrow{\mathrm{Pt(111)}} \mathrm{Au(hcp)}$	11-a	$\mathrm{OH^* + H^* \xrightarrow{Cu(100)} H_2O^*}$
2	$\mathrm{CO(fcc)} \xrightarrow{\mathrm{Pd(111)}} \mathrm{CO(hcp)}$	11-b	$\mathrm{CO^* + O^* \xrightarrow{Cu(100)} CO_2^*}$
3-a	$\mathrm{CO + O \xrightarrow{Pd(111)} CO_2^*}$	12-a	$\mathrm{OH^* + H^* \xrightarrow{Cu(111)} H_2O^*}$
3-b	$\mathrm{CO_2^* \xrightarrow{Pd(111)} CO_2(g)}$	12-b	$\mathrm{CO^* + O^* \xrightarrow{Cu(111)} CO_2^*}$
4-a	$\mathrm{CO + O \xrightarrow{Ru(0001)} CO_2^*}$	13-a	$\mathrm{OH^* + H^* \xrightarrow{Cu(110)} H_2O^*}$
4-b	$\mathrm{CO_2^* \xrightarrow{Ru(0001)} CO_2(g)}$	13-b	$\mathrm{CO^* + O^* \xrightarrow{Cu(110)} CO_2^*}$
5-a	$\mathrm{H(fcc)} \xrightarrow{\mathrm{Ni(111)}} \mathrm{H(fcc)}$	14-a	$\mathrm{H_2S^* \xrightarrow{W(111)} HS^* + H^*}$
5-b	$\mathrm{H(fcc)} \xrightarrow{\mathrm{Ni(111)}} \mathrm{H(hcp)}$	14-b	$\mathrm{HS^* \xrightarrow{W(111)} S^* + H^*}$
5-c	$\mathrm{H(hcp)} \xrightarrow{\mathrm{Ni(111)}} \mathrm{H(fcc)}$	14-c	$\mathrm{H^* + H^* \xrightarrow{W(111)} H_2(g)}$
6-a	$\mathrm{Cu(bridge)} \xrightarrow{\mathrm{Cu(110)}} \mathrm{Cu(hollow)}$	15-a	$\mathrm{H_2O^* \xrightarrow{W(111)} HO^* + H^*}$
6-b	$\mathrm{Cu(hollow)} \xrightarrow{\mathrm{Cu(110)}} \mathrm{Cu(hollow)}$	15-b	$\mathrm{HO^* \xrightarrow{W(111)} O^* + H^*}$
6-c	$\mathrm{Cu} \xrightarrow{\mathrm{Cu(110)}} \mathrm{Cu(atom\ swap)}$	16-a	$\mathrm{CH_2{=}CHCH_2OH^* + O^* \xrightarrow{Au(111)} CH_2{=}CHCH_2O^* + OH^*}$
7-a	$\mathrm{CH_3CH_2COOH^* \xrightarrow{Pd(111)} CH_3CH_2CO^* + OH^*}$	16-b	$\mathrm{CH_2{=}CHCH_2O^* + O^* \xrightarrow{Au(111)} CH_2{=}CHCH{=}O^* + OH^*}$
7-b	$\mathrm{CH_3CH_2CO^* \xrightarrow{Pd(111)} CH_3CH_2^* + CO^*}$	17-a	$\mathrm{CH_3OH^* \xrightarrow{Cu(110)} CH_3O^* + H^*}$
8-a	$\mathrm{CH_2CH_2^* + H^* \xrightarrow{Pd(111)} CH_3CH_2^*}$	17-b	$\mathrm{CH_3O^* \xrightarrow{Cu(110)} H_3C{=}O^* + H^*}$
8-b	$\mathrm{CH_3CH_2^* + H^* \xrightarrow{Pd(111)} CH_3CH_3^*}$	18-a	$\mathrm{CN^* + H^* \xrightarrow{Pt(111)} CNH^*}$
9-a	$\mathrm{NH_3^* \xrightarrow{RuO_2(110)} NH_2^* + H^*}$	18-b	$\mathrm{CNH^* + H^* \xrightarrow{Pt(111)} CNH_2^*}$
9-b	$\mathrm{NH_2^* \xrightarrow{RuO_2(110)} NH^* + H^*}$	19-a	$\mathrm{NH_3(g) \xrightarrow{Si(111)-Cl} NH_3^*}$
9-c	$\mathrm{2\,N^* \xrightarrow{RuO_2(110)} N_2^*}$	19-b	$\mathrm{NH_3^* \xrightarrow{Si(111)-Cl} NH_2^* + HCl(g)}$
9-d	$\mathrm{N^* + O^* \xrightarrow{RuO_2(110)} NO^*}$	20	$\mathrm{H_2O(g) \xrightarrow{Si(111)-H} OH^* + H_2(g)}$
10-a	$\mathrm{CO^* + H^* \xrightarrow{Ni(111)} COH^*}$	21	$\mathrm{CH_4(g) \xrightarrow{Ir(111)} CH_3^* + H^*}$
10-b	$\mathrm{COH^* + H^* \xrightarrow{Ni(111)} C^* + H_2O(g)}$		

Table 2.1: Elementary step test cases for GSM. Asterisks designate the surface species.

2.6.2 Comparison of Reaction Paths from DE-GSM and CI-NEB

To further demonstrate the reliability and robustness of DE-GSM, it will now be compared in more detail to the CI-NEB method. As shown in Figure A.3, the activation energies predicted by these two methods are similar, with a linear regression of slope 0.995 and R^2 of 0.989. This correlation shows that DE-GSM with its exact TS search produces similar barriers compared to those from CI-NEB. Some deviation is expected, however, because GSM performs an exact saddle point search, while CI-NEB provides an approximate TS. The maximum difference in calculated activation energies by two double-ended methods occurs for water dissociation on W(111) (Reaction 15-a), which differs by 4.6 kcal/mol between DE-GSM and CI-NEB. In section A.5 of SI we show that this difference occurs because the two methods find distinct reaction pathways, which should not have the same barrier.

There are two cases where CI-NEB did not compute a realistic RP, specifically two copper surface rearrangements, [110] reactions 6-b and 6-c. Such reaction steps are known to be important for copper-promoted graphene growth [140] and silicon device production, [218] and otherwise represent standard reactions that should be resolvable by double-ended methods. In Reaction 6-c, two Cu atoms exchange positions on Cu(110) surface, but the linear Cartesian interpolation of the CI-NEB's initial RP causes the moving atoms to sit directly on top of one another (section A.5 of SI). From this geometry, convergence of the DFT density and energy fails, and optimization cannot proceed. DE-GSM, by incrementally adding and optimizing nodes, never reaches such problematic structures and optimizes smoothly to the desired RP. Reaction 6-b similarly has a problem with the initial interpolation in CI-NEB, which is discussed in section A.5 of SI.

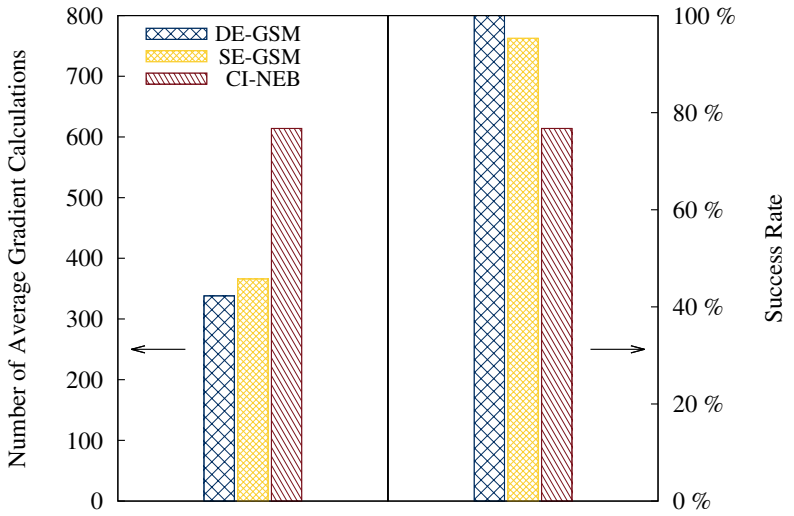


Figure 2.2: Average number of gradient calculations and success rate for each method. (Calculations with more than 1,800 gradient calculations are not included in the average gradient calculation).

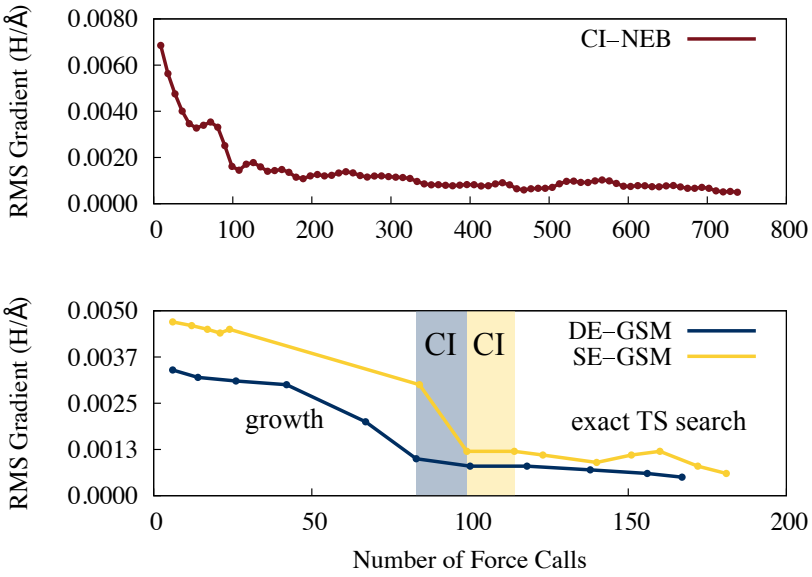


Figure 2.3: Convergence behavior of the methods plotted for reaction 8-b. CI-NEB has a higher initial RMS gradient compared to GSM in addition to larger RMSD of initial and final RPs, and therefore more force calls are required to reach convergence. The gradient calls required for each phase of GSM calculations are labeled in the bottom plot.

2.6.3 Comparison of DE-GSM and SE-GSM

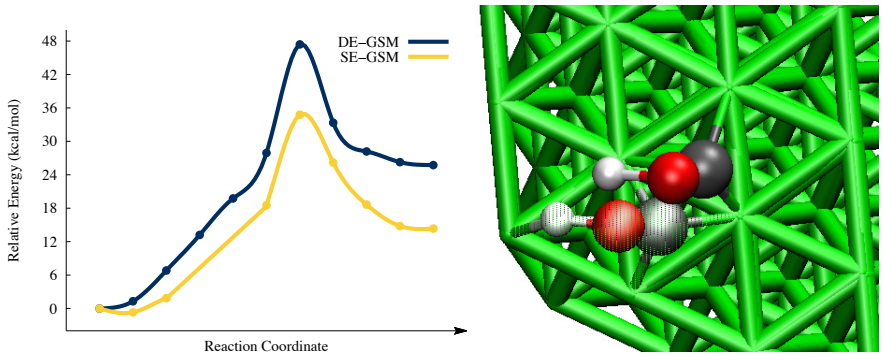
		DE-GSM		SE-GSM	
Bonds (Å) & angle	Reactant	Transition State	Product	Transition State	Product
C5-Ni1	2.037	1.965	1.892	1.940	1.884
C5-Ni2	2.082	1.939	1.949	3.287	3.187
C5-Ni3	1.856	1.800	1.803	1.870	1.849
C5-Ni4	3.114	2.974	2.991	1.880	1.863
C5-O6	1.189	1.273	1.339	1.280	1.348
O6-H7	3.557	1.349	0.981	1.311	0.983
∠ C5-O6-H7	60.3	88.2	111.0	96.7	107.3

Table 2.2: Bond lengths and angles for the reactant, TSs, and products of Reaction 10-a calculated by DE-GSM and SE-GSM. Colored values indicate bonds and angles that are different in structures calculated by the methods. Both methods result in the same product while the product's position on the surface is different.

In the case of SE- and DE-GSM, the reaction tangent definitions are different and can lead to unique RPs being found for the same qualitative reaction. This can occur in reactive systems with more than a few degrees of freedom, where there are often multiple pathways from a given initial state to a single product structure. [187] Usually, TS finding methods locate only one such path at a time, and thus offer no guarantee that all connecting TSs will be found. Cases where two different RPs were found by SE- and DE-GSM are discussed in this section.

The activation energies computed by SE-GSM and DE-GSM (Figure A.8) are usually similar, but less closely related than DE-GSM compared to CI-NEB. The comparison of SE- to DE-GSM yields a slope of 0.899 and R^2 value of 0.875. Because the optimization process is identical for the two methods after the string endpoints are connected, this slight dissimilarity is due to differences in the initial reaction tangent and RP. Specifically, since DE-GSM uses curvilinear interpolation in ICs between the two frontier nodes to estimate a RP, it does not generally have the same tangent as SE-GSM, where the tangent consists of a few specific ICs used as driving coordinates.

This difference in tangent definition and its influence on the outcome of a calculation is most pronounced in Reactions 10-a, 14-c, 15-a, and 16-b. Reaction 10-a is a representative example that describes addition of a hydrogen atom to oxygen of CO on Ni(111) surface to release H_2O and deposit C on the surface. SE-GSM's initial RP is formed under a more



(a) (b)

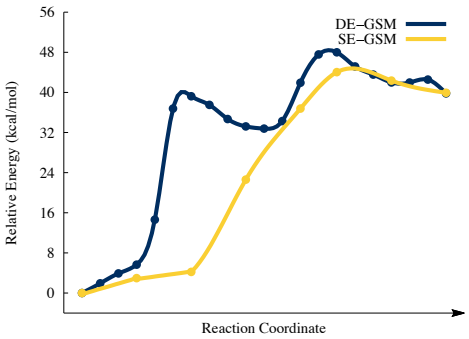
Figure 2.4: (a) Reaction path calculated by DE-GSM (blue) and SE-GSM (yellow) for COH formation on Ni(111). (b) TS structures calculated by DE-GSM (opaque) and SE-GSM (translucent). CO molecule is not stationary in the case of SE-GSM.

free reaction tangent compared to DE-GSM, because its reaction tangent consists of only one driving coordinate (addition of hydrogen and oxygen).

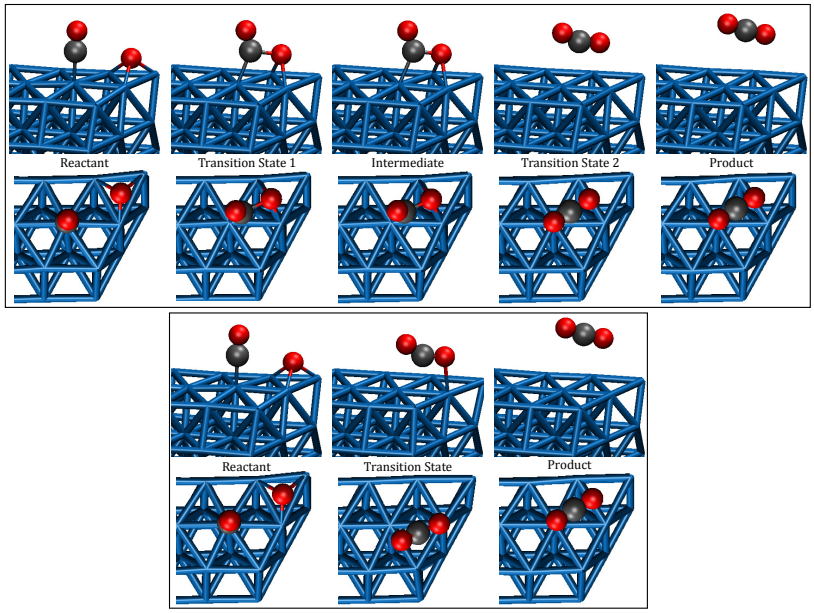
This freedom of movement in SE-GSM ultimately results in variations in energies and chemical structures of the TSs. In this example, SE-GSM results in a lower activation barrier and a more stable product (Figure 2.4). This occurs because the CO molecule is stationary in DE-GSM, while it moves from its starting binding site to a neighboring fcc site in the SE-GSM case (Figure A.9). Chemical structures of this example are quantitatively compared in Table 2.2. A similar situation happens in other cases (reactions 14-c, 15-a, 16-b) where the products form on different binding sites or in different relative positions on the surface.

In addition to different single elementary step transformations, we observed that reactions 4 and 5-b proceed in different number of elementary steps through the two GSM growth strategies. For example in reaction 4, CO and O combine on a Ru(0001) surface to release carbon dioxide. For this case the DE-GSM's RP consists of two elementary steps, in which a CO−O complex is formed on the surface followed by its desorption. On the other hand, the SE-GSM's RP proceeds through a single elementary step that combines CO_2 formation and desorption (Figure 2.5) through the asymmetric dissociation of Ru−O bonds (Figure A.10 and TS structures of Figure 2.5b).

When a system has many degrees of freedom, multiple pathways connecting the same two qualitative chemical structures can be present. SE- and DE-GSM provide two varying growth methods due to their tangent definitions, which enables exploring alternative paths for a given reaction. This will be especially the case if multiple SE-GSM trials are attempted, which is a subject of future research and will be reported subsequently.



(a)



(b)

Figure 2.5: (a) Reaction path calculated by DE-GSM (blue) and SE-GSM (yellow) for CO_2 formation on Ru(0001). (b) Reactant, TS, and product structures for reaction (4) calculated by DE-GSM (top) and SE-GSM (bottom). Reaction proceeds in one and two elementary steps via SE-GSM and DE-GSM, respectively.

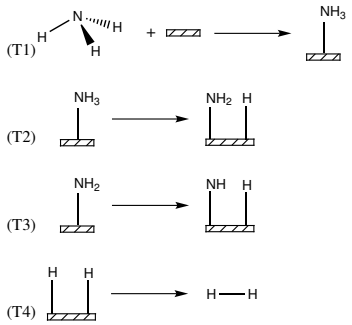


Figure 2.6: Proposed network of reactions for dehydrogenation of NH_3 on Cu(111).

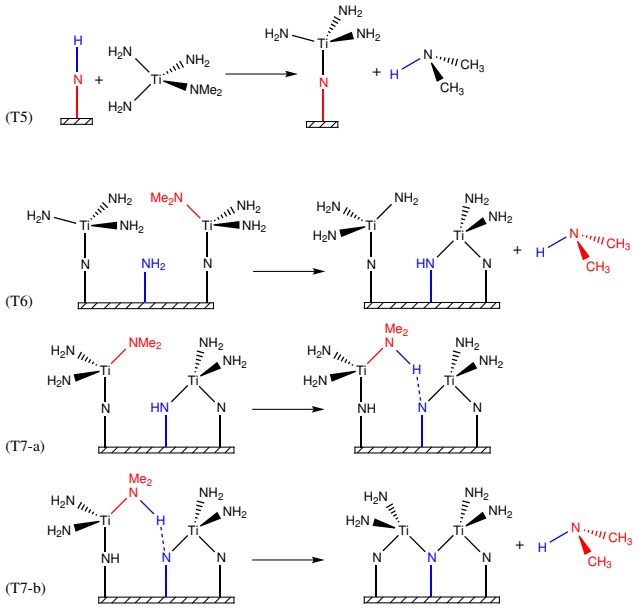


Figure 2.7: Proposed reactions during first TDMAT cycle.

2.6.4 Atomic Layer Deposition of TiN on Cu(111)

Titanium Nitride (TiN) has many desirable properties that make it a good candidate as a wear-resistant coating or copper diffusion barrier in microelectronics. [152] To build TiN layers of controlled thickness, ALD is an especially useful technique. In ALD, alternating cycles of two self-limiting and complementary reactions utilize gaseous precursors to form ultrathin, conformal, and uniform films with monolayer control over the thickness. [225] Experimental studies [62,127] have shown that tetrakis(dimethylamido)titanium (TDMAT) and ammonia (NH_3) are good precursors for ALD of TiN. By first reacting NH_3 onto the surface in the form of NH_x units and their derivatives (N, N_2), the subsequent ALD cycle using TDMAT precursors attaches Ti-containing species to these surface sites. Upon repeating these cycles, TiN layers can be formed in a controlled fashion.

Little mechanistic information is available for this ALD reaction, hindering our ability to extend the scope and availability of new precursors and conditions for reaction. While some mechanistic information for related processes are available, [52,63] these fail to capture any specific details of the TiN ALD mechanism. Given this lack of information, SE-GSM is ideally suited for investigating this process because it starts from a single initial state and locates the TS, RP, and the product in one computation. This capability enables systematic exploration of the reactive space, without requiring a guess transition structure close to the saddle point or even a complete set of reactive intermediates. Studying ALD of TiN will therefore serve to demonstrate the capabilities and advantages of SE-GSM for

reactions that are not already well-known.

In particular, the initiation steps to form three atomic layers of TiN on Cu(111) will be studied. The reactive process proposed here proceeds through three general steps: Step 1.) Addition of ammonia to nucleate surface sites and release H_2 gas. Step 2.) Deposition of titanium via a ligand-exchange with TDMAT extruding dimethylamine gas. Step 3.) Addition of ammonia to the titanium-terminated surface. Repeating steps 2 and 3 provides access to additional layers of TiN. In order to reduce the computation complexity, $-N(CH_3)_2$ ligands are truncated to $-NH_2$ except when the ligand is involved directly in a reaction. A summary of activation energies, proposed reactions, and chemical structures are shown in Table 2.3 and Figures 2.6 to 2.9 and A.11 to A.14. The asterisk (*) on chemical moieties means they are adsorbed on surface.

2.6.4.1 Nitrogen nucleation during first NH_3 cycle

During the first deposition cycle, molecular NH_3 is chemisorbed on the surface in a barrierless transformation that is exothermic by 10.6 kcal/mol (T1, Figure 2.6). Three different orientations of adsorbed NH_3 are possible on surface which yield slightly different activation energies for the subsequent reactions. Reaction T1 is followed by progressive dehydrogenation of NH_3 to form NH_x* species (x=2, 1, 0) and subsequent release of H_2 gas from hydrogen adatoms present on the surface (Reactions T2-T4 of Figure 2.6) as suggested by experimental studies. [127] During the first dehydrogenation step (T2, Figure 2.6 and Figure A.12a), NH_3* moves from atop position to form NH_2* in a higher-coordinated bridge site and H* in fcc through a barrier of 31.4 kcal/mol and is endothermic by 16.8 kcal/mol. A second hydrogen dissociation and migration from NH_2* proceeds through a very similar process with a barrier of 30.5 kcal/mol and is also endothermic by 27.6 kcal/mol (T3, Figure 2.6 and Figure A.12b).

Although the barriers for fragmentation of surface bound ammonia are too high to proceed at room temperature, the experimental conditions can exceed 150 °C, making these barriers surmountable. The barrier for reductive-coupling of two surface-bound hydrogen adatoms is 22.6 kcal/mol, which allows release of hydrogen gas and provides an entropic driving force for these reactions (T4, Figure 2.6 and Figure A.12c).

2.6.4.2 Ti layer formation

The second ALD cycle introduces TDMAT to the reaction. A previous report [62] on this process proposes that NH* moieties on surface likely serve as nucleation sites for TDMAT deposition, rather than open Cu surface sites. This step is therefore driven by the

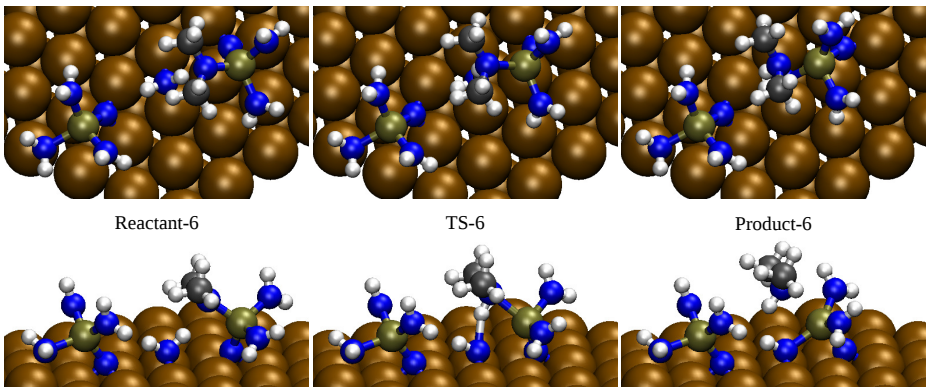


Figure 2.8: Reactants, TS, and products of Reaction (T6). After adsorption of two TDMAT molecules on surface, they connect by a bridging N* that comes from an NH_2 species adsorbed on surface. In this reaction, one of the adsorbed tris(dimethylamido)titanium species reacts with NH_2*. Atoms are N (blue), H (white), C (gray), Ti (tan green), and Cu (ochre).

electron-rich dimethylamido ligands on TDMAT which serve as strong H-bond acceptors for NH* groups. This characteristic also means that its dimethylamido ligands make TDMAT preferable to other titanium precursors such as $Ti(NH_2)_4$. As a result, deposition of titanium on the surface (T5, Figure 2.7) is initiated by the gradual formation of a strong H-bond between one of the dimethylamido ligands and an NH*. As the reaction approaches the transition state, the hydrogen from NH* is formally transferred to dimethylamine, resulting in elongation of the titanium-amino bond by 0.20 Å. Additionally, the distance between the titanium center and the N*, 2.95 Å , is too long to be a covalent bond, highlighting the importance of a strong H-bond interaction to stabilize these types of species (Figure A.11). The resulting $Ti(N(CH_3)_2)_3(NH(CH_3)_2)$ intermediate is stabilized by interaction between one of the dimethylamido ligands on TDMAT and a surface amine, in this case N*. The final step of tethering titanium onto the surface, via ligation of N*, proceeds through a concerted ligand-exchange with a barrier of 7.9 kcal/mol, displacing one of the dimethylamido ligands via a dissociative transformation (Figure A.13a). This mechanism is in agreement with the experimental observation of the build-up of dimethylamine gas during this process. [62, 127]

Once the titanium is surface-bound through N*, it is plausible that further hydrogen-transfer/ligand-exchange reactions could lead to the formation of complex Ti−N bonding-networks. Such networks are suggested by the crystal structure of TiN, where titanium is coordinated to six nitrogens. Specifically, a dimethylamido ligand on titanium can undergo a hydrogen-transfer reaction with its neighboring unreacted NH_2* fragments (Reaction T6, Figure 2.7) through concerted hydrogen-transfer/ligand-exchange. This transformation,

which is similar to the initial TDMAT attachment step, adds another tethering site for titanium to bind to the surface. This second N-ligand-exchange with titanium has a barrier of 19.3 kcal/mol and is endothermic by 4.4 kcal/mol with a thermodynamic driving-force via release of the gaseous dimethylamine. Similar to Reaction T5, formation of N−H H-bonds in the reactant and TS structures facilitates proton transfer from NH_2* to $N(CH_3)_2$ ligand in a dissociative concerted mechanism. After completion of proton transfer, a new bond between Ti and N* starts to form while breaking the Ti dimethylamine bond, resulting in a five coordinated titanium center before desorption of dimethylamine gas.

Although ligand-exchange pathways described so far have all been concerted, the formation of N*-bridged complex connecting two adjacent titanium species proceeds through a two-step process. The observation of a step-wise hydrogen transfer followed by ligand-exchange is likely a result of increasing steric demand of the incoming NH*. Nevertheless, the first step in this transformation is a hydrogen-transfer from NH* to the dimethylamido of titanium (Reaction T7-a, TS = 24.0 kcal/mol) resulting in a dimethylamine ligand on titanium (Reaction T7a, Figure A.13b). The second step which is formation of Ti−N−Ti chain proceeds through a facile (Reaction T7-b, TS = 9.7 kcal/mol) associative ligand-exchange releasing the dimethylamine gas (Reaction T7b, Figure A.13c).

ID	E_a (kcal/mol)	$\Delta E_{reaction}$ (kcal/mol)
T1	-	-10.6
T2	31.4	16.8
T3	30.5	27.6
T4	22.6	5.4
T5	7.9	-13.5
T6	19.3	4.4
T7-a	24.0	22.8
T7-b	9.7	8.9
T8	29.9	10.0
T9	29.1	20.2

Table 2.3: Activation energies and heat of reactions for the elementary steps of ALD of TiN on Cu(111).

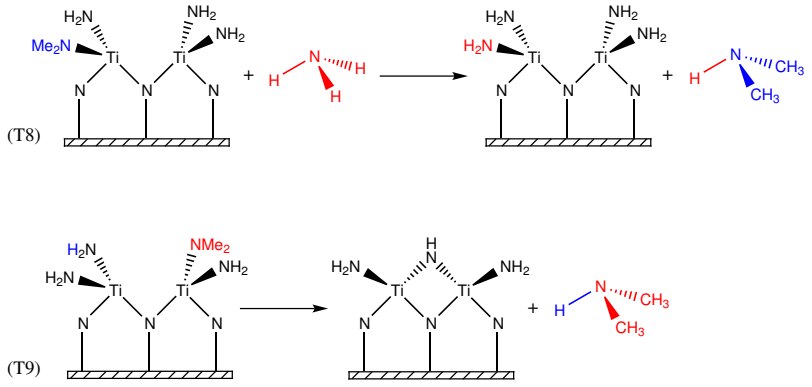


Figure 2.9: Second NH_3 cycle.

2.6.4.3 Second NH_3 cycle

The alternating cycles of the ALD process require a third step of NH_3 exposure. Through a transamination reaction, SE-GSM shows that the topmost fragments of the deposited layers and the incoming NH_3 react through a step-wise process with an activation energy of 29.9 kcal/mol (Reaction T8, Figure 2.9). Initially, hydrogen-transfer from NH_3* to one of the dimethylamido ligands of Ti results in a dimethylamine ligand and formation of NH_2*. This step is followed by addition of NH_2* to one of the Ti centers and cleavage of the Ti−$NH(CH_3)_2$ bond to replace a dimethylamido ligand with an amido group (Figure A.14a). Similar to Reaction T5, networks of H-bonds stabilize NH_3 over surface during this reaction. Reaction T9 of Figure 2.9 is the final step in forming the third atomic layer, where a binding site for the incoming TDMAT of the fourth cycle is available. Hydrogen-transfer from the amido group to a dimethylamido ligand of a nearby Ti results in a bridged NH group and desorption of dimetylamine gas, with a barrier of 29.1 kcal/mol (Figure A.14b).

These calculations suggest nucleation of the first Cu-N sites is rate-limiting. After this event, formation of H-bonds between ligands, moieties in the gas phase, and intramolecular H-bonds stabilize the various reactive intermediates and allow the deposition to proceed. Overall, the computed activation barriers are feasible given the high temperature reaction conditions, but desorption of gaseous products is a necessary step for most of these reactions to be favorable thermodynamically.

2.7 Conclusions

Surface reactions cover an important branch of chemistry that contains a wide variety of interesting processes. In this area, GSM is found to be a powerful method for the study of

reactions due to its accuracy, reliability, fast convergence, and relative ease of use. The four components for success, strategies for quickly approaching the vicinity of the saddle point, estimation of the direction of negative curvature, optimizer, and coordinate system, were carefully considered, and together integrated into GSM for surfaces to make a method that is highly proficient at reaction path finding.

GSM's efficacy was confirmed by comparison with CI-NEB on an extensive set of reactions characteristic of modern surface chemistry studies. In these cases, GSM reduces the computational cost (in terms of gradient computations) by about 45% on average over CI-NEB.

In addition to high efficiency, GSM has the advantage of operating in single-ended way to enable explorative study of chemical reactions. The strength of the SE-GSM for the study of novel reactions was demonstrated in this article via the first study of ALD of TiN on Cu(111), which provided a wealth of details about the operating mechanism for deposition. In the future, the use of a combinatorial set of driving coordinates in surface SE-GSM to guide a reaction to many different outcomes will be possible through systematic graphical methods. [234, 236]

CHAPTER 3

Uncovering Reaction Sequences on Surfaces through Graphical Methods

This Chapter largely based upon published work:
Reproduced with permission from M. Jafari and P.M. Zimmerman, Phys. Chem. Chem. Phys., 2018, 20, 7721

3.1 Abstract

ZStruct is a graph-based model that generates an ensemble of plausible reaction pathways starting from a given initial state, without requiring prior knowledge of reaction intermediates. In this article, a surface-reaction oriented implementation (S-ZStruct) is introduced for unimolecular and bimolecular reactions, including sampling of different binding sites and adsorption orientations. To test the unimolecular reaction generation ability and angle sampling feature of S-ZStruct, the propanoic acid to ethylene dissociation network is studied. Starting from multiple initial orientations of two key intermediates, multiple unique reaction pathways, each with different activation energies, were discovered. Atomic layer deposition of TiN on Si(100)–involving a challenging reaction mechanism–is also studied as an example of bimolecular reactions. In addition to locating a number of expected pathways, S-ZStruct found that the little-understood step of reduction of Ti(IV) to Ti(III) likely occurs through β-hydride transfer between diamido ligands on neighboring Ti centers. These results suggest that S-ZStruct is a powerful tool for exploring surface chemistry, which will permit discoveries of reaction mechanisms in a wide variety of environments.

3.2 Introduction

Reaction mechanisms form a basis for designing experiments and optimizing chemical processes, and therefore are a centerpiece of modern reaction development. Quantum chemical simulation can provide this information, which includes the atomistic details of reaction pathways as well as thermodynamic and kinetic parameters, but standard paradigms require extensive chemical intuition and much computational effort. This type of mechanistic evaluation leverages existing knowledge and standard computational tools for reaction path (RP) finding, [19, 84, 91, 94, 101, 210] and is best applied to reasonably well-understood reactive processes. For unknown chemistries or complex chemical networks [122, 156, 163], alternative strategies are required to thoroughly explore the reaction landscape.

The high dimensionality of quantum chemistry's underlying potential energy surfaces (PES) precludes exhaustive searches, making brute-force computations impractical for all but the smallest system sizes. To overcome this challenge, new approaches must examine low-dimensional subspaces of the full landscape. The choice of these subspaces largely determines the efficacy of the approach, as tighter restrictions lead to lower costs, yet higher probability of missing key RPs. This conundrum has been examined in the context of metadynamics [65, 100, 119] and other molecular dynamics [66, 100, 151] (MD) techniques where specific collective variables comprise the core reactive subspace. In related work, the artificial force induced reaction method [135, 136] searches a subspace of interatomic distances that depends on alignment of two substrates. In another example, a surrogate model is trained to predict significant RPs, each of which is later examined using DFT to compute activation barriers. [211] In all of these approaches, attempts are made to refine reactive coordinates to the minimum needed to locate the most plausible RPs. Many related works also exist where the reaction coordinates are predefined. For instance, methods which use reduced gradients [97, 98, 175, 177] (RGF) and reduced potential energy surfaces [29] (RPES) follow the direction of a few specified driving coordinates, while the activation-relaxation technique [18] (ART) follows a modified force vector in the PES to adjacent saddle points. Despite the utility of these approaches for locating individual reaction pathways, most cannot be used to systematically sample reaction pathways of surface reactions because they refine paths for specified reaction coordinates, rather than locate new ones.

In many reactive systems, the core goal of RP finding is to locate networks of elementary steps where changes in chemical bonding occur. This focus allows invocation of graphs, where changes in connections between atoms become the dominant characteristics of chemical reactions. This strategy has been used with substantial success in molecular

systems, where the graph representation is easily invoked. [58, 69, 164, 200, 234, 237, 238] These tools have been applied to a number of organometallic reactions [58, 122, 132, 155, 156, 163] and even shown to be able to uncover unexpected elementary steps and intermediates. While graph theories cannot assure complete examination of reactive space–just like other exploration methods–their abilities to inform chemical intuition about new types of reaction steps has made them especially valuable tools. The present work introduces a novel graphical approach for reaction mechanism exploration on surfaces (Figure 3.1), which have substantial complications beyond finite cluster/molecular systems.

The complications of surface reaction discovery are delineated as follows. (1) Efficient RP and TS finding for surfaces requires designing a new coordinate system to handle periodicity and high coordination number of surface atoms. [89, 101, 184] A molecule on a surface can interact with multiple surface atoms and binding sites, resulting in reactive modes that are unique to surfaces and thus not trivially handled by gas-phase reaction finding techniques. This aspect will be drawn out below in the examination of propanoic acid decomposition, where multiple binding sites will be shown to be important. (2) Any reaction discovery tool will need to treat unimolecular and bimolecular reactions taking place on surfaces, as in the deposition reaction of the penultimate section of this paper. (3) Another complication is in sampling the initial binding sites to which the reactants become connected and undergo chemical transformation. This aspect requires aligning adsorbate molecules and binding sites in a variety of adsorption orientations to find reasonable starting structures for elementary reactions, a feature missing in many methods developed for cluster/molecular systems.

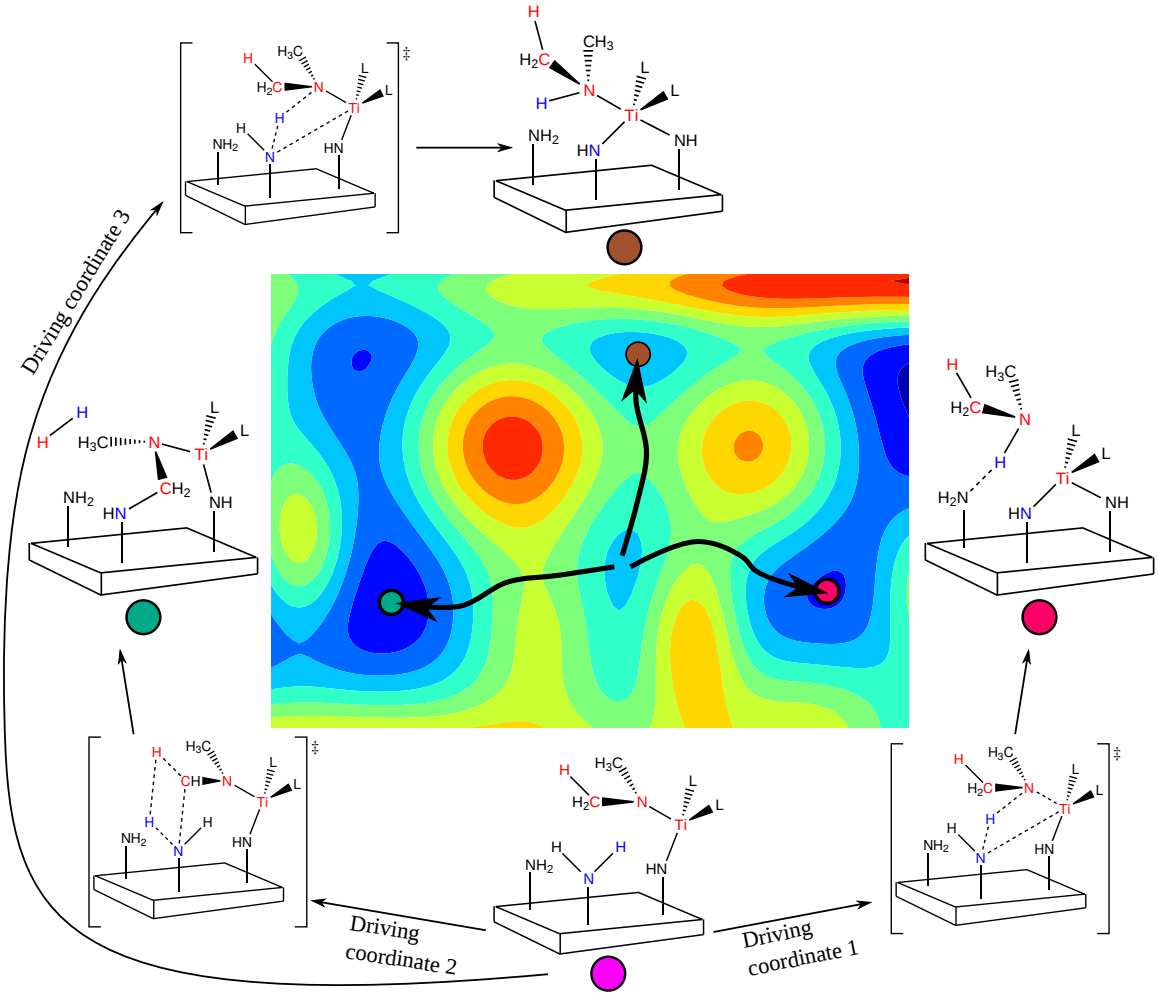


Figure 3.1: Conceptualization of S-ZStruct. An arbitrary PES with different sets of driving coordinates–enabling combinatorial search of the PES–are shown here.

This article presents a reaction mechanism search method capable of examining the fundamentals of surface chemistry by discovering a variety of plausible elementary steps, beyond those available to chemical intuition. The proposed method systematically generates all the combinations of driving coordinates based on reactive sites identified by the user and (computationally tractable) connectivity rules. Starting from a single initial state, this method finds all the single elementary step reactions and products, and the process may be repeated to construct full reaction networks. In the Methods section of this article, this technique will be introduced in detail, and two challenging surface reactions will be discussed in the Results section.

3.3 Methods

3.3.1 Single-ended growing string method

The growing string method (GSM) [101, 169, 233, 235, 238] is a computational protocol for finding individual TSs and RPs. GSM forms a RP by iteratively adding new chemical structures (nodes) until a TS with stable intermediates on each side of the RP (string) is present. GSM has been shown to be more efficient than other competing methods [101] mainly due to its growing nature, i.e. incremental addition of new nodes which results in avoiding high energy regions of the PES.

GSM is implemented for molecular and surface chemistry reactions in gas phase and has three main phases: growth, optimization, and exact TS search. These three phases and their components work synergistically to locate TSs and form RPs efficiently and reliably. Among GSM's building blocks, the reaction tangent is an important component since it determines the growth direction and the constraint for geometry optimization. This tangent allows single-ended growing string method (SE-GSM) to find a RP, TS, and reaction product in one single run. Starting from a local minimum, the Cartesian coordinates of the initial state are converted to internal coordinates and desired changes in connectivity (bond breaks and formations and changes in angles and torsions) are added to the internal coordinate set. Based on this information, the reaction tangent for SE-GSM is defined, and SE-GSM is reliable in finding RPs in the direction of the tangent. Full details on GSM can be found in previous articles. [101, 233, 235, 238] Importantly, a chemical space of interest can be explored using a variety of reaction tangents, allowing a driving coordinate generator to provide the search directions for SE-GSM.

3.3.2 S-ZStruct method

To overcome the complexity of exploring a given chemical system of high dimensionality, a systematic, low-cost search method is needed. As chemical reactions change chemical bonding, one class of reaction discovery method uses changes in connectivity as the means for its exploration [57]. While this could be done over the entire internal coordinate space of the system, exploring the whole chemical space without restrictions would be computationally intractable.

In S-ZStruct, atoms that are allowed to undergo changes in connectivity (e.g. with bonds to them being broken or formed) are identified by the user as reactive centers. To determine their connectivity to other atoms, the distance between the reactive center and each atom is compared with the sum of their atomic radii. If the distance between two atoms is

smaller than the sum of their atomic radii, they are deemed connected. After determining the connectivity, the atomic coordination number is trivially available. Based on these two concepts–chemical bond and coordination number–ZStruct [234, 237] encodes this information into a graph where new connections may be formed between unlinked atoms, and connections broken between linked atoms. To ensure that only reasonable chemical steps are sampled, coordination number restrictions are applied, for instance, such that C does not become 5 coordinate, or H zero coordinate. The internal coordinate system of GSM provides a direct link between a RP optimization method and the driving coordinates of ZStruct. The integration of GSM with S-ZStruct results in relatively low costs for discovering TSs: less than 1,000 gradients are required per TS. This strategy therefore is highly efficient compared to MD techniques, where ~10,000 gradient calls are needed to locate a single reaction pathway.

The original ZStruct method, however, has no ability to handle surface sites, where attachment occurs not necessarily at a single atom, but often at bridge or hollow sites between groups of atoms. Therefore S-ZStruct was implemented and will be described in the following sections. Similar to ZStruct, it uses connectivity and coordination number rules with some modifications that make it suitable for the study of surface chemistry reactions. S-ZStruct is made of five main stages: (1) binding site finding, (2) aligning and sampling reactant structures, (3) geometry optimization, (4) unique structure identification, and (5) isomer generation and SE-GSM invocation. These steps are outlined in Figure 3.2.

(1) Binding site identification. Binding sites on surfaces are the active sites to which adsorbate molecules connect. The available binding sites on each surface depend on the slab termination, and these are identified using standard definitions on flat surfaces. The Atomic Simulation Environment [10] (ASE) is used to create the initial surface structures, and S-ZStruct then identifies standard binding sites (atop, bridge, fcc, hcp, and hollow) on typical flat surfaces such as FCC, BCC, and HCP crystal structures with 100, 110, 111, and 0001 terminations. For unconventional surfaces, the user can input binding site coordinates manually.

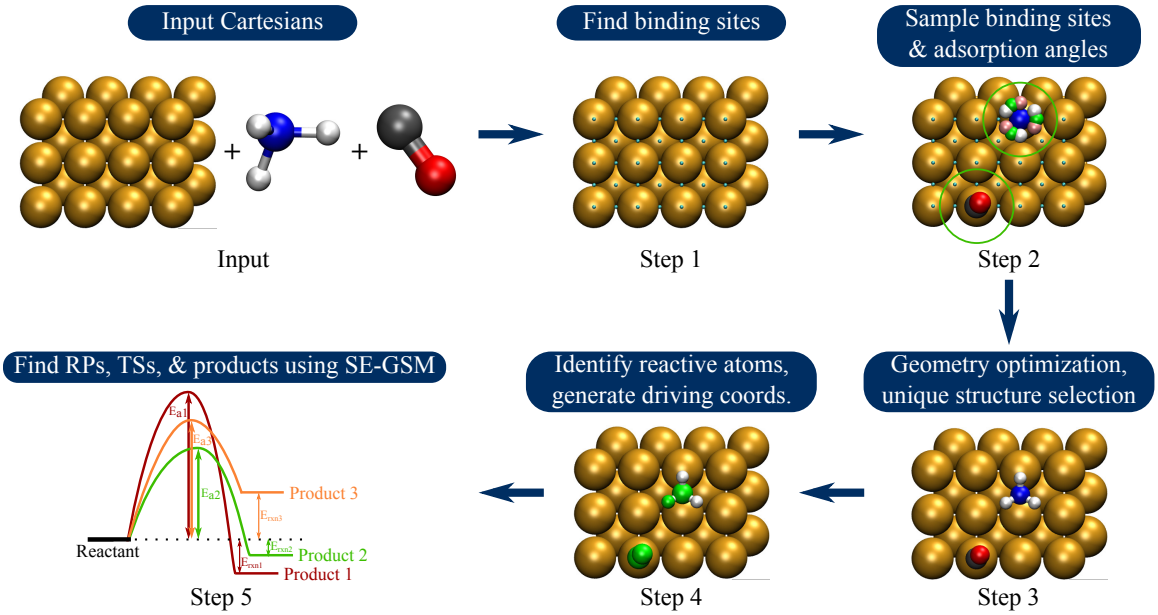


Figure 3.2: Process flow for S-ZStruct. (Input) Cartesian coordinates of the slab and adsorbates are input, separately, (Step 1) Binding sites are identified, (Step 2) Adsorbates are added to selected sites with sampling of adsorption orientations, (Steps 3 & 4) Geometries are optimized and unique structures selected, and (Step 5) Driving coordinates are created for SE-GSM.

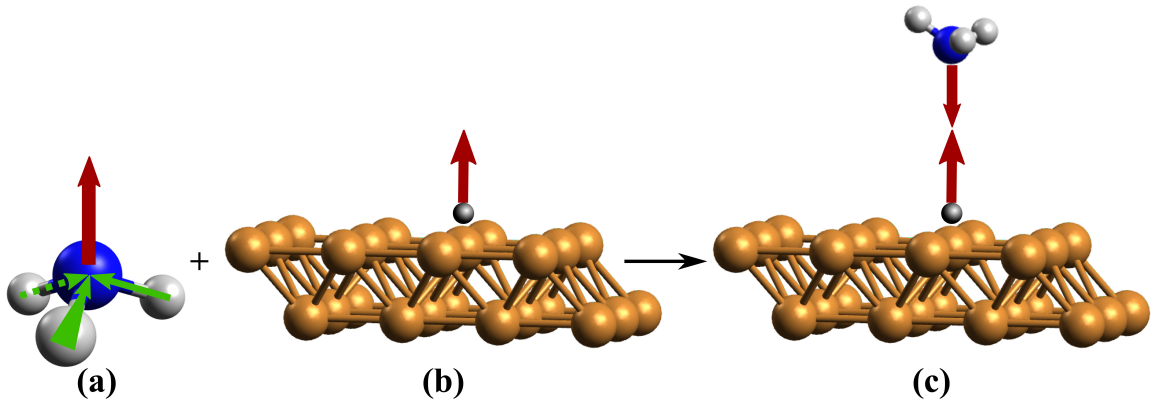


Figure 3.3: (a) Each bond is treated as a vector and the averaged bond vector determines the less crowded side of the molecule. (b) The surface normal originating from the selected binding site (black sphere) is calculated. (c) Surface and adsorbate are aligned using the calculated averaged bond vector and surface normal.

(2) Aligning and sampling chemical structures. The relative position of the moieties involved in a surface reaction is an important factor that determines its feasibility. S-ZStruct therefore aligns chemical structures onto the surface by taking the adsorbate molecule, facing its binding atom downward towards the surface, and moving the molecule into contact

with the surface binding site (see Figure 3.3). Since the rotational orientation of the adsorbate also matters, the adsorbate is rotated along the binding axis at specified angles to produce multiple starting conformations. To ensure low energy starting structures for TS finding step are found, S-ZStruct repeats this sampling within a radius around the chosen binding site (Figure 3.2, step 2).

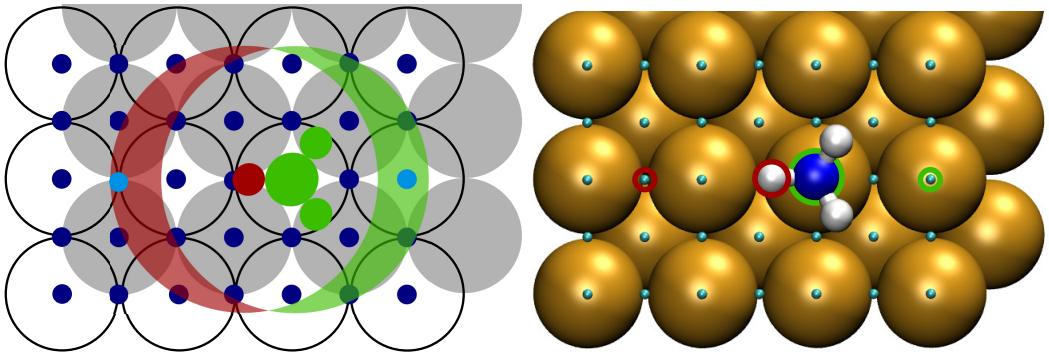
(3) Geometry optimization. Structures generated in stage 3 are optimized using DFT methods and the output is used in stage 4. This is the first computationally costly step.

(4) Unique structure identification. When geometry optimization is complete, the root mean square deviation (RMSD) of atomic positions for each pair of structures is compared to filter out duplicate structures. Structures with low RMSD value are considered identical. Even after removing duplicate structures, it is possible to end up with a large number of initial structures for next stage. Unique optimized structures are thus sorted by energy and lowest energy chemical structures are selected for next stage.

(5) Assigning driving coordinates for SE-GSM. When stages 1 through 4 are complete, the final step for generating driving coordinates and locating TSs is initiated. Since S-ZStruct is capable of exploring unimolecular and bimolecular reactions, the driving coordinate generation procedure depends on the type of reaction being studied. For unimolecular reactions (primarily dissociative) the bonds between the connected reactive centers are broken and each new fragment is driven towards a nearby binding site. All the empty binding sites within a 3.0 Å vicinity of each reactive center are considered, and the pair of binding sites that are farthest apart are selected as the fragment destinations (Figure 3.4). This category of reactions will be tested by propanoic acid dissociation on Pd(111) surface in the Results section.

Bimolecular reactions are generally more complex than the first category. To sample these reactions, combinations of connectivity breaks are generated for reactive atoms first. The coordination number for all the other reactive atoms is determined to avoid unreasonable chemical reactions before generating bond-forming driving coordinates. Then, combinations of add and break moves are generated, with a limit of four simultaneous add and break moves (two of each). This category of reactions has been tested on the atomic layer deposition (ALD) reaction to create TiN on Si(100).

After driving coordinates are generated for each optimized initial structure, TS finding commences using SE-GSM. Since SE-GSM produces single elementary step reactions, all intermediates/products of each step will be available for subsequent iterations of S-ZStruct to determine full RPs.



(a) (b)

Figure 3.4: (a) The bond between green and red fragments will be broken. Available binding sites for each fragment are shaded by the fragment's color. The binding sites marked in light blue are the selected binding sites where each fragment will end up on. (b) NH_3 dissociating to form NH and H fragments. N and H are marked in green and red, respectively, and will be driven to the binding sites marked in the same color.

3.4 Computational Details

Initial geometries were created using ASE. [10] Reactant structures were optimized using the BFGS Line Search optimizer and were considered converged when the maximum force on each atom was below 0.05 eV/Å. All energy and gradient calculations are performed in a plane wave basis set under periodic boundary conditions as implemented in the Vienna Ab Initio Simulation Package [115–118] (VASP). The PBE functional and projector-augmented wave methods are used to describe the exchange-correlation energy and electron-ion interactions, respectively. An energy cutoff of 300 eV and a smearing parameter of 0.2 eV were used for the plane waves. The Brillouin zone is sampled using a 1×1×1 Monkhorst-Pack mesh. A Pd slab with 2 layers and 12 atoms in each layer is used for propanoic acid dissociation network, while the bottom layer is kept fixed. The Si slab has six layers, with 16 atoms in each layer. Surface GSM is implemented in C++ and invokes ASE to provide the quantum mechanical gradients through VASP. S-ZStruct is implemented in C++ and Python and uses Open Babel [1, 159] to determine interatomic connectivity. To compare structure similarity in Stage 4, a RMSD value of 0.01 Å was chosen. For elements in the test examples, the maximum and minimum coordination numbers were fixed such that hydrogen maintains single coordination, carbon must be 1-4 coordinate, oxygen is 1-2 coordinate, and titanium has a maximum coordination number of 6.

3.5 Results and Discussion

3.5.1 Propanoic Acid Dissociation Network on Pd(111)

S-ZStruct's capacity for uncovering unimolecular reactions was tested on key reactions involved in the decomposition of propanoic acid to ethane on Pd(111). [130] This reaction network has been extensively investigated by the Heyden group [21, 129, 129–131, 133] for its role in biomass hydrodeoxygenation, [130] where this step is believed to be rate limiting for the overall process. [130] Of specific interest is whether the sampling capabilities of S-ZStruct could (at minimum) reproduce the previously reported initial elementary steps along this pathway (i.e. Figure B.1), which were all generated using chemical intuition. Because this test case involves a large number of plausible RPs and intermediates, S-ZStruct's ability to sample additional degrees of freedom, automatically set up calculations, and provide single-ended searches is expected to allow a rapid, thorough examination of the reaction mechanism.

As it is our intent to test S-ZStruct on nontrivial reactions, two intermediates (derived from propanoic acid) in the center of the reported reaction network on Pd(111) (Figure B.1) were considered rather than propanoic acid itself. Figure 3.5 shows the elementary steps found by S-ZStruct for intermediate 1, ethylidene-1-ol-1-olate. Figure 3.6 likewise gives the reactive steps immediately connected to 2, carbonylethylidene. Using the reactive atom selection shown in red, all previously identified RPs were found for 1 and 2, showing good enumeration over the possible unimolecular reactions. Even more so, one pathway from each intermediate (P1, P6) was not previously reported, and these are highlighted in blue. This step, methyl fragmentation, was not previously studied because the authors had assumed it was not important to the reaction network. S-ZStruct, on the other hand, samples any pathway consistent with the reactive atoms, and therefore located these two possibilities.

S-ZStruct additionally provides the ability to sample adsorption orientations on the surface. For the carbonylethylidene geometry (Figure 3.7), S-ZStruct created three initial adsorption orientations (0°, 30°, and 60°) that were optimized to find the three stable adsorption structures. The previously reported initial structure [130] matches the current 0° geometry, and the 30° structure was 4.0 kcal/mol above the 0° and 60° geometries, which were degenerate. Interestingly, the specific details of the reaction mechanisms vary with adsorption orientations, as summarized in Table 3.1. Figure 3.7 shows that as the initial structures decomposed, different adsorption sites were preferred by products of each elementary step. However, there are still some similarities between the results, for example CO fragments prefer high coordination binding sites while CH_3 is more stable on low

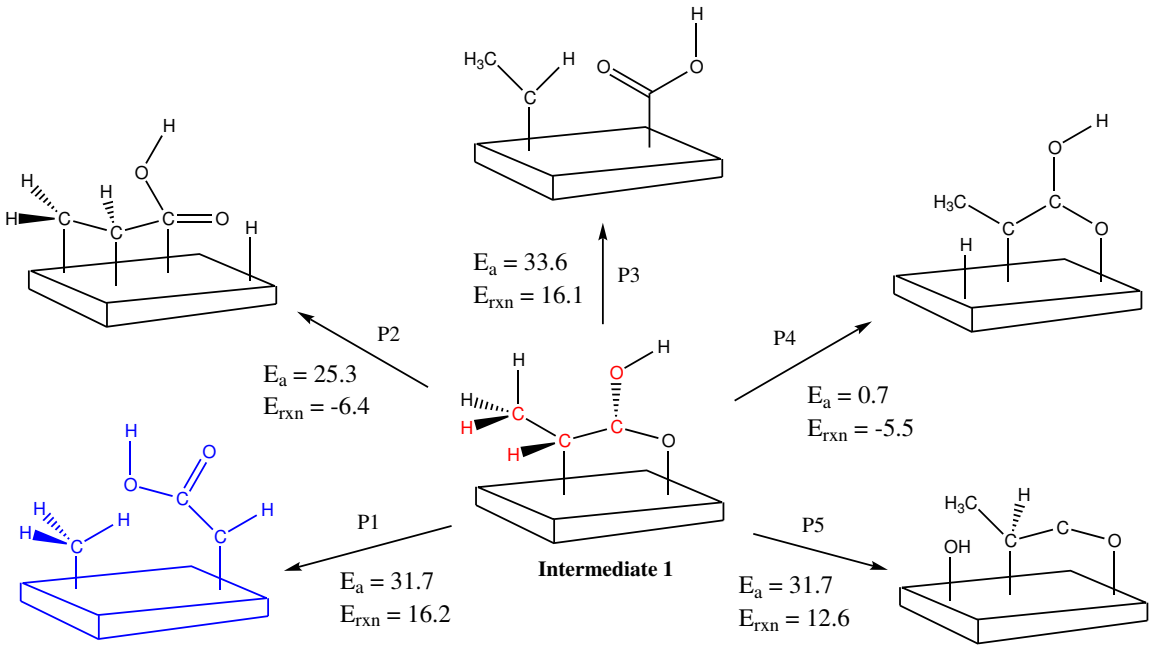


Figure 3.5: Reaction paths for intermediate 1 of figure B.1 as calculated by S-ZStruct. Reactive atoms are marked in red. Energies are in kcal/mol.

coordination bridge position. Computed activation energies for each reaction also vary, depending on how much the neighboring binding sites of a dissociating fragment help facilitate the reaction. Vitally, the activation energies at the 60° geometries were uniformly lower than the 0° geometry, demonstrating how important it is to sample these degrees of freedom when computing realistic RPs.

3.5.2 Atomic Layer Deposition of TiN using TDMAT and Ammonia on Si(100)

Titanium nitride (TiN) is one of the widely used diffusion barriers in modern inorganic semiconductors. [152] TiN can be deposited on Si surfaces using techniques such as physical vapor deposition (PVD), chemical vapor deposition (CVD), and atomic layer deposition (ALD). [107] ALD is especially useful because it leads to conformal layers with molecular-level control over their thickness, which is especially useful as the size of structural features continues to shrink. ALD achieves this using alternating cycles of two self-limiting and complementary reactions that use gaseous precursors, overall resulting in layer-by-layer growth. Two common precursors for ALD of TiN are tetrakis(dimethylamido)titanium (TDMAT) and ammonia (NH_3). [225]

Despite its popularity, the reaction mechanism for ALD of TiN is not completely un-

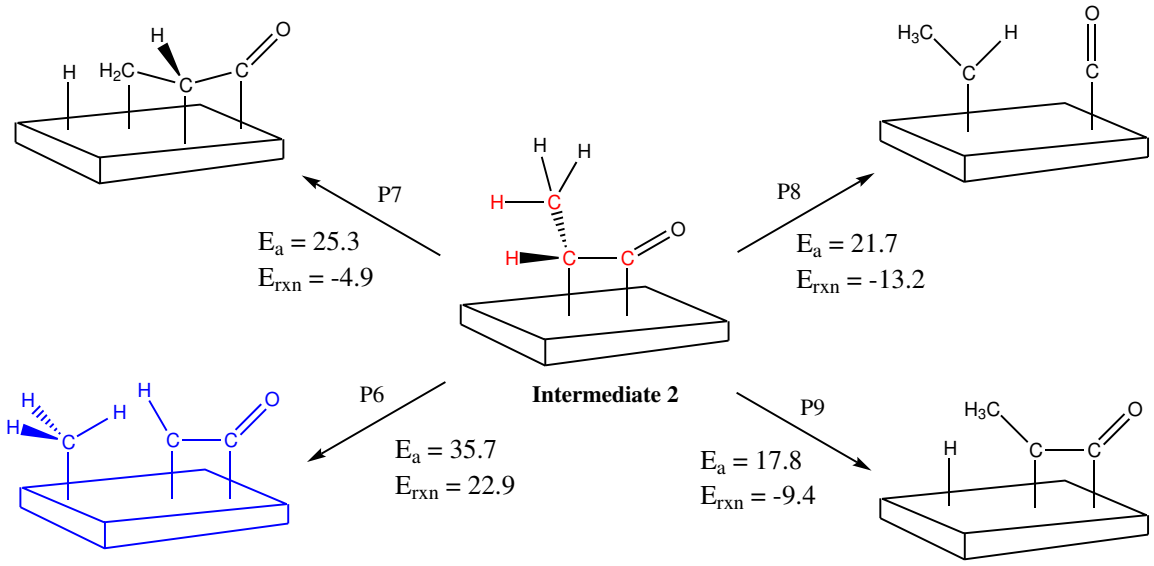


Figure 3.6: Reaction paths for intermediate 2 of figure B.1 as calculated by S-ZStruct. Reactive atoms are marked in red. Energies are in kcal/mol.

		0°			30°			60°		
ID	Atom†	E_a	E_{rxn}	Binding Site*	E_a	E_{rxn}	Binding Site*	E_a	E_{rxn}	Binding Site*
P6	C	35.7	22.9	bridge	46.5	11.7	bridge	32.3	10.1	bridge
P7	H	25.3	-4.9	fcc	36.8	11.1	bridge	15.9	5.8	bridge
P8	C	21.7	-13.2	hcp	15.5	-25.7	fcc	16.6	-21.5	hcp
P9	H	17.8	-9.4	hcp	13.5	-19.1	fcc	14.6	-2.7	hcp

*The binding site where the dissociating fragment ends up on.
† Atom of the fragment directly connected to the binding site.

Table 3.1: Final binding sites of the dissociated fragments and activation energies calculated for adsorption orientations of intermediates 1 and 2 of propanoic acid reaction at 0°, 30°, and 60°. All energies are in kcal/mol

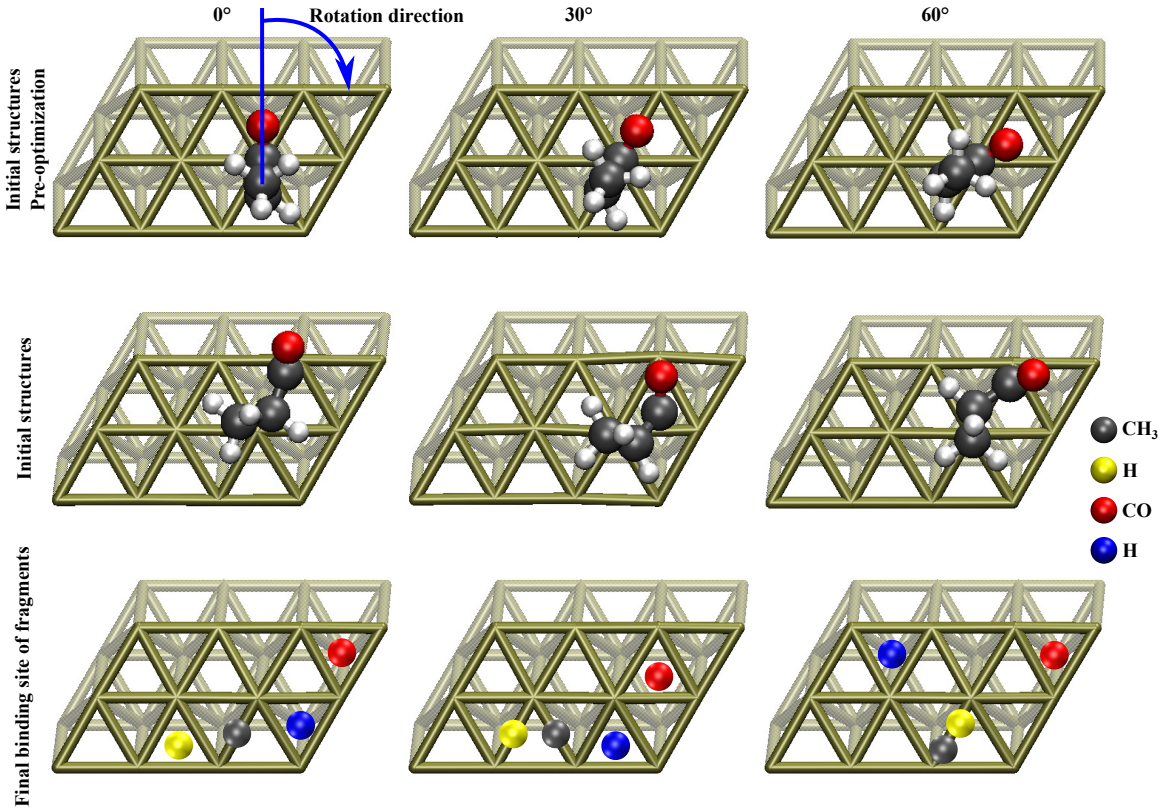


Figure 3.7: Details for reaction of carbonylethylidene and Pd(111). First row: stage 2 of S-ZStruct generates geometries after aligning input structures at three different sampling orientations. Second row: structures after geometry optimization. Last row: each sphere indicates the final binding site for fragments produced during one of four dissociation reactions.

derstood, limiting the ability to tune the precursors and conditions for improved reactivity. Computational studies, however, can investigate this process at molecular level to provide deep atomistic insight into its chemistry. The first few steps of TiN ALD include transamination, ligand elimination as dimethylamine gas, and insertion of ammonia as NH_2. While these particular steps are qualitatively well-understood, S-ZStruct can be useful for providing 3-dimensional and energetic information to better inform these steps. The later steps of reaction, however, involve reduction of Ti(IV) to Ti(III) and are poorly understood. These reductive pathways will serve as a strong test for S-ZStruct's power to discover new reactions, and therefore inform chemical intuition.

After a sequence of S-ZStruct runs, the TiN ALD process was found to proceed through: Step (1) Addition of ammonia to replace surface hydroxides by NH_x species, with concurrent release of H_2O. Step (2) Deposition of titanium via ligand exchange between NH_2 and TDMAT, extruding dimethylamine gas. Steps (3 and 4) Bonding of Ti to neighboring NH_2* groups to form N bridges that are observed in the TiN unit cell. Step (5) Ti reduction and addition of ammonia to the Ti-terminated surface and preparing it for addition of further TDMAT species. Repeating steps 2 to 5 provides access to additional layers of TiN. A summary of activation energies and elementary step reactions are shown in Table B.2 and Figures 3.8 to 3.10 and B.2 to B.4.

Step (1): Nitrogen nucleation during first NH_3 cycle. To initiate nitride growth on the surface, NH_3 molecules are attached onto the OH-terminated Si(100) surface by proton exchange between OH* and NH_3(g). After this step, the surface becomes NH_x terminated (x=0-2) and water is produced. This step is shown in Figure B.2.

Step (2): Deposition of Ti layer. Ti is incorporated onto the surface NH_2 centers through transamination [62, 127] during the first TDMAT cycle (Figure B.3). Activation energies and heat of reactions for this step and competing reactions are summarized in Table B.2.

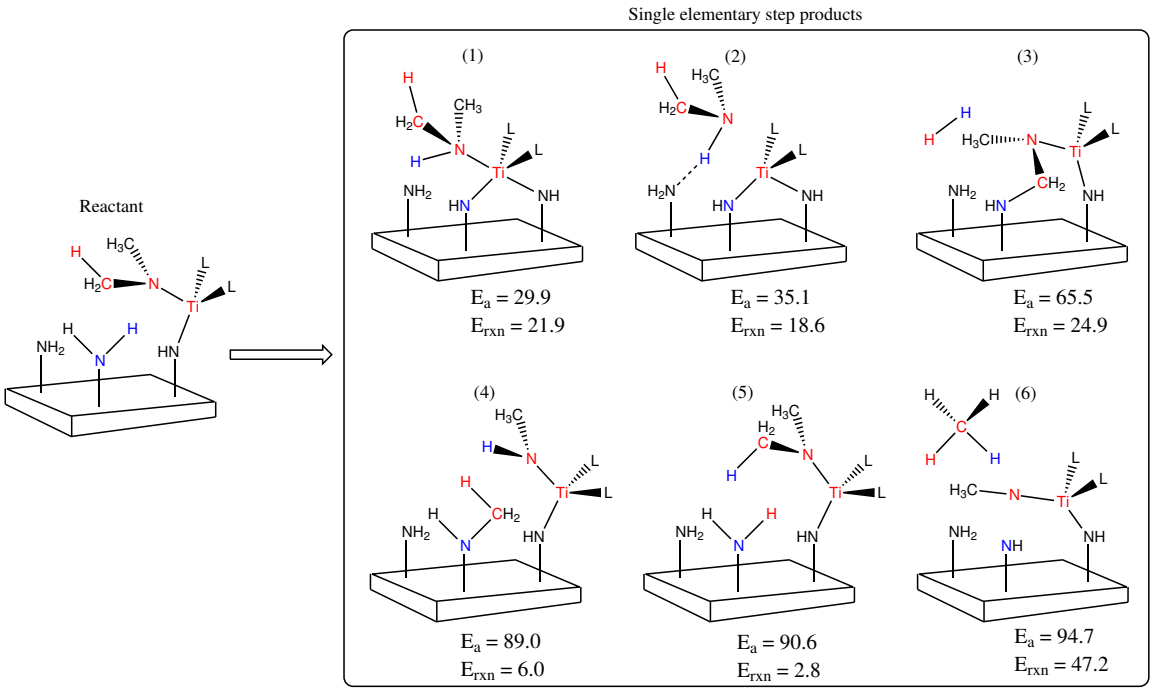


Figure 3.8: RPs for the reaction between Ti* and subsequent NH_2* species (*Step 3*). Reactive atoms are marked in red and blue. Energies are in kcal/mol.

Steps (3 and 4): Ti-bridging steps. From the single Ti-N bond of *Step (2)*, proton-transfer/ligand-exchange reactions rearrange the bonding to be closer to the expected bonding-network of TiN(111), where each Ti is connected to six nitrides in an octrahedral configuration. Two reactions happen in this stage, where diamido ligands on Ti undergo proton-transfer followed by dimethylamine release, where in each step a Ti−N bond is formed. While these steps are all endothermic, the increase in entropy from gas release makes these reactions thermodynamically favorable (Figures 3.8 and 3.9). After *Steps 1 to 4*, Ti on the surface remains Ti(IV), as the ligand-exchange reactions leave the oxidation state unchanged.

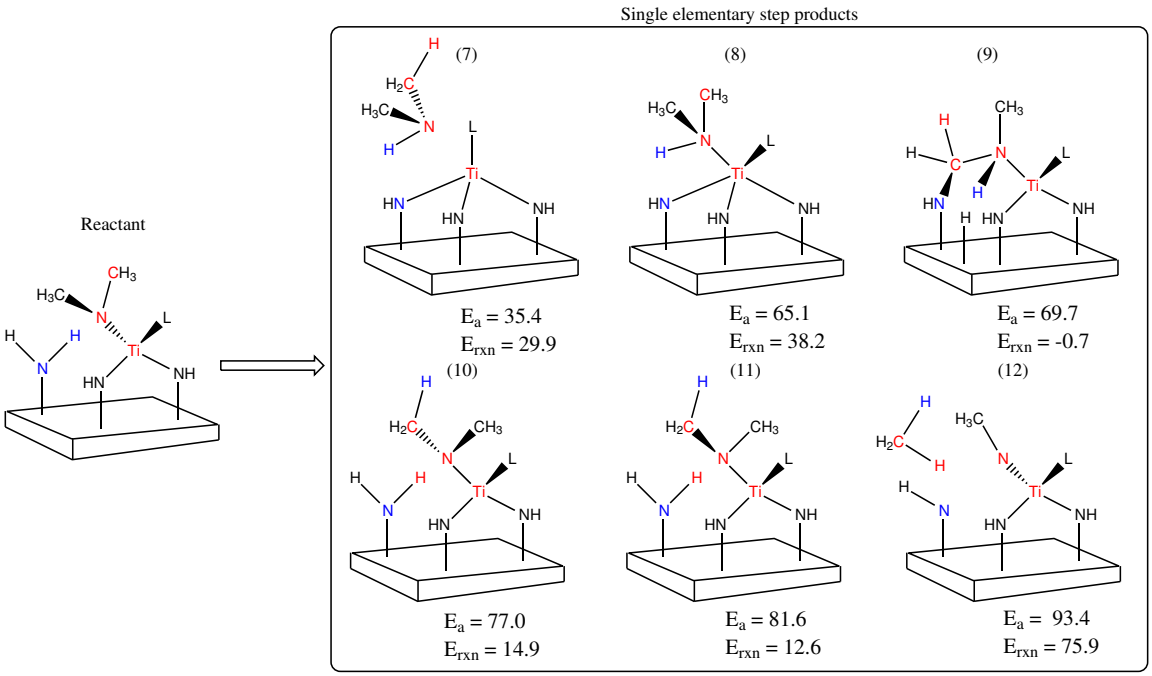


Figure 3.9: Reaction paths for the reaction between Ti* and subsequent NH_2* species (*Step 4*). Reactive atoms are marked in red and blue. Energies are in kcal/mol.

Step (5): Ti reduction. The reduction of Ti(IV) to Ti(III) is a vital, yet little understood step of TiN ALD. Two proposed mechanisms have been suggested, one involving reductive elimination of hydrazine derivatives [51–53] and the other β-hydride elimination involving dimethylamido ligands. [31, 32] Byproducts suggestive of each of these reactions have been observed experimentally. [31, 32] It is unclear, however, from which intermediates these reactions may take place, or their relative kinetic viability. Therefore the starting geometries for S-ZStruct represent intermediates generated prior to (reactions 15 to 17) or after (reactions 13 and 14) the second ammonia cycle. Activation energies and heat of reactions from S-ZStruct for some unfeasible elementary reactions of Ti reduction at the reaction conditions (T = 150 °C - 200 °C) are summarized in Figures B.5-B.7.

Three reactions were found to release hydrazine derivatives and reduce Ti(IV) to Ti(III), reactions 13, 14 and 15 of Figure 3.10. Two of these reactions (13, 14) have large activation barriers, while the intermediate of reaction 15 with two alkyl amido groups has favorable electronics and thus a relatively low barrier of 40.3 kcal/mol. The most viable reduction pathways involve β-hydride transfer from one of the methyl groups of a diamido ligand to the N of a neighboring ligand (reactions 16 and 17). The barrier is substantially reduced in 17 compared to 16 due to the high valence of the Ti centers.

The starting structure of reaction 17–the lowest barrier pathway–has two Ti centers,

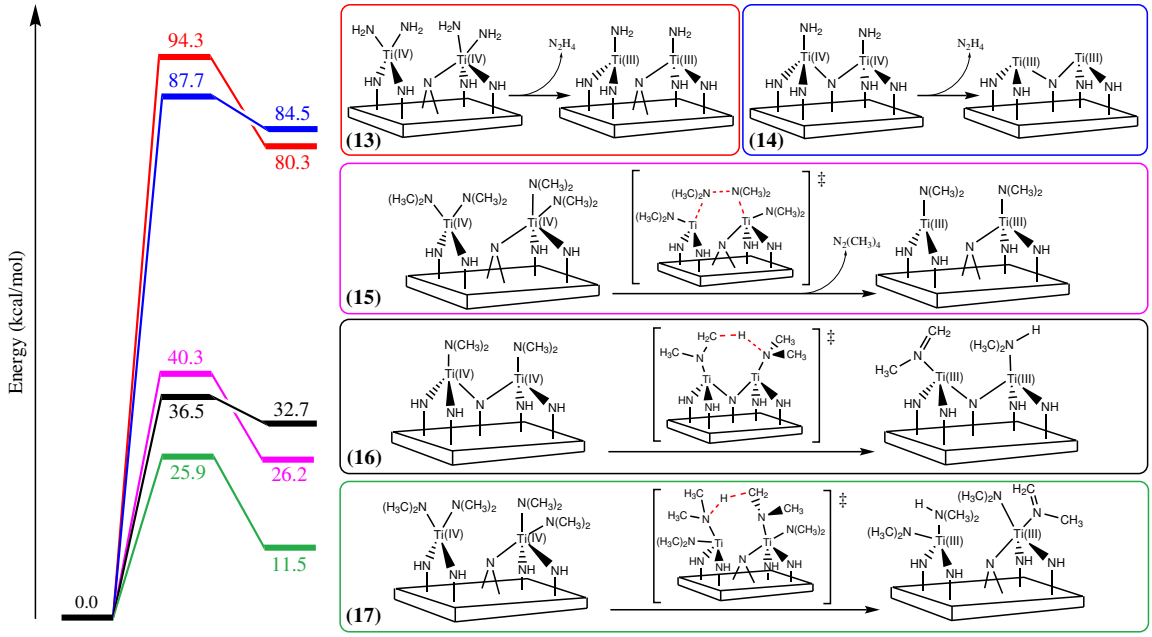


Figure 3.10: Alternative Ti reduction pathways. Energies are in kcal/mol.

one with a tetrahedral and the other a distorted trigonal bipyramidal geometry, whereas in reaction 16 both metal centers start with a tetrahedral geometry. Due to these steric effects, the distance between the hydride-accepting N and transferring H is 3.5 Å in reaction 16, while this distance is 3.1 Å in 17 at the reactant geometry. The change in the bond length between the Ti center and the H-accepting nitrogen is the same in both reactions (1.9 Å to 2.2 Å), but the angle changes are quite different. The C−H−N angle between H-donating carbon, transferring hydrogen, and H-accepting nitrogen changes from 104.3° to 170.2° in reaction 16 from the reactant intermediate to the TS, and 116.7° to 161.4° for reaction 17. In other words, the starting geometry of reaction 17 is more predisposed for reduction compared to that of 16, as 17 has key geometric parameters that start closer to its product than 16. Overall, the extra ligand of 17 is responsible for steric distortions that strongly promote reductive elimination via β-hydride transfer.

Putting *Steps 1 through 5* together, Figure 3.11 shows a possible mechanism for the overall reaction cycle of TiN deposition on Si(100) surface. Two primary reaction types comprise the entire cycle: ligand exchange and β-hydride transfer, where the former is rate-limiting through step 4. Details of this reaction cycle are expected to help design new ALD precursors for TiN deposition.

The proposed reaction sequence is related to, but distinct from, ALD of TiN on Cu(111). [101] On Si, the ALD process initiates by N nucleation onto atop or bridge sites, which occurs through proton-exchange with the OH-terminated slab. In contrast, on Cu(111), N

is placed in fcc or hcp sites via N−H bond dissociation to form dihydrogen. After N sites form, the two surfaces undergo similar ligand exchange reactions with the TiN precursor, and formation of intraligand H-bonding networks stabilizes the reaction intermediates to accelerate proton-transfer. [101] The reduction step was not studied on the Cu(111) surface, but we surmise it may occur through a similar β-hydride transfer mechanism.

3.5.3 Atomic Layer Deposition of W from WF_6 and SiH_4 on Si(100)

In this study, tungsten ALD takes place on hydroxyl-terminated Si(100) slab with WF_6 and SiH_4 precursors acting as as sources of W and reducing reagent, respectively. The reaction can initiate in two ways, either WF_6 molecules reacting with −OH to form HF or HOF (2 and 3), or silane displacing the −OH through hydrogen gas or water formation (4 and 5). The lowest energy path for this step has an activation energy of 7.5 kcal/mol, however it produces structure 3 that has oxygen impurities in the W layer. The water displacement pathway using silane results in a more stable intermediate (-13.2 kcal/mol) with a higher activation energy (E_a=36.2 kcal/mol). Another step of this reaction between −OH and SiH_4 removes −OH groups from the surface and results in SiH_2 species that are doubly-coordinated to Si surface atoms (structure 6).

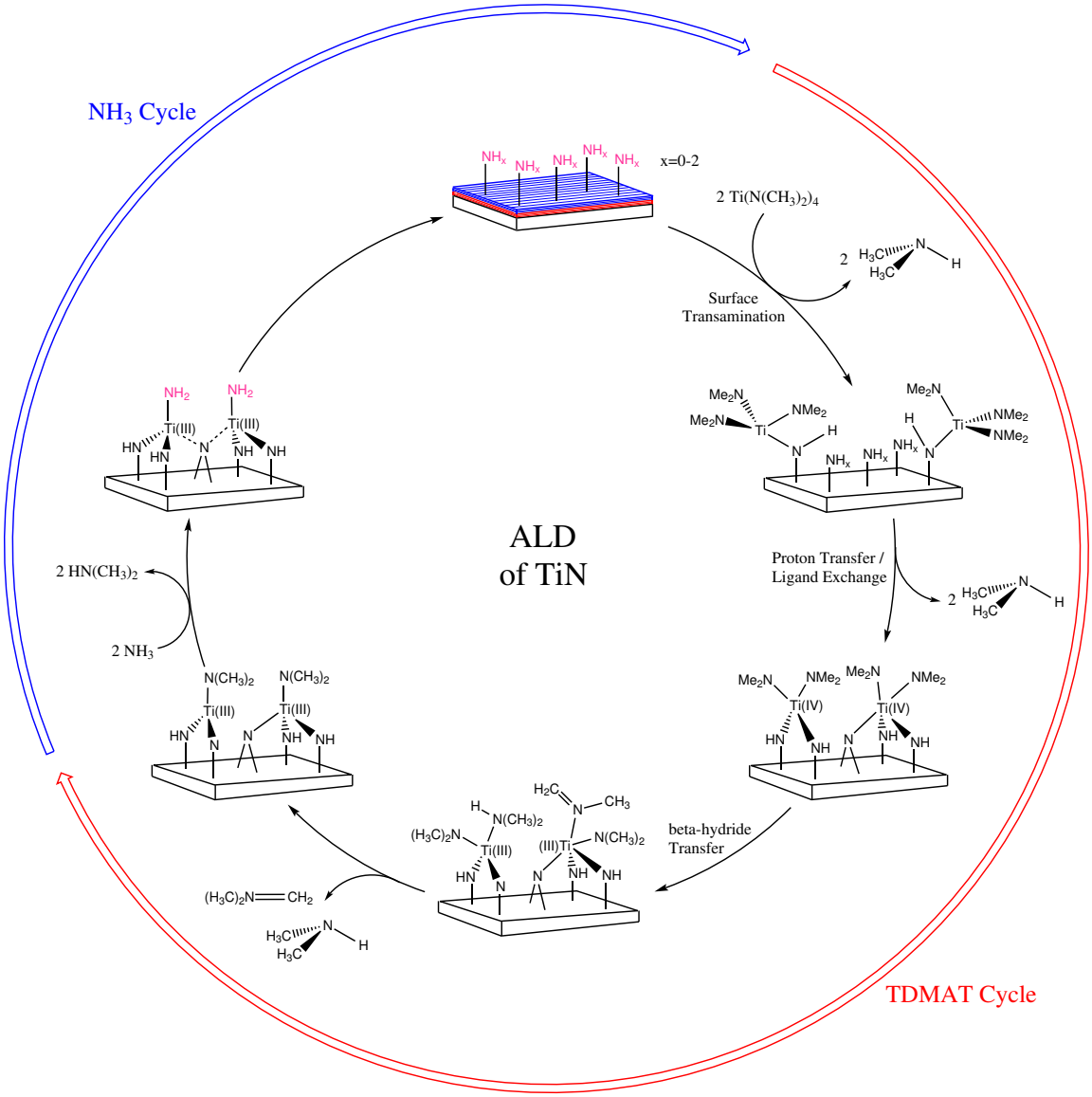


Figure 3.11: Overall reaction cycle for TiN ALD. The reaction starts by replacing terminating OH group by NH_2, marked in pink. TDMAT is instoduced to the reaction in the next steps, followed by NH_3 cycle which displaces diamido ligands.

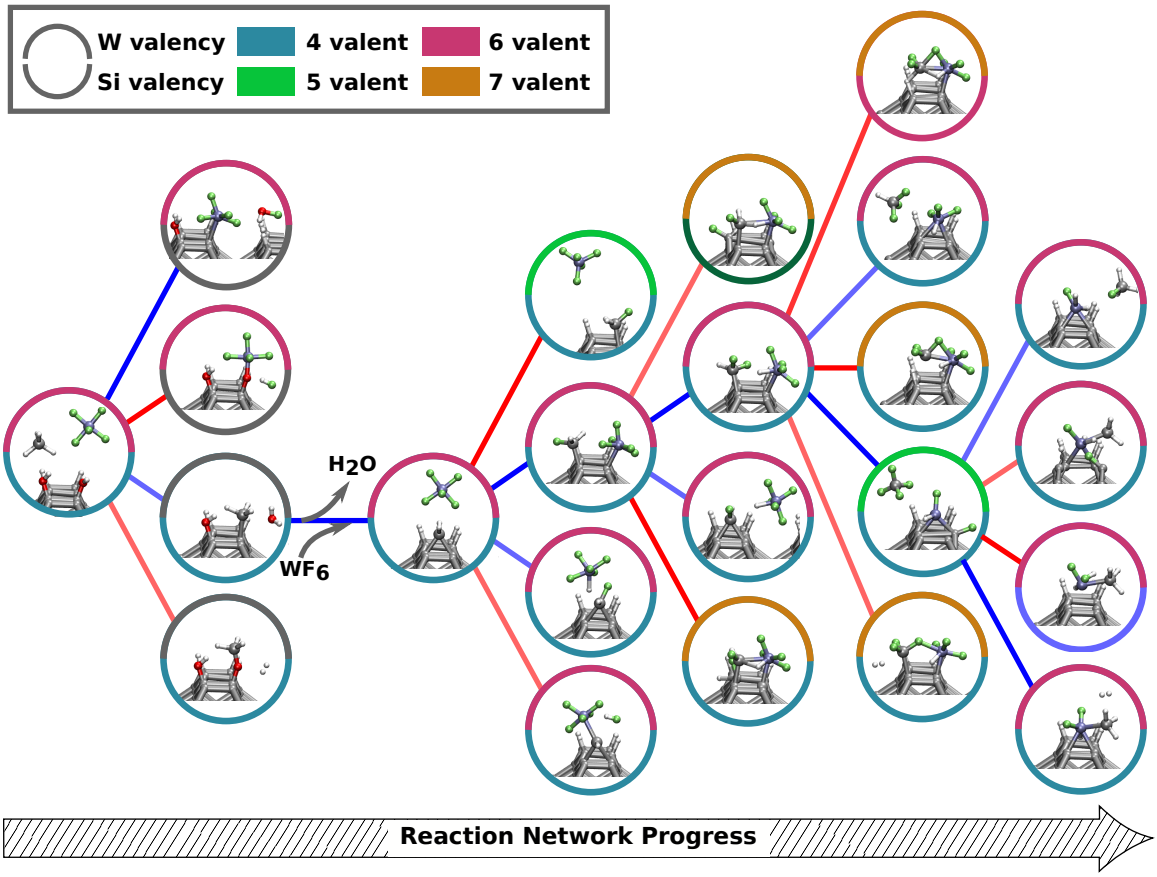


Figure 3.12: The partial reacion network for ALD W calculated by surface-ZStruct.

At this stage, the other precursor, WF_6, is introduced to the reaction where it transfers one of its F ligands to the Si center of deposited silane, removing one of the coordinations between Si and the surface (double to single coordination), and opening up a new surface site for W attachment (E_a=15.4 kcal/mol). At this point, a ligand-exchange reaction followed by ligand-transfer results in extruding SiF_4 with small activation energies of 2.6 and 7.5 kcal/mol, respectively. Alternatively, a ligand-transfer from WF_5H to SiF_2H with an activation energy of 24.8 kcal/mol could produce trifluorosilane and adsorbed WF_3H (ΔE=3.3 kcal/mol). In the next step, another round of silane addition is carried out, removing a hydrogen atom from W center of intermediate 17 through ligand-exchange with silane, releasing hydrogen gas (E_a=21.9 kcal/mol). Repeating silane reaction cycles will reduce W(IV) in 23 to W(0) through HF, F_2, and $SiF_xH_{(4-x)}$ production and deposit layers of tungsten on silicon slab.

3.6 Conclusions

Elucidating a reaction cycle, including its relevant intermediates and connecting TSs, is one of the major challenges in computational chemistry. While understanding a chemical process gives enormous control, creating the complete reaction network of a chemical process is usually computationally expensive and requires much manual labor. This is especially the case for surface reactions, which typically have many intermediates and a host of reactive possibilities.

To overcome these difficulties, the S-ZStruct method is herein developed and tested for the systematic study of surface reactions. The five main stages of S-ZStruct: binding site finding, reactant binding and orientation sampling, geometry optimization, unique structure identification, and creating driving coordinates, were specifically designed for investigating reactivity on surfaces, and are distinct from the original ZStruct method. S-ZStruct was found competent for studying unimolecular and bimolecular reactions, and has been tested using propanoic acid and TiN ALD reaction networks as examples in each category, respectively. Multiple adsorption orientations for the intermediates of propanoic acid reaction were found, each yielding different activation energies and reaction mechanism. Not only was S-ZStruct able to reproduce routes known in the literature and to chemical intuition, it also located realistic, previously unexplained pathways for reduction of Ti in the ALD reaction. In summary, S-ZStruct–which leverages the power of the SE-GSM–is now expected to unearth more reaction possibilities involved in a wide variety of challenging surface chemistries.

CHAPTER 4

Multivariate Linear Analysis of Thermodynamic and Steric Parameters for Characterizing KtnC Catalyzed Biaryl Coupling

4.1 Abstract

This report describes using steric and thermodynamic linear relationships to investigate and understand the reaction mechanism and site-selectivity for KtnC catalyzed cross-coupling of coumarin biaryls. Multiple reaction mechanisms are proposed for activation and coupling of biaryl reactants while there is not enough evidence to suggest which mechanism is operating for this specific cytochrome P450 enzyme. There are four different possible reaction mechanism: 1) hydride abstraction 2) proton coupled electron transfer 3) hydrogen abstraction and 4) double hydrogen abstraction. To this end, the size of substituents and ΔH values for elementary step reactions of each mechanism were calculated and examined for a linear relationship with yield. The highest correlation across all molecule classes was observed between the double H-abstraction (di-radical) mechanism and the yield, with substituent size playing a significant role in some molecule classes. Based on correlation values, the results provide evidence for the di-radical pathway for cross-coupling of coumarins. Additionally, the yield of the reaction for substrates with structural similarities could be predicted using bond dissociation energies and size, as parameters.

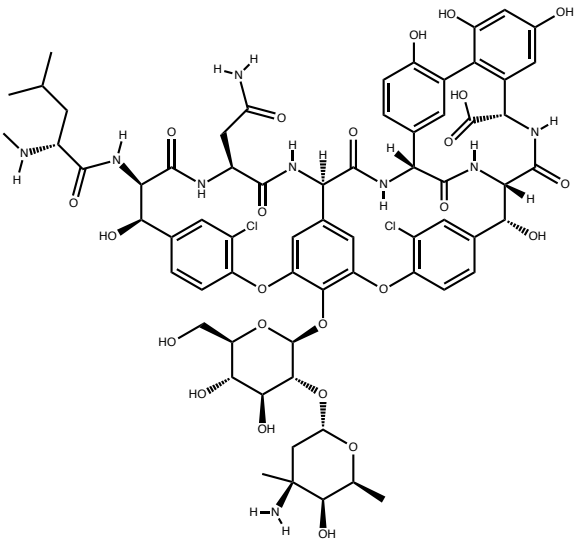
4.2 Introduction

Chemists have always tried to formulate the reactivity and selectivity of chemical reactions using electronic and steric parameters that are developed based on physical organic principles [125, 219] in order to predict reactivity in similar classes of reactions and design optimal reagents. These parameters contain physical and chemical meaning, such as

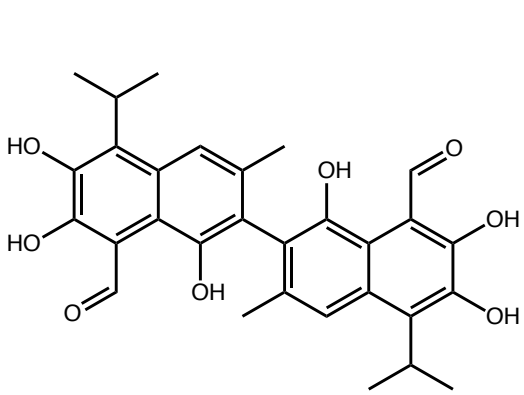
electronegativity, that can be tuned by modifying the substituent groups on a structural framework like a benzene ring. There are a number of quantitative parameters available for characterizing chemical reactivity and selectivty, such as Hammett σ [87], Tolman cone angle [206], and Charton [46] and Taft [202] parameters. An early example is the Hammett equation [87, 103] that gives information on the dependency of electron donating/withdrawing properties and reaction rate, an example of relationships that are broadly described as linear free energy relationships (LFER) [88, 125, 146, 219]. LFER relate the logarithm of a reaction rate constant for one series of reactions with the logarithm of the rate for a related series of reactions. Using this principle, the behavior of a reaction can be predicted by characterizing the reaction mechanism of a similar class of reaction.

Quantum chemical calculations can provide hundreds of chemical descriptors such as charge, spin multipilicity, and orbital occupancy that can be used as parameters for building a model like the Hammett equation to describe a chemical process. These descriptors can be used as features in data analysis approaches like machine learning algorithms to unravel chemical interactions and describe relationships that are more complex than the examples given above. However, machine learning techniques need at least hundreds of data points to operate reliably. With experimental reaction results, especially when the reactions are hard to preform or time consuming, collecting hundreds of data points is impractical and the analysis of the reaction's controlling factors is possible only through LFER analysis. Another limitation of machine learning techniques is their interpretability, meaning the degree to which a human can understand the cause of a decision. In the context of chemical problems, it is important to understand why a decision is made because the goal of model development is to explain experimental observations and alter the reaction conditions or reactant structures to achieve the desired outcome.

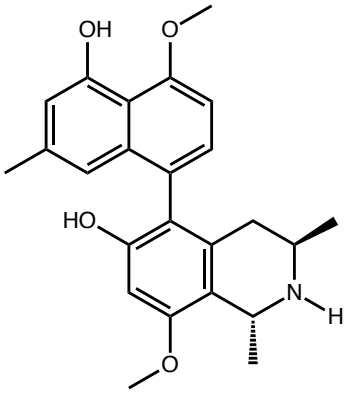
Before diving into the data analysis approach, the chemical problem at hand and its significance needs to be explained. Designing synthetic pathways for compounds containing stereogenic centers has posed a great challenge to synthetic organic chemists [35, 220]. An important class of chiral molecules, axially chiral biaryl compounds (CBC), are rotationally hindered biaryls (atropisomers) where the configuration at the biaryl axis can dictate bioactivity in natural products and pharmaceuticals [35]. Examples include antibiotic heptapeptide vancomycin [72, 128, 143], antimalarial korupensamine E [128, 220], and gossypol [220] that is used as a male contraceptive in some parts of the world and it is also an antimalarial drug (Figure 4.1). Synthetic axially chiral biaryls, such as BINOL [224] and BINAP [5], are used as chiral auxiliaries or ligands by chiral catalysts for asymmetric synthesis.



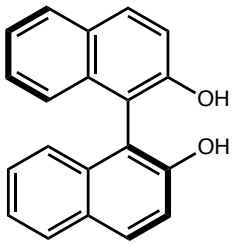
(a)



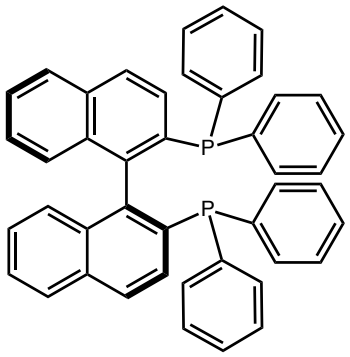
(b)



(c)



(d)



(e)

Figure 4.1: Structures of natural products, pharmaceuticals, and synthetic compounds that share the axially chiral biaryl structure. (a) vancomycin, (b) gossypol, (c) korupensamine E, (d) BINOL, and (e) BINAP.

Despite the significance and extensive studies on CBCs, they are still challenging to access through chemical synthesis [34]. Established approaches for the synthesis of biaryl compounds [34,35,113,182,188], including oxidative cross-coupling of phenols [128] with transition metal catalysts [], can be atroposelective if chiral ligands are utilized, however, it remains challenging to achieve high site- and enantio-selectivity across a broad range of substrates.

Designing a synthetic route that is atom economical, environmentally benign, and proceeds under mild reaction conditions with high site- and stereo-selectivity is an active research topic in this area. Identification and expression of KtnC and DesC [143] which are fungal cytochrome P450 enzymes that catalyze regio- and stereo-selective intermolecular biaryl couplings opens new avenues for asymmetric biosynthesis of chirally active biaryls. The native substrate for these enzymes is 7-demethylsiderin, (**1**), which is catalyzed to 6,8' (**2**) and 8,8' (**3**) isomers by DesC and KtnC, respectively (Figure 4.2). 7-demethylsiderin has a coumarin scaffold with three potential C−C bond formation sites, making the products regio- and stereo-isomers (total of 9 regio-isomers with 2 stereo-isomer for each regio-isomer). The goal of this project is to develop KtnC as a general biocatalyst for the asymmetrical coupling of aromatic compounds. At this stage, the goal is to understand and predict the reactivity and selectivity of the reaction for courmarin containing substrates, and use this information to selectively synthesize the desired regio- and stereo-isomer of any aromatic compound.

4.3 Results and Discussion

Despite numerous studies on using cytochrome P450 enzymes to catalyze coupling reactions, to date, no in-depth structural or mechanistic studies on KtnC have been reported due to limited information about the structure and active site of the enzyme. Given the lengthy and cumbersome procedure of synthesis, purification, and identification of biosynthesis products, using alternative approaches like multiparameter analysis could help in identifying the reaction mechanism, the rate limiting step, predicting site-selectivity and the yield, and eventually substrate engineering to achieve asymmetric coupled products with high yields.

4.3.1 Experimental Results

The goal of this project is to synthesize a variety of atroposelective biaryls with high yield and site-selectivity where a range of substrates are screened for this purpose. For each set

of reactions in Table C.1, the column molecule (**A** partner) is combined with each of the row molecules (**B** partner) in 1:10 or 1:20 ratio. It was observed that increasing the ratio of **B** increases the yield of A−B heterocoupling product until it reaches a plateau. The reactions are run in vivo using P. pastoris and the products are then separated for concentration and yield analysis using LC-MS. The remaining concentration of **A** and **B** and the concentration of A−A, A−B, and B−B products were measured using 1,3,5-trimethoxybenzene as the internal standard. The area under each peak in LC-MS corresponds to the concentration of that product and it could be converted to percent yield conversion using the starting concentrations and **A** partner as the limiting reagent. The yields measured using this method are reported in Table C.1.

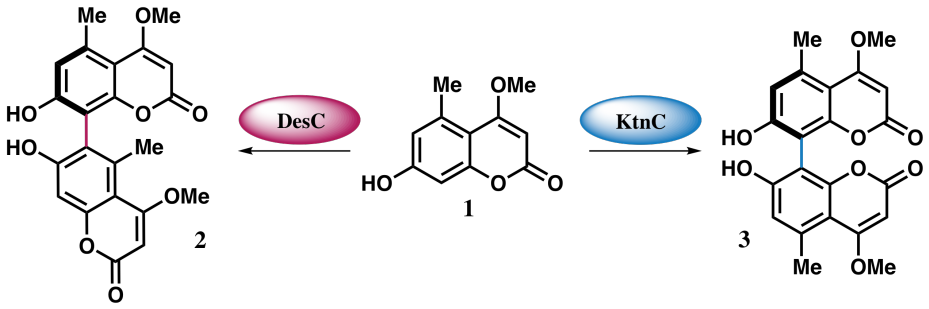


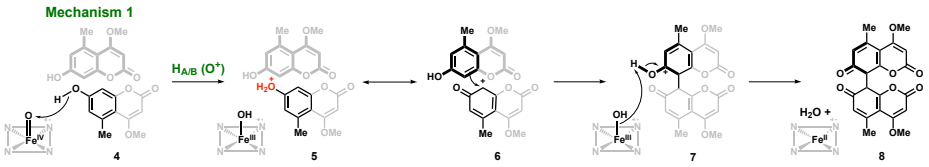
Figure 4.2: DesC and KtnC enzymes biocatalyze M-desertorin A, structure **2**, and P-orlandin, structure **3**, products starting from 7-demethylsiderin, **1**.

4.3.2 Proposed Reaction Mechanisms

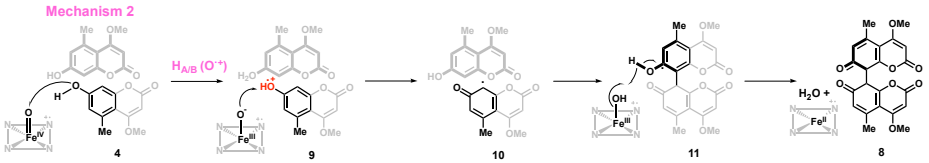
Multiple mechanisms [40, 80, 231] are proposed in the literature for cytochrome P450 class of enzymes that are briefly discussed here. The first mechanism (Figure 4.3a) involves hydride abstraction to form an electrophile at C8 position of the **A** partner at **6** that attacks the C8 carbon of **B** partner to form intermediate **7**. Proton abstraction from intermediate **7** results in structure **8**. The hydride abstraction mechanism is suggested to consist of an initial H-atom abstraction followed by a very fast electron transfer [43, 203]. This mechanism is proposed based on a combination of mass spectrometry and DFT studies on models of enzyme heme-centers in gas-phase. It is also observed that the reaction mechanism could change by changing the substrate, for example cycloheptatriene proceeds through hydride transfer, whereas cyclohexadiene and toluene react through hydrogen atom abstraction [43].

Another possible pathway, proton coupled electron transfer (PCET) [178, 199, 230], starts with an electron abstraction from the phenolic oxygen that results in a radical cation activated substrate **9**. This structure undergoes a proton transfer to form an activated radical

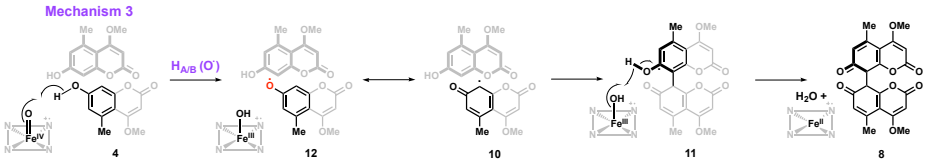
structure. A resonance form of the activated radical substrate, **10**, (Figure 4.3c) attacks the closed-shell partner to form a C−C bond and yield **11**, and then the final product **8**. PCET could proceed as a concerted single-electron, single-proton transfer or in could happen consecutively.



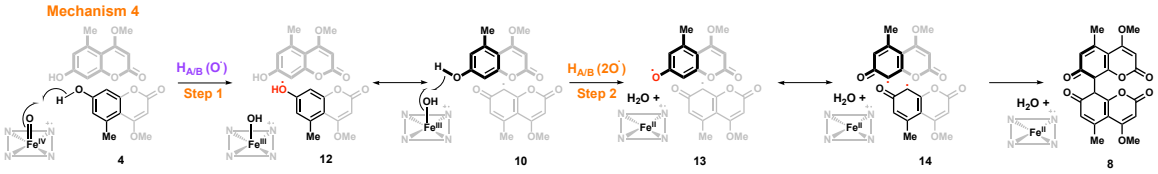
(a) Mechanism 1, hydride abstraction, starts with tranfer of hydride species from one of the substrate partners to yield structure **5**. Structure **6** depicts the resonance structure where the positive charge is on C8. The hydride-abstracted substrate acts as a nucleophile and attacks the cationic carbon center to form a C−C bond and product **8**.



(b) Mechanism 2, PCET, consists of consecutive reaction steps of electron and proton transfer. Structure **10** is a resonance form of the activated substrate. The activated radical substrate attacks the second substrate to yield intermediate **11**, followed by another H-abstraction to give product **8**.



(c) Mechanism 3 shares intermediates **10** and **11** with mechanism 2, however the activation step is different. In this mechanism, the π-stacked complex **4** undergoes an H-atom abstraction to generate intermediate **12** and resonance form **10**.



(d) Mechism 4 consists of 2 elementary steps. Step 1 is identical to mechanism 3, followed by a second H-atom abstraction to form a bi-radical intermediate **14** that would eventually recombine to form product **8**.

Figure 4.3: Proposed reaction mechanisms for biaryl coupling catalyzed by KtnC enzyme.

Alternatively, we can have a single or double H-abstraction steps leading to mechanisms 3 and 4 of Figure 4.3. In mechanism 3, the heme active site of the enzyme abstracts a H atom and is reduced to iron(III), while forming an activated free radical substrate **12**.

The resonance form of radical **10**, with the single electron at C8 can attack the closed-shell molecule to form a C−C bond (**11**). This H-abstraction can occur one more time, **13**, to form two free radical partners and produce water and iron(II), followed by two radicals combining to form the 8,8' product **8**. This reaction is reported in the literature [80, 198, 231], but there is no concrete evidence on the operating mechanism.

4.3.3 Linear Regression Analysis of Thermodynamic and Steric Properties

Multiple physical properties are tested for a relationship with the yield in order to identify the reaction mechanism and predict the yield of the reaction. For the modelings in this section, linear regression is used because of (1) the limited number of datapoints available in each subgroup of reactions and (2) the importance of interpretability of the model for explaining observations using chemical concepts. Having only tens of datapoints restricts the number of mathematical models that could be applied to the data while also avoiding overfitting.

Based on physical organic principles, steric and electronic effects play a significant role in controlling the reactivity of given sets of reactants. In order to quantify these effects, subsets of coumarin molecules with methyl group on the C5 position and varying R groups on C4 position are selected. For steric effects, the sterimol parameters [33] are used to quantify changes in the size of substituents while no electronic parameter is available to measure the effects of R groups on neighboring rings in biaryl systems. Sterimol parameters consists of three elements that measure the minimum and maximum widths of a substituent and the total distance following the primary axis of attachment of R groups. It should be emphasized that electronic and steric effects are inseparable in this case (with the %yield results available to us) because the reactions are catalyzed by an enzyme and while changes in ether groups increase electron donating/withdrawing property, it also makes the substrate bulkier, interfering with enzyme's activity. Because the measured yield values are only available for a limited number of substrates, the comparison of electronic effects for R groups with similar size (like ethyl, methoxy, fluorine or chlorine) or similar electronic effects but varying size (such as tertiary butyl and methyl) is not feasible. Thus, identifying the importance of sterics or electronic effects without contributions from the accompanying effect is not possible yet.

Evans-Polanyi principle states that the activation energy and the enthalpy of reaction are proportional for reactions in the same family. Based on this, it is assumed the rate limiting step for each mechanism is the activation step and the ΔH values for each mechanism's

activation pathway are calculated in order to distinguish between different reaction mechanisms. Some of the mechanisms share common intermediates, but they all have different activation steps and intermediates. It was also observed that the thermodynamic properties of the π-stacked complexes between the reaction partners play an important role in activating the substrates. The stacking is the result of attractive interactions between aromatic rings and their π bonds. π-stacking is responsible for many important structural phenomena such as self-assembly and protein ligand binding as well as controlling stereoselectivity in some classes of reactions [114]. Thermodynamic and steric properties of individual molecules (non-complex form) show no correlation with the yield while these properties in the complex form show relatively significant correlation with reactivity, given that the experiments are performed in biological systems that makes the measurement highly complex and the reaction results very dependent on small changes in reaction conditions.

Another objective was to predict the yield of the reaction with just a few features that could be easily obtained from quantum chemical calculations. To achieve this, multivariate linear regression with size and bond dissociation energies (BDE) as variables is also used to predict the yield of reactions while the error is measured using mean square error loss function (Figure 4.4).

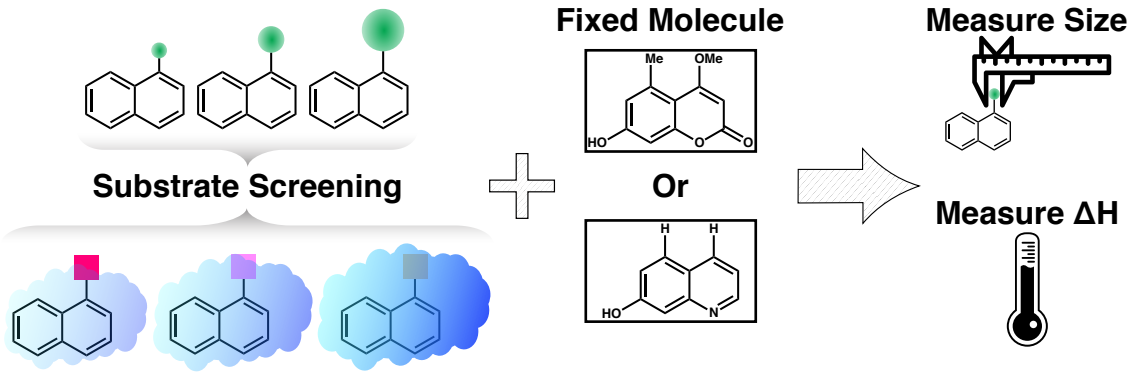


Figure 4.4: BDEs and sterimol parameters for a range of substrates are calculated and the linear relationship between the calculated parameters and reported experimental yields is examined to identify any significant correlation and reaction mechanism.

4.3.3.1 Identifying the Reaction Mechanism

As mentioned before, there are four different reaction mechanisms proposed for this reaction. In an attempt to identify the operating pathway, enthalpies of each elementary step are calculated and compared with the reported experimental yield. The following reactions are considered:

$\Delta H_A(O^+)$: AO–H BOH $\longrightarrow$ AO^+ + BOH + H^- Mech. 1

$\Delta H_B(O^+)$: AOH BO–H $\longrightarrow$ AOH + BO^+ + H^- Mech. 1

$\Delta H_A(O^{\cdot+})$: AO–H BOH $\longrightarrow$ $AOH^{\cdot+}$ + BOH + e^- Mech. 2

$\Delta H_B(O^{\cdot+})$: AOH BO–H $\longrightarrow$ AOH + $BOH^{\cdot+}$ + e^- Mech. 2

$\Delta H_A(O^{\cdot})$: AO–H BOH $\longrightarrow$ $AO^{\cdot}$ + BOH + $H^{\cdot}$ Mech. 3

$\Delta H_B(O^{\cdot})$: AOH BO–H $\longrightarrow$ AOH + $BO^{\cdot}$ + $H^{\cdot}$ Mech. 3

$\Delta H_A(2O^{\cdot})$: $AO^{\cdot}$ BO–H $\longrightarrow$ $AO^{\cdot}$ + $BO^{\cdot}$ + $H^{\cdot}$ step 2 of Mech. 4

$\Delta H_B(2O^{\cdot})$: AO–H $BO^{\cdot}$ $\longrightarrow$ $AO^{\cdot}$ + $BO^{\cdot}$ + $H^{\cdot}$ step 2 of Mech. 4

ΔH of each step was calculated for molecule class I (see Figure C.1), if a significant correlation ($R^2 > 0.5$) was observed for molecule class I, the ΔH for the same step was calculated for other molecule classes to see if the correlation holds. Molecule class I was selected for correlation checking because it has the highest number of datapoints (8 points). No correlation was observed between the cationic pathway and yields for molecule class I (Figure C.2c), so this pathway was ruled out. Another possibility was proton coupled electron transfer, mechanism 2, where coefficient of determination was 0.55 for molecule class I. However, repeating this analysis for molecule classes I and V gave $R^2 < 0.1$ for both sets. So, both mechanisms 1 and 2 are ruled out.

As mentioned before, radical and di-radical pathways are also suggested in the literature and recently further evidence for the radical pathway (direct attack of the radical) is provided in a computational study for an intermolecular oxidative cyclization of griseofulvin [80]. ΔH of reactions 5-8 and the sterimol parameters of the R group at C5 position were calculated and their association with $\log_{10}$(yield) was examined. The measured percent yield values are very sensitive to cell growth state, temperature, and separation of yeast debries from the product. To minimize the effect of experimental errors on yield and our analyses, the $\log_{10}$ of yield was used for correlations, this way only significant changes in yield are reflected in correlation analysis results. In data 1 with 8 datapoints (Figure C.3a), where the **A** partner is the native substrate, we see a positive correlation between yield and ΔH of AB $\longrightarrow$ $A^{\cdot}B^{\cdot}$ and a negative correlation between the size and yield. As the R groups are getting bulkier, their electron donating property is also increasing, so it is not possible to separate these two effects.

In molecule classes II and III, with 7 and 6 datapoints respectively, more endothermic reactions result in lower yield while the dependency in size is not consistent between the two sets. Like set 1, bulkier groups decrease the yield for set 2 (Figure C.3b) while this effect is the opposite for set 3 (Figure C.3c). In set 3, both variables have small correlations that could be attributed to low reactivity in this set (highest reported yield is 7.1%). The correlation between size and $\log_{10}$(yield) is data 5 is less than 0.1, that could be explained

by the relatively large size of −OEt group at C4 position of the **A** partner. The increase in the size of C4 substituent with respect to data 1 could increase the distance between two partners such that the size of the substituents on the other partner does not affect the yield significantly, because there is enough space between two partners to accommodate bulkier groups. Another difference between this set and the other three is the elementary step that has the highest correlation with $\log_{10}$(yield); in this case, ΔH of AB$^{\cdot}$ ⟶ A$^{\cdot}$B$^{\cdot}$ becomes more significant (Figure C.3d).

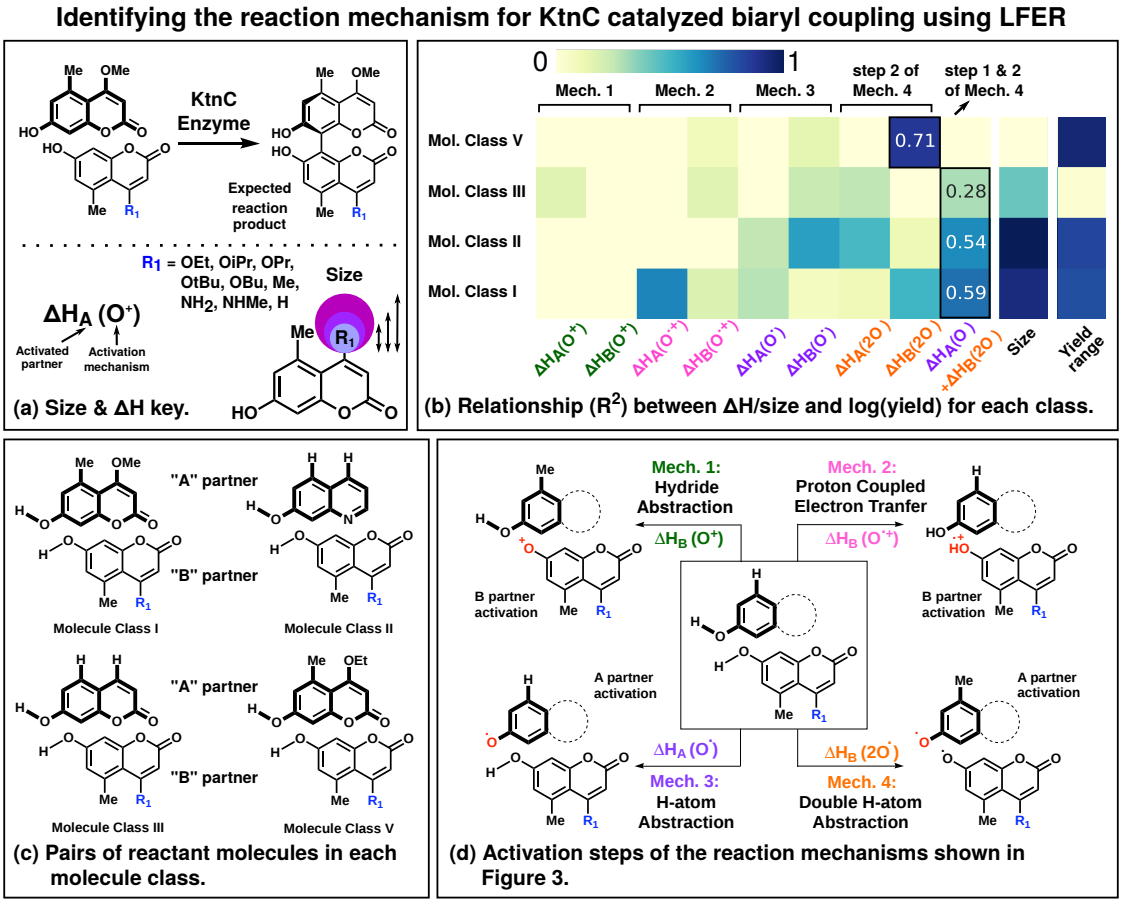


Figure 4.5: Schematic representation of dependency between changes in ΔH and size vs. the yield. ΔH values are also shown in Section 4.3.3.1.

Here, we are effectively measuring the BDE for homolytic O−H bond cleavage. It should be mentioned that no correlation between ΔH/BDEs and $\log_{10}$(yield) was observed for isolated molecules (meaning **A** or **B** by itself, not in the complex form), so formation of the AB complex is playing a role is activating each substrate for H-abstraction. Based on the calculated correlation values for each reaction mechanism for molecule class I, the di-radical pathway seems more plausible because of its high correlation with yield. In

addition to the relatively high correlation values (considering this is a biological system and the correlation target is the yield, not the reaction rate), this pathway is consistently the one with the highest correlation between all four molecule classes.

4.3.3.2 Identifying Site-selectivity

Due to limitations in accessing the crystal structure and active site geometry of the KtnC enzyme, modeling the reaction in enzyme's active site is not possible, which makes studying kinetic effects unachievable. However, we can calculate the thermodynamics of different C−C bond formations and compare them with experimental site-selectivity findings to find a pattern. Experimental site-selectivity results are gathered by comparing the LC-MS retention times of authentic standards with the products from enzyme. Based on the values reported in Table 4.1, for molecule ID 2 and 33 where there is a major and minor product, the reported major product matches the connectivity of the lowest calculated ΔH. For molecule ID 3 and 5, there are two equally proportional products, with one of them having the 8,8' connectivity. In these cases, the lowest ΔH corresponds to a different connectivity, but the second lowest ΔH matches the 8,8' site-selectivity. In order to make sure this prediction is not random, we need to have access to experimental site-selectivity results of another molecule class to compare experimental reports with calculated ΔH values for that set as well.

		A−B - $\dot{A}\dot{B}$			$\dot{A}$−B - $\dot{A}B$		A−$\dot{B}$ - $A\dot{B}$		Exp.	Ratio
Mol. ID	R1 Group	6,8'	8,6'	8,8'	3,3'	8,3'	3,3'	3,8'	Ma - Mi	Ma:Mi
2	R1=OEt	-9.08	-8.83	**-10.94**	27.39	24.06	25.87	27.11	8,8'-Unk	12:1
3	R1=OiPr	-4.62	**-8.06**	-6.04	25.80	25.51	25.45	31.64	8,8'-Unk	1:1
4	R1=OtBu	-10.02	-9.35	**-11.60**	23.33	21.82	23.17	23.42		
5	R1=OBu	-3.85	**-7.91**	-7.50	29.94	26.70	29.49	29.63	8,8'-Unk	1:1
6	R1=NH2	-9.41	-9.10	**-9.60**	21.08	20.78	24.17	28.94		
7	R1=NHMe	-9.74	-10.12	**-10.47**	25.84	26.05	26.36	34.63		
8	R1=H	-12.90	-10.31	**-14.03**	18.08	16.08	24.89	24.52		
9	R1=CH3	-10.89	-10.91	**-12.39**	18.76	17.14	20.44	20.70		
33	R1=OPr	-11.44	-10.01	**-14.91**	25.03	21.92	26.28	22.83	8,8'-Unk	3:1

Table 4.1: Comparison of the observed experimental connectivities and their calculated ΔH values. The major product for data 1 is 8,8' and it matches the lowest calculated ΔH values. The lowest ΔH of each row is in bold.

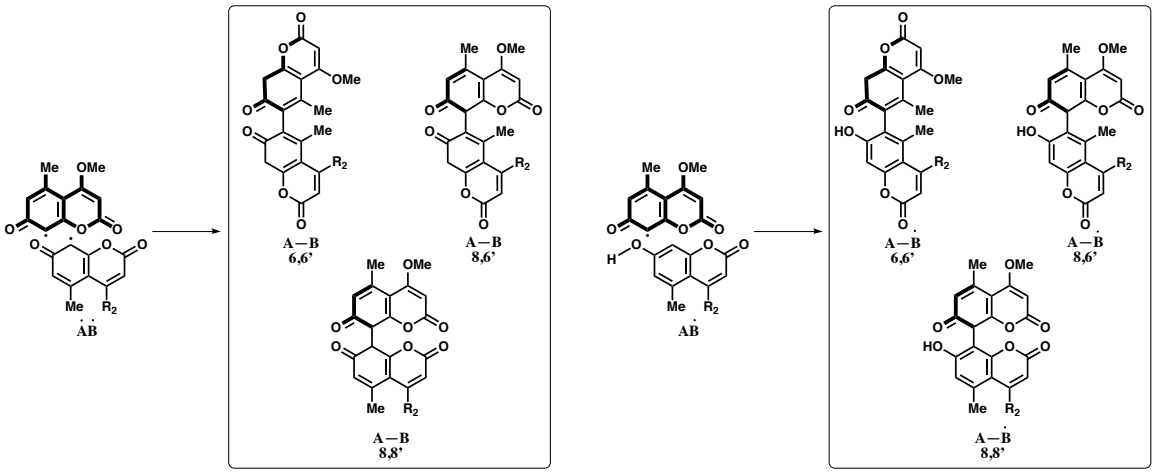


Figure 4.6: Connectivity of different products resulting from AB or AB˙ starting complexes. ΔH of each product is calculated for data set 1 where the column molecule is the native substrate and the row molecule has a varying C4-R group.

4.3.3.3 Predicting the Yield and Substrate Engineering

For datatsets 1, 2, and 5, multivariate linear regression could be used to predict the yield. The table below summarizes the Leave-1-Out (L1O) cross-validation results of C5-Me substituted yields for each of these sets. For substituents from the same group of chemicals, i.e. amines or ethers, the test score will increase from the reported one in Table 4.2, consistent with our ability to predict the behavior of substituents from the same chemical family. Another method for predicting the yield is using natural charge analysis of the activated complexes. In this case, the natural charge for atoms of each **A** or **B** partner is calculated and the highest correlation between $\log_{10}$(yield) and charge is selected. A summary of the correlation value and the atom index is given in Table 4.3 and Figure 4.7. We have not been able to explain the variation in atomic indices between different molecule classes.

Mol. Class	# of datapoints	Test R^2	Avg. coef. 1 (ΔH)	Avg. coef. 2 (size)	Avg. intercept	Avg. abs. error
1	8	0.738	52.66	-0.65	-10.72	4.75
2	7	0.777	-1.46	-0.29	51.66	2.28
3	6	-5.742	-0.33	-0.07	11.73	-11.0
5	7	0.558	-0.26	-	22.5	4.46

Table 4.2: C5-Me cross-validated predicted yield values.

4.4 Conclusions

Regio- and stereo-selective synthesis of chemical compounds has always been challenging using chemical synthetic methods, whereas nature has been able to selectively produce complex chemicals with high yield and selectivity. The problem is that enzymes that are nature's way for synthesis are evolved to catalyze a specific substrate or class of chemicals. In this work, a variety of substrates are screened for their reactivity in an attempt to develop KtnC, a cytochrome P450 enzyme, as a general catalyst for cross-coupling of biaryls.

Detailed mechanistic investigation of KtnC enzyme is challenging due to limited information available on the structure and geometry of the enzyme's active site. The crystal structure for the active site is not available yet because part of enzyme's protein chain is embedded in the living organism's membrane. As a result, performing computational kinetic studies and modeling the reaction is not feasible at this point. So, the linear free energy relationships are used to find a relationship between sterics and electronic properties of substrates, and reactivity.

There are multiple reaction mechanisms proposed for this class of enzymes as shown in Figure 4.3. Our results suggest that the operating mechanism is mechanism 4, where a di-radical is formed through double H-abstraction and the activated di-radical complex is combined to form the desired product. Based on the results, the enzyme plays a role in pre-organizing the substrate pair into an activated complex that makes bond dissociation easier for both substrates involved in the reaction. The governing rules for each molecule set are slightly different, meaning the dependency on size or BDE is not the same across the classes. There is not enough evidence to identify the rate limiting step for each molecule class, but there is high correlation (R^2 = 0.71) between AB$^{\cdot}$ $\longrightarrow$ A$^{'}$B$^{'}$ BDE step and the yield in molecule class V. Using BDEs and sterimol parameters for size, it is also possible to predict the yield within a reasonable accuracy for compounds that are structurally similar, meaning they have only one varying substituent. Obtaining more conclusive results on

Molecule Class	Highest R^2	Atomic index	A or B
1	0.802	C5	B in AB$^{\cdot}$
2	0.821	O1	B in A$^{\cdot}$B
3	0.895	C5	B in A$^{\cdot}$B$^{\cdot}$
5	-	-	-

Table 4.3: Highest correlation between charge on at atom and yield.

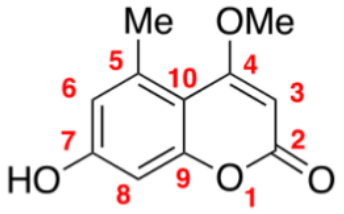


Figure 4.7: Numbering of the coumarin structure.

the rate limiting step for each molecule class, effect of heteroatoms in the aromatic biaryl structure, and the role of each substituent requires having access to a larger number of experimental reactivity results.

In the current experimental setting (whole cell experiments), performing mechanistic studies is not possible because such studies requires having access to the purified enzyme for in vitro investigations, either isolated from yeast or expressed in E. Coli. With that, the crystal structure of the enzyme could be obtained and QM/MM simulations could provide more concrete evidence on the operating reaction mechanism. Experimentally, mutagenesis of amino acid residues in the enzyme provides information on which residues are catalytically important and why. Combining theoretical and experimental results would definitely clarify and explain the details of biaryl coupling reaction mechanism.

CHAPTER 5

Conclusions

5.1 Summary of Contributions

While computational reaction discovery has made great progress in the case of organic and inorganic small molecules due to their relatively small number of atoms, it is still challenging and expensive to study surface reactions computationally due to the number of electrons and atoms in these systems. Additionally, the high coordination number and multiple binding sites of surface atoms combined with multiple potential adsorption sites in molecules results in multiple starting structures that should be optimized and screened for reaction path finding. The limitations of DFT methods for the study of periodic systems coupled with multiple binding possibilities in surface structures makes exploring heterogeneous systems a complex task.

The ultimate goal of reaction discovery methods is to guide experiments, rather than being a complementary tool for explaining experimental observations. Developing computational methods capable of doing so requires designing and implementing methods that can overcome the complexity of PES by performing efficient, reliable, and thorough searches on designated areas of the PES. However, the widespread adoption of computational reaction discovery methods, especially in heterogeneous catalysis, is currently obstructed by two major hurdles. First, methods that can perform systematic searches in the desired chemical space with minimal input from the user are not available. Second, RP finding methods for heterogeneous catalysis are expensive due to inefficiencies in their design and the large number of atoms in such reactions. Even when efficient computational methods are available, there are cases where simulating the reaction mechanism using RP finding tools is not feasible due to lack of information on the initial chemical structure of the reactants or the catalyst. Information gathered from quantum mechanical studies provides a means to shed light on mechanistic details in these situations. The efficient treatment of first and second problems and devising new methods for the third case are the main areas

to which this dissertation has contributed. This dissertation presents two reaction path finding methods that can operate with minimal information from experiments. Surface-GSM is a method for finding a RP, transition state structure, and reaction product is one single run. Surface-ZStruct is the automation of surface-GSM where reactant alignment, geometry optimization setup, binding site and adsorption angle sampling, and RP finding is done automatically by the program.

A more efficient and reliable reaction path finding method for the study of surface reactions In Chapter 2, algorithmic details for implementing the surface growing string method (surface-GSM) is detailed. This method is built upon molecular-GSM, but uses a completely different coordinate system. The surface reactions pose a challenge in designing a coordinate system because they are a combination of high coordination transition metals with low coordination main-group elements. While including the surface atoms in the coordinate system is important due to their role in supporting the structure of the slab and adsorbate molecules, it is also not possible to convert them to internal coordinate system. Treating the whole system using internal coordinates will impose a high computation burden and make the conversion from Cartesians to internals and vice versa and geometry optimization very slow and inefficient. To overcome this problem, a new coordinate system that is a combination of Cartesians and internal coordinates is designed and implemented. In this new system, surface atoms that are not involved in the reaction are treated using Cartesian coordinates while those attached to adsorbate molecules are included in internals. Adsorbate atoms are also treated in internals as conventionally done. To ensure efficiency and reliability, surface-GSM was benchmarked against nudged elastic band on a test set of more than 40 elementary step reactions. The results indicate surface-GSM is at least 2 times faster than nudged elastic band and converges even in cases that nudged elastic band fails.

Uncovering Reaction Sequences on Surfaces through Graphical Methods Even with efficient reaction path and transition state finding methods, sampling different conformers or adsorption orientations, preparing geometry optimizations, and setting up different reaction path finding calculations is cumbersome, when performed manually. Additionally, it is possible to miss some important pathways because they are not obvious based on chemical intuition. This motivates implementation and testing of a new method for automated reaction discovery of surface reactions. The surface-ZStruct is tested for the study of unimolecular and bimolecular reactions, where the propanoic acid dissociation pathway is an example of unimolecular reactions and is completely elucidated using this new

method. The resulting reaction network is compared with numerous studies on this reaction to confirm accuracy. Another studied reaction is atomic layer deposition of titanium nitride, where ammonia and tetrakis(dimethylamido)titanium (TDMAT) react on Si(100) surface to deposit TiN and extrude dimethylamine and other gaseous products. Repeating the following 5 steps results in TiN deposition: Step (1) Addition of ammonia to replace surface hydroxides by NH_x species, with concurrent release of H_2O. Step (2) Deposition of titanium via ligand exchange between NH_2 and TDMAT, extruding dimethylamine gas. Steps (3) and (4) Bonding of Ti to neighboring NH_2* groups to form N bridges that are observed in the TiN unit cell. Step (5) Ti reduction and addition of ammonia to the Ti-terminated surface and preparing it for addition of further TDMAT species. The complete reaction cycle for TiN ALD was successfully uncovered using surface-ZStruct.

Multivariate Linear Analysis of Thermodynamic and Steric Parameters for Characterizing KtnC Catalyzed Biaryl Coupling Modeling enzymatic reactions is a complex task because there are multiple factors at play in biological systems and the geometry of the active site, the protein scaffold, and solution parameters are important factors affecting the outcome of modeling studies. In this Chapter, the reactivity and site-selectivity of KtnC enzyme was investigated using the experimental yield values and structures of substrates involved in the reaction. KtnC is a cytochrome P450 enzyme that catalyzes the homocoupling of 7-demethylsiderin (native substrate) in nature with high yield and stereoselectivity. The goal is to develop KtnC as a general catalyst for cross-coupling of any desired aromatic compound. The objective of the first stage is to understand the reaction mechanism, factors controlling the site-selectivity, and to predict the yield. This information could later be used to manipulate reaction conditions or substrates to achieve high yield and selectivity for other biaryls. Two substrates are organized in gas-phase and are held together by dispersion interactions to form an activated complex. Bond dissociation energies of the complex and the size of substituents on coumarin structures are measured and correlated with experimental yields. The results show a higher correlation between the enthalpies of di-radical pathway and the yield compared to enthalpies of other suggested mechanisms. Also, the site-selectivity for cross-coupling of the native substrate with other coumarins could be predicted using ΔH of different C$-$C bond formations.

5.2 Future Directions

Surface growing string method has shown to be very efficient and reliable compared to similar competing methods, three most important advantages of the growing string method

are: 1) its speed 2) its ability to find a transition state, reaction product, and reaction path in one single run and 3) its ability to find a reaction path and product by just knowing the starting structure. Despite these advantages, growing string method could benefit from code optimization to make it even faster. Also, multiple versions of the code available for different types if chemical processes could be merged into one single well-written program for maintenance and distribution purposes. In terms of chemistry, the algorithm could be extended to potential bias studies to model and investigate electrochemical reactions with significance in many industrial processes such as coating, corrosion, photosynthesis, fuel cells, and many other important electrochemical transformations.

Another potential direction for the work presented in this dissertation is improving the surface-ZStruct method. ZStruct methods are great tools for exploring chemical networks, especially the reactions that are not well-studied. However, computing hundreds of reaction paths even for small systems requires powerful computational resources. Another shortcoming is the number of duplicate and failed reaction paths calculated by ZStruct relative to productive reaction paths. The reaction discovery also needs to be set up for each new intermediate, in a step-by-step procedure. Finally, analyzing hundreds of reaction paths for meaningful chemical insight is a tedious process manually. To address the first two problems, incorporation of data on thermodynamics and kinetics of similar reactions might help to reduce the pathways that need to be calculated and thus the number of failed reaction path computations. Another future direction could be a new program that analyses ZStruct outputs for changes in connectivity and compares new structures with desired products based on the driving coordinates and also compares the final geometry with a database of typical chemical intermediates and activation energies to prune unproductive pathways.

The results of Chapter 4 could be greatly improved if more datapoints for various substrates with only small variations in one structural parameter were available. For this study, the role of each substrate pair and their interaction to activate each other needs to be understood. In addition, the effect of each R group attached to the coumarin scaffold and its effect on sterics and electronics of the substrate and the other partner should be investigated. This information could provide enough evidence to describe the reaction mechanism in full detail.

APPENDIX A

Supplementary Information for Chapter 2

A.1 Hybrid coordinate system

The B matrix in primitive coordinates is formed using standard techniques [16]

$$B_{ij} = \frac{\partial q_i}{\partial x_j} \tag{A.1}$$

$$\Delta q = B\Delta X \tag{A.2}$$

where q are the primitive (hybrid) coordinates and X are Cartesians. The G matrix is formed and diagonalized as described below to produce a set of 3N non-redundant (linearly independent) vector space, U.

$$G = BB^{\top} \tag{A.3}$$

$$G(U \;\; R) = (U \;\; R)\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} \tag{A.4}$$

The B matrix in non-redundant (NR) (hybrid) coordinates is formed based on

$$B^{NR} = U^{\top}B \;,\; U \in \mathbb{R}^{3N} \tag{A.5}$$

For the constraint optimization, the constraint vector, U_C, is formed by projecting the unit vector C corresponding to the constant primitive coordinates onto the full non-redundant subspace

$$U_C = \sum_{k=1}^{3N} \langle C|U_k\rangle U_k \tag{A.6}$$

The constraint vector, U_C, is normalized and the set V with 3N+1 vectors is formed by concatenating vectors U_k and the vector U_C

$$V = \{U_C, U_k; k = 1, \cdots, 3N\} \tag{A.7}$$

Schmidt orthonormalization is carried out to form a new set (V^*) with 3N-1 vectors U_k and the vector U_C

$$V_k^* = \alpha_k \left(V_k - \sum_{l=1}^{k-1} \langle V_k | V_l^* \rangle V_l^* \right) \tag{A.8}$$

where α_k is a normalization constant, V_k are the vectors from the set V, and vectors V_l^* compose the new orthonormal basis, V^*. [15, 235]

A.2 Hessian construction and update at each node

An initial diagonal Hessian in primitive coordinates is constructed from bonds, angles, and torsions and maintained to build a new non-redundant coordinates Hessian after each update and reparameterization step. This procedure is enforced because non-redundant coordinates change as reparameterization proceeds. The non-redundant coordinates Hessian, H, at each node is created by applying change of basis to the Hessian in primitive coordinates (H^{prim})

$$H = U^\top H^{prim} U \tag{A.9}$$

where U is the non-redundant coordinates matrix. Both Hessians are updated using the BFGS [36, 68, 77, 189] scheme

$$\Delta H_{BFGS} = \frac{\Delta g \Delta g^\top}{\Delta g^\top \Delta x} - \frac{H_{i-1} \Delta x \Delta x^\top H_{i-1}}{\Delta x^\top H_{i-1} \Delta x} \tag{A.10}$$

where H_{i-1} is the Hessian of the previous step, and Δg and Δx are changes in current and previous gradient and coordinates, respectively. Note that Δg and Δx are in their respective non-redundant coordinate or primitive coordinate basis for each corresponding Hessian matrix.

A.3 Hessian construction and update at TS node

After completion of CI, the exact TS search starts by constructing a Hessian with desired eigenvalue structure from TS node's existing Hessian. The curvature, C, along the reaction path at the TS node is approximated using the two neighboring nodes to estimate the TS eigenvector [7]

$$C = \frac{2E_{TS-1}}{a(a+b)} - \frac{2E_{TS}}{ab} + \frac{2E_{TS+1}}{b(a+b)} \tag{A.11}$$

where E_{TS-1} and E_{TS+1} are the energies of the nodes prior to and following the TS node, E_{TS} is the energy of the TS node, a is the distance between the TS and the previous node, and b is the distance to the following node.

This modification is applied by subtracting the curvature along the reaction tangent from C, and multiplying it by a symmetric matrix with proper size, $U_C U_C^\top$,

$$\Delta H = \left(C - U_C^\top H U_C\right) U_C U_C^\top \tag{A.12}$$

The new Hessian is updated using the Bofill [28] method, which allows negative eigenvalues

$$\Delta H_{Bofill} = \phi \Delta H_{MS} + (1-\phi)\Delta H_{PSB} \tag{A.13}$$

$$\Delta H_{MS} = \frac{(\Delta g - H_{i-1}\Delta x)(\Delta g - H_{i-1}\Delta x)^\top}{(\Delta g - H_{i-1}\Delta x)^\top \Delta x} \tag{A.14}$$

$$\Delta H_{PSB} = \frac{(\Delta g - H_{i-1}\Delta x)\Delta x^\top + \Delta x(\Delta g - H_{i-1}\Delta x)^\top}{\Delta x^\top \Delta x} - \frac{\Delta x^\top (\Delta g - H_{i-1}\Delta x)\Delta x \Delta x^\top}{(\Delta x \Delta x^\top)^2} \tag{A.15}$$

$$\phi = \frac{\left((\Delta g - H_{i-1}\Delta x)^\top \Delta x\right)^2}{|\Delta g - H_{i-1}\Delta x|^2 |\Delta x|^2} \tag{A.16}$$

where H_{i-1}, Δg, and Δx are the same variables as described for Equation (A.10).

A.4 Additional Computational Details

The computational details for GSM are similar to those described in previous GSM papers. [234–236] There are slight differences in the convergence criteria for single-ended and double-ended GSMs during the growth phase. During the growth phase of SE-GSM, new nodes are optimized until the root mean square (RMS) gradient is below 10 times the convergence threshold (0.005 Hartree/Å), or 30 steps. In this method, the number of nodes is not specified in the input and is actively determined during the growth phase, so two criteria are used to determine whether to continue growing: (1) the frontier node is lower in energy than the previous node (by a threshold value), or (2) the frontier node's constraint gradient is positive. In DE-GSM, at most two nodes can be added during the growth phase, and each node undergoes two optimization steps per optimization iteration. The new nodes are added only when the perpendicular gradient magnitudes on the frontier nodes fall below 0.1 Hartree/Å. For double-ended methods (DE-GSM & CI-NEB), the number of nodes is

predefined as part of the input parameters.

In both single-ended and double-ended GSM, the CI search commences after the sum of the perpendicular gradient magnitudes over all nodes, F, is converged to $F < 0.3$ Hartree/Å . The reaction path is considered converged when RMS gradient on the TS node is below 0.0005 Hartree/Å. The exact TS search is initiated when one of the three sets of conditions is met: 1.) the total gradient is < 0.2 Hartree/Å, the TS node is converged to within 10 times the nodal convergence tolerance, and the constraint force is < 0.01 Hartree/Å, 2.) the total gradient is < 0.1 Hartree/Å, the TS node is converged to within 10 times the convergence tolerance and the constraint force is < 0.02 Hartree/Å, or 3.) the TS node is within five times the convergence tolerance. In SE-GSM, after completion of the exact TS search and string convergence, the last node on the string which is the predicted reaction product is optimized in all directions (without constraint) to a local minimum.

A.5 DE-GSM vs. CI-NEB RPs

RPs for Reaction 6-c is shown in Figure A.5c in detail. As shown in the snapshots (4) to (8), the two swapping Cu atoms collide on the interpolated path which leads to CI-NEB's failure immediately after the initial interpolation for placing the nodes. This issue arises from linear interpolation in Cartesians and CI-NEB's strategy in node placement while the growing nature of DE-GSM prevents this problem. In the path calculated by DE-GSM (Figure A.5b) the top left translucent Cu atom passes over the other moving Cu atom and avoids collision.

Similar to Reaction 6-c, Reaction 6-b has a problem with the initial interpolation in NEB, as shown in Figure A.4c. Here, a Cu atom moves from within the surface to a hollow site on the top layer. Instead of moving upward then sideways, NEB's interpolation places the Cu atom between two surface Cu atoms, resulting in a problematic initial RP that never recovers to reach a realistic path. Specifically, optimization from this initial condition leads to deformation of the slab (Figure A.4b, snapshots (3) and (5)), which is not desired or required for this reaction. Improvements to the interpolation algorithm in NEB, for instance linear and quadratic synchronous transit, [20, 84, 166] may be able to remedy these issues. DE-GSM, however, takes advantage of curvilinear interpolation for placing nodes and incremental node addition, and is able to refine path 6-c without a challenge (Figure A.4a).

In Figure A.3, the maximum difference in activation energies occurs for water dissociation on W(111) (15-a), which differs by 4.6 kcal/mol between DE-GSM and CI-NEB. To understand why the barriers are so different, the corresponding RPs are shown in Fig-

ure A.6b and the TS structures quantitatively compared in Table A.2. While many aspects of the O−H dissociation are similar between DE-GSM and CI-NEB, two unique reaction paths were found. One proceeds by directly moving the transferring hydrogen to its final binding site and the other moves hydrogen in a half-circle before reaching its final location. The disagreement in activation barriers between DE-GSM and CI-NEB, therefore, is not due to failure of either method. In general, there is no guarantee that double-ended methods will converge to the same RP and TS, though it appears to be a relatively infrequent event in our test set.

A representative example of reactions that proceed through very different pathways when calculated by DE-GSM or CI-NEB is reaction 5-b (H diffusion in Ni(111) from fcc site to second layer) which is an example of hydrogen embrittlement in metals (Figure A.7). Fracturing in metals due to hydrogen adsorption and then its diffusion into the bulk material is particularly important in high strength steels and nickel fabrication [195]. In this reaction, DE-GSM finds a path that involves diffusion of the H atom first directly into the surface (Figure A.7b, snapshots (1) to (6)), and then its migration from fcc to hcp site underneath the top slab layer. The calculated TS for this reaction has an activation energy of 18.5 kcal/mol. The path calculated by CI-NEB first moves the H atom from fcc to hcp site over the top slab layer and then moves it in between the first and second layers to end up at the same product structure as DE-GSM. In cases with two distinct RPs from different methods, increasing the number of nodes in order to achieve a more converged path did not result in similar paths or activation barriers.

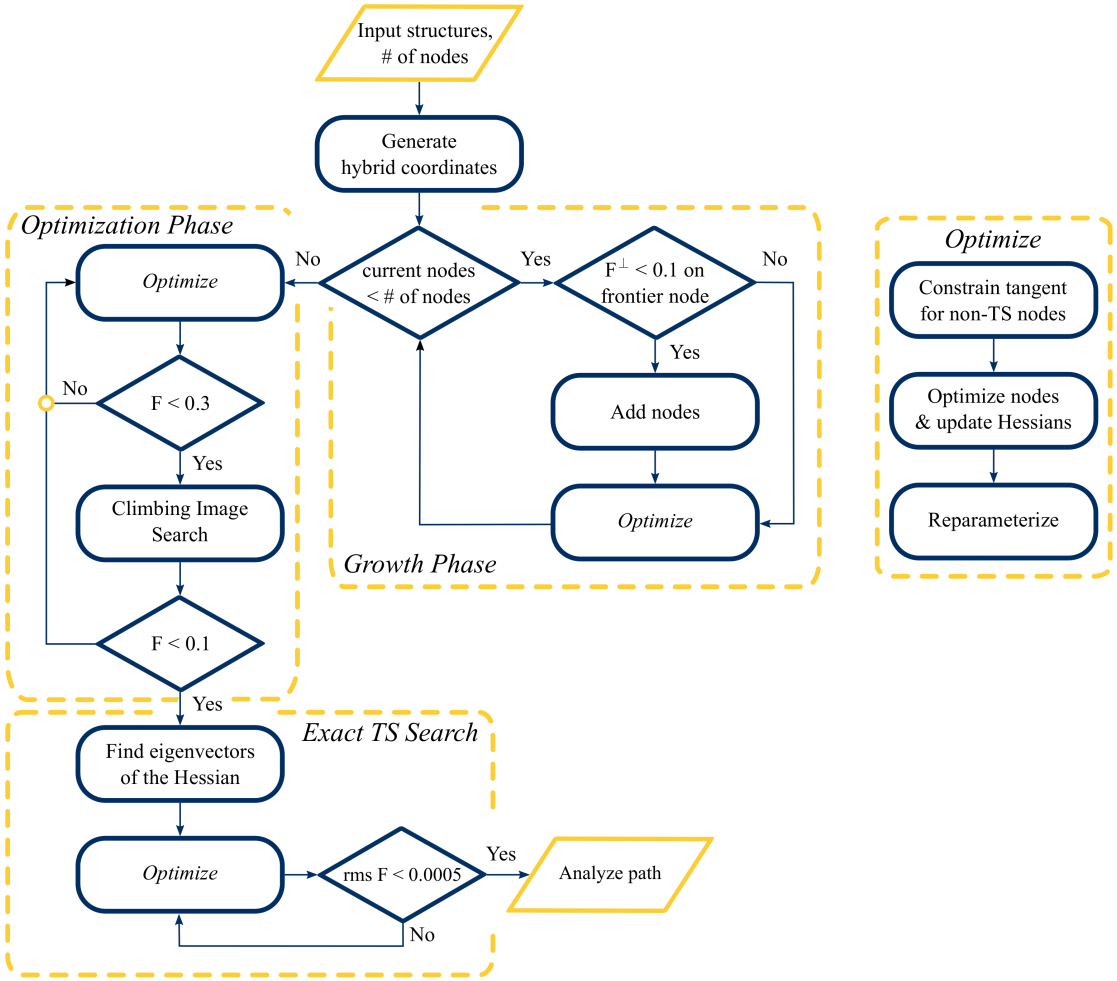


Figure A.1: Process flow for DE-GSM.

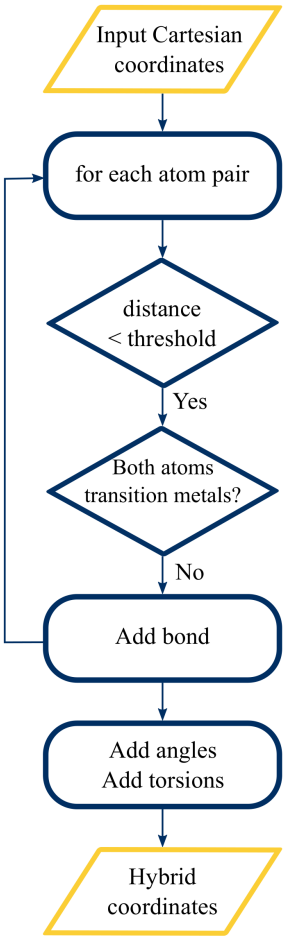


Figure A.2: Process flow for generation of hybrid coordinate system. Threshold value is determined based on atomic radii of atoms.

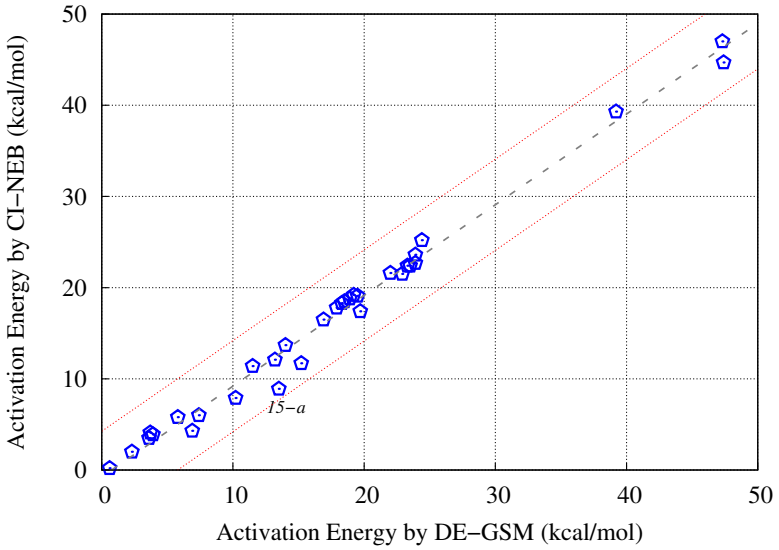
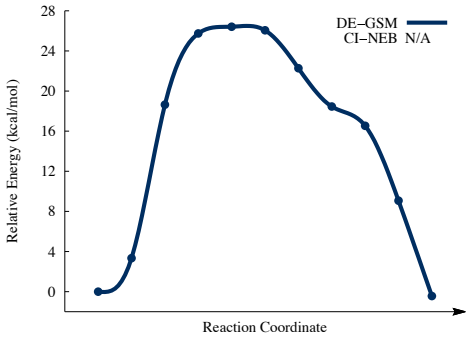
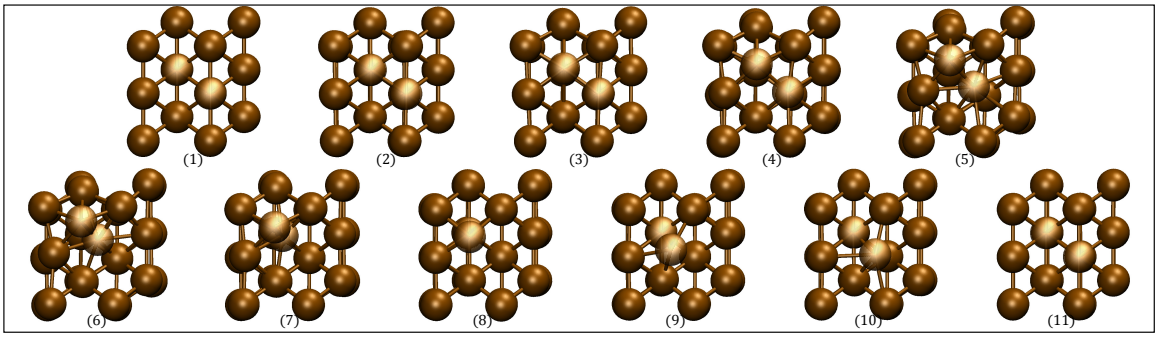


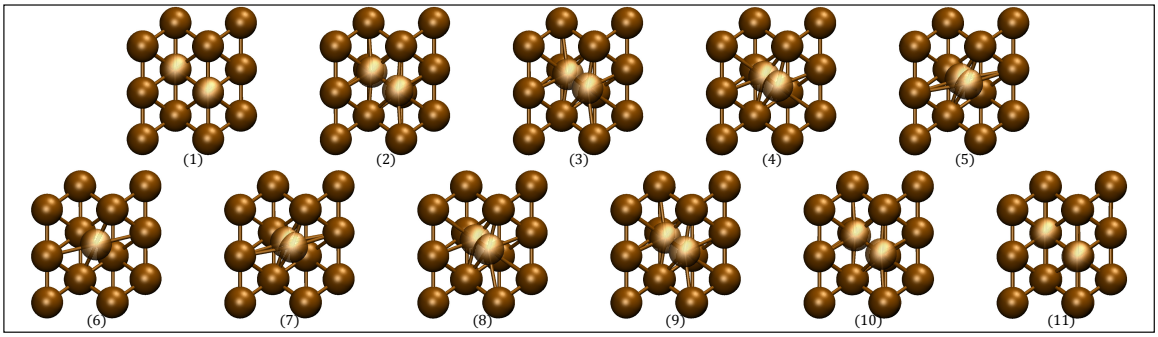
Figure A.3: Comparison of the activation energies calculated by DE-GSM and CI-NEB methods. The area between the two dotted red lines confines ±5 kcal/mol deviation from the best fitted line ($R^2 = 0.989$ and $y = 0.995x - 0.75$).



(a) Reaction path calculated by DE-GSM (blue) for Cu atoms swapping on Cu(110).

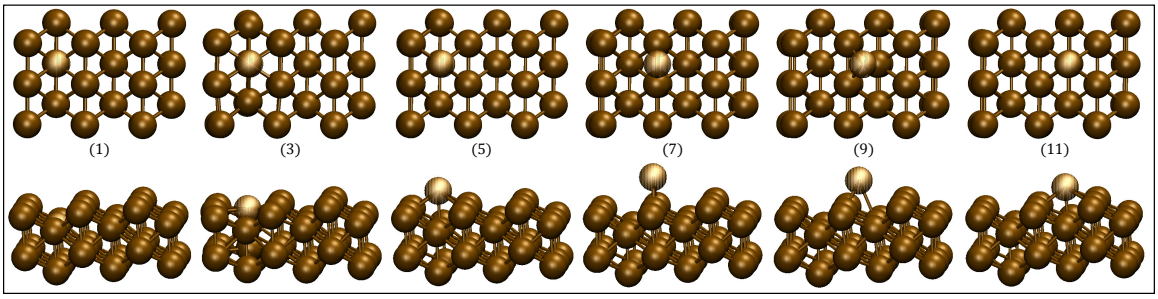


(b) Snapshots of Cu atoms swapping on Cu(110) calculated by DE-GSM. Number 5 is the transition state. The top left translucent atom moves over the other atom during the reaction.

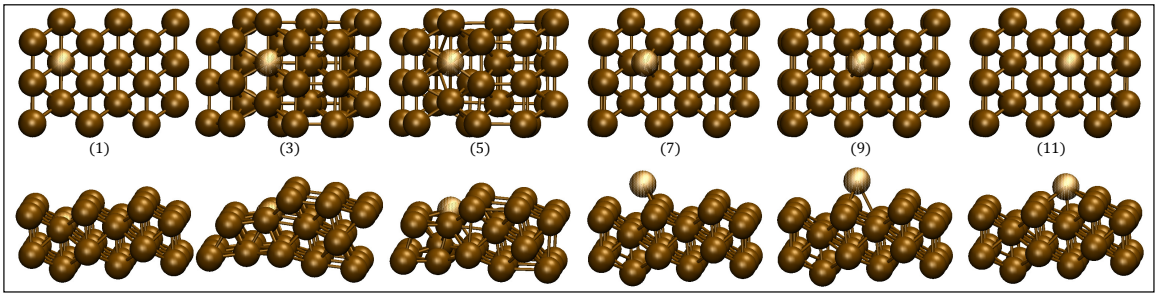


(c) Snapshots of Cu atoms swapping on Cu(110) calculated by CI-NEB. Calculation fails immediately after the initial interpolation due to the collision of the Cu atoms (image number (6)).

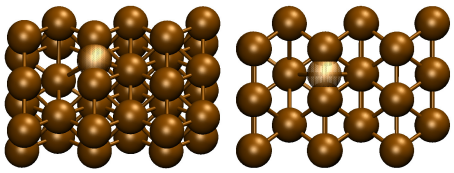
Figure A.5: Reaction 6-c. Reaction paths for swapping of Cu atoms on Cu(110).



(a) Snapshots of a Cu atom moving on Cu(110) calculated by DE-GSM. Number 5 is the TS.

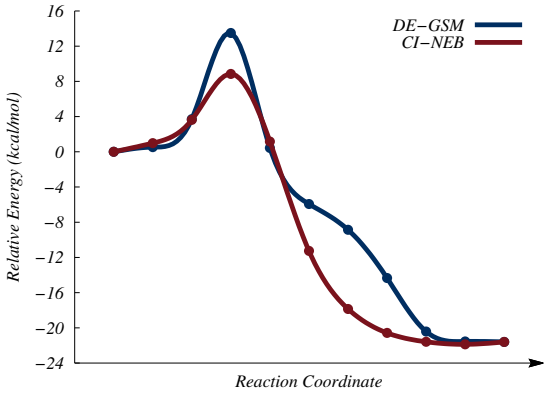


(b) Snapshots of a Cu atom moving on Cu(110) calculated by CI-NEB. The calculated path involves deformation of the slab.

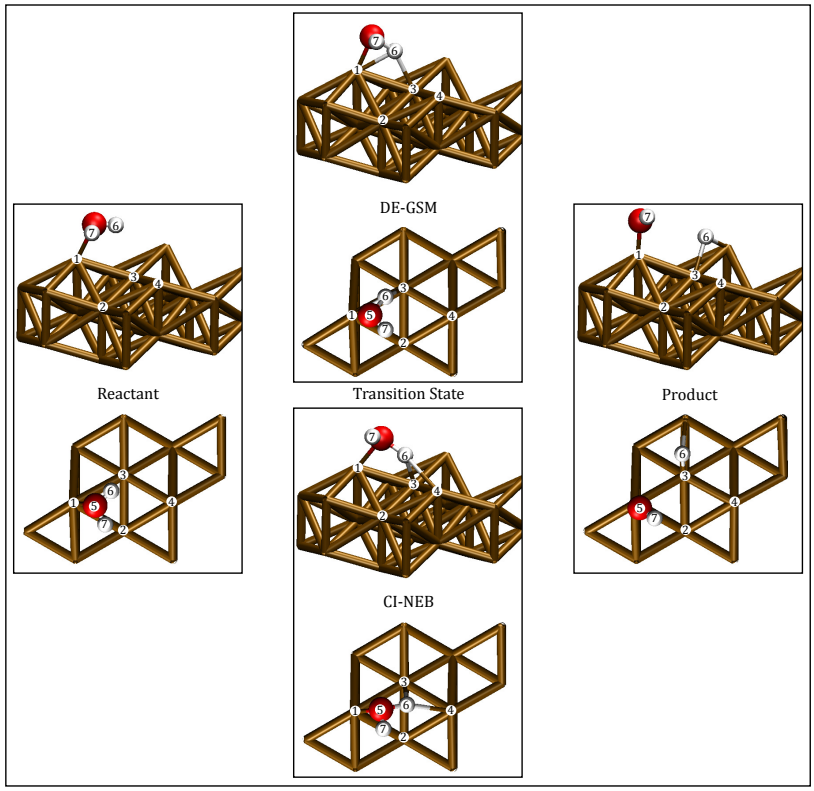


(c) Moving Cu atom intersects the bond between two other surface atoms in predicted TS along the initial interpolated path by CI-NEB.

Figure A.4: Reaction 6-b. Reaction paths for a Cu atom diffusing to a hollow site on Cu(110).

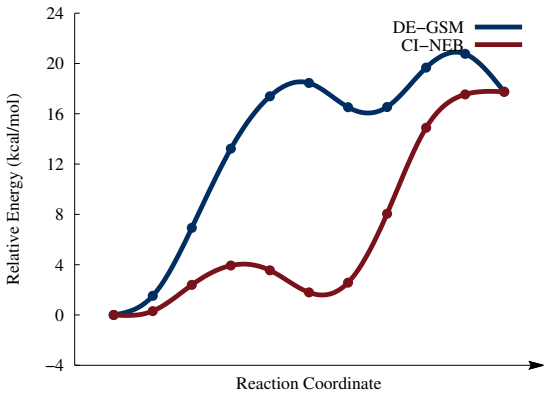


(a) Reaction paths calculated by DE-GSM (blue) and CI-NEB (red) for dissociation of H_2O atom on W(111).

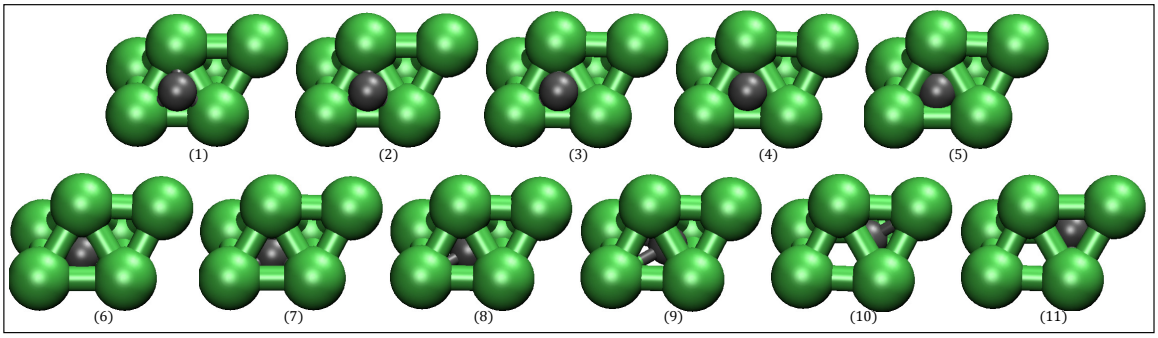


(b) Reactant, TS, and product structures for reaction 15-a calculated by DE-GSM (top) and CI-NEB (bottom).

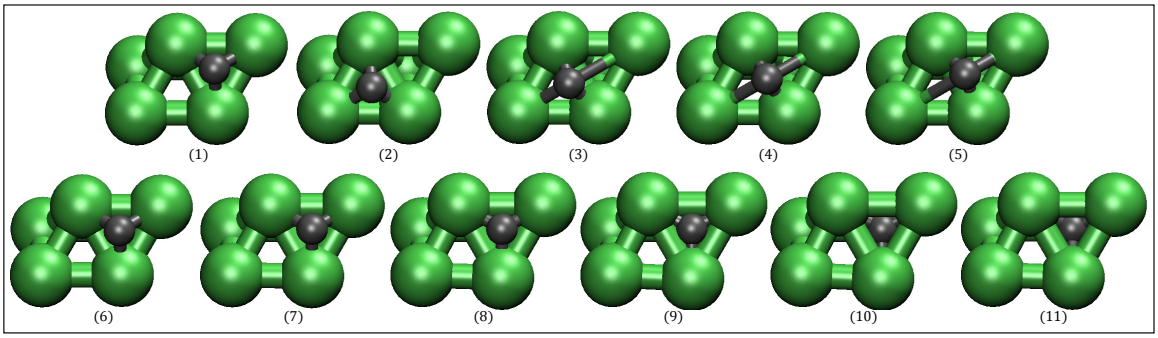
Figure A.6: Reaction paths for dissociation of H_2O on W(111).



(a) Reaction paths calculated by DE-GSM (blue) and CI-NEB (red) for H diffusion in Ni(111).



(b) Snapshots of H (gray) diffusion in Ni(111) calculated by DE-GSM. There is a TS for diffusing from fcc to hcp site and another TS (node number 10) for diffusion beneath the first latyer.



(c) Snapshots of H (gray) diffusion in Ni(111) calculated by CI-NEB. The H atom first diffuses on the surface and there is no barrier for diffusion from hcp site to the layer below.

Figure A.7: Reaction 5-b. Reaction paths for diffusion of H atom in Ni(111).

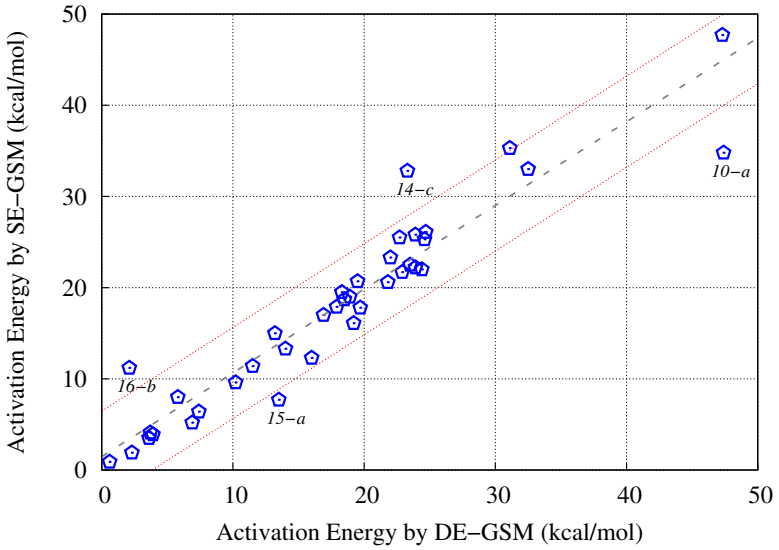


Figure A.8: Comparison of the activation energies calculated by DE-GSM and SE-GSM methods. The area between the two dotted red lines confines ±5 kcal/mol deviation from the best fitted line ($R^2 = 0.875$ and $y = 0.899x + 1.92$).

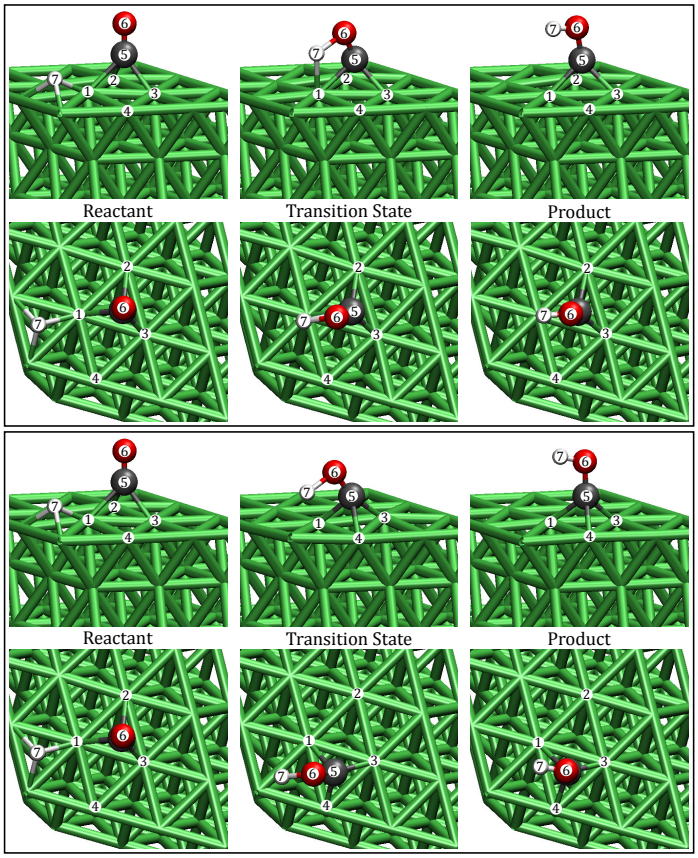


Figure A.9: Reactant, TS, and product structures for reaction (10-a) calculated by DE-GSM (top) and SE-GSM (bottom). The CO molecule is not stationary in the case of SE-GSM.

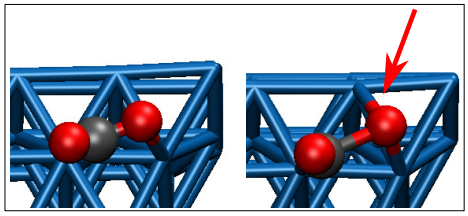


Figure A.10: Complexes formed right before TS for SE-GSM (left) and DE-GSM (right). Note the asymmetric bond cleavage in SE-GSM case.

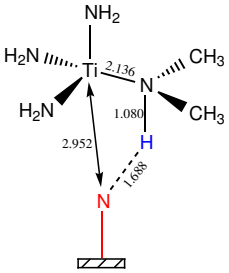
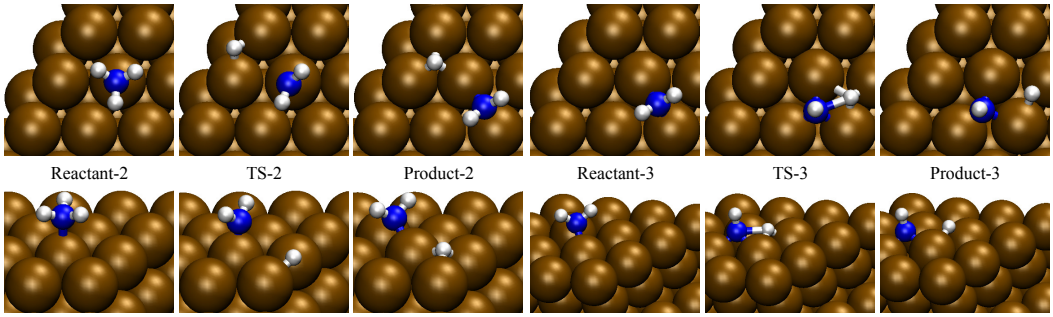
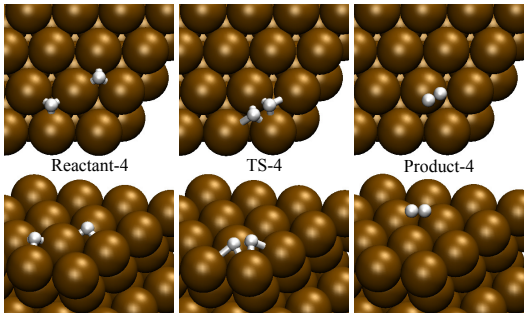


Figure A.11: TS structure of reaction T5. Bond lengths are in Ångstroms.



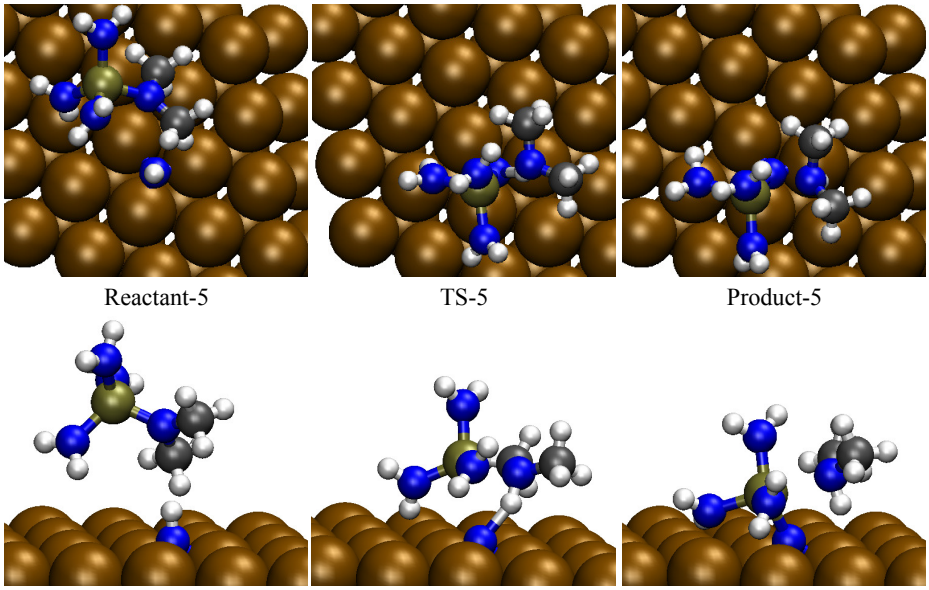
(a) Reaction-T2. Dissociation of NH_3.

(b) Reaction-T3. Dissociation of NH_2 to NH and H.

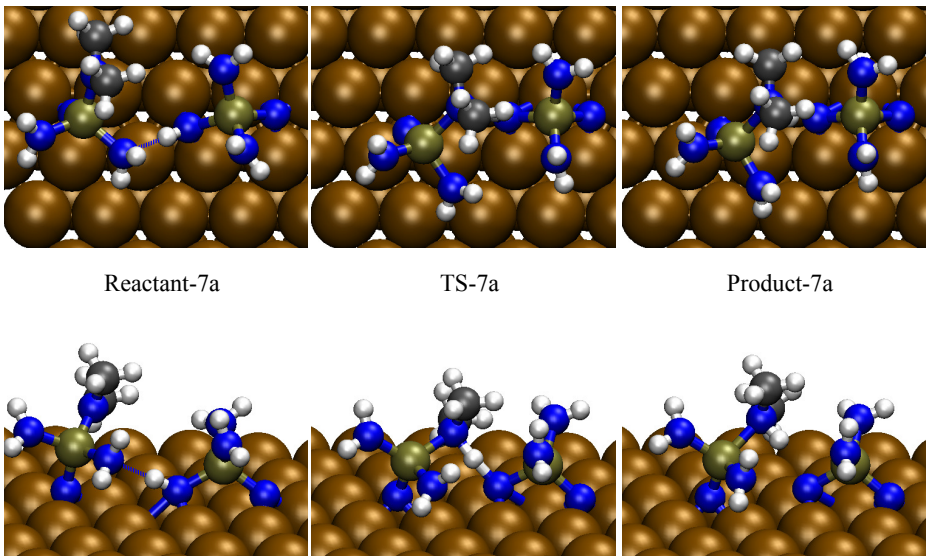


(c) Reaction-T4. Formation of H_2 from two H atoms.

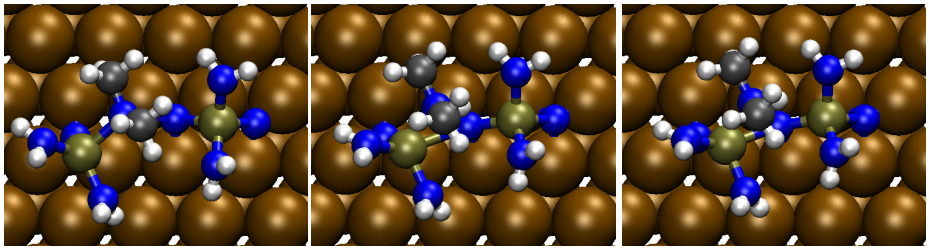
Figure A.12: Reactants, TSs, and products of the first deposition cycle. Atoms are N (blue), H (white), C (gray), Ti (tan green), and Cu (ochre).



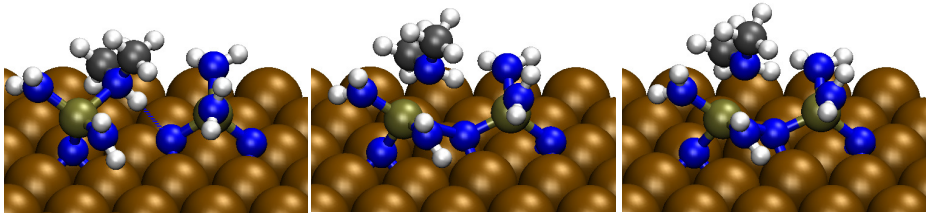
(a) Reaction-T5. Adsorption of first TDMAT molecule on the surface through ligand-exchange with NH*.



(b) Reaction-T7a. Following reaction T6, the second tris(dimethylamido)titanium species reacts with the remaining H on N to form a new atomic layer. This reaction takes place in two steps, the first step which is reaction (T7-a) proceeds through a hydrogen-transfer from NH_2* to dimethylamido ligand.

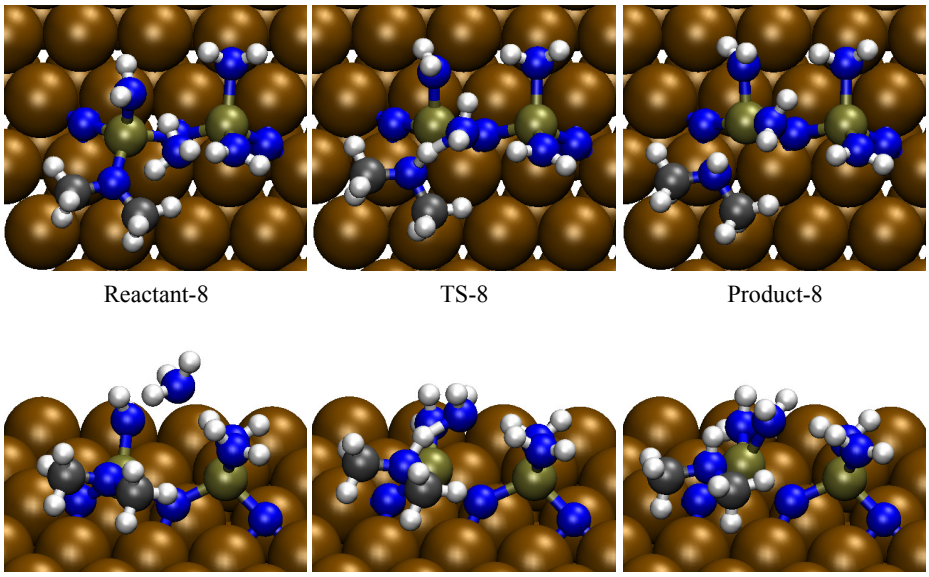


Reactant-7b TS-7b Product-7b

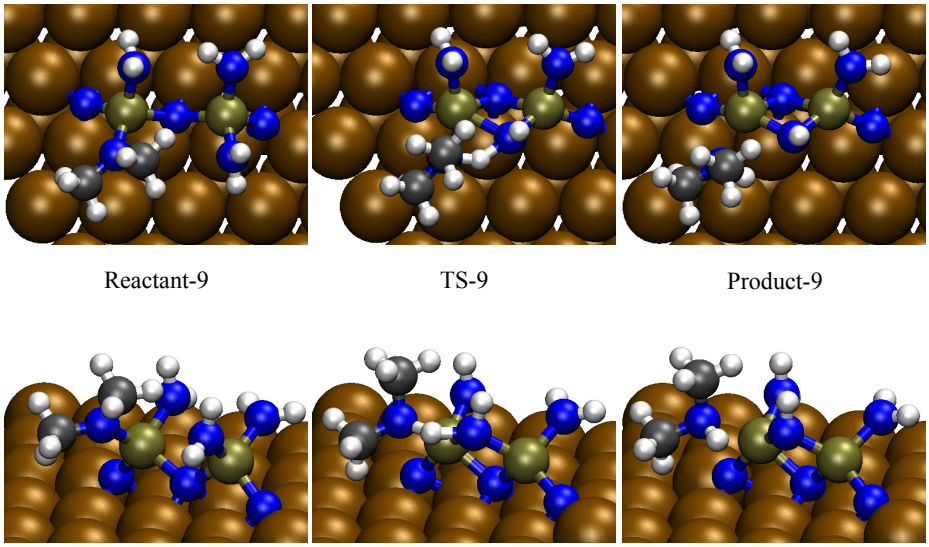


(c) Reaction-7b. In the second step of reaction (T7), dimethylamine desorbs from the surface while the second Ti connects to N* in an associative ligand-exchange reaction.

Figure A.13: Reactants, TSs, and products of the second deposition cycle. Atoms are N (blue), H (white), C (gray), Ti (tan green), and Cu (ochre).

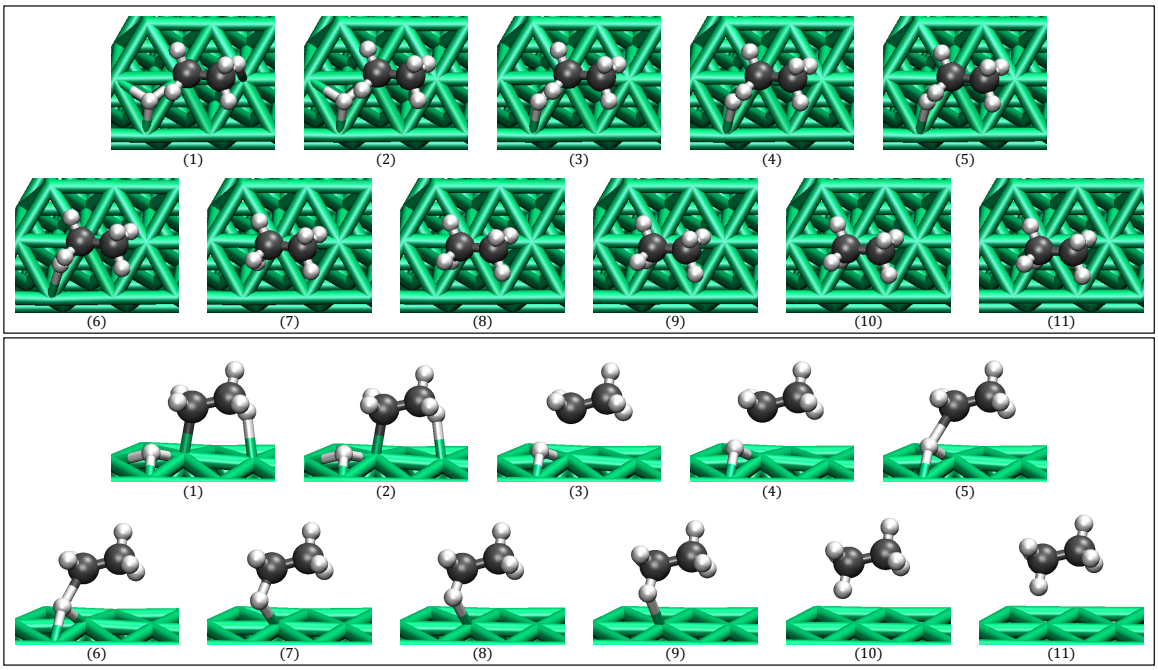


(a) Reaction-8. Hydrogen-transfer occurs between NH_3^* and dimethylamido ligand of titanium while NH_2^* connects to titanium through an associative process.

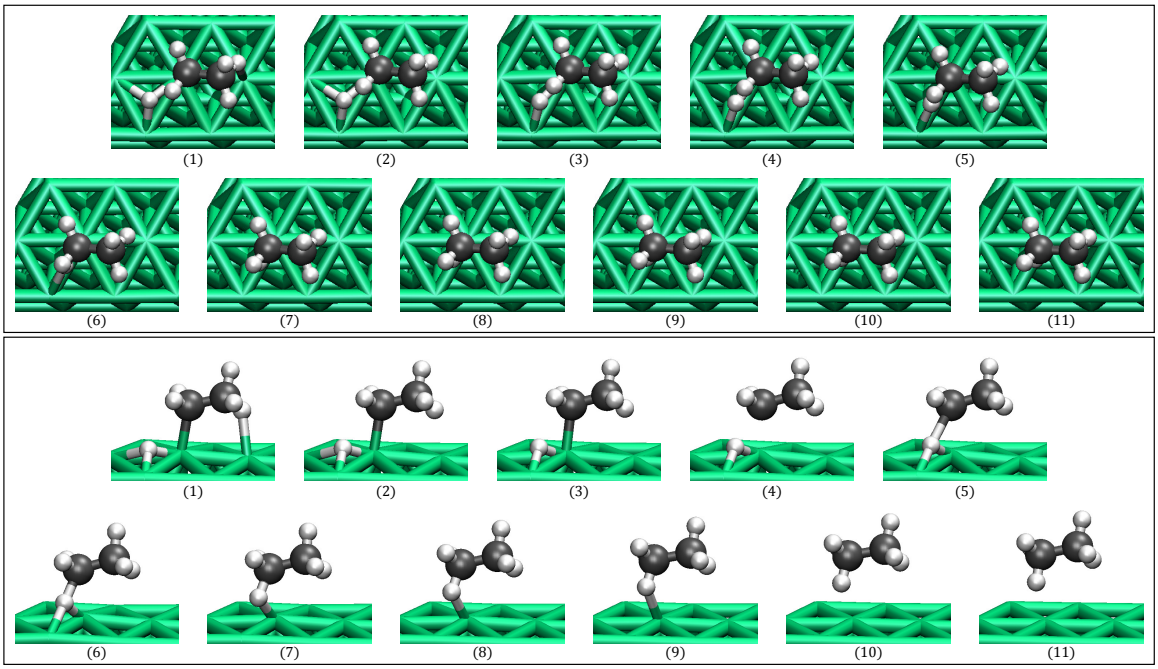


(b) Reaction-9. In this reaction, dimethylamido of one species reacts with amido of a nearby fragment which was replaced in reaction T8. The third atomic layer is formed and dimethylamine is released. The bridging NH can act as a binding site for the incoming TDMAT in the next TDMAT cycle.

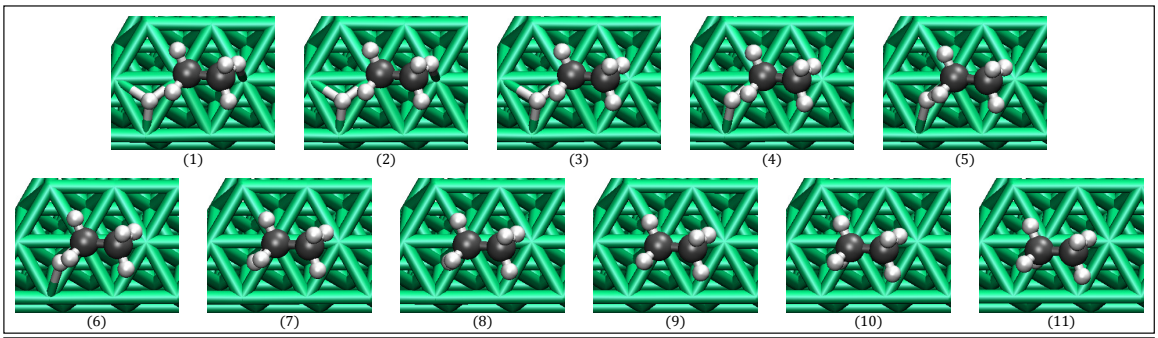
Figure A.14: Reaction paths for ALD of TiN on Cu(111). Atoms are N (blue), H (white), C (gray), Ti (tan green), and Cu (ochre).

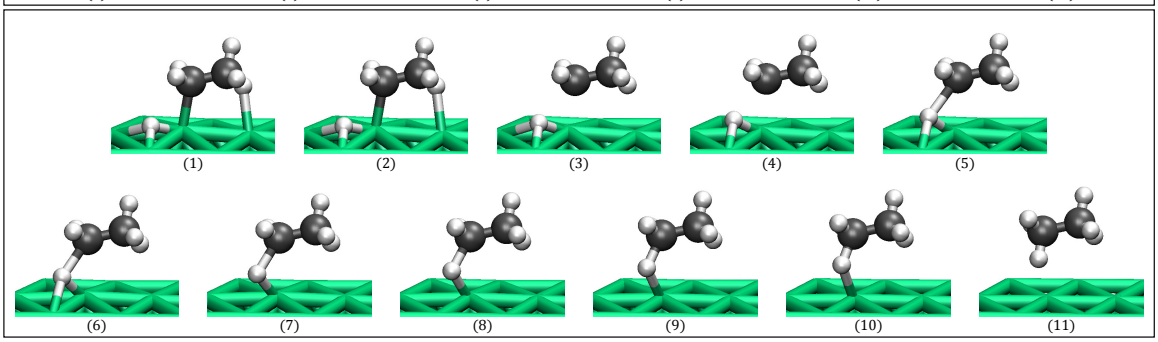


(a) Snapshots of the initial path of ethane formation on Pd(111) calculated by DE-GSM.

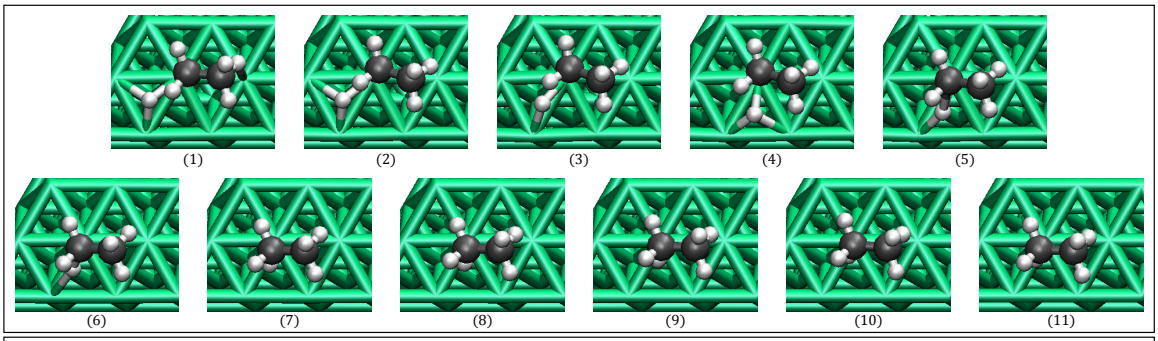


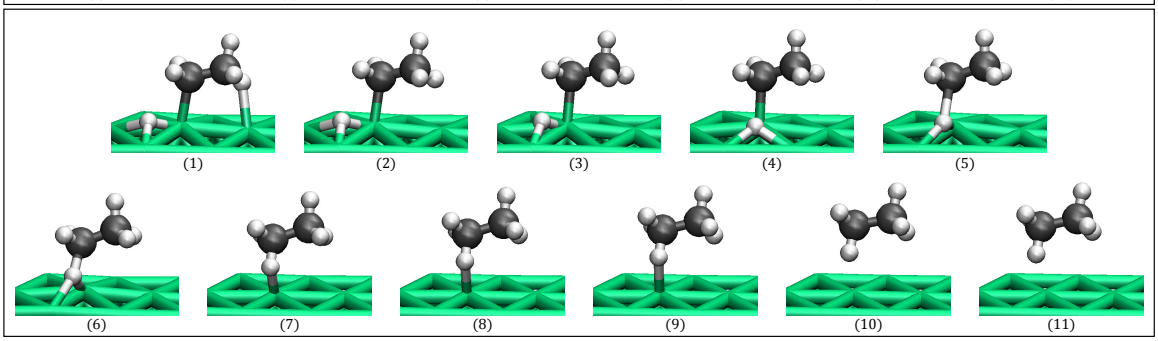
(b) Snapshots of the final converged path of ethane formation on Pd(111) calculated by DE-GSM.





(c) Snapshots of the initial path of ethane formation on Pd(111) calculated by CI-NEB.





(d) Snapshots of the final converged path of ethane formation on Pd(111) calculated by CI-NEB.

Figure A.15: Reaction 8-b. Reaction paths for ethane formation on Pd(111) from ethyl and hydrogen.

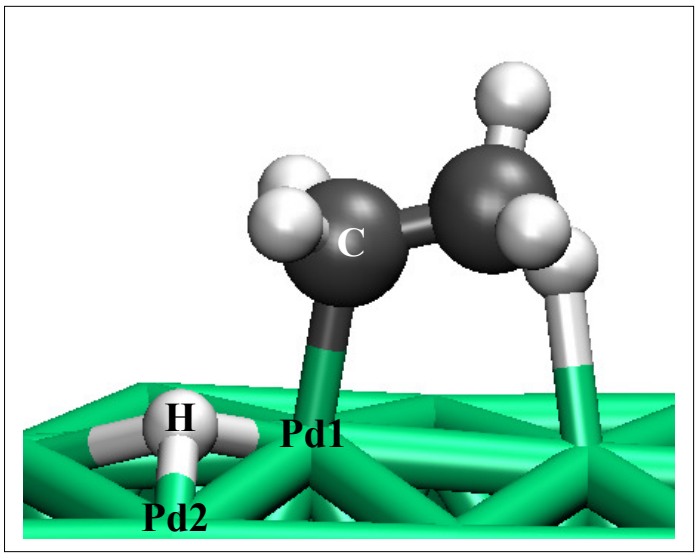


Figure A.16: Atomic labels for Table A.1.

	C-Pd1			H-Pd2		
Node	CI-NEB	DE-GSM	Diff. (GSM - NEB)	CI-NEB	DE-GSM	Diff. (GSM - NEB)
1	2.088	2.088	0.000	1.698	1.698	0.000
2	2.088	2.103	0.015	1.698	1.720	0.022
3	2.179	2.150	-0.029	1.776	1.767	-0.009
4	2.275	2.228	-0.047	1.864	1.846	-0.017
5	2.374	2.297	-0.077	1.959	1.917	-0.042
6	2.476	2.391	-0.084	2.061	2.023	-0.038
7	2.580	2.584	0.004	2.169	2.261	0.092
8	2.687	2.633	-0.054	2.282	2.377	0.095
9	2.795	2.782	-0.013	2.399	2.516	0.117
10	2.906	2.943	0.037	2.520	2.649	0.129
11	3.131	3.131	0.000	2.771	2.771	0.000

Table A.1: Bond lengths for initial and final RPs of Reaction 8-b calculated by CI-NEB and DE-GSM. The difference is measured by subtracting CI-NEB's value from DE-GSM's. All values are in Ångstroms.

		DE-GSM	CI-NEB	
Bonds (Å) & angle	Reactant	Transition State	Transition State	Product
W1-O5	4.684	2.087	2.115	1.891
O5-H6	0.994	1.358	1.347	3.343
O5-H7	0.986	0.981	0.982	0.972
W3-H6	5.297	2.118	2.115	1.906
W4-H6	3.504	4.950	2.389	4.805
W2-H7	3.570	6.157	6.153	6.664
∠ O5-H6-W3	103.979	136.967	114.971	85.325
∠ O5-H6-W4	121.762	131.643	144.183	91.637
∠ H6-O5-H7	104.059	95.586	102.853	96.772

Table A.2: Bond lengths and angles for the reactant, TSs, and product of Reaction 15-a calculated by DE-GSM and CI-NEB. Colored values indicate bonds and angles that are different in structures calculated by the methods.

	Reaction 4			
	TS1	Intermediate	TS2	Product
DE-GSM	39.2	32.8	48.0	39.8
SE-GSM	44.0	-	-	39.9

Table A.3: Energies of TSs, intermediate, and products of Reaction 4. All the energies are in kcal/mol and referenced to the reactant structure of each reaction.

A.6 Tables of number of gradient calculations and E_a

ID	DE-GSM	CI-NEB	SE-GSM
1	50	45	535
2	116	135	46
3-a	205	T	131
3-b	144	162	69
4-a	154	216	426
4-b	157	882	-
5-a	135	117	238
5-b	-	297	147
5-b-1	201	-	-
5-b-2	307	-	-
5-c	169	432	306
6-a	339	216	1282
6-b	-	F	F
6-b-1	277	-	-
6-b-2	180	-	-
6-c	639	F	F
7-a	826	T	1944
7-b	210	T	270
8-a	154	513	142
8-b	167	738	181
9-a	286	504	155
9-b	474	810	378
9-c	1572	459	336
9-d	1042	459	183
10-a	190	1422	257
10-b	932	1773	592
11-a	288	T	207
11-b	258	288	195
12-a	276	T	91
12-b	260	T	158
13-a	172	342	123
13-b	124	342	141
14-a	144	549	112
14-b	175	675	139
14-c	89	216	126
15-a	145	1548	120
25-b	182	513	263
16-a	152	477	494
16-b	512	T	92
17-a	255	882	184
17-b	166	558	183
18-a	181	585	167
18-b	222	477	340
19-a	518	1449	1133
19-b	757	549	1667
20	1158	T	884
21	258	1062	200

(a) Number of gradient calculations required by each method for the reactions in the test set. F and T stand for failed and terminated reactions. Terminated reactions are the ones that required more than 1,800 gradient calculations for 9 active images.

ID	DE-GSM	CI-NEB	SE-GSM
1	3.6	3.5	3.5
2	3.9	3.9	3.9
3-a	32.5	T	33.0
3-b	3.7	4.1	4.1
4-a	39.2	39.3	44.0
4-b	15.2	11.7	-
5-a	18.5	18.5	18.7
5-b	-	3.9	25.5
5-b-1	18.5	-	-
5-b-2	6.4	-	-
5-c	19.7	17.4	17.8
6-a	17.9	17.8	17.9
6-b	-	F	F
6-b-1	20.6	-	-
6-b-2	22.7	-	-
6-c	26.4	F	F
7-a	21.8	T	20.6
7-b	24.6	T	25.3
8-a	19.5	19.1	20.7
8-b	13.2	12.1	15.0
9-a	11.5	11.4	11.4
9-b	16.9	16.5	17.0
9-c	18.9	18.8	19.0
9-d	23.5	22.4	22.5
10-a	47.4	44.7	34.8
10-b	47.3	47.0	47.7
11-a	24.7	T	26.1
11-b	18.3	18.3	19.5
12-a	22.7	T	25.5
12-b	16.0	T	12.3
13-a	22.0	21.6	23.3
13-b	7.4	6.0	6.4
14-a	0.6	0.2	0.9
14-b	5.8	5.8	8.0
14-c	23.3	22.4	32.8
15-a	13.5	8.9	7.7
25-b	19.2	19.2	16.1
16-a	2.3	2.0	1.9
16-b	2.1	T	11.2
17-a	24.4	25.2	22.0
17-b	23.9	23.6	25.8
18-a	14.0	13.7	13.3
18-b	23.9	22.7	22.2
19-a	22.9	21.5	21.7
19-b	6.9	4.3	5.2
20	31.1	T	35.3
21	10.2	7.9	9.6

(b) Activation energies in kcal/mol calculated by each method for the reactions in the test set. F and T stand for failed and terminated reactions. Terminated reactions are the ones that required more than 1,800 gradient calculations for 9 active images.

APPENDIX B

Supplementary Information for Chapter 3

B.1 Tables of activation energies and reaction paths

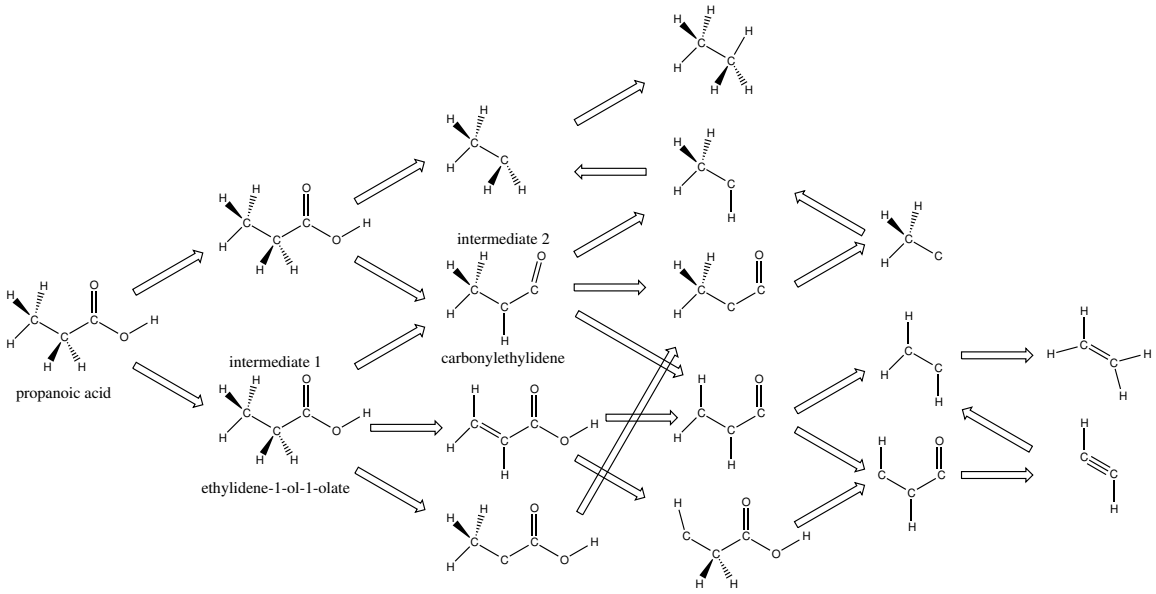


Figure B.1: Propanoic acid dissociation reaction network derived from reference [[130]].

		S-ZStruct		Original Study	
ID	Reaction	E_a	ΔE	E_a	ΔE
P1	$CH_3CHCOOH^* \longrightarrow CH_3^* + CHCOOH^*$	31.7	16.2	-	-
P2	$CH_3CHCOOH^* \longrightarrow CH_2CHCOOH^* + H^*$	25.3	-6.4	-	-
P3	$CH_3CHCOOH^* \longrightarrow CH_3CHCO^* + OH^*$	33.6	16.1	-	-
P4	$CH_3CHCOOH^* \longrightarrow CH_3CH^* + COOH^*$	0.7	-5.5	-	-
P5	$CH_3CHCOOH^* \longrightarrow CH_3CCOOH^* + H^*$	31.7	12.6	-	-
P6	$CH_3CHCO^* \longrightarrow CH_3^* + CHCO^*$	35.7	22.9	-	-
P7	$CH_3CHCO^* \longrightarrow CH_2CHCO^* + H^*$	25.3	-4.9	-	-
P8	$CH_3CHCO^* \longrightarrow CH_3CH^* + CO^*$	21.7 (6.9%)	-13.2 (26.9%)	23.3	-18.2
P9	$CH_3CHCO^* \longrightarrow CH_3CCO^* + H^*$	17.8	-9.4	-	-

Table B.1: Activation energies and heat of reactions for the portion of the propanoic acid dissociation network calculated by S-ZStruct. For reactions P6 to P9 the energies for the initial structure at 0° are reported. Only reaction P8 is run with the same parameters in the original study [130]. All energies are in kcal/mol. Asterisk (*) designates surface species.

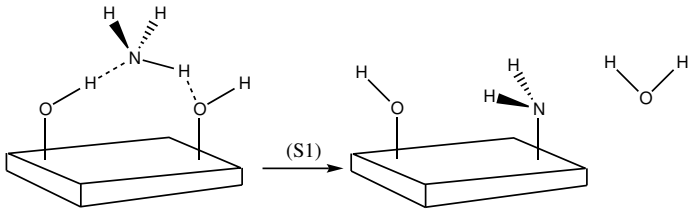


Figure B.2: Adsorption of an ammonia molecule on OH-terminated Si(100) surface. E_a = 3.9 kcal/mol, E_{rxn} = -8.1 kcal/mol.

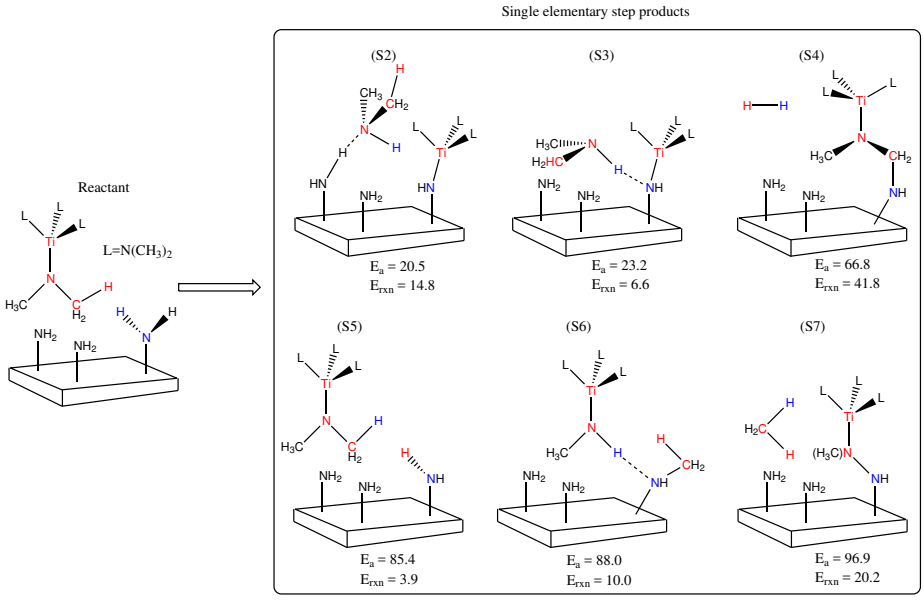


Figure B.3: Reaction paths for the reaction between TDMAT and adsorbed amido ligands. Reactive atoms are marked in red and blue. All energies are in kcal/mol.

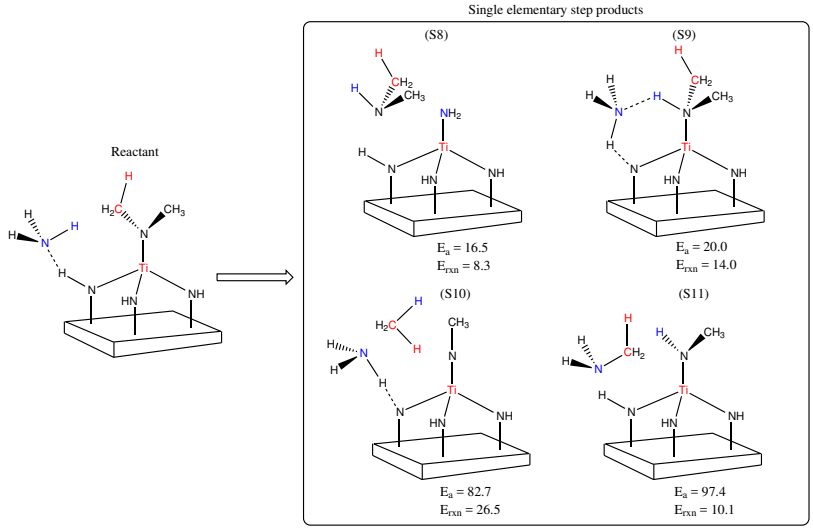


Figure B.4: Reaction paths for the second ammonia cycle. NH_3 is stabilized over the surface through formation of H-bonds. Reaction S8 is the lowest barrier pathway. Reactive atoms are marked in red and blue. All energies are in kcal/mol.

Step (1)			Step (2)			Step (3)			Step (4)			Step (5)			Ti Reduction		
ID	E_a	ΔE	ID	E_a	ΔE	ID	E_a	ΔE	ID	E_a	ΔE	ID	E_a	ΔE	ID	E_a	ΔE
S1	21.2	15.4	S2	20.5	14.8	1	29.9	21.9	7	35.4	29.9	S8	16.5	8.3	13	94.3	80.3
			S3	23.2	6.6	2	35.1	18.6	8	65.1	38.2	S9	20.0	14.0	14	87.7	84.5
			S4	66.8	41.8	3	65.5	24.9	9	69.7	-0.7	S10	82.7	26.5	15	40.3	26.2
			S5	85.4	3.9	4	89.0	6.0	10	77.0	14.9	S11	97.4	10.1	16	36.5	32.7
			S6	88.0	10.0	5	90.6	2.8	11	81.6	12.6				17	25.9	11.5
			S7	96.9	20.2	6	96.3	48.8	12	93.4	75.9						

Table B.2: Activation energies and heat of reactions for ALD of TiN network. All energies are in kcal/mol.

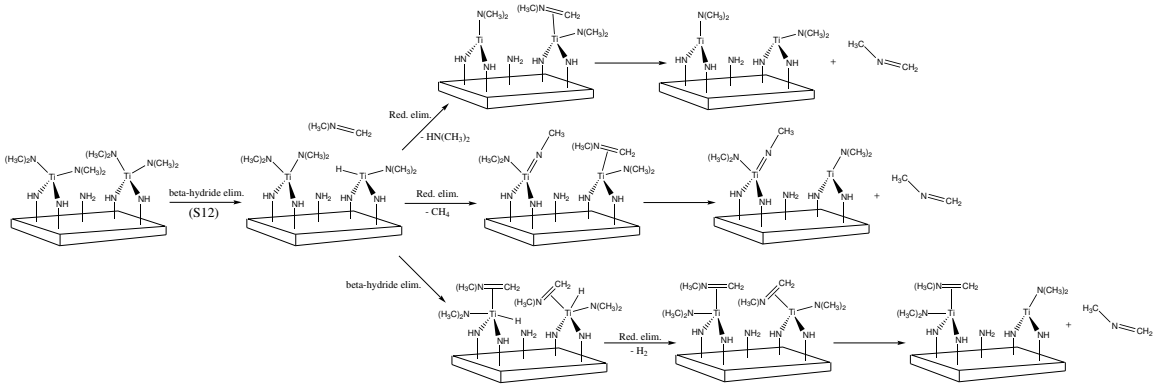


Figure B.5: Reaction paths for β-hydride elimination and Ti reduction.

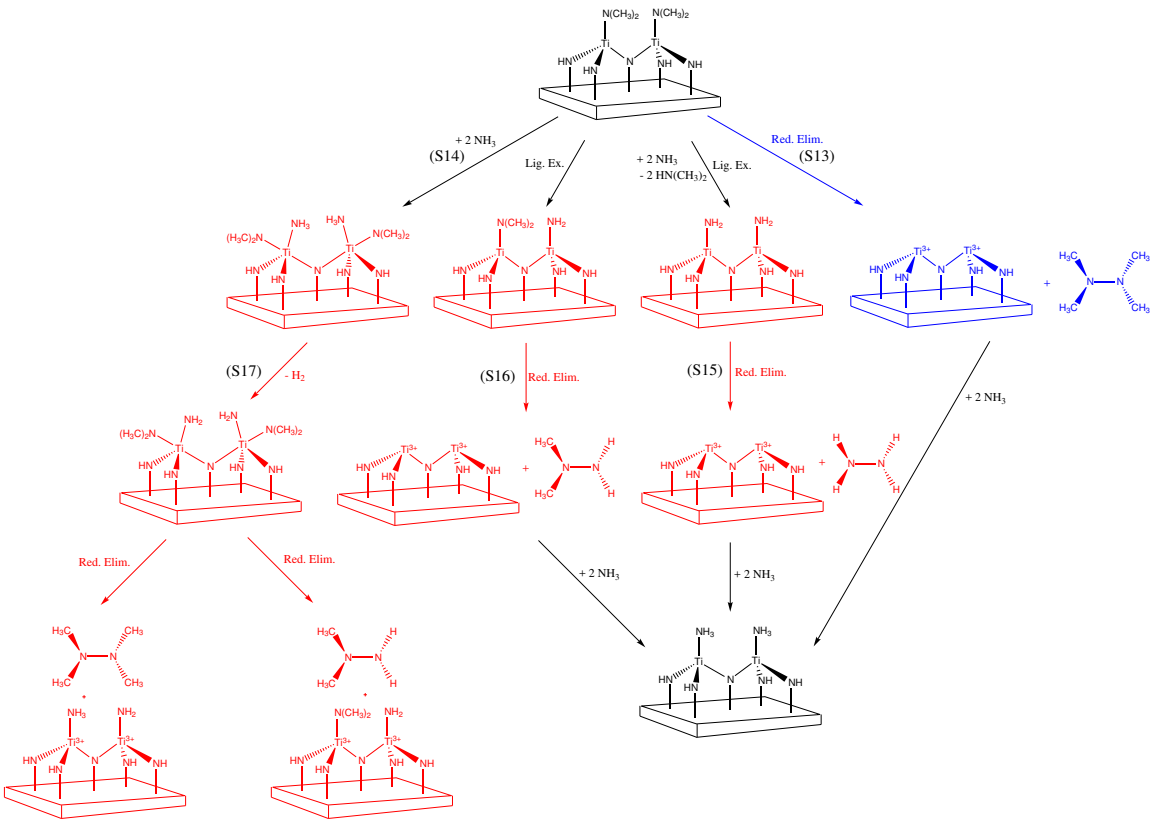
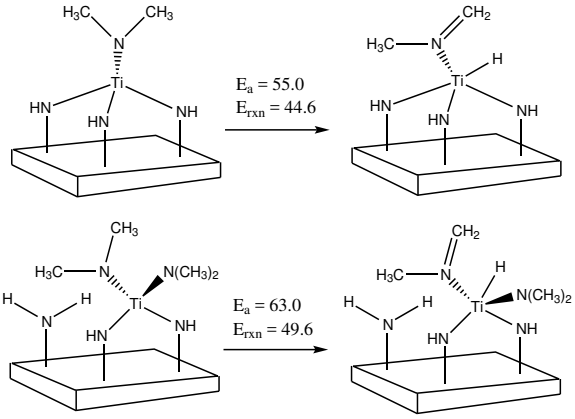


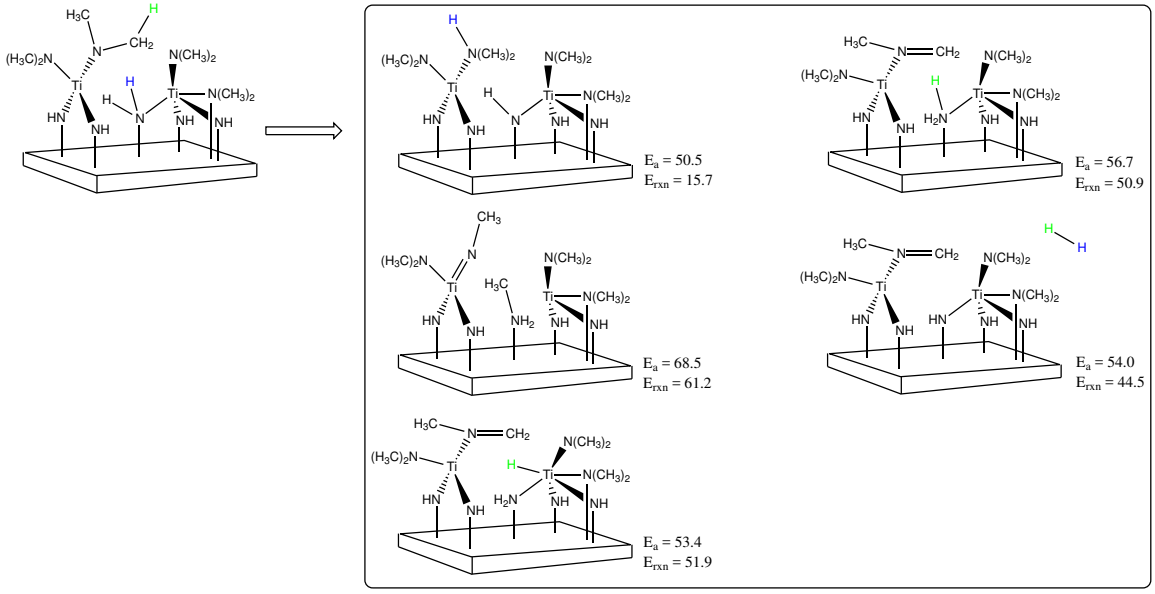
Figure B.6: Summary of the proposed reaction paths by the literature for reduction of Ti(IV) to Ti(III). Blue: reductive elimination of tetramethylhydrazine after the third ligand-exchange reaction. Red: reductive elimination of hydrazine or 1,1-dimethylhydrazine after the third ligand-exchange and ammonia cycle. Reaction paths following a high barrier are considered infeasible and not simulated.

ID	E_a	ΔE
S12	58.6	37.8
S13	74.4	71.8
S14	7.8	-0.1
S15	87.7	84.5
S16	76.6	76.0
S17	91.5	74.6

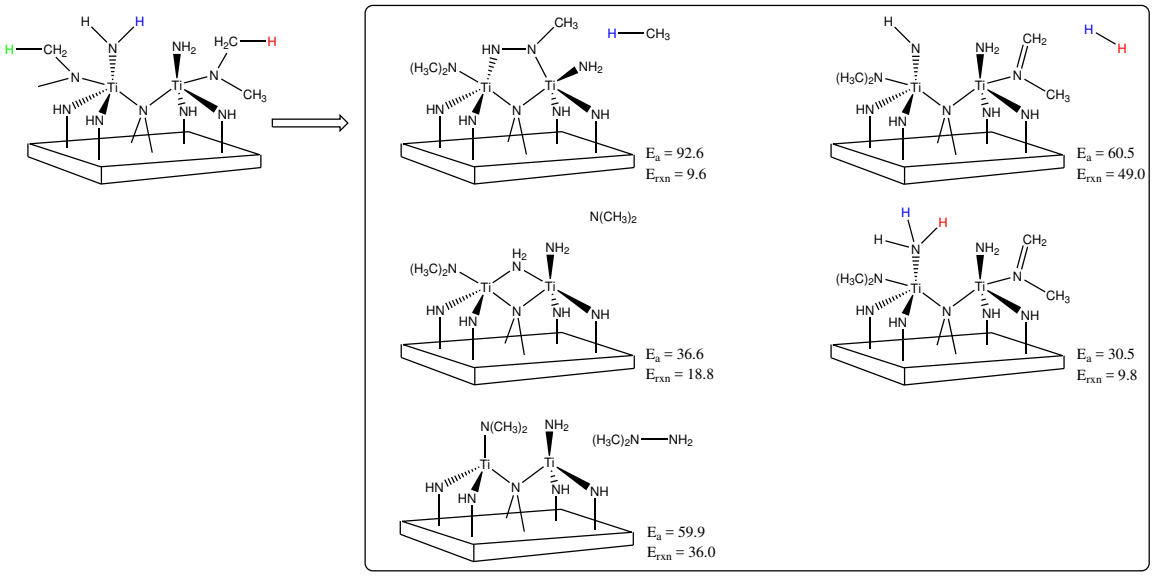
Table B.3: Activation energies and heat of reactions Ti reduction. All energies are in kcal/mol.


E_a = 55.0
E_{rxn} = 44.6
E_a = 63.0
E_{rxn} = 49.6

(a)


E_a = 50.5
E_{rxn} = 15.7
E_a = 56.7
E_{rxn} = 50.9
E_a = 68.5
E_{rxn} = 61.2
E_a = 54.0
E_{rxn} = 44.5
E_a = 53.4
E_{rxn} = 51.9

(b)



(c)

Figure B.7: Some reaction pathways calculated by S-ZStruct that have high activation energies but demonstrate capabilities of S-ZStruct to examine different possibilities. β-hydride elimination is followed by formation of N-methylmethanimine, H_2, methane, or reductive elimination of dimethylamine. All energies are in kcal/mol.

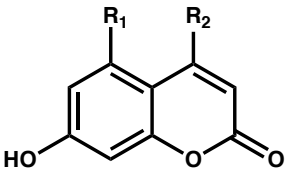
APPENDIX C

Supplementary Information for Chapter 4

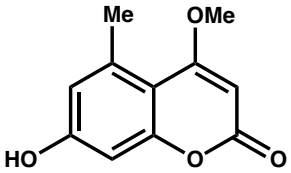
C.1 Supplementary Figures and Tables

Row Mol. or B partner		Col. Mol. or A partner (Mol. ID)				
Mol. ID	R Groups	Data 1 (1)	Data 2 (32)	Data 3 (20)	Data 4 (26)	Data 5 (2)
1	R1=Me, R2=OMe	99.2%	11.7%	2.1%	2.9%	66.3%
2	R1=Me, R2=OEt	66.3%	7.5%	6.0%	-	74.0%
3	R1=Me, R2=OiPr	40.6%	0.7%	0.8%	-	12.6%
4	R1=Me, R2=OtBu	12.9%	0%	-	-	-
5	R1=Me, R2=OBu	34.7%	3.0%	1.2%	-	39.3%
6	R1=Me, R2=NH_2	13.2%	0%	-	-	-
7	R1=Me, R2=NHMe	70.4%	25.7%	7.1%	4.7%	83.6%
8	R1=Me, R2=H	54.2%	68.1%	1.7%	2.4%	27.1%
9	R1=Me, R2=Me	4.4%	12.1%	trace	-	6.7%
33	R1=Me, R2=OPr	14.6%	-	-	-	14.6%
16	R1=H, R2=OMe	39.6%	10.1%	0.6%	0.5%	7.9%
17	R1=H, R2=NMe(OMe)	1.8%	-	0.8%	-	-
18	R1=H, R2=Me	21.7%	4.9%	-	-	0.0%
19	R1=H, R2=Ph	0.9%	0%	-	-	-
20	R1=H, R2=H	4.0%	10.6%	-	-	36.6%
1	R1=Me, R2=OMe	-	3.8%	2.1%	2.9%	66.3%
12	R1=Cl, R2=OMe	14.5%	3.8%	0.9%	-	15.2%
14	R1=Br, R2=OMe	19.5%	10.1%	-	-	15.7%
16	R1=H, R2=OMe	39.6%	0%	0.6%	0.5%	7.9%
34	R1=Et, R2=OMe	28.9%	-	-	-	-

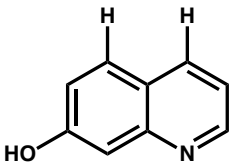
Table C.1: Reported experimental yields for heterocoupling (AB products) reactions of C5-Me, C5-H, and C4-OMe substituted coumarins. The number in parentheses is the molecule ID of the column molecule.



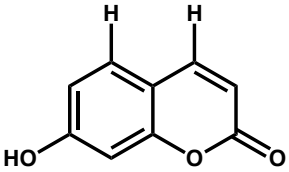
(a) The structure of the row molecules. R_1 and R_2 groups for each molecule are listed in Table C.1.



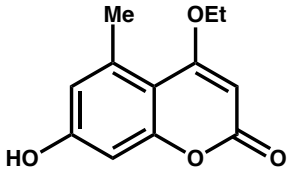
(b) Column molecule for data set 1. Molecule ID 1.



(c) Column molecule for data set 2. Molecule ID 32.

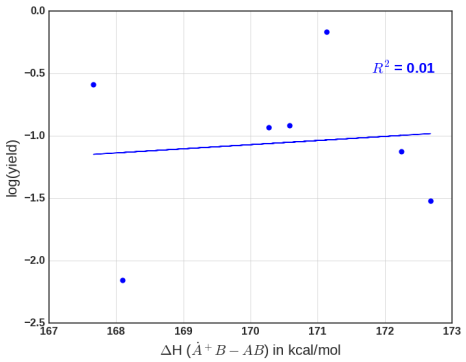


(d) Column molecule for data set 3. Molecule ID 20.

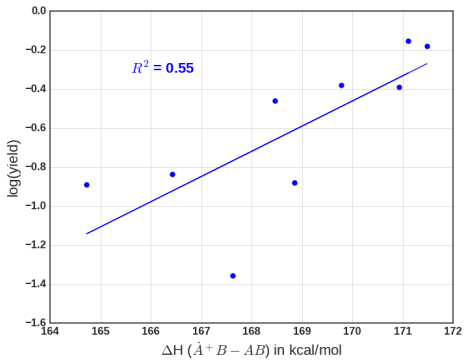


(e) Column molecule for data set 5. Molecule ID 2.

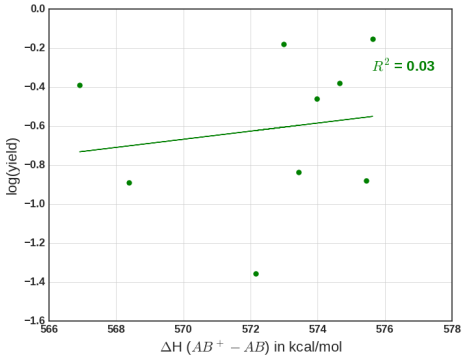
Figure C.1: Structures of row and column molecules of Table C.1.



(a) The highest correlation between ΔH and $\log_{10}$(yield) for mechanism 2 in data 2.

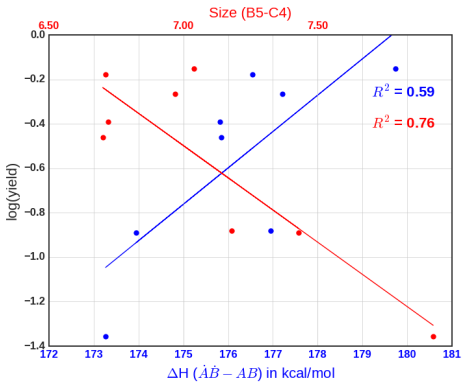


(b) The highest correlation between ΔH and $\log_{10}$(yield) for mechanism 2 in data 1.

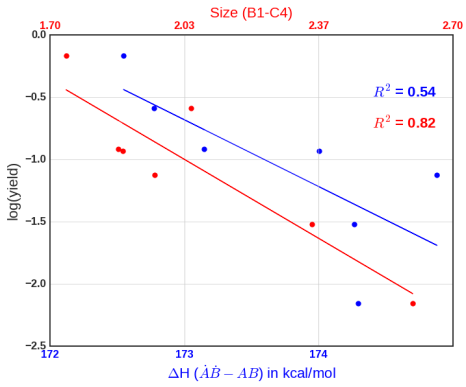


(c) The highest correlation between ΔH and $\log_{10}$(yield) for mechanism 1 in data 1.

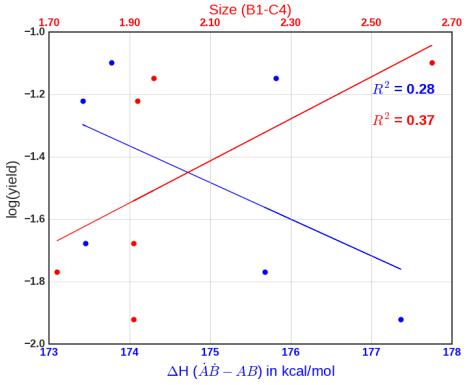
Figure C.2: Correlation between $\log_{10}$(yield) and ΔH of elementary steps for pathways 1 and 2 of Figure 4.3.



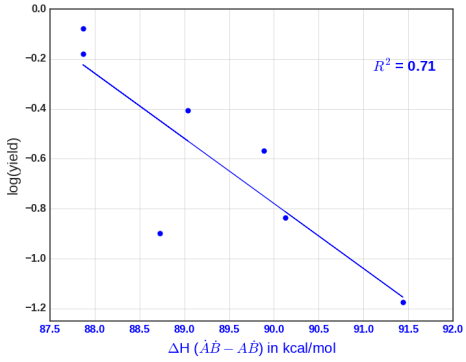
(a) Data 1



(b) Data 2



(c) Data 3



(d) Data 5

Figure C.3: Correlation between $\log_{10}$(yield) and ΔH of reaction steps for di-radical pathway of Figure 4.3.

BIBLIOGRAPHY

[1] Open Babel: The Open Source Chemistry Toolbox. http://openbabel.org/wiki/Main_Page. [Online; accessed 30-December-2017].

[2] STRUCT.PDF.

[3] Ruben Abagyan, Maxim Totrov, and Dmitry Kuznetsov. ICM - A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.*, 15(5):488–506, 1994.

[4] F. Abild-Pedersen and M.P. Andersson. CO adsorption energies on metals with correction for high coordination adsorption sites A density functional study. *Surf. Sci.*, 601(7):1747–1753, apr 2007.

[5] Susumu Akutagawa. Asymmetric synthesis by metal BINAP catalysts. *Appl. Catal. A Gen.*, 128(2):171–207, aug 1995.

[6] Sudhakar V. Alapati, J. Karl Johnson, and David S. Sholl. Using first principles calculations to identify new destabilized metal hydride reactions for reversible hydrogen storage. *Phys. Chem. Chem. Phys.*, 9(12):1438, 2007.

[7] Elena V. Anoshkina, Alexander G. Belyaev, and Hans-Peter Seidel. Asymtotic analysis of three-point approximations of vertex normals and curvatures. In *Proceedings of the Vision, Modeling, and Visualization Conference 2002 (VMV 2002), Erlangen, Germany, November 20-22, 2002*, pages 211–216, 2002.

[8] Aaron M Appel, John E Bercaw, Andrew B Bocarsly, Holger Dobbek, Daniel L. DuBois, Michel Dupuis, James G Ferry, Etsuko Fujita, Russ Hille, Paul J A Kenis, Cheryl A Kerfeld, Robert H Morris, Charles H F Peden, Archie R Portis, Stephen W Ragsdale, Thomas B Rauchfuss, Joost N H Reek, Lance C Seefeldt, Rudolf K Thauer, and Grover L Waldrop. Frontiers, Opportunities, and Challenges in Biochemical and Chemical Catalysis of CO 2 Fixation. *Chem. Rev.*, 113(8):6621–6658, aug 2013.

[9] Philippe Y Ayala and H Bernhard Schlegel. A combined method for determining reaction paths, minima, and transition state geometries. *J. Chem. Phys.*, 107(2):375, 1997.

[10] S.R. Bahn and K.W. Jacobsen. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.*, 4(3):56–66, 2002.

[11] J. Baker and P. M. W. Gill. An algorithm for the location of branching points on reaction paths. *J. Comput. Chem.*, 9(5):465–475, 1988.

[12] Jon Baker. An algorithm for the location of transition states. *J. Comput. Chem.*, 7(4):385–395, aug 1986.

[13] Jon Baker. Techniques for geometry optimization: A comparison of cartesian and natural internal coordinates. *J. Comput. Chem.*, 14(9):1085–1100, 1993.

[14] Jon Baker and Fora Chan. The location of transition states: A comparison of Cartesian, Z-matrix, and natural internal coordinates. *J. Comput. Chem.*, 17(7):888–904, may 1996.

[15] Jon Baker, Alain Kessi, and Bernard Delley. The generation and use of delocalized internal coordinates in geometry optimization. *J. Chem. Phys.*, 105(1):192, 1996.

[16] Vebjørn Bakken and Trygve Helgaker. The efficient optimization of molecular geometries using redundant internal coordinates. *J. Chem. Phys.*, 117(20):9160, 2002.

[17] Ajit Banerjee, Noah Adams, Jack Simons, and Ron Shepard. Search for stationary points on surfaces. *J. Phys. Chem.*, 89(1):52–57, jan 1985.

[18] G.T Barkema and Normand Mousseau. The activationrelaxation technique: an efficient algorithm for sampling energy landscapes. *Comput. Mater. Sci.*, 20(3-4):285–292, mar 2001.

[19] Andrew Behn, Paul M. Zimmerman, Alexis T. Bell, and Martin Head-Gordon. Efficient exploration of reaction paths via a freezing string method. *J. Chem. Phys.*, 135(22):224108, 2011.

[20] Andrew Behn, Paul M. Zimmerman, Alexis T. Bell, and Martin Head-Gordon. Incorporating Linear Synchronous Transit Interpolation into the Growing String Method: Algorithm and Applications. *J. Chem. Theory Comput.*, 7(12):4019–4025, dec 2011.

[21] Sina Behtash, Jianmin Lu, Christopher T. Williams, John R. Monnier, and Andreas Heyden. Effect of Palladium Surface Structure on the Hydrodeoxygenation of Propanoic Acid: Identification of Active Sites. *J. Phys. Chem. C*, 119(4):1928–1942, 2015.

[22] Maike Bergeler, Carmen Herrmann, and Markus Reiher. Mode-tracking based stationary-point optimization. *J. Comput. Chem.*, 36(19):1429–1438, 2015.

[23] Salomon R. Billeter, Alexander J. Turner, and Walter Thiel. Linear scaling geometry optimisation and transition state search in hybrid delocalised internal coordinates. *Phys. Chem. Chem. Phys.*, 2(10):2177–2186, 2000.

[24] Adam B. Birkholz and H. Bernhard Schlegel. Coordinate reduction for exploring chemical reaction paths. *Theor. Chem. Acc.*, 131(3):1170, mar 2012.

[25] Adam B. Birkholz and H. Bernhard Schlegel. Path optimization by a variational reaction coordinate method. I. Development of formalism and algorithms. *J. Chem. Phys.*, 143(24):244101, dec 2015.

[26] Adam B Birkholz and H Bernhard Schlegel. Using bonding to guide transition state optimization. *J. Comput. Chem.*, 36(15):1157–1166, jun 2015.

[27] Adam B. Birkholz and H. Bernhard Schlegel. Exploration of some refinements to geometry optimization methods. *Theor. Chem. Acc.*, 135(4):84, apr 2016.

[28] Josep Maria Bofill. Updated Hessian matrix and the restricted step method for locating transition structures. *J. Comput. Chem.*, 15(1):1–11, jan 1994.

[29] Josep Maria Bofill and Josep Maria Anglada. Finding transition states using reduced potential-energy surfaces. *Theor. Chem. Acc.*, 105(6):463–472, may 2001.

[30] Peter G Bolhuis, David Chandler, Christoph Dellago, and Phillip L Geissler. TRANSITION PATH SAMPLING: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.*, 53(1):291–318, oct 2002.

[31] Menno Bouman and Francisco Zaera. The Surface Chemistry of Atomic Layer Deposition (ALD) Processes for Metal Nitride and Metal Oxide Film Growth. *ECS Trans*, 33(2):291–305, 2010.

[32] Menno Bouman and Francisco Zaera. Reductive Eliminations from Amido Metal Complexes: Implications for Metal Film Deposition. *J. Electrochem. Soc.*, 158(8):D524, 2011.

[33] Alexandre V Brethomé, Stephen P Fletcher, and Robert S. Paton. Conformational Effects on Physical-Organic Descriptors the Case of Sterimol Steric Parameters. *ACS Catal.*, 2019.

[34] Gerhard Bringmann, Tanja Gulder, Tobias A. M. Gulder, and Matthias Breuning. Atroposelective Total Synthesis of Axially Chiral Biaryl Natural Products. *Chem. Rev.*, 111(2):563–639, feb 2011.

[35] Gerhard Bringmann, Anne J. Price Mortimer, Paul A. Keller, Mary J. Gresser, James Garner, and Matthias Breuning. Atroposelective Synthesis of Axially Chiral Biaryl Compounds. *Angew. Chemie Int. Ed.*, 44(34):5384–5427, aug 2005.

[36] C. G. BROYDEN. The Convergence of a Class of Double-rank Minimization Algorithms. *IMA J. Appl. Math.*, 6(3):222–231, 1970.

[37] Steven K Burger and Paul W Ayers. Methods for finding transition states on reduced potential energy surfaces. *J. Chem. Phys.*, 132(23):234110, 2010.

[38] Steven K. Burger and Weitao Yang. Quadratic string method for determining the minimum-energy path based on multiobjective optimization. *J. Chem. Phys.*, 124(5):054109, 2006.

[39] Steven K Burger and Weitao Yang. Sequential quadratic programming method for determining the minimum energy path. *J. Chem. Phys.*, 127(16):164107, 2007.

[40] Ralph A. Cacho, Yit-Heng Chooi, Hui Zhou, and Yi Tang. Complexity Generation in Fungal Polyketide Biosynthesis: A Spirocycle-Forming P450 in the Concise Pathway to the Antifungal Drug Griseofulvin. *ACS Chem. Biol.*, 8(10):2322–2330, oct 2013.

[41] Caitlin Callaghan, Ilie Fishtik, Ravindra Datta, Michael Carpenter, Matthew Chmielewski, and Anibal Lugo. An improved microkinetic model for the water gas shift reaction on copper. *Surf. Sci.*, 541(1-3):21–30, sep 2003.

[42] E Cancès, F Legoll, M-C Marinica, K Minoukadeh, and F Willaime. Some improvements of the activation-relaxation technique method for finding transition pathways on potential energy surfaces. *J. Chem. Phys.*, 130(11):114711, 2009.

[43] Fabián G. Cantú Reinhard, Simonetta Fornarini, Maria Elisa Crestoni, and Sam P. de Visser. Hydrogen Atom vs. Hydride Transfer in Cytochrome P450 Oxidations: A Combined Mass Spectrometry and Computational Study. *Eur. J. Inorg. Chem.*, 2018(17):1854–1865, may 2018.

[44] Charles J. Cerjan. On finding transition states. *J. Chem. Phys.*, 75(6):2800, 1981.

[45] Hugh Chaffey-Millar, Astrid Nikodem, Alexei V Matveev, Sven Krueger, and Notker Roesch. Improving Upon String Methods for Transition State Discovery. *J. Chem. Theory Comput.*, 8(2):777–786, 2012.

[46] Marvin Charton. Steric effects. I. Esterification and acid-catalyzed hydrolysis of esters. *J. Am. Chem. Soc.*, 97(6):1552–1556, mar 1975.

[47] Hsin-tsung Chen, Djamaladdin G Musaev, and M C Lin. Adsorption and Dissociation of H 2 O on a W(111) Surface: A Computational Study. *J. Phys. Chem. C*, 111(46):17333–17339, nov 2007.

[48] Ming Chen, Michel A. Cuendet, and Mark E. Tuckerman. Heating and flooding: A unified approach for rapid generation of free energy surfaces. *J. Chem. Phys.*, 137(2):024102, 2012.

[49] Gui-Juan Cheng, Xinhao Zhang, Lung Wa Chung, Liping Xu, and Yun-Dong Wu. Computational Organic Chemistry: Bridging Theory and Experiment in Establishing the Mechanisms of Chemical Reactions. *J. Am. Chem. Soc.*, 137(5):1706–1725, feb 2015.

[50] Jhih-Wei Chu, Bernhardt L. Trout, and Bernard R. Brooks. A super-linear minimization scheme for the nudged elastic band method. *J. Chem. Phys.*, 119(24):12708, 2003.

[51] Jason B. Cross and H. Bernhard Schlegel. Molecular Orbital Studies of Titanium Nitride Chemical Vapor Deposition: Gas Phase Complex Formation, Ligand Exchange, and Elimination Reactions. *Chem. Mater.*, 12(8):2466–2474, aug 2000.

[52] Jason B. Cross and H.Bernhard Schlegel. Molecular orbital studies of titanium nitride chemical vapor deposition: gas phase β-elimination. *Chem. Phys. Lett.*, 340(3-4):343–347, jun 2001.

[53] Jason B. Cross, Stanley M. Smith, and H. Bernhard Schlegel. Molecular Orbital Studies of Titanium Nitride Chemical Vapor Deposition: Imido Dimer Formation and Elimination Reactions. *Chem. Mater.*, 13(3):1095–1100, mar 2001.

[54] Ernest R. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *J. Comput. Phys.*, 17(1):87–94, jan 1975.

[55] Jorge M Del Campo and Andreas M Köster. A hierarchical transition state search algorithm. *J. Chem. Phys.*, 129(2):024107, 2008.

[56] Amanda L. Dewyer, Alonso J. Argüelles, and Paul M. Zimmerman. Methods for exploring reaction space in molecular systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, page e1354, nov 2017.

[57] Amanda L. Dewyer and Paul M. Zimmerman. Finding reaction mechanisms, intuitive or otherwise. *Org. Biomol. Chem.*, 15(3):501–504, 2017.

[58] Amanda L. Dewyer and Paul M. Zimmerman. Simulated Mechanism for Palladium-Catalyzed, Directed γ-Arylation of Piperidine. *ACS Catal.*, 7(8):5466–5477, aug 2017.

[59] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Phys. Rev. B*, 66(5):052301, aug 2002.

[60] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Finite Temperature String Method for the Study of Rare Events . *J. Phys. Chem. B*, 109(14):6688–6693, apr 2005.

[61] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.*, 126(16):164103, 2007.

[62] J.W Elam, M. Schuisky, J.D Ferguson, and S.M George. Surface chemistry and film growth during TiN atomic layer deposition using TDMAT and NH3. *Thin Solid Films*, 436(2):145–156, jul 2003.

[63] Simon D Elliott. Atomic-scale simulation of ALD chemistry. *Semicond. Sci. Technol.*, 27(7):074008, jul 2012.

[64] Bernd Ensing, Marco De Vivo, Zhiwei Liu, Preston Moore, and Michael L Klein. Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions. *Acc. Chem. Res.*, 39(2):73–81, feb 2006.

[65] Bernd Ensing, Marco De Vivo, Zhiwei Liu, Preston Moore, and Michael L. Klein. Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions. *Acc. Chem. Res.*, 39(2):73–81, feb 2006.

[66] Bernd Ensing, Marco De Vivo, Zhiwei Liu, Preston Moore, and Michael L. Klein. Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions. *Acc. Chem. Res.*, 39(2):73–81, feb 2006.

[67] Patton L Fast and Donald G Truhlar. Variational reaction path algorithm. *J. Chem. Phys.*, 109(10):3721, 1998.

[68] R. Fletcher. A new approach to variable metric algorithms. *Comput. J.*, 13(3):317–322, mar 1970.

[69] Connie W. Gao, Joshua W. Allen, William H. Green, and Richard H. West. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Comput. Phys. Commun.*, 203:212–225, jun 2016.

[70] Qingfeng Ge, Rickmer Kose, and David A. King. Adsorption energetics and bonding from femtomole calorimetry and from first principles theory. volume 45, pages 207–259. 2000.

[71] Qingfeng Ge, Rickmer Kose, and David A. King. Adsorption energetics and bonding from femtomole calorimetry and from first principles theory. In *Advances in Catalysis*, volume 45, pages 207–259. 2000.

[72] Nina Geib, Odile Meyer, Katja Zerbe, Katharina Woithe, Francesca Vitali, Dong Bo Li, Severine Fournier-Rousset, Nobuatsu Matoba, John A Robinson, Khaled Abou-Hadeed, and Markus Heck. Oxidative Phenol Coupling Reactions Catalyzed by OxyB: A Cytochrome P450 from the Vancomycin Producing Organism. Implications for Vancomycin Biosynthesis. *J. Am. Chem. Soc.*, 129(21):6887–6895, may 2007.

[73] Steven M. George. Atomic Layer Deposition: An Overview. *Chem. Rev.*, 110(1):111–131, jan 2010.

[74] S Alireza Ghasemi and Stefan Goedecker. An enhanced splined saddle method. *J. Chem. Phys.*, 135(1):014108, 2011.

[75] Amit a. Gokhale, James a. Dumesic, and Manos Mavrikakis. On the Mechanism of Low-Temperature Water Gas Shift Reaction on Copper. *J. Am. Chem. Soc.*, 130(4):1402–1414, jan 2008.

[76] Amit A. Gokhale, Shampa Kandoi, Jeffrey P. Greeley, Manos Mavrikakis, and James A. Dumesic. Molecular-level descriptions of surface chemistry in kinetic models using density functional theory. *Chemical Engineering Science*, 59(22-23):4679–4691, nov 2004.

[77] Donald Goldfarb. A family of variable-metric methods derived by variational means. *Math. Comput.*, 24(109):23–23, jan 1970.

[78] Anthony Goodrow, Alexis T Bell, and Martin Head-Gordon. Development and application of a hybrid method involving interpolation and ab initio calculations for the determination of transition states. *J. Chem. Phys.*, 129(17):174109, 2008.

[79] Anthony Goodrow, Alexis T. Bell, and Martin Head-Gordon. Transition state-finding strategies for use with the growing string method. *J. Chem. Phys.*, 130(24):244108, 2009.

[80] Jessica M. Grandner, Ralph A. Cacho, Yi Tang, and K. N. Houk. Mechanism of the P450-Catalyzed Oxidative Cyclization in the Biosynthesis of Griseofulvin. *ACS Catal.*, 6(7):4506–4511, jul 2016.

[81] Rebecca Granot and Roi Baer. A spline for your saddle. *J. Chem. Phys.*, 128(18):184111, 2008.

[82] Axel Groß. Dynamics of Reactions at Surfaces. In *Modeling and Simulation of Heterogeneous Catalytic Reactions*, volume 45, pages 39–70. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, dec 2011.

[83] Moritz P Haag, Alain C Vaucher, Maël Bosson, Stéphane Redon, and Markus Reiher. Interactive chemical reactivity exploration. *Chemphyschem*, 15(15):3301–19, 2014.

[84] Thomas A. Halgren and William N. Lipscomb. The synchronous-transit method for determining reaction pathways and locating molecular transition states. *Chem. Phys. Lett.*, 49(2):225–232, jul 1977.

[85] B Hammer, L B Hansen, and J. K. Nørskov. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys. Rev. B*, 59(11):7413–7421, mar 1999.

[86] B. Hammer and J.K. Nørskov. Theoretical surface science and catalysiscalculations and concepts. In *Experimental Neurology*, volume 144, pages 71–129. apr 2000.

[87] Louis P. Hammett. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.*, 59(1):96–103, jan 1937.

[88] Kaid C Harper and Matthew S Sigman. Three-Dimensional Correlation of Steric and Electronic Free Energy Relationships Guides Asymmetric Propargylation. *Science (80-.).*, 333(6051):1875–1878, sep 2011.

[89] John D Head. Partial optimization of large molecules and clusters. *J. Comput. Chem.*, 11(1):67–75, jan 1990.

[90] Graeme Henkelman, Gísli Jóhannesson, and Hannes Jónsson. Methods for Finding Saddle Points and Minimum Energy Paths. In *Theor. Methods Condens. Phase Chem.*, pages 269–302. Kluwer Academic Publishers, Dordrecht, 2002.

[91] Graeme Henkelman and Hannes Jónsson. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.*, 111(15):7010, 1999.

[92] Graeme Henkelman and Hannes Jónsson. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.*, 113(22):9978, 2000.

[93] Graeme Henkelman and Hannes Jónsson. Theoretical Calculations of Dissociative Adsorption of <math display="inline"> <mrow> <msub> <mrow> <mi>CH</mi> </mrow> <mrow> <mn>4</mn> </mrow> </msub> </mrow> </math> on an Ir(111) Surface. *Phys. Rev. Lett.*, 86(4):664–667, jan 2001.

[94] Graeme Henkelman, Blas P. Uberuaga, and Hannes Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.*, 113(22):9901, 2000.

[95] Carmen Herrmann and Markus Reiher. Direct targeting of adsorbate vibrations with mode-tracking. *Surf. Sci.*, 600(9):1891–1900, 2006.

[96] Andreas Heyden, Alexis T Bell, and Frerich J Keil. Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method. *J. Chem. Phys.*, 123(22):224101, 2005.

[97] Michael Hirsch and Wolfgang Quapp. Improved RGF method to find saddle points. *J. Comput. Chem.*, 23(9):887–894, jul 2002.

[98] Michael Hirsch, Wolfgang Quapp, and Dietmar Heidrich. The set of valleyridge inflection points on the potential energy surface of water. *Phys. Chem. Chem. Phys.*, 1(23):5291–5299, 1999.

[99] Xiaofeng Hu, Jun Yin, Randall J. Meyer, and Michael Trenary. Kinetics of Aminocarbyne Formation on Pt(111). *J. Phys. Chem. C*, 119(26):14506–14512, jul 2015.

[100] Marcella Iannuzzi, Alessandro Laio, and Michele Parrinello. Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics. *Phys. Rev. Lett.*, 90(23):238302, jun 2003.

[101] Mina Jafari and Paul M. Zimmerman. Reliable and efficient reaction path and transition state finding for surface reactions with the growing string method. *J. Comput. Chem.*, 38(10):645–658, apr 2017.

[102] Mina Jafari and Paul M. Zimmerman. Uncovering reaction sequences on surfaces through graphical methods. *Phys. Chem. Chem. Phys.*, 20(11):7721–7729, 2018.

[103] H. H. Jaffe. A Reëxamination of the Hammett Equation. *Chem. Rev.*, 53(2):191–261, oct 1953.

[104] Yanju Ji, Xiaocun Liu, Gang Fu, Ying Chen, Fengxiang Wang, and Junqing Zhao. Density function theory study on structures and electronic properties of 2Meq 2 AlOPh and its derivatives. *IOP Conf. Ser. Mater. Sci. Eng.*, 87(1):012104, jul 2015.

[105] Hannes Jónsson, Greg Mills, and Karsten W. Jacobsen. Nudged elastic band method for finding minimum energy paths of transitions. Citeseer, 1998.

[106] Jaewoon Jung, Suyong Re, Yuji Sugita, and Seiichiro Ten-no. Improved constrained optimization method for reaction-path determination in the generalized hybrid orbital quantum mechanicalmolecular mechanical calculations. *J. Chem. Phys.*, 138(4):044106, 2013.

[107] Marika Juppo, Mikko Ritala, and Markku Leskelä. Use of 1,1-Dimethylhydrazine in the Atomic Layer Deposition of Transition Metal Nitride Thin Films. *J. Electrochem. Soc.*, 147(9):3377, 2000.

[108] Johannes Kästner and Paul Sherwood. Superlinearly converging dimer method for transition state search. *J. Chem. Phys.*, 128(1):014106, 2008.

[109] Ilja V Khavrutskii, Jason B Smith, and Anders Wallqvist. Exploring chemical reaction mechanisms through harmonic Fourier beads path optimization. *J. Chem. Phys.*, 139(16):165104, 2013.

[110] Sung Youb Kim, In-Ho Lee, and Sukky Jun. Transition-pathway models of atomic diffusion on fcc metal surfaces. II. Stepped surfaces. *Phys. Rev. B*, 76(24):245408, dec 2007.

[111] John R. Kitchin. Modeling materials using density functional theory. pages 1–298, 2012.

[112] Jií Klimeš, David R Bowler, and Angelos Michaelides. A critical assessment of theoretical methods for finding reaction pathways and transition states of surface processes. *J. Phys. Condens. Matter*, 22(7):074203, feb 2010.

[113] Marisa C. Kozlowski, Barbara J. Morgan, and Elizabeth C. Linton. Total synthesis of chiral biaryl natural products by asymmetric biaryl coupling. *Chem. Soc. Rev.*, 38(11):3193, 2009.

[114] Elizabeth H Krenske and K N Houk. Aromatic interactions as control elements in stereoselective organic reactions. *Acc. Chem. Res.*, 46(4):979–89, 2013.

[115] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.*, 6(1):15–50, jul 1996.

[116] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54(16):11169–11186, oct 1996.

[117] G. Kresse and J. Hafner. Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B*, 48(17):13115–13118, nov 1993.

[118] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59(3):1758–1775, jan 1999.

[119] A Laio and M Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci.*, 99(20):12562–12566, oct 2002.

[120] C Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Natl. Bur. Stand. (1934).*, 45(4):255, oct 1950.

[121] K. Lejaeghere, G. Bihlmayer, T. Bjorkman, P. Blaha, S. Blugel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, S. de Gironcoli, T. Deutsch, J. K. Dewhurst, I. Di Marco, C. Draxl, M. Du ak, O. Eriksson, J. A. Flores-Livas, K. F. Garrity, L. Genovese, P. Giannozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Granas, E. K. U. Gross, A. Gulans, F. Gygi, D. R. Hamann, P. J. Hasnip, N. A. W. Holzwarth, D. Iu an, D. B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepernik, E. Kucukbenli, Y. O. Kvashnin, I. L. M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordstrom, T. Ozaki, L. Paulatto, C. J. Pickard, W. Poelmans, M. I. J. Probert, K. Refson, M. Richter, G.-M. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K. Schwarz, S. Sharma, F. Tavazza, P. Thunstrom, A. Tkatchenko, M. Torrent, D. Vanderbilt, M. J. van Setten, V. Van Speybroeck, J. M. Wills, J. R. Yates, G.-X. Zhang, and S. Cottenier. Reproducibility in density functional theory calculations of solids. *Science*, 351(6280):aad3000–aad3000, mar 2016.

[122] Maxwell W. Li, Ian M. Pendleton, Alex J. Nett, and Paul M. Zimmerman. Mechanism for Forming B,C,N,O Rings from NH 3 BH 3 and CO 2 via Reaction Discovery Computations. *J. Phys. Chem. A*, 120(8):1135–1144, mar 2016.

[123] Xiaosong Li and Michael J. Frisch. Energy-Represented Direct Inversion in the Iterative Subspace within a Hybrid Geometry Optimization Method. *J. Chem. Theory Comput.*, 2(3):835–839, may 2006.

[124] Wenkel Liang, Haitao Wang, Jane Hung, Xiaosong Li, and Michael J. Frisch. Eigenspace Update for Molecular Geometry Optimization in Nonredundant Internal Coordinate. *J. Chem. Theory Comput.*, 6(7):2034–2039, 2010.

[125] Ching Yeh Lin, Sylvain R. A. Marque, Krzysztof Matyjaszewski, and Michelle L. Coote. Linear-Free Energy Relationships for Modeling StructureReactivity Trends in Controlled Radical Polymerization. *Macromolecules*, 44(19):7568–7583, oct 2011.

[126] Yuli Liu, Steven K Burger, and Paul W Ayers. Newton trajectories for finding stationary points on molecular potential energy surfaces. *J. Math. Chem.*, 49(9):1915–1927, oct 2011.

[127] Delphine Longrie, Davy Deduytsche, Jo Haemers, Philippe F. Smet, Kris Driesen, and Christophe Detavernier. Thermal and Plasma-Enhanced Atomic Layer Deposition of TiN Using TDMAT and NH 3 on Particles Agitated in a Rotary Reactor. *ACS Appl. Mater. Interfaces*, 6(10):7316–7324, may 2014.

[128] Pauline Loxq, Eric Manoury, Rinaldo Poli, Eric Deydier, and Agnès Labande. Synthesis of axially chiral biaryl compounds by asymmetric catalytic reactions with transition metals. *Coord. Chem. Rev.*, 308(Part 2):131–190, feb 2016.

[129] Jianmin Lu, Sina Behtash, Muhammad Faheem, and Andreas Heyden. Microkinetic modeling of the decarboxylation and decarbonylation of propanoic acid over Pd(1 1 1) model surfaces based on parameters obtained from first principles. *J. Catal.*, 305:56–66, 2013.

[130] Jianmin Lu, Sina Behtash, and Andreas Heyden. Theoretical Investigation of the Reaction Mechanism of the Decarboxylation and Decarbonylation of Propanoic Acid on Pd(111) Model Surfaces. *J. Phys. Chem. C*, 116(27):14328–14341, jul 2012.

[131] Jianmin Lu, Muhammad Faheem, Sina Behtash, and Andreas Heyden. Theoretical investigation of the decarboxylation and decarbonylation mechanism of propanoic acid over a Ru(0 0 0 1) model surface. *J. Catal.*, 324(27):14–24, apr 2015.

[132] Jacob R. Ludwig, Paul M. Zimmerman, Joseph B. Gianino, and Corinna S. Schindler. Iron(III)-catalysed carbonylolefin metathesis. *Nature*, 533(7603):374–379, apr 2016.

[133] Yuliana K. Lugo-José, John R. Monnier, Andreas Heyden, and Christopher T. Williams. Hydrodeoxygenation of propanoic acid over silica-supported palladium: effect of metal particle size. *Catal. Sci. Technol.*, 4:3909–3916, 2014.

[134] Satoshi Maeda, Erika Abe, Miho Hatanaka, Tetsuya Taketsugu, and Keiji Morokuma. Exploring potential energy surfaces of large systems with artificial force induced reaction method in combination with ONIOM and microiteration. *J. Chem. Theory Comput.*, 8(12):5058–5063, 2012.

[135] Satoshi Maeda, Yu Harabuchi, Makito Takagi, Tetsuya Taketsugu, and Keiji Morokuma. Artificial Force Induced Reaction (AFIR) Method for Exploring Quantum Chemical Potential Energy Surfaces. *Chem. Rec.*, pages 1–17, jun 2016.

[136] Satoshi Maeda and Keiji Morokuma. Finding reaction pathways of type A + B X: Toward systematic prediction of reaction mechanisms. *J. Chem. Theory Comput.*, 7:2335–2345, 2011.

[137] Satoshi Maeda, Koichi Ohno, and Keiji Morokuma. An Automated and Systematic Transition Structure Explorer in Large Flexible Molecular Systems Based on Combined Global Reaction Route Mapping and Microiteration Methods. *J. Chem. Theory Comput.*, 5(10):2734–2743, oct 2009.

[138] Shaama Mallikarjun Sharada, Paul M. Zimmerman, Alexis T. Bell, and Martin Head-Gordon. Automated Transition State Searches without Evaluating the Hessian. *J. Chem. Theory Comput.*, 8(12):5166–5174, dec 2012.

[139] Y. Matsumoto, T. Amano, T. N. Kato, and M. Hoshino. Stochastic electron acceleration during spontaneous turbulent reconnection in a strong shock wave. *Science (80-.).*, 347(6225):974–978, feb 2015.

[140] Cecilia Mattevi, Hokwon Kim, and Manish Chhowalla. A review of chemical vapour deposition of graphene on copper. *J. Mater. Chem.*, 21(10):3324–3334, 2011.

[141] Reinhard J. Maurer, Victor G. Ruiz, Javier Camarillo-Cisneros, Wei Liu, Nicola Ferri, Karsten Reuter, and Alexandre Tkatchenko. Adsorption structures and energetics of molecules on metal surfaces: Bridging experiment and theory. *Prog. Surf. Sci.*, 91(2):72–100, may 2016.

[142] Joseph W. May, Jeremy D. Lehner, Michael J. Frisch, and Xiaosong Li. Transition State Search Using a Guided Direct Inversion in the Iterative Subspace Method. *J. Chem. Theory Comput.*, 8(12):5175–5179, 2012.

[143] Laura S. Mazzaferro, Wolfgang Hüttel, Alexander Fries, and Michael Müller. Cytochrome P450-Catalyzed Regio- and Stereoselective Phenol Coupling of Fungal Natural Products. *J. Am. Chem. Soc.*, 137(38):12289–12295, 2015.

[144] Simon J. Meek, Catherine L. Pitman, and Alexander J. M. Miller. Deducing Reaction Mechanism: A Guide for Students, Researchers, and Instructors. *J. Chem. Educ.*, 93(2):275–286, feb 2016.

[145] Donghai Mei, Lijun Xu, and Graeme Henkelman. Potential Energy Surface of Methanol Decomposition on Cu(110). *J. Phys. Chem. C*, 113(11):4522–4537, mar 2009.

[146] Jeremie J. Miller and Matthew S. Sigman. Quantitatively Correlating the Effect of Ligand-Substituent Size in Asymmetric Catalysis Using Linear Free Energy Relationships. *Angew. Chemie Int. Ed.*, 47(4):771–774, jan 2008.

[147] Greg Mills and Hannes Jónsson. Quantum and thermal effects in H2 dissociative adsorption: Evaluation of free energy barriers in multidimensional quantum systems. *Phys. Rev. Lett.*, 72(7):1124–1127, feb 1994.

[148] V Milman, B Winkler, J. A. White, C. J. Pickard, M. C. Payne, E V Akhmatskaya, and R H Nobes. Electronic structure, properties, and phase stability of inorganic crystals: A pseudopotential plane-wave study. *Int. J. Quantum Chem.*, 77(5):895–910, 2000.

[149] Noritaka Mizuno and Makoto Misono. Heterogeneous Catalysis. *Chem. Rev.*, 98(1):199–218, feb 1998.

[150] Gregory M. Mullen, Liang Zhang, Edward J. Evans, Ting Yan, Graeme Henkelman, and C. Buddie Mullins. Oxygen and Hydroxyl Species Induce Multiple Reaction Pathways for the Partial Oxidation of Allyl Alcohol on Gold. *J. Am. Chem. Soc.*, 136(17):6489–6498, apr 2014.

[151] E. Matthias Müller, Armin de Meijere, and Helmut Grubmüller. Predicting unimolecular chemical reactions: Chemical flooding. *J. Chem. Phys.*, 116(3):897–905, jan 2002.

[152] J. Musschoot, Q. Xie, D. Deduytsche, S. Van den Berghe, R.L. Van Meirhaeghe, and C. Detavernier. Atomic layer deposition of titanium nitride from TDMAT precursor. *Microelectron. Eng.*, 86(1):72–77, jan 2009.

[153] E. Matthias Muller, Armin de Meijere, and Helmut Grubmuller. Predicting unimolecular chemical reactions: Chemical flooding. *J. Chem. Phys.*, 116(3):897, 2002.

[154] Haruyuki Nakano, Tohru Nakajima, and Shigeru Obara. Efficient and stable method of searching for optimum structures of molecules containing cyclic parts. *Chem. Phys. Lett.*, 177(4-5):458–462, mar 1991.

[155] Alex J. Nett, John Montgomery, and Paul M. Zimmerman. Entrances, Traps, and Rate-Controlling Factors for Nickel-Catalyzed CH Functionalization. *ACS Catal.*, 7(10):7352–7362, oct 2017.

[156] Alex J. Nett, Wanxiang Zhao, Paul M. Zimmerman, and John Montgomery. Highly Active Nickel Catalysts for CH Functionalization Identified through Analysis of Off-Cycle Intermediates. *J. Am. Chem. Soc.*, 137(24):7636–7639, jun 2015.

[157] J. K. Norskov, F. Abild-Pedersen, F. Studt, and T. Bligaard. Density functional theory in surface chemistry and catalysis. *Proceedings of the National Academy of Sciences*, 108(3):937–943, jan 2011.

[158] J.K. Nørskov, M Scheffler, and H. Toulhoat. Density Functional Theory in Surface Science and Heterogeneous Catalysis. *MRS Bulletin*, 31(09):669–674, sep 2006.

[159] Noel M. O'Boyle, Michael Banck, Craig A. James, Chris Morley, Tim Vandermeersch, and Geoffrey R. Hutchison. Open Babel: An open chemical toolbox. *J. Cheminform.*, 3(1):33, 2011.

[160] R. A. Olsen, G. J. Kroes, G. Henkelman, A. Arnaldsson, and H. Jónsson. Comparison of methods for finding saddle points without knowledge of the final states. *J. Chem. Phys.*, 121(20):9776, 2004.

[161] Bela Paizs, Jon Baker, Sandor Suhai, and Peter Pulay. Geometry optimization of large biomolecules in redundant internal coordinates. *J. Chem. Phys.*, 113(16):6566, 2000.

[162] Andreas Pedersen, Sigurdur F. Hafstein, and Hannes Jónsson. Efficient Sampling of Saddle Points with the Minimum-Mode Following Method. *SIAM J. Sci. Comput.*, 33(2):633–652, jan 2011.

[163] Ian M. Pendleton, Mónica H. Pérez-Temprano, Melanie S. Sanford, and Paul M. Zimmerman. Experimental and Computational Assessment of Reactivity and Mechanism in C(sp 3)N Bond-Forming Reductive Elimination from Palladium(IV). *J. Am. Chem. Soc.*, 138(18):6049–6060, may 2016.

[164] Ian M. Pendleton, Mónica H. Pérez-Temprano, Melanie S. Sanford, and Paul M. Zimmerman. Experimental and Computational Assessment of Reactivity and Mechanism in C(sp 3)N Bond-Forming Reductive Elimination from Palladium(IV). *J. Am. Chem. Soc.*, 138(18):6049–6060, may 2016.

[165] Chunyang Peng, Philippe Y Ayala, H Bernhard Schlegel, and Michael J Frisch. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *J. Comput. Chem.*, 17(1):49–56, jan 1996.

[166] Chunyang Peng and H. Bernhard Schlegel. Combining Synchronous Transit and Quasi-Newton Methods to Find Transition States. *Isr. J. Chem.*, 33(4):449–454, 1993.

[167] Shih-Feng Peng and Jia-Jen Ho. Theoretical Study of H 2 S Dissociation and Sulfur Oxidation on a W(111) Surface. *J. Phys. Chem. C*, 114(45):19489–19495, nov 2010.

[168] Baron Peters, Andreas Heyden, Alexis T. Bell, and Arup Chakraborty. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *J. Chem. Phys.*, 120(17):7877, 2004.

[169] Baron Peters, Andreas Heyden, Alexis T. Bell, and Arup Chakraborty. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *J. Chem. Phys.*, 120(17):7877, 2004.

[170] Baron Peters, WanZhen Liang, Alexis T Bell, and Arup Chakraborty. Biasing a transition state search to locate multiple reaction pathways. *J. Chem. Phys.*, 118(21):9533, 2003.

[171] Simone Piccinin and Michail Stamatakis. CO Oxidation on Pd(111): A First-Principles-Based Kinetic Monte Carlo Study. *ACS Catal.*, 4(7):2143–2152, jul 2014.

[172] P. Plessow. Reaction Path Optimization without NEB Springs or Interpolation Algorithms. *J. Chem. Theory Comput.*, 9(3):1305–1310, mar 2013.

[173] Dieter Poppinger. On the calculation of transition states. *Chem. Phys. Lett.*, 35(4):550–554, oct 1975.

[174] P. Pulay and G. Fogarasi. Geometry optimization in redundant internal coordinates. *J. Chem. Phys.*, 96(4):2856, 1992.

[175] W Quapp, M Hirsch, and D Heidrich. Bifurcation of reaction pathways: the set of valley ridge inflection points of a simple three-dimensional potential energy surface. *Theor. Chem. Acc.*, 100(5-6):285–299, 1998.

[176] Wolfgang Quapp. A growing string method for the reaction pathway defined by a Newton trajectory. *J. Chem. Phys.*, 122(17):174106, 2005.

[177] Wolfgang Quapp, Michael Hirsch, Olaf Imig, and Dietmar Heidrich. Searching for saddle points of potential energy surfaces by following a reduced gradient. *J. Comput. Chem.*, 19(9):1087–1100, jul 1998.

[178] Steven Y. Reece, Justin M. Hodgkiss, JoAnne Stubbe, and Daniel G. Nocera. Proton-coupled electron transfer: the mechanistic underpinning for radical transport and catalysis in biology. *Philos. Trans. R. Soc. B Biol. Sci.*, 361(1472):1351–1364, aug 2006.

[179] Ruichao Ren and G Orkoulas. Parallel Markov chain Monte Carlo simulations. *J. Chem. Phys.*, 126(21):211102, 2007.

[180] Weiqing Ren and Eric Vanden-Eijnden. A climbing string method for saddle point search. *J. Chem. Phys.*, 138(13):134105, 2013.

[181] Amit Samanta and Weinan E. Atomistic simulations of rare events using gentlest ascent dynamics. *J. Chem. Phys.*, 136(12):124104, mar 2012.

[182] Carlo Sambiagio, Stephen P. Marsden, A. John Blacker, and Patrick C. McGowan. Copper catalysed Ullmann type chemistry: from mechanistic aspects to modern development. *Chem. Soc. Rev.*, 43(10):3525–3550, 2014.

[183] Benjamin Sanchez-Lengeling and Alán Aspuru-Guzik. Inverse molecular design using machine learning: Generative models for matter engineering. *Science (80-.).*, 361(6400):360–365, jul 2018.

[184] H Bernhard Schlegel. A comparison of geometry optimization with internal, cartesian, and mixed coordinates. *Int. J. Quantum Chem.*, 44(S26):243–252, mar 1992.

[185] H. Bernhard Schlegel. Exploring potential energy surfaces for chemical reactions: An overview of some practical methods. *J. Comput. Chem.*, 24(12):1514–1527, sep 2003.

[186] H Bernhard Schlegel. Optimization of Equilibrium Geometries and Transition Structures. *Adv. Chem. Phys.*, 3(2):249–286, 2007.

[187] H. Bernhard Schlegel. Geometry optimization. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 1(5):790–809, sep 2011.

[188] Hadas Shalit, Anna Libman, and Doron Pappo. meso -Tetraphenylporphyrin Iron Chloride Catalyzed Selective Oxidative Cross-Coupling of Phenols. *J. Am. Chem. Soc.*, 139(38):13404–13413, sep 2017.

[189] D. F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Math. Comput.*, 24(111):647–647, sep 1970.

[190] Shaama Mallikarjun Sharada, Alexis T. Bell, and Martin Head-Gordon. 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.*, 140(16):164115, apr 2014.

[191] Daniel Sheppard, Rye Terrell, and Graeme Henkelman. Optimization methods for finding minimum energy paths. *J. Chem. Phys.*, 128(13):134106, 2008.

[192] Daniel Sheppard, Penghao Xiao, William Chemelewski, Duane D. Johnson, and Graeme Henkelman. A generalized solid-state nudged elastic band method. *J. Chem. Phys.*, 136(7):074103, 2012.

[193] David S. Sholl and Janice A. Steckel. *Density Functional Theory*. John Wiley & Sons, Inc., Hoboken, NJ, USA, mar 2009.

[194] Jack Simons, Poul Joergensen, Hugh Taylor, and Judy Ozment. Walking on potential energy surfaces. *J. Phys. Chem.*, 87(15):2745–2753, jul 1983.

[195] Jun Song and W.A. Curtin. A nanoscale mechanism of hydrogen embrittlement in metals. *Acta Mater.*, 59(4):1557–1569, feb 2011.

[196] Federico a. Soria, Eduardo M. Patrito, and Patricia Paredes-Olivera. Tailoring the Surface Reactivity of Silicon Surfaces by Partial Halogenation. *J. Phys. Chem. C*, 117(35):18021–18030, sep 2013.

[197] Sérgio Filipe Sousa, Pedro Alexandrino Fernandes, and Maria João Ramos. General Performance of Density Functionals. *J. Phys. Chem. A*, 111(42):10439–10452, oct 2007.

[198] R. Stadler and M. H. Zenk. The purification and characterization of a unique cytochrome P-450 enzyme from Berberis stolonifera plant cell cultures. *J. Biol. Chem.*, 268(2):823–831, 1993.

[199] JoAnne Stubbe, Daniel G. Nocera, Cyril S. Yee, and Michelle C. Y. Chang. Radical Initiation in the Class I Ribonucleotide Reductase: Long-Range Proton-Coupled Electron Transfer? *Chem. Rev.*, 103(6):2167–2202, jun 2003.

[200] Yury V. Suleimanov and William H. Green. Automated Discovery of Elementary Chemical Reaction Steps Using Freezing String and Berny Optimization Methods. *J. Chem. Theory Comput.*, 11(9):4248–4259, sep 2015.

[201] Daniel P. Tabor, Loïc M. Roch, Semion K. Saikin, Christoph Kreisbeck, Dennis Sheberla, Joseph H. Montoya, Shyam Dwaraknath, Muratahan Aykol, Carlos Ortiz, Hermann Tribukait, Carlos Amador-Bedolla, Christoph J. Brabec, Benji Maruyama, Kristin A. Persson, and Alán Aspuru-Guzik. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.*, 3(5):5–20, may 2018.

[202] Robert W. Taft. Linear Steric Energy Relationships. *J. Am. Chem. Soc.*, 75(18):4538–4539, sep 1953.

[203] Laleh Tahsini, Mojtaba Bagherzadeh, Wonwoo Nam, and Sam P. de Visser. Fundamental Differences of Substrate Hydroxylation by High-Valent Iron(IV)-Oxo Models of Cytochrome P450. *Inorg. Chem.*, 48(14):6661–6669, jul 2009.

[204] Peng Tao, Milan Hodošček, Joseph D Larkin, Yihan Shao, and Bernard R Brooks. Comparison of Three Chain-of-States Methods: Nudged Elastic Band and Replica Path with Restraints or Constraints. *J. Chem. Theory Comput.*, 8(12):5035–5051, 2012.

[205] W John Thomas, John Meurig and Thomas. *Principles and practice of heterogeneous catalysis*. John Wiley \& Sons, 2014.

[206] Chadwick A. Tolman. Steric effects of phosphorus ligands in organometallic chemistry and homogeneous catalysis. *Chem. Rev.*, 77(3):313–348, jun 1977.

[207] Maxim Totrov and Ruben Abagyan. Efficient parallelization of the energy, surface, and derivative calculations for internal coordinate mechanics. *J. Comput. Chem.*, 15(10):1105–1112, 1994.

[208] D G Truhlar and B C Garrett. Variational Transition State Theory. *Annu. Rev. Phys. Chem.*, 35(1):159–189, oct 1984.

[209] Donald G Truhlar and Bruce C Garrett. Variational transition-state theory. *Acc. Chem. Res.*, 13(12):440–448, dec 1980.

[210] Semen A. Trygubenko and David J. Wales. A doubly nudged elastic band method for finding transition states. *J. Chem. Phys.*, 120(5):2082, 2004.

[211] Zachary W. Ulissi, Andrew J. Medford, Thomas Bligaard, and Jens K. Nørskov. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nat. Commun.*, 8:14621, 2017.

[212] Eric Vanden-Eijnden and Maddalena Venturoli. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.*, 130(19):194103, 2009.

[213] Arthur F. Voter. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Phys. Rev. Lett.*, 78(20):3908–3911, 1997.

[214] David J Wales. Basins of attraction for stationary points on a potential-energy surface. *J. Chem. Soc. Faraday Trans.*, 88(5):653, 1992.

[215] Chia-ching Wang, Ya-Jen Yang, Jyh-chiang Jiang, Dah-Shyang Tsai, and Horng-Ming Hsieh. Density Functional Theory Study of the Oxidation of Ammonia on RuO 2 (110) Surface. *J. Phys. Chem. C*, 113(40):17411–17417, oct 2009.

[216] Gui-Chang Wang and Junji Nakamura. Structure Sensitivity for Forward and Reverse Water-Gas Shift Reactions on Copper Surfaces: A DFT Study. *J. Phys. Chem. Lett.*, 1(20):3053–3057, oct 2010.

[217] Guichang Wang, Ling Jiang, Cai, Pan, Zhao, Wei Huang, Xie, Li, Sun, and Bing Zhong. Surface Structure Sensitivity of the WaterGas Shift Reaction on Cu(h kl) Surfaces: A Theoretical Study. *J. Phys. Chem. B*, 107(2):557–562, jan 2003.

[218] Shi-Qing Wang. Barriers Against Copper Diffusion into Silicon and Drift Through Silicon Dioxide. *MRS Bull.*, 19(08):30–40, aug 1994.

[219] Peter R. Wells. Linear Free Energy Relationships. *Chem. Rev.*, 63(2):171–219, apr 1963.

[220] J. Wencel-Delord, A. Panossian, F. R. Leroux, and F. Colobert. Recent advances and new concepts for the synthesis of axially stereoenriched biaryls. *Chem. Soc. Rev.*, 44(11):3418–3430, 2015.

[221] Ian H. Williams and Gerald M. Maggiora. Use and abuse of the distinguished-coordinate method for transition-state structure searching. *J. Mol. Struct. THEOCHEM*, 89(3-4):365–378, oct 1982.

[222] Joost Wintterlin. Scanning tunneling microscopy studies of catalytic reactions. volume 45, pages 131–206. 2000.

[223] Penghao Xiao, Juliana Duncan, Liang Zhang, and Graeme Henkelman. Ridge-based bias potentials to accelerate molecular dynamics. *J. Chem. Phys.*, 143(24):244104, 2015.

[224] Jin-Fei Yang, Rong-Hua Wang, Yin-Xia Wang, Wei-Wei Yao, Qi-Sheng Liu, and Mengchun Ye. Ligand-Accelerated Direct CH Arylation of BINOL: A Rapid One-Step Synthesis of Racemic 3,3-Diaryl BINOLs. *Angew. Chemie Int. Ed.*, 55(45):14116–14120, nov 2016.

[225] Francisco Zaera. The surface chemistry of thin film atomic layer deposition (ALD) processes for electronic device manufacturing. *J. Mater. Chem.*, 18(30):3521, 2008.

[226] Nikolai A. Zarkevich and Duane D. Johnson. Nudged-elastic band method with two climbing images: Finding transition states in complex energy landscapes a). *J. Chem. Phys.*, 142(2):024106, jan 2015.

[227] Yi Zeng, Penghao Xiao, and Graeme Henkelman. Unification of algorithms for minimum mode optimization. *J. Chem. Phys.*, 140(4):044115, jan 2014.

[228] Qingfan Zhang, Bo Han, Xiaowei Tang, Kevin Heier, Jimmy X Li, John Hoffman, Minfa Lin, Stephanie L Britton, Agnes Derecskei-Kovacs, and Hansong Cheng. On the Mechanisms of Carbon Formation Reaction on Ni(111) Surface. *J. Phys. Chem. C*, 116(31):16522–16531, aug 2012.

[229] Xiao-Jie Zhang and Zhi-Pan Liu. Variable-Cell Double-Ended Surface Walking Method for Fast Transition State Location of Solid Phase Transitions. *J. Chem. Theory Comput.*, 11(10):4885–4894, oct 2015.

[230] Xiaoqian Zhang, Xiao-Xi Li, Yufang Liu, and Yong Wang. Suicide Inhibition of Cytochrome P450 Enzymes by Cyclopropylamines via a Ring-Opening Mechanism: Proton-Coupled Electron Transfer Makes a Difference. *Front. Chem.*, 5(January):1–10, jan 2017.

[231] B. Zhao, F. P. Guengerich, A. Bellamine, D. C. Lamb, M. Izumikawa, L. Lei, L. M. Podust, M. Sundaramoorthy, J. A. Kalaitzis, L. M. Reddy, S. L. Kelly, B. S. Moore, D. Stec, M. Voehler, J. R. Falck, T. Shimada, and M. R. Waterman. Binding of Two Flaviolin Substrate Molecules, Oxidative Coupling, and Crystal Structure of Streptomyces coelicolor A3(2) Cytochrome P450 158A2. *J. Biol. Chem.*, 280(12):11599–11607, mar 2005.

[232] Hongjuan Zhu and Tom Ziegler. Probing the Influence of Trans and Leaving Ligands on the Ability of Square-Planar Platinum(II) Complexes to Activate Methane. A Theoretical Study. *Organometallics*, 28(9):2773–2777, may 2009.

[233] Paul Zimmerman. Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. Theory Comput.*, 9(7):3043–3050, jul 2013.

[234] Paul M. Zimmerman. Automated discovery of chemically reasonable elementary reaction steps. *J. Comput. Chem.*, 34(16):1385–1392, jun 2013.

[235] Paul M. Zimmerman. Growing string method with interpolation and optimization in internal coordinates: Method and examples. *J. Chem. Phys.*, 138(18):184102, 2013.

[236] Paul M. Zimmerman. Navigating molecular space for reaction mechanisms: an efficient, automated procedure. *Mol. Simul.*, 41(1-3):43–54, feb 2015.

[237] Paul M. Zimmerman. Navigating molecular space for reaction mechanisms: an efficient, automated procedure. *Mol. Simul.*, 41(1-3):43–54, feb 2015.

[238] Paul M Zimmerman. Single-ended transition state finding with the growing string method. *J. Comput. Chem.*, 36(9):601–611, apr 2015.

[239] Paul M. Zimmerman, Franziska Bell, Matthew Goldey, Alexis T. Bell, and Martin Head-Gordon. Restricted active space spin-flip configuration interaction: Theory and examples for multiple spin flips with odd numbers of electrons. *J. Chem. Phys.*, 137:1–12, 2012.