

# Molecular Volcano Plots: Tools for Rationalizing and Predicting the Performance of Homogeneous Catalysts

Présentée le 9 avril 2021

Faculté des sciences de base  
Laboratoire de design moléculaire computationnel  
Programme doctoral en chimie et génie chimique

pour l'obtention du grade de Docteur ès Sciences

par

**Boodsarin SAWATLON**

Acceptée sur proposition du jury

Prof. J. Waser, président du jury  
Prof. A.-C. Corminboeuf, directrice de thèse  
Dr A. Poblador-Bahamonde, rapporteuse  
Dr T. Gensch, rapporteur  
Prof. R. Buonsanti, rapporteuse



*If you want to live and work outside the 'system',  
then you have to live on your wits.  
But that's ok.*

— David Weininger



# Acknowledgements

First of all, I address my gratitude to my advisor, Prof. Clémence Corminboeuf for giving me an opportunity to be a part of LCMD and for her instantly responses, supports and inspiration. During my PhD life, I had many chances to attend conferences, schools, and seminars thanks to her “yes” answers every time I asked for permission and she also encouraged me to take an opportunity to present/talk in front of audiences. Besides, she is very careful and concerned about the life of LCMD crews outside work, which is a bonus.

Secondly, for my colleagues in LCMD, I specifically thank Dr. Matthew D. Wodrich for being a great mentor and collaborator and for his generous help of critical reading all my works. As we both have worked on molecular volcanoes, he became my second advisor who I had discussed and argued with all the time. Many ideas and works have been done together during these 4 years. The second person I would express my gratitude is Dr. Benjamin Meyer who is the first collaborator I had worked with in LCMD. Thanks to him, I have learned a lot about the job, bash script, and also general things outside work. Knowing him, his personality, and his energetic activities inspire me to be more active and easy-going person. I feel lucky to meet and work with him here. Also, I would address my appreciation to Dr. Alberto Fabrizio, Dr. Kun-Han Lin, Raimon Fabregat, and Veronika Juraskova for their instant helps no matter how busy they were and for being understandable as we were all in the same situation as doctoral students for years. I would love to mention here some of former and current LCMD members who spent valuable time with me during my PhD life: Dr. Antonio Prlj, Dr. Ganna (Any) Gryn’ova, Dr. Laurent Vannay, Dr. Sergio (Sergi) Vela and Dr. Maria Fumanal. Moreover, I listed LCMD members I have been interacted with during four years here as a personal memorial that our time in LCMD was overlapped: Dr. Stepan Ruzicka, Dr. Jian-Hao Li, Dr. Giulia Mangione, Thierry Tran, Daniel Hollas, Théo Inizan, Manuel Cordova, Edurne Avellanal, Dr. Shubhajit Das, Sinjini Bhattacharjee, Terence (Terry) Blaskovits, Alan Scheidegger, Ksenia Briling, Simone Gallarati, Dr. Marc Garner, Dune Andre, Pit Steinbach, Dr. Rubén Laplaza, Dr. Frédéric Celérse and Puck Van Gerwen. Also, my doctoral period will not be this smooth if I do not have a hand from Véronique Bujard and Daniel Jana for administration stuff and cluster management.

Being a doctoral student at EPFL is a big step far from my home but everything went very well thanks to my trusted and caring family. I am lucky that they accept and support my decisions without any doubt. During the time in Switzerland, I am happy to be a part of the association of Thai Students in Switzerland (ATSS) where I have met many new Thai friends and did activities together. I also thank Thai community at EPFL who are very thoughtful and supportive,

## Acknowledgements

---

especially Teerawat Songsichan (P’ Bank), Nukorn Plainpan (Ice), and Surapat Ek-in (Arm). Notably, I have met a lot of people here who become an important part of my life. I am grateful to Livio Di Tria who always listens to my annoying stories and supports me during a doctoral crisis and to Dimitri Vaudroz and his family for always staying by my side and encouraging me. My life would be different without him. I hope he has a happy life and enjoys every moment in the future.

I am also extremely grateful to Prof. Jérôme Waser, Prof. Raffaella Buonsanti, Dr. Amalia I. Poblador-Bahamonde and Dr. Tobias Gensch for having accepted to be my jury and for providing a fruitful discussion during the oral exam. Finally, the National Centre of Competence in Research (NCCR) “Materials’ Revolution: Computational Design and Discovery of Novel Materials (MARVEL)”, the Swiss National Science Foundation (SNSF) (Grant No. 200020\_175496) and the EPFL are acknowledged for financial support.

*Lausanne, 16 March 2021*

# Abstract

The generation of catalyst design tools is important for developing both economically and environmentally friendly reactions. This thesis focuses on the use and development of such a tool, molecular volcano plots, which have the ability to estimate catalytic performance while also providing an ameliorated understanding of the intricate chemistry that influence reactions. The original work presented here is divided into three parts that cover: (1) exploration of volcano functions, (2) acceleration of catalyst-screening process and (3) extraction of the detailed chemistry for interesting chemical reactions from volcano results.

The first examines additional functions of volcano plots in which, for the first time, substrate scopes can be screened using volcano plots as similar as catalysts. This type of volcanoes can estimate the reactivity of each substrate and also provide an enhanced understanding of substrate scope, an important facet of chemistry that is often overlooked by computation.

The second section focuses on accelerating the catalyst-screening process through the use of machine-learning models. Specifically, we illustrate how machine-learning models can be used to predict the value of a descriptor variable that can be related to catalyst performance through volcano plots. This procedure allows us to expand the number of catalyst being screened to tens of thousands of species which, in turn, provides a much broader picture of catalyst behavior. The use of big-data techniques highlights the specific manner in which metals and ligands can be combined to identify tuned catalyst for a given chemical reaction.

In the final section, we use molecular volcanoes to probe the catalytic hydrogenation of carbon dioxide to formate using transition metals paired with pincer ligands. This work ultimately identified a combination of group 9 metals with  $\pi$ -acidic pincer-ligands as the best catalysts for this reaction. Using this same reaction, we also developed a molecular volcano variant that directly predicts an experimental observable, the turnover frequency (TOF), from the value of a descriptor variable, thus establishing “TOF volcanoes”.

Overall, this thesis demonstrates how molecular volcanoes can be used as a rational catalyst design tool in the field of homogeneous catalysis as well as route to uncovering chemical trends that provide greater fundamental understanding about why specific catalysts and/or substrates exhibit high functionality for a particular reaction.

## Abstract

---

**Keywords:** molecular volcano plots, linear free energy scaling relationships, catalyst design, homogeneous catalysts, metal-ligand combinations, chemical intrinsic, effects of catalysts and substrates



# Résumé

La génération d'outils pour le design de catalyseurs est importante pour développer des réactions à la fois économiques et respectueuses de l'environnement. Cette thèse se concentre sur l'utilisation et le développement d'un tel outil, les volcano plot moléculaires, qui permettent d'estimer les performances catalytiques d'une molécule tout en fournissant une meilleure compréhension de la chimie qui influence les réactions. Le travail original présenté ici est divisé en trois parties qui couvrent : (1) l'exploration des fonctions du volcan, (2) l'accélération du processus de sélection des catalyseurs et (3) l'analyse des transformations chimiques pour des réactions intéressantes à partir des résultats du volcan.

La première partie examine les fonctions additionnelles des volcano plots dans lesquelles un seul volcan peut être utilisé pour sélectionner et estimer les performances de catalyseurs étant caractérisées par différents ligands, états d'oxydation et états de spin. En outre, pour la première fois, nous montrons comment les volcano plots peuvent être utilisés pour avoir une meilleure compréhension du domaine d'application du substrat, un aspect important de la chimie souvent négligé dans les calculs.

La deuxième section se concentre sur l'accélération du processus de sélection des catalyseurs grâce à l'utilisation de modèles d'apprentissage automatique. Plus précisément, nous montrons comment les modèles d'apprentissage automatique peuvent être utilisés pour prédire la valeur d'un descripteur directement lié aux performances du catalyseur via les volcano plots. Cette procédure nous permet d'augmenter le nombre de catalyseurs potentiels à des dizaines de milliers d'espèces qui, à leur tour, nous permettent d'établir une image beaucoup plus complète des caractéristiques du cycle catalytique. L'utilisation de techniques d'analyse sur des grandes bases de données nous a permis de découvrir comment les métaux et les ligands peuvent être combinés afin d'obtenir le catalyseur idéal pour chaque réaction chimique.

Dans la dernière section, nous appliquons les volcano plot moléculaires au problème de l'hydrogénation catalytique pour transformer le dioxyde de carbone en formiate avec des métaux de transition en combinaison avec des ligands pinces. Ce travail a finalement identifié une combinaison de métaux du groupe 9 avec des ligands pinces  $\pi$ -acides comme les meilleurs catalyseurs pour cette réaction. En utilisant cette même réaction, nous avons également développé une variante de volcan moléculaire qui prédit directement une observable expérimentale, la fréquence de turnover (TOF), à partir de la valeur d'une variable descriptive, établissant ainsi les «volcanos TOF».

## Résumé

---

Dans l'ensemble, cette thèse démontre comment les volcans moléculaires peuvent être utilisés comme outil de conception rationnelle de catalyseurs dans le domaine de la catalyse homogène. En outre, cette thèse montre que ces outils peuvent être utilisés comme moyen de découvrir les tendances chimiques et obtenir une meilleure compréhension des raisons pour lesquelles certains catalyseurs et / ou substrats présentent une activité élevée pour une réaction particulière.

**Mots clefs:** les volcano plot moléculaires, libre énergie linéaire, relation de mise à l'échelle, conception de catalyseur, homogène catalyseur, combinaisons des ligands métalliques, chimie intrinsèque, effets des catalyseurs et substrats

# Contents

<b>Acknowledgements.....</b>	<b>v</b>
<b>Abstract.....</b>	<b>vii</b>
<b>Résumé .....</b>	<b>ix</b>
<b>Contents .....</b>	<b>xi</b>
<b>1 Introduction .....</b>	<b>1</b>
<b>2 The Genesis of Molecular Volcano Plots .....</b>	<b>5</b>
2.1 History and The Origin of Volcano Plots .....	5
2.2 Thermodynamic Volcano Plots – Proof-of-Concept .....	6
2.3 Kinetic Volcano Plots – Proof-of-Concept.....	9
<b>3 Probing Substrate Scope with Molecular Volcanoes.....</b>	<b>13</b>
3.1 Introduction .....	13
3.2 Computational Details .....	14
3.3 Results and Discussion .....	15
3.4 Conclusions .....	20
<b>4 Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts .....</b>	<b>21</b>
4.1 Introduction .....	21
4.2 Computational Details .....	24
4.3 Methods .....	25
4.3.1 Database .....	25
4.3.2 Training .....	26
4.4 Results and Discussion .....	27
4.4.1 Machine Learning.....	27
4.4.2 Catalyst Prediction .....	29
4.5 Conclusions .....	31

<b>5</b>	<b>Data Mining the C–C Cross-Coupling Genome.....</b>	<b>33</b>
5.1	Introduction.....	33
5.2	Computational Details.....	35
5.2.1	DFT Computations and Machine Learning .....	35
5.2.2	Sketch-Maps .....	36
5.3	Results and Discussion .....	36
5.4	Conclusions.....	46
<b>6</b>	<b>Unraveling Metal/Pincer Ligand Effects in the Catalytic Hydrogenation of Carbon Dioxide to Formate .....</b>	<b>47</b>
6.1	Introduction.....	47
6.2	Computational Details.....	49
6.3	Results and Discussion .....	49
6.3.1	Linear Free Energy Scaling Relationships.....	49
6.3.2	Molecular Volcano Plots .....	50
6.4	Conclusions.....	53
<b>7</b>	<b>Activity-Based Screening of Homogeneous Catalysts through the Rapid Assessment of Theoretically Derived Turnover Frequencies .....</b>	<b>55</b>
7.1	Introduction.....	55
7.2	Computational Details.....	57
7.3	Results and Discussion .....	58
7.3.1	Screening with TOF Volcanoes .....	61
7.3.2	Influence of Temperature on TOF Volcanoes .....	63
7.3.3	Strengths and Weaknesses of the Model .....	64
7.4	Conclusions.....	66
<b>8</b>	<b>General Conclusions and Outlook.....</b>	<b>67</b>
<b>A</b>	<b>Supplementary for Probing Substrate Scope with Molecular Volcanoes .....</b>	<b>71</b>
A.1	Construction of Substrate Volcano Plots .....	71
A.2	Correlation between the Charge on Pd, Substrate Sterics and the Energy of Transmetalation .....	75
<b>B</b>	<b>Supplementary for Machine Learning Meets Volcano Plots and Data Mining of the C–C Cross-Coupling.....</b>	<b>77</b>
B.1	Ligand Database.....	77
B.2	Distribution of Binding Energies in the Training Set .....	82
B.3	Sketch-Maps (Energy Descriptors vs HOMO Energies) of Complex 1.....	83
B.4	The Shifts of Ligands from P–M–O References .....	85

---

<b>C</b>	<b>Supplementary for Volcano Plots of the Hydrogenation of Carbon Dioxide to Formate</b>	<b>87</b>
C.1	Comparison of Reaction Mechanism .....	87
C.2	Construction of the Molecular Volcano Plot .....	88
C.3	Linear Free Energy Scaling Relationships on Different Temperatures .....	90
C.4	Temperature Influence on the TOF Volcano .....	91
	<b>Bibliography .....</b>	<b>93</b>
	<b>Curriculum Vitae .....</b>	<b>117</b>



# 1 Introduction

Catalysts are essential components of chemical reactions, as over 90% of industrial chemical products include catalytic reactions as part of their production process.<sup>1</sup> The development of new catalysts is, therefore, necessary for further improving reaction efficiency and selectivity, as well as for identifying cheaper and/or milder operating conditions that are highly beneficial on an industrial scale. To accomplish this, various tools have been developed for screening and predicting the performance of prospective new catalysts including: multidimensional modeling using a design of experiments (DoE),<sup>2</sup> high throughput screening,<sup>3</sup> linear scaling relationships (LSRs) and volcano plots,<sup>4-7</sup> as well as others.<sup>8-11</sup> High throughput screening and LSRs/volcano plots are regularly used in electrocatalysis,<sup>12-13</sup> heterogeneous<sup>14-15</sup> and homogeneous catalysis<sup>16-18</sup> whereas multidimensional modeling using DoE has mainly been limited to the study involving asymmetric homogeneous catalysis.<sup>2, 19-25</sup> While a strict requirement, the increased use of computational modeling has led to the creation of the significant amounts of data that has further aided in the development and application of these tools. On one hand, computational modeling is a practical path for uncovering the details of a specific catalytic process.<sup>11</sup> Information such as the active form of a catalyst, the details of reaction mechanisms, and the relative energetic cost of completing reaction steps can all be elucidated, which can each be used to guide the development of future reactions. The principal drawback, however, is that individual computations on a single catalytic system are time-consuming and not necessarily transferable to other (even related) systems. Thus, a clear need exists for further developing tools that combine large quantities of data about different catalysts and establish trends that predict catalytic properties (*i.e.*, reaction energies, turnover frequencies).

The notion and application of linear scaling relationships (LSRs) are peppered throughout the history of chemistry. Pertinent examples include the Hammett equation<sup>26-29</sup> and Bell–Evans–Polanyi principle<sup>30-31</sup> from physical organic chemistry, the Brønsted catalysis equation<sup>32</sup> from homogeneous catalysis, and the general use of such relationships in electrocatalysis.<sup>33-34</sup> By plotting LSRs, a relationship between two parameters, *e.g.*, energies, rate constants, and bond distances, can be established. Linearity within the plot indicates that an association between two states is universal for a certain condition. Importantly, LSRs allow reaction behavior to be defined in terms of a single descriptor variable, which can be used to construct volcano plots (*vide infra*). Because the relative energies of different species in the catalytic cycle depend on one another, their relative stabilities can be determined through the use of the LSRs when the value of the descriptor variable is known. In other words, the energy

profiles that dictate catalytic behavior are concealed within the LSRs and can be unraveled by knowing this single descriptor. Moreover, the mathematical equations obtained from LSRs can be processed further to construct volcano plots, which can relate intrinsic properties of the catalyst to a descriptor variable.

Volcano plots are built upon Sabatier's principle, which states that the interaction between the substrate and an ideal catalyst should be neither too strong nor too weak.<sup>35-36</sup> This concept can be depicted in a volcano-shaped plot, where the two volcano slopes that define the overall shape correspond to catalyst/substrate interactions being either "too strong" (left slope) or "too weak" (right slope). The plateau region located on the top of the volcano represents a balanced interaction, which is a hallmark of ideal catalysts. Formerly, volcano plots were exclusively used for applications in electrocatalysis and heterogeneous catalysis. In 2015, however, the concept of volcano plots was successfully transferred to homogeneous catalysis by our group.<sup>5</sup> The first "molecular volcano plot" depicted catalyst performance for a model Suzuki–Miyaura cross-coupling reaction. Following this proof-of-principle study, the molecular volcano concept has been used to identify new and promising catalysts for various chemical reactions and, by incorporating other algorithms/tools such as machine learning represents a capable tool for the large-scale screening of catalysts and rationalizing the behavior of homogeneous catalysis.

This overarching theme of this thesis involves developing and utilizing molecular volcano plots as tools for fine-tuning catalytic reactions. The material is organized following the three specific aspects we wanted to examine: (1) explore the capability and functionality of molecular volcano plots, (2) develop ways for reducing the computational cost associated with screening catalysts, and (3) apply volcano plots to study various chemically relevant reactions to understand the chemical information hidden underneath.

**Chapter 2** provides the reader a brief history as well as discussion on the state-of-the-art of molecular volcano plots, which has also been published in *Accounts of Chemical Research*.<sup>37</sup>

Since the beginning, volcano plots are often used to investigate the behavior of catalysts; specifically, for homogeneous catalysis, the effects of each component of the catalysts *i.e.*, transition metals and ligands have been examined using molecular volcano plots. However, volcanoes can also be used to better understand the impact of other reaction components. The reacting substrates are a key element that influence the catalytic process, where modifying functional groups or core structures can result in significant changes to overall activity. **Chapter 3** shows how molecular volcano plots are used to provide an enhanced picture of reaction substrate scope by examining a series of electrophile substrates (allyl bromide) for a Suzuki–Miyaura C–C cross-coupling reaction. This work clearly demonstrates that the ability of volcano plots is not limited to examining only catalysts and unlocks a door that links the energetics of the catalyst and substrate worlds. This communication has been published in *Organic Letters*.<sup>38</sup>

Molecular volcanoes are computational based tools, yet the process of catalyst screening that requires computing the value of the descriptor variable remains costly when large numbers of catalysts must be analyzed. As such, finding alternative approaches that reduce the computational burden during screening is paramount to fully explore catalyst space. Machine-learning (ML) models, statistical based approaches that rapidly analyze huge amounts of data, have previously



---

been developed and exploited to solve many chemical problems. In **Chapter 4**, we demonstrate how a ML model can be combined with molecular volcano plots to rapidly predict the descriptor variables and ascertain the species having the best thermodynamics from a pool of over twenty-thousand catalyst candidates for a Suzuki–Miyaura cross-coupling reaction. This article has been published in *Chemical Science*.<sup>39</sup>

The results obtained from the aforementioned machine-learning/volcano plot tandem not only can be used to estimate the descriptor values for a large number of catalysts but also to extract chemical trends using methods borrowed from big-data analysis. In **Chapter 5**, we employ dimensionality reduction algorithms (*i.e.*, Sketch-Maps)<sup>40-41</sup> to illustrate and decipher hidden patterns in the chemistry of cross-coupling reaction that reflect the different behavior and the magnitude to which altering the metals and ligands changes the catalytic cycle energy profiles. This article has been published in *ChemCatChem*.<sup>42</sup>

The third objective of this thesis aims at utilizing volcano plots to study interesting chemical reactions. In **Chapter 6**, we use molecular volcanoes to elucidate the effects of metal–pincer ligands for the hydrogenation of carbon dioxide to formate, an important feedstock product that can be used in numerous chemical applications. The volcano plots focus on how coupling certain metals with specific ligands can lead to ideally tuned catalysts possessing superior energetic profiles. This article has been published in *Organometallics*.<sup>43</sup>

Typically, molecular volcanoes distinguish “good” from “bad” catalysts by plotting the most energetically costly reaction step in the catalytic cycle as a function of the descriptor variable. However, evaluating catalysts using this direct energy criterion provides no direct relation to measurements of catalytic activity obtained from experiment. In **Chapter 7**, we use the same hydrogenation of carbon dioxide to formate reaction to create a volcano variant that directly relates an experimental observable, the turnover frequency (TOF), with the value of the descriptor variable. This article has been published in *ACS Catalysis*.<sup>44</sup>

Finally, **Chapter 8** concludes the thesis by summarizing the main findings of molecular volcano plots with respect to the three overarching objectives. Future work is suggested regarding potential further developments that will further raise the applicability and utility of volcano plots for homogeneous catalysis.



## 2 The Genesis of Molecular Volcano Plots

This chapter is partly based on following publication:

Wodrich M. D., Sawatlon B., Busch M., and Corminboeuf C. The Genesis of Molecular Volcano Plots, *Acc. Chem. Res.* **2021**, 54, 1107-1117.

### 2.1 History and The Origin of Volcano Plots

Volcano plots are efficient tools for optimizing catalytic reactions and, correspondingly, have found widespread use within many areas of catalysis. Historically, these plots are derived from an over century old idea, Sabatier's principle,<sup>35-36</sup> which states that an ideal catalyst should neither bind the products too strongly nor the reactants too weakly. This concept was put into practice in the 1950s independently by Gerischer<sup>45</sup> and Parsons<sup>46</sup> as a means to visualize the activity of metals towards electrochemical H<sub>2</sub> evolution. As in modern volcano plots, these early works utilize a descriptor variable (*e.g.*, the binding energy of hydrogen) plotted along the *x*-axis and a measurement of catalytic activity (*e.g.*, the experimental current density) plotted on the *y*-axis. In line with Sabatier's principle, the resulting volcano shape (Figure 2.1) demonstrates a clear relationship between the descriptor variable value and catalytic activity. The volcano shape can be subdivided into three regions: the left slope where the catalyst binds intermediates too strongly (*i.e.*, the "strong-binding" side of the volcano), the right slope where catalysts that bind intermediates too weakly are found, (*i.e.*, the "weak-binding" side of the volcano) and finally an ideal binding region on the volcano plateau where catalysts bind intermediates neither too strongly nor too weakly. It is catalysts that fall into this region (*i.e.*, those located at or near the volcano plateau or peak) that satisfy Sabatier's principle, and are likely to be amongst the best catalysts for a given reaction.

Despite the appealing simplicity of the volcano concept, only a handful applications appeared in the literature in the following decades.<sup>47-48</sup> Beginning in the early 2000s, they began appearing with increasing frequency, largely due to the work of Nørskov. In part, this growth in applications can be ascribed to the rise of density functional theory (DFT) computations that provided direct access to the key energetic quantities needed to construct the plots, specifically the energy of the descriptor intermediate and a measure of activity such as the thermodynamic overpotential.<sup>49-51</sup> The subsequent realization that linear free energy scaling relationships (LFESRs) govern the relative stabilities of different catalytic cycle intermediates in the form of binding energies<sup>52-53</sup> as well as transition state barriers<sup>51, 54</sup> further provided a simply route for obtaining the overall volcano shape for any given reaction mechanism. Correspondingly, the global limitations for many reactions of interest were obtained, as exemplified by Rossmeisl for the water oxidation reaction over transition metal oxides.<sup>55-56</sup> Other, more recent advancements, include employing

volcano plots to examine the stability of materials<sup>57</sup> or accounting for competing reaction mechanisms.<sup>58-59</sup>

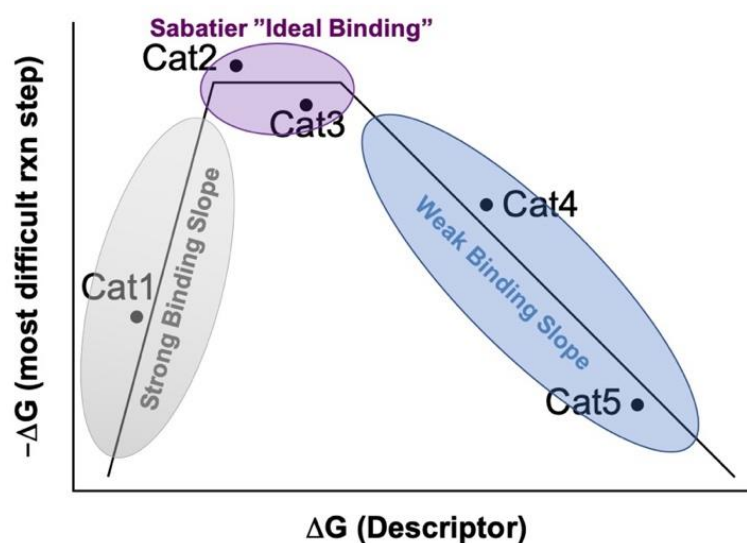


Figure 2.1 Schematic overview of a volcano plot. The descriptor value is plotted along the  $x$ -axis and the negative of the free energy of the most difficult reaction step of the catalytic cycle (or another measure of catalytic activity) along the  $y$ -axis. Catalysts with the best thermodynamic/kinetic profiles (*e.g.*, Cat2 and Cat3) appear at or near the volcano plateau (or peak) in the Sabatier ideal binding region (purple), catalysts having overly strong catalyst/substrate interactions (*e.g.*, Cat1) appear along the left “strong binding” slope (gray), while catalysts with overly weak catalyst/substrate interactions (*e.g.*, Cat4 and Cat5) appear along the “weak binding” right slope (blue).

In spite of their simplicity and predictive power, volcano plots and LFESRs have remained nearly exclusively the province of electrocatalysis and heterogeneous catalysis. Although proposed in an abstract manner by Swiegers in 2008,<sup>60</sup> in 2015 our group first constructed “molecular volcano plots”<sup>55</sup> by examining a prototypical homogenous catalysis reaction, Suzuki–Miyaura cross-coupling. This initial work opened a new research line aimed at, broadly speaking, identifying new and improved catalysts for relevant chemical problems by developing and expanding a computational framework based on creating and applying volcano plots to the study of homogeneous catalysis. This chapter explores the key developments and chemical findings that constitute the emerging field of molecular volcano plots.

## 2.2 Thermodynamic Volcano Plots – Proof-of-Concept

Our initial work<sup>5</sup> began by identifying a suitable homogeneously catalyzed reaction with available experimental data that would serve as an important “check” on the validity of the results obtained from our yet unproven molecular volcanoes. As a proof-of-principle example, we examined a Suzuki–Miyaura cross-coupling reaction of a vinylbromide and a vinyl boronic acid to form butadiene (Equation 2.1). Invoking a protocol typically used in heterogeneous/electrocatalysis, we restricted ourselves to analyzing only thermodynamic aspects of the catalytic cycle (*i.e.*, intermediates only) for a series of 36 catalysts comprised of six metal centers (Ni, Pd, Pt, Cu, Ag, Au) appended with a set of six electronically diverse ligands. From this catalyst set, we identified

robust linear free energy scaling relationships (LFESRs) between the relative free energies of the different catalytic cycle intermediates. This finding was a key requirement that allowed us to proceed with constructing molecular volcano plots. Ultimately, the relative energy of intermediate **2** (equivalent to oxidative addition, Rxn A, in Figure 2.2) was found to be the best descriptor variable. Post-processing the LFESRs<sup>i</sup> led to the Figure 2.3a volcano plot, where the descriptor variable is plotted along the *x*-axis and the negative of the most difficult reaction step in the catalytic cycle (by convention referred to as the potential determining step, pds) is given on the *y*-axis. Adding points for the 36 catalysts to the volcano showed that the results aligned well with experimental trends; palladium catalysts appear near the plateau while most other group 10 metal catalysts (nickel and platinum) are located along the volcano's strong-binding left slope (where the free energy of Rxn C is the most energetic-costly catalytic cycle step) of the volcano. On the other hand, all coinage metal catalysts were found along the volcano's weak binding right slope (where the free energy of Rxn A is the most energetic-costly catalytic cycle step), indicative of generally worse thermodynamic profiles than the group 10 catalysts. Overall, the existence of unambiguous LFESRs, the reconstitution of a typical volcano shape, and the location of experimentally verified catalysts atop the volcano validates that the volcano concept is indeed applicable to homogeneous catalysis.

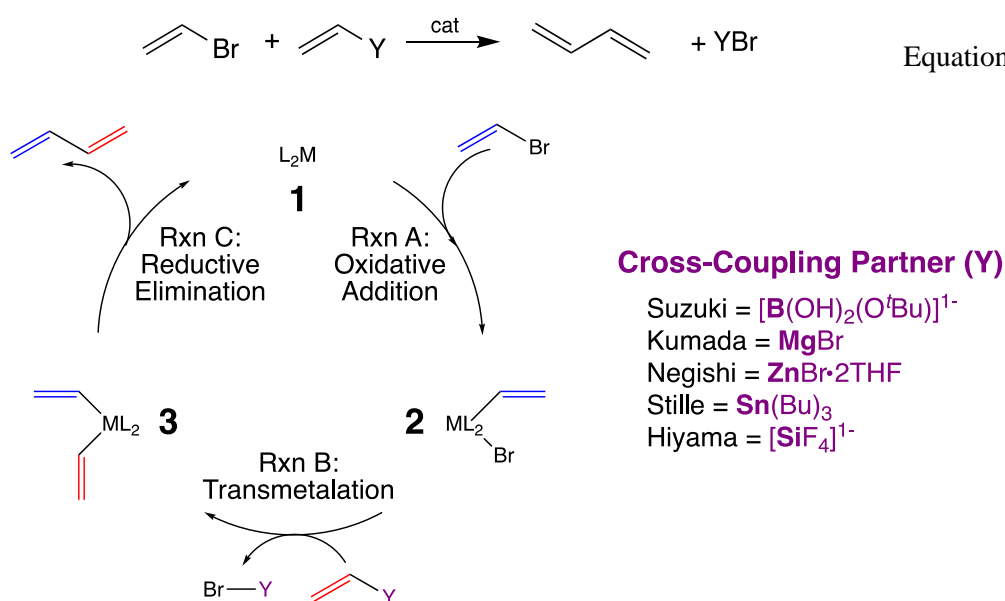


Figure 2.2 Abbreviated catalytic cycle depicting the key steps for C–C cross-coupling reactions.

While our initial work established the viability of molecular volcano plots, a number of questions concerning their overall robustness and transferability persisted. In cross-coupling reactions, for instance, some catalysts (*e.g.*, those bulky ligands) likely transit the catalytic cycle in a monoligated rather than bisligated (as examined in our original work) state. Moreover, it is also conceivable that some catalysts exist in higher oxidation and/or alternative spin states, which would also influence reactivity. This raises the question: do these alterations lead to separate sets of LFESRs and unique volcano plots? To answer this question, we reexamined Suzuki cross-coupling, placing emphasis on how changes in the ligation, spin, or oxidation state influence both the accuracy of the underlying LFESRs and the volcano shape.<sup>61</sup> We found that while the overall

<sup>i</sup> The example on how to construct volcano plots is shown in **Appendix A** and **C**.

thermodynamic profiles of the catalysts change when the ligation, spin, or oxidation states are altered, that the same sets of LFESRs remain. As a result, a single volcano capably describes catalysts with any (or all) of the aforementioned changes. This is nicely illustrated by the monoligated Ni catalysts shown in Figure 2.3b, which possess good thermodynamic profiles when traversing the catalytic cycle in Ni(I)/Ni(III) oxidation state, but are less active in their Ni(0)/Ni(II) state. Indeed, Figure 2.3b shows that the change in oxidation state from Ni(0)/Ni(II) to Ni(I)/Ni(III) causes a weakening of interaction between the catalyst and the reaction intermediates that results in a rightward shift in the volcano plot that corresponds to improved thermodynamic profiles.

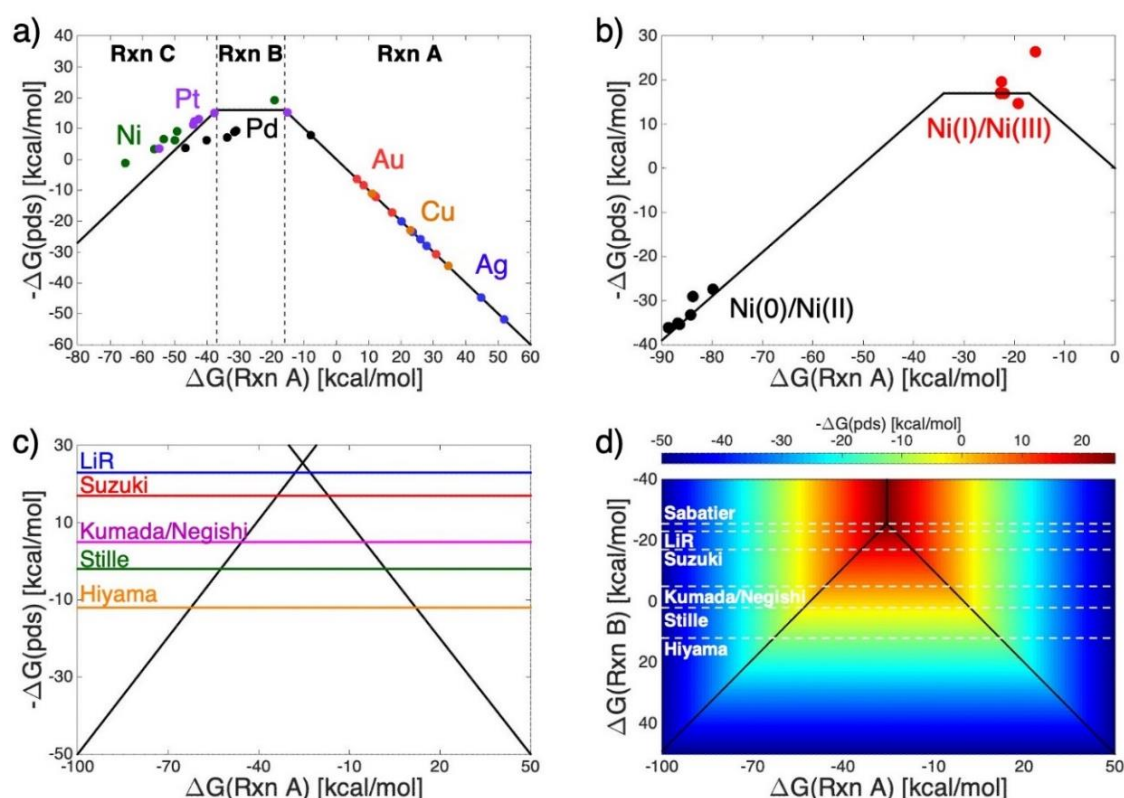


Figure 2.3 (a) Thermodynamic volcano plot for a C–C cross-coupling reaction. (b) Volcano plot highlighting the thermodynamic influence induced by changes in oxidation state. Note that the same linear free energy scaling relationships and volcano plot describe both sets of species. (c) Volcano plot depicting changes in the transmetalation energy from different named chemical reactions and (d) a three-dimensional volcano showing the energetic relationship of the different cross-coupling variants.

The previously described works applied molecular volcanoes to a Suzuki–Miyaura cross-coupling reaction. However, Suzuki–Miyaura is only one variant of a number of well-known “named” cross-coupling reactions (*e.g.*, Kumada, Negishi, Stille, Hiyama) that differ only by the chemical agent employed during transmetalation (“Y” in Rxn B, Figure 2.2). Indeed, the free energy associated with transmetalation is often the largest thermodynamic barrier encountered in the catalytic cycle for “good” catalysts (*i.e.*, this reaction defines the volcano plateau). Thus, altering the cross-coupling partner influences the energy of transmetalation that dictates the height of the volcano plateau (Figure 2.3c), where more reactive coupling partners (*e.g.*, LiR)

raise and less reactive partners (*e.g.*,  $\text{SiF}_4^{1-}$ , Hiyama) lower the plateau height in the corresponding volcano plots.

Establishing the relationship between each of these individual cross-coupling volcanoes allows a unified thermodynamic picture of cross-coupling reactions to be formed, which can be displayed as a three-dimensional volcano plot (Figure 2.3d).<sup>62</sup> In essence, this three-dimensional volcano is most easily understood as a series of standard two-dimensional volcanoes (*e.g.*, Figure 2.3a) pasted together, where the descriptor variable is displayed on the *x*-axis and  $-\Delta G(\text{pds})$  on the *z*-axis. The three-dimensional volcano also uses the transmetalation energy as a second descriptor variable (*y*-axis). Knowing the value of the transmetalation step associated with a specific chemical reagent (*e.g.*,  $\text{SiF}_4^{1-}$  for Hiyama coupling) reveals the location of the specific “slice” that gives the corresponding two-dimensional volcano. In this case, as the transmetalation step becomes easier (*i.e.*, moves toward more exergonic transmetalation energies), the value of the pds decreases (seen in Figure 2.3d as a shift from blue to green to red). This increase in the reaction’s thermodynamic drive is accompanied by a narrowing of the volcano plateau, meaning that fewer catalysts will have ideal thermodynamic profiles. Ultimately, the transmetalation becomes so energetically facile that the volcano plateau disappears entirely (represented by the Sabatier line, Figure 2.3d), leaving only a peak where the energies associated with either reductive elimination (left slope) or oxidative addition (right slope) dictate the catalytic cycle thermodynamics. Overall, this generalized picture of C–C cross-coupling provides routes to improving the thermodynamics of any catalyst by identifying the cross-coupling variant that leads to the most energetically balanced catalytic cycle. The three-dimensional volcano concept can also be used to examine other facets of homogeneous catalytic reactions, such as the energetic role played by the electrophilic coupling component in cross-coupling reactions.<sup>63</sup>

## 2.3 Kinetic Volcano Plots – Proof-of-Concept

The molecular volcanoes described in the first section considered only catalytic cycle thermodynamics, however, we knew that kinetic aspects must also be considered to have a meaningful impact in homogeneous catalysis. With this in mind, we quickly went to work on “kinetic volcanoes”<sup>64</sup> by examining a model hydroformylation reaction (Equation 2.2 by following the catalytic cycle in Figure 2.4) with a catalyst database created by combining eight metal centers taken from groups 8-10 with four monodentate phosphine ligands having different steric parameters ( $\text{PH}_3$ ,  $\text{PMe}_3$ ,  $\text{PPh}_3$ ,  $\text{PCy}_3$ ). This work showed that transition state barriers could be predicted with acceptable accuracy directly from the descriptor variable and reproduced experimental trends with Rh catalysts being found near the volcano peak (Figure 2.5a). A closer examination of the linear free energy scaling relationships (LFESRs) revealed that differences in ligand bulkiness led to slightly different sets of scaling relationships. By constructing separate volcanoes for each ligand, we demonstrated that increasing steric bulk reduces the key transition state barriers and leads to more active catalysts (Figure 2.5b). Further, establishing relationships between the peaks of the different volcanoes and a measure of steric bulk led to simple structure-activity relationships that allowed the height of the volcano peak (corresponding to anticipated catalytic activity) for any phosphine ligand to be directly estimated from the Tolman cone angle.

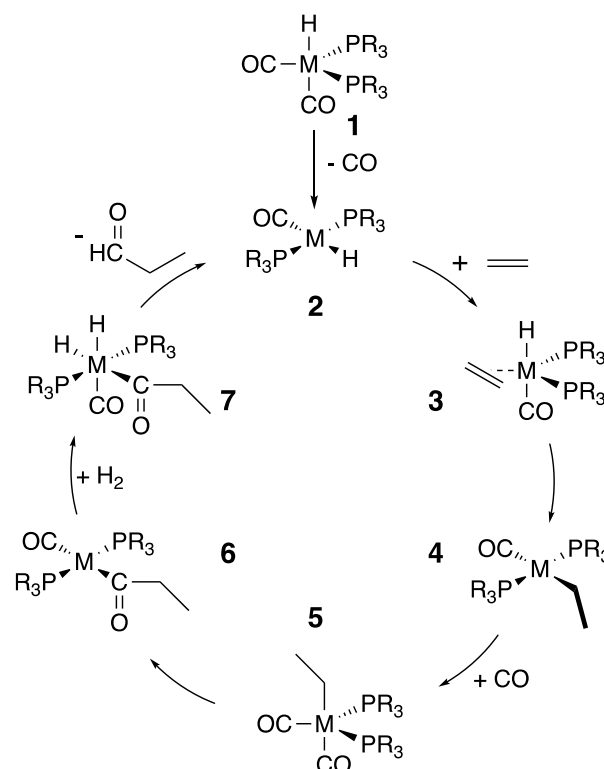
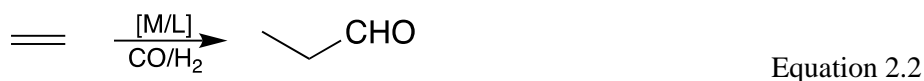
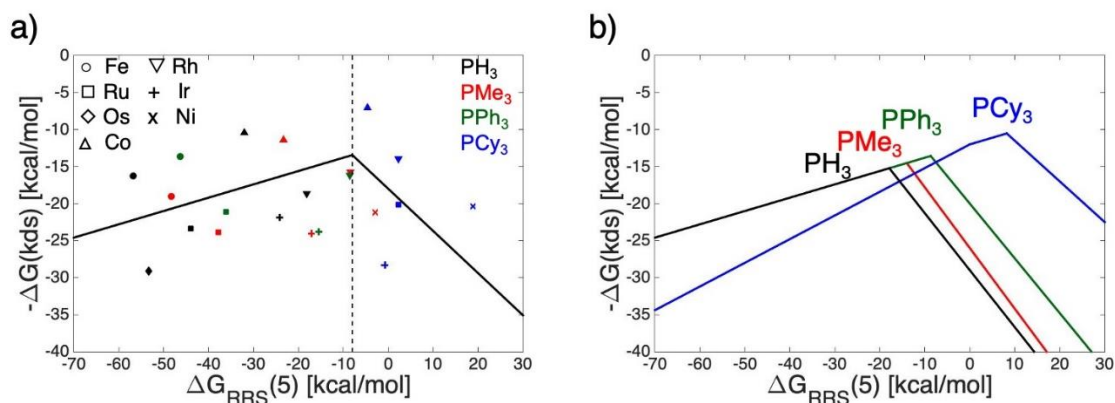


Figure 2.4 Proposed catalytic cycle for the hydroformylation reaction with ethylene substrate.

Figure 2.5 (a) Kinetic volcano plot for the hydroformylation reaction using ethylene as a substrate. Left slope corresponds to  $3 \rightarrow \text{TS}_{3,4}$  and right slope to  $6 \rightarrow \text{TS}_{6,7}$ . (b) Ligands separated based on their sterics reveal increasing bulk yields a higher volcano corresponding to a catalytic cycle with lower free energy barriers.

Eager to expand upon the kinetic volcano concept to increase their usefulness to address problems relevant to synthetic chemists, we revisited the hydroformylation reaction by examining the ability of volcanoes to predict product regioselectivity.<sup>65</sup> Our initial application used ethylene as a substrate which leads to only one product (Equation 2.2). To introduce regioselectivity, we chose 2-methylpropene as a substrate (Equation 2.3), which leads to the formation of two



regioisomers, a branched (2,2-dimethylpropanal) and a linear species (3-methylbutanal). In addition to identifying active catalysts (*i.e.*, those with low transition state barriers) directly from a descriptor variable, we also wanted to accurately predict the regiomer excess (*re*) values. Obviously, this represents a challenging task given the sensitivity of the *re* to small changes in the free energy differences between product distinguishing transition states. Since identifying highly accurate LFESRs was going to be key, we restricted our analysis to rhodium catalysts bearing bidentate phosphine ligands. Ultimately, two volcano plots were constructed [one for the linear (blue) and one for the branched (green) product shown in Figure 2.7a while their mechanism are shown in Figure 2.6] that estimated both catalytic activity and selectivity through the use of two descriptor variables. Remarkably, of ten catalysts predicted to be highly selective (*i.e.*, *re* > 95), nine had similar *re* values when the free energy differences of the transition states that dictate regioselectivity were directly computed. After construction, the volcano plots were used to screen a database of 68 additional ligands to search for active species that selectively form the more elusive branched regioisomer. This procedure ultimately identified several promising ligands (Figure 2.7b).

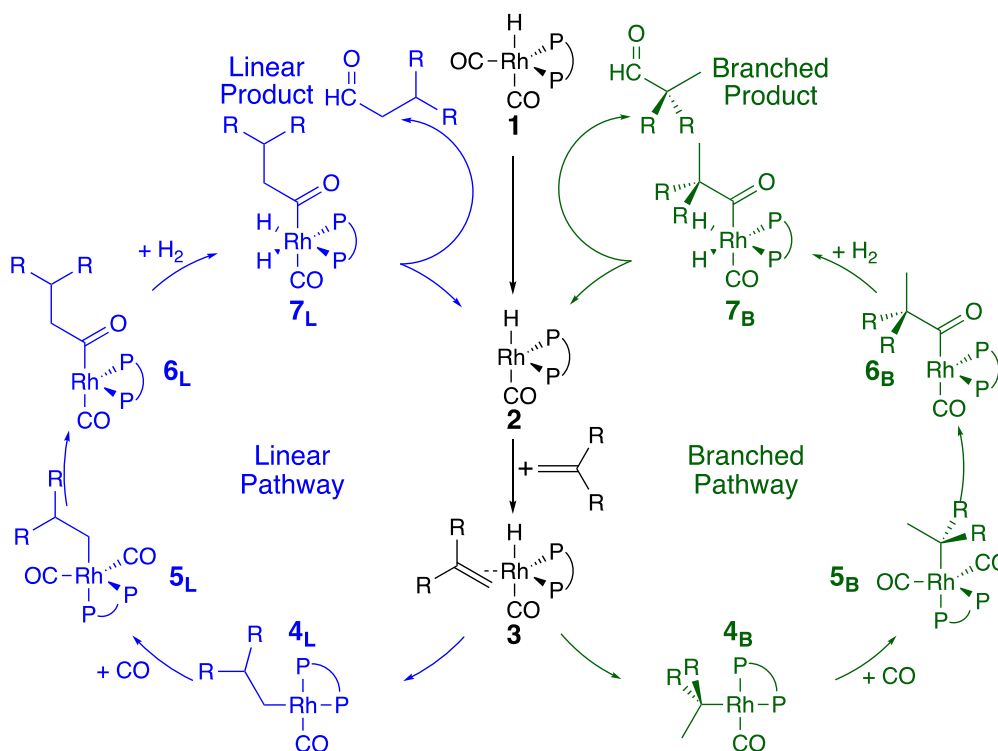
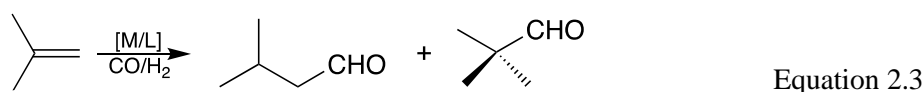


Figure 2.6 Proposed catalytic cycle for the hydroformylation reaction with a substituted substrate leading to two regioisomers.

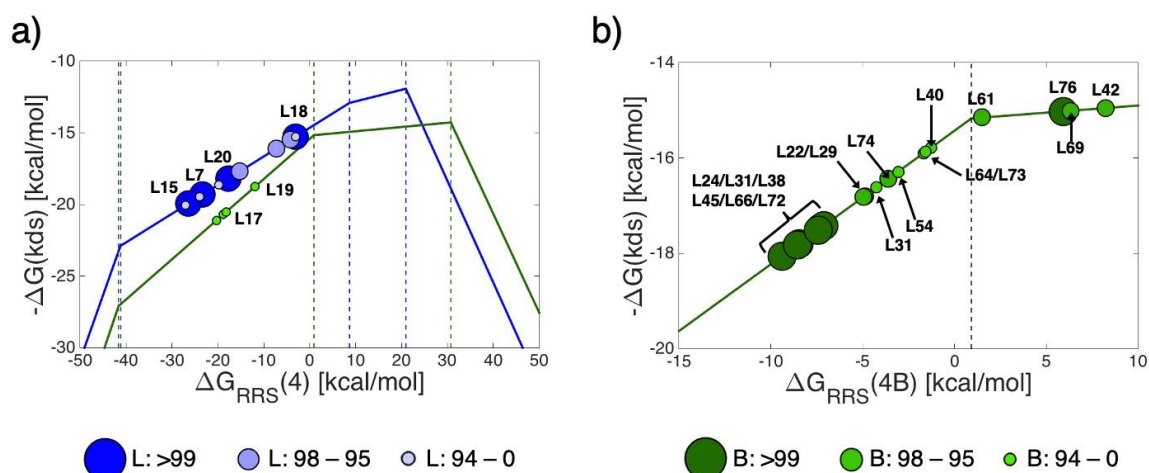


Figure 2.7 (a) Regioselective molecular volcano plots [linear product (blue), branched product (green)] for hydroformylation. Circle size indicates the range of  $r_e$  values for each catalyst according to the provided color codes. (b) Screening for catalysts that selectively form the branched product revealed several ligand structures with predicted  $r_e$  values greater than 90.

Overall, the molecular volcanoes described above represented the state-of-the-art at the time this thesis was started, which primarily focused on using volcanoes to estimate the catalytic performance of organometallic species by probing the thermodynamic or kinetic free energy profiles of various metal/ligand combinations. However, these early works represent only a small part of a potentially much larger story, as numerous possibilities exist to extend and apply these tools to a wide array of problems in homogeneous catalysis. The proceeding chapters further explore the functionality of molecular volcanoes, while also developing and refining the framework to permit new aspects of homogeneous catalysis to be examined.

## 3 Probing Substrate Scope with Molecular Volcanoes

This chapter is based on following publication:

Sawatlon B., Wodrich M. D., and Corminboeuf C. Probing Substrate Scope with Molecular Volcanoes, *Org. Lett.* **2020**, 22, 7936–7941.

### 3.1 Introduction

Improving the activity and selectivity of chemical reactions on both laboratory and industrial scales often focuses on catalyst design, where the constituent metal atom(s) and accompanying ligands represent key components. Correspondingly, the chemistry of metal/ligand combinations and their environments has been the subject of in-depth studies aimed at gaining insight into catalytic efficacy. Whereas the catalyst choice clearly affects the reactivity, so do different chemical reactants, additives, as well as changes in reaction conditions. As an example, for Suzuki–Miyaura cross-coupling,<sup>66–68</sup> countless substrates and other conditions have been reported with large variations in overall efficiency.

Regarding changes to the electrophilic substrate specifically, bromobenzene and aryl boronic acid effectively undergo cross-coupling in an 89% yield. This reactivity, however, can be enhanced by adding electron-withdrawing or electron-donating substituents to the para position. Conversely, substitutions to the ortho position produce lower isolated yields.<sup>69–71</sup> Moreover, changing the location of the halide on the electrophile (*e.g.*, 4-bromopyridine is more inert than 2-bromopyridine and 3-bromopyridine)<sup>72</sup> or heteroatom substitution (*e.g.*, the reaction with dibenzofuran proceeds in 12 h at room temperature,<sup>73</sup> whereas dibenzothiophene requires 24 h at 80 °C)<sup>74</sup> further influences reactivity. Clearly, different substrates are governed by unique free energy profiles that can differ profoundly from one another.

Today, computational-based approaches represent important tools in mechanism discovery, catalyst design, and the extraction of chemical information.<sup>2, 6, 8–11</sup> Yet computational studies (and tools built upon them) often place greater emphasis on understanding a specific reaction mechanism or characterizing the energetics of a series of catalysts while much less frequently compiling a picture of how different reactant substrates influence the overall energetics. Among the tools used for catalyst identification are volcano plots, which relate an energy-based descriptor variable to the thermodynamics/kinetics of a catalytic process. Using these plots, the activity of any prospective catalyst can be estimated by knowing the value of the corresponding descriptor variable. In 2015, our research group first built molecular volcano plots depicting the

thermodynamics of a model Suzuki–Miyaura cross-coupling reaction.<sup>5, 61</sup> This proof-of-principle study succeeded in transferring the concept from heterogeneous/electrocatalysis<sup>4, 75-76</sup> to homogeneous catalysis and was followed by additional theoretical refinements<sup>39, 44, 62, 64</sup> as well as applications to interesting chemical problems.<sup>43, 77</sup>

Whereas molecular volcanoes have been applied numerous times to estimate catalytic performance, they have never been used to examine the substrate scope. The purpose of this Chapter is to present a tool that facilitates the investigation of catalytic cycle energetics brought about by changes in the reactant substrate. Specifically, we evaluate Suzuki–Miyaura cross-coupling to study the impact caused by different electrophiles (blue species in Figure 3.1). The results obtained by using these plots, in turn, both illuminate the range of substrates that can be accommodated by a single catalyst and reveal strategic routes toward identifying better catalysts for any individual substrate. It is important to emphasize that whereas the chemical trends concerning the reaction studied here are generally well known, the underlying computational tool is broadly applicable and can be used to analyze any desired reaction.

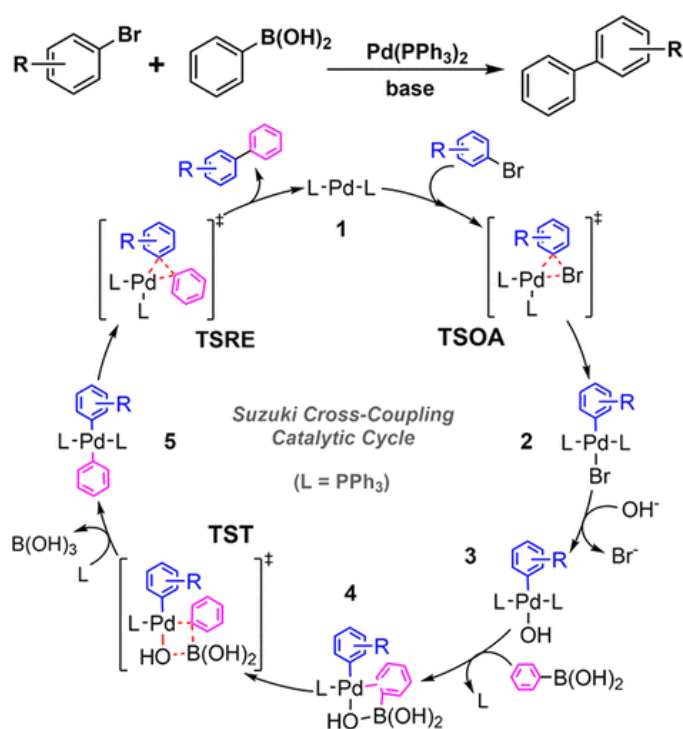


Figure 3.1 Catalytic cycle of Suzuki–Miyaura cross-coupling using Pd(PPh<sub>3</sub>)<sub>2</sub> as a catalyst.

## 3.2 Computational Details

All intermediates, transition states and chemical reactants were optimized using PBE0<sup>78-80</sup> density functional with Grimme's dispersion correction (–D3)<sup>81</sup> along with the def2-SVP basis set<sup>82</sup> in implicit THF solvent using the SMD solvation model<sup>83</sup> in Gaussian16 program.<sup>84</sup> The “ultrafine” integration grid was used to prevent an error on the computed free energies resulting from the lack of rotational invariance.<sup>85</sup> The analysis of the vibrational frequencies confirmed all geometries as stationary points on the potential energy surface (zero imaginary frequencies for intermediates and one imaginary frequency for transition states). As non-covalent interactions between aromatic rings can stabilize the catalytic cycle intermediates, final structures were

checked to ensure the presence of similar orientations of the aromatic rings that correspond to the global minima (Figure 3.2).

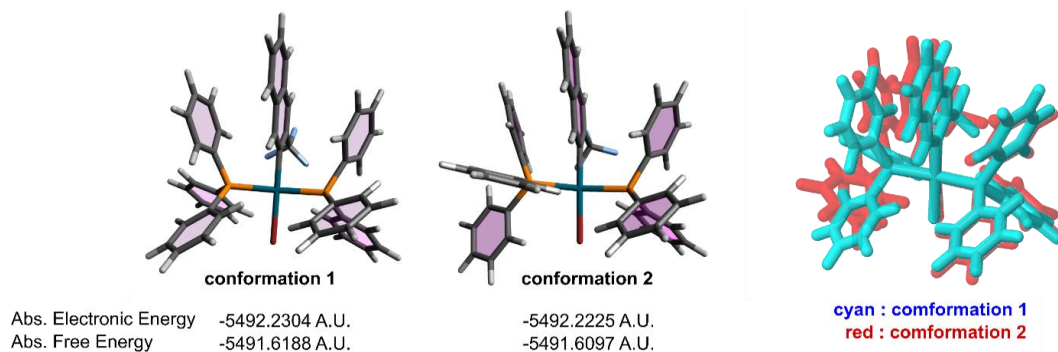


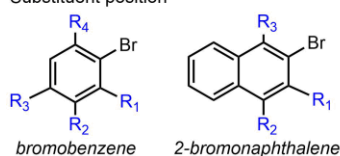
Figure 3.2 Comparison of conformation of intermediate **2** with 2-bromo-3-(trifluoromethyl)naphthalene substrate.

### 3.3 Results and Discussion

Suzuki–Miyaura cross-coupling has three main mechanistic steps: oxidative addition, transmetalation, and reductive elimination (Figure 3.1). Whereas the transition-state geometries associated with oxidative addition (**TSOA**) and reductive elimination (**TSRE**) are unambiguous, the transmetalation (**TST**) process has historically been the subject of controversy.<sup>86-92</sup> Here we follow the experimentally verified mechanism proposed by Denmark *et al.*,<sup>93-94</sup> which reveals that the Pd–O–B linkages (**4** in Figure 3.1) are formed prior to transmetalation, which is followed by a four-membered ring transition state (Pd–O–B–Ar, **TST**) and finally, the transmetalation product (**5** in Figure 3.1). Because our objective is to examine the substrate scope, the catalytic cycle always employs Pd(PPh<sub>3</sub>)<sub>2</sub> as the catalyst and phenylboronic acid as the nucleophile. Figure 3.3 shows the electrophile substrates being studied, which can be separated into two groups based on their skeletal structures as substrates containing: (1) six-membered rings (6MRs, 57 species) and (2) five-membered rings (5MRs, 24 species).

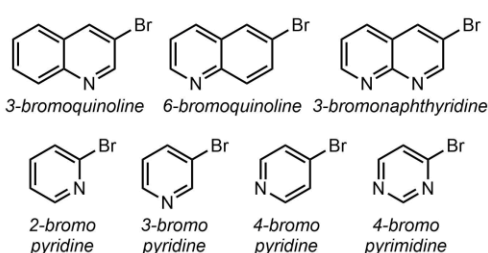
#### 6-membered rings

Substituent position



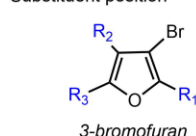
Substituent group

-NO<sub>2</sub> -CF<sub>3</sub> -OMe -Me -<sup>i</sup>Bu -Cl



#### 5-membered rings

Substituent position



Substituent group

-NO<sub>2</sub> -CF<sub>3</sub> -OMe -<sup>i</sup>Bu

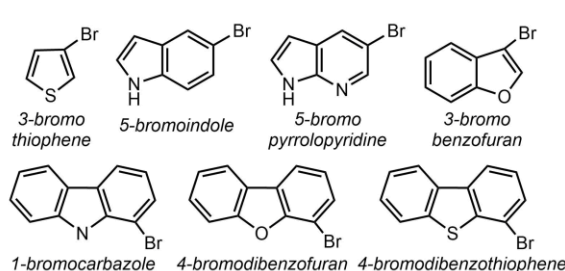
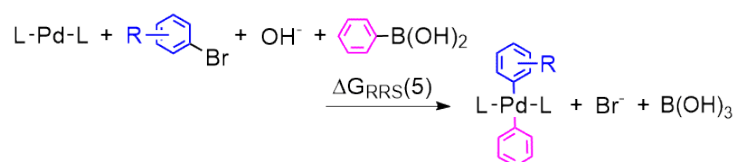


Figure 3.3 Substrates studied.



Equation 3.1

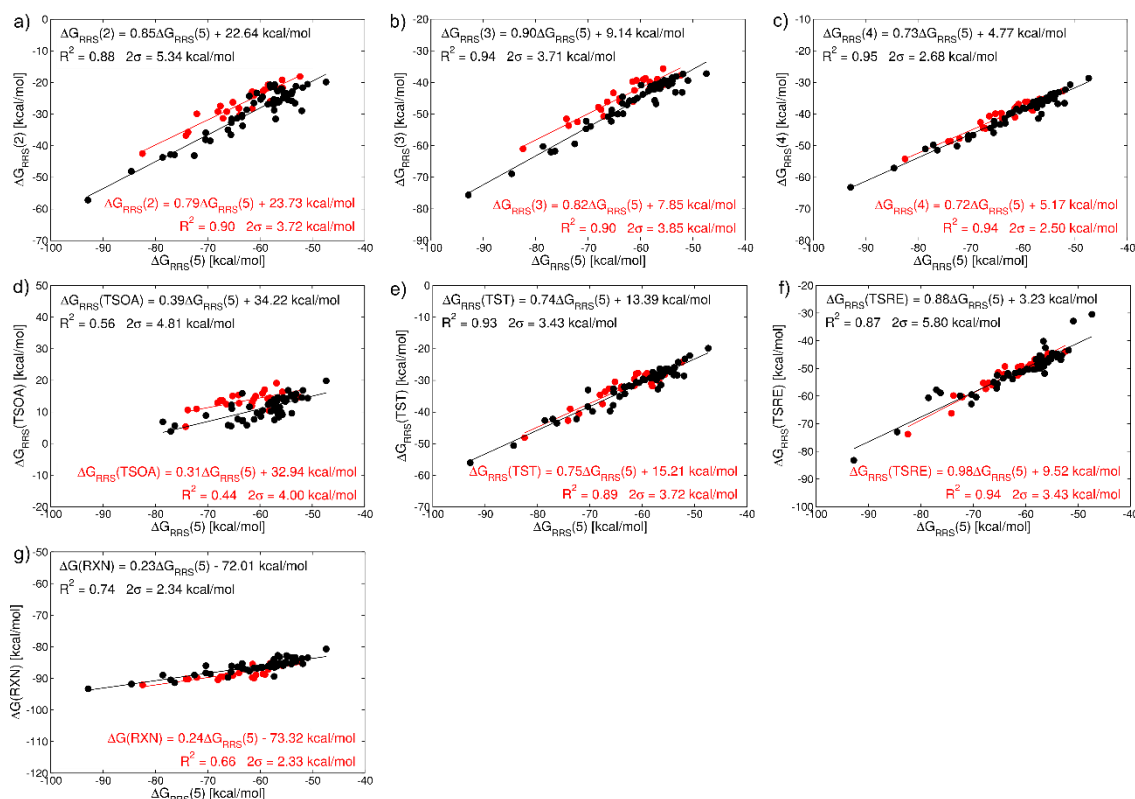


Figure 3.4 Linear free energy scaling relationships of 6-membered rings (black) and 5-membered rings (red). X-axis is the chosen descriptor  $\Delta G_{\text{RRS}}(5)$ ; y-axis represents (a)  $\Delta G_{\text{RRS}}(2)$ , (b)  $\Delta G_{\text{RRS}}(3)$ , (c)  $\Delta G_{\text{RRS}}(4)$ , (d) total reaction energy [ $\Delta G(\text{RXN})$ ], (e)  $\Delta G_{\text{RRS}}(\text{TSOA})$ , (f)  $\Delta G_{\text{RRS}}(\text{TST})$  and (g)  $\Delta G_{\text{RRS}}(\text{TSRE})$ .

Linear free energy scaling relationships (LFESRs) form the bases of volcano plots by showing the relationship between various catalytic cycle intermediates/transition states and a descriptor variable. On the basis of the quality of the linear fits (Figure 3.4),  $\Delta G_{\text{RRS}}(5)$ , as defined by Equation 3.1, was chosen to be the descriptor. The LFESRs show a clear distinction between the behavior of the 6MR and 5MR substrates, which requires a unique volcano plot to be constructed for each.<sup>i</sup> After post-processing the LFESRs (see **Appendix A** for details), the two volcano curves shown in Figure 3.5 are obtained. The axes of the volcano plot correspond to the value of the descriptor variable (x-axis) and the negative of the energy span of the catalytic cycle (y-axis). In essence, the energy span is a measurement of the difference between the lowest energy intermediate [*i.e.*, the catalyst resting state, termed the turnover determining intermediate (TDI)] and the highest transition-state barrier [*i.e.*, the turnover determining transition state (TDTS)].<sup>95-97</sup>

<sup>i</sup> Note that the LFESRs of both the 5MR and 6MR substrate groups are clearly separate for some intermediates and transition state but not for others. A different parsing of the substrate database (*i.e.*, using alternative criterion to separate the substrates into groups) might also represent a route to gain additional insight into substrate effects using volcano plots.

As shown in the Figure 3.5 plot, the turnover determining states for both the 6MR and 5MR substrates are identical. The volcano's left slopes are governed by the energy necessary to separate the formed product from the catalyst (TDI: **5**, TDTS: **1**), the right slopes are governed by the energy associated with the oxidative addition (TDI: **1**, TDTS: **TSOA**), and the central regions are governed by the energy associated with transmetalation (TDI: **3**, TDTS: **TST**).<sup>ii</sup> Comparing the overall shapes, the plateau region for the 5MR substrates is flatter and slightly higher than that for the 6MR substrates, implying that a wide range of 5MR substrates traverse the catalytic cycle with roughly equivalent energetics, whereas moving away from the highest point of the volcano invokes a more significant energy penalty for the 6MR substrates.

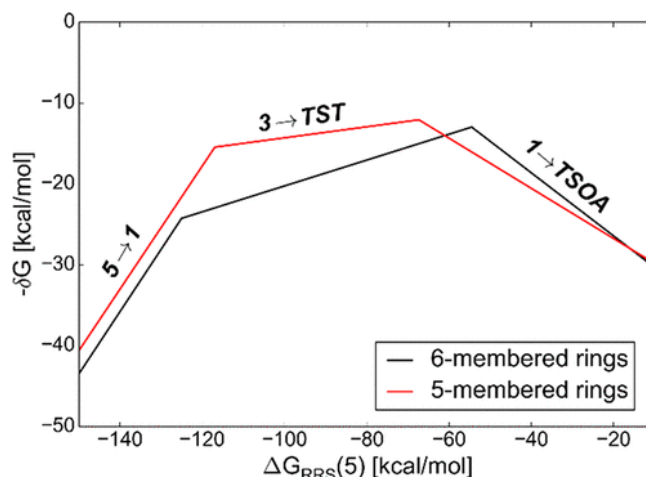


Figure 3.5 Molecular volcano plots of six-membered rings (6MRs, black) and five-membered rings (5MRs, red).

To examine the influence of substituent groups, we placed each of the Figure 3.3 substrates into the appropriate volcano plot (Figure 3.6a for substituted bromobenzenes, Figure 3.6b for substituted bromonaphthalenes, and Figure 3.6c for substituted bromofurans) based on the value of its descriptor variable. The location of each substrate defines its most energetically costly reaction steps, which (for the range of descriptor variables for the substrates) is either transmetalation (left slope) or oxidative addition (right slope). Both plots representing the 6MR (Figure 3.6a and b) display similar trends, with most substrates clustering near the peak, including unsubstituted compounds (gray inverted triangle, Figure 3.6a; gray square, Figure 3.6b). However, substrates with multiple (star) and ortho substitutions (triangle) have more exergonic descriptor values and are, correspondingly, shifted far from the top along the left slope. This indicates that they have difficulty with transmetalation. Substrates with multiple functional groups at both ortho positions suffer from additional steric hindrance around the location of bond breaking/formation, which further suppresses the transmetalation rate as steric bulk impedes the coordination of phenylboronic acid to the palladium center (Figure 3.1). Notably, independent of the substitution position, the greater the electron-withdrawing ability of the substituent, the larger the leftward shift away from the volcano peak.

Electronic effects on the metal center are known to influence the transmetalation rate; for instance, the presence of electrophilic palladium causes an acceleration.<sup>94</sup> When bromobenzene

<sup>ii</sup> Note that the location and the nature of these reaction steps are equivalent to our previously derived thermodynamic volcano plots for Suzuki–Miyaura cross-coupling.



binds to the metal center, the electron-withdrawing substituent groups on the substrates enhance the electrophilicity of palladium, yielding a faster process. However, substrates with electron-withdrawing groups can also become sluggish during transmetalation when they are ortho-substituted. Hence, the actual reason for slow transmetalation is steric hindrance and not electronic effects (for a more detailed discussion as well as correlation plots between electronic and steric parameters with transmetalation energies, see **Appendix A**), as noted by Harvey *et al.*<sup>98</sup> as well as by Denmark *et al.*, who stated, “the steric parameters are twice as important as the electronic parameters in the transmetalation.”<sup>94</sup> This statement is consistent with our substrate distributions that show that the substitution position plays a larger role in dictating catalytic cycle energetics than the electronic nature of the functional group.

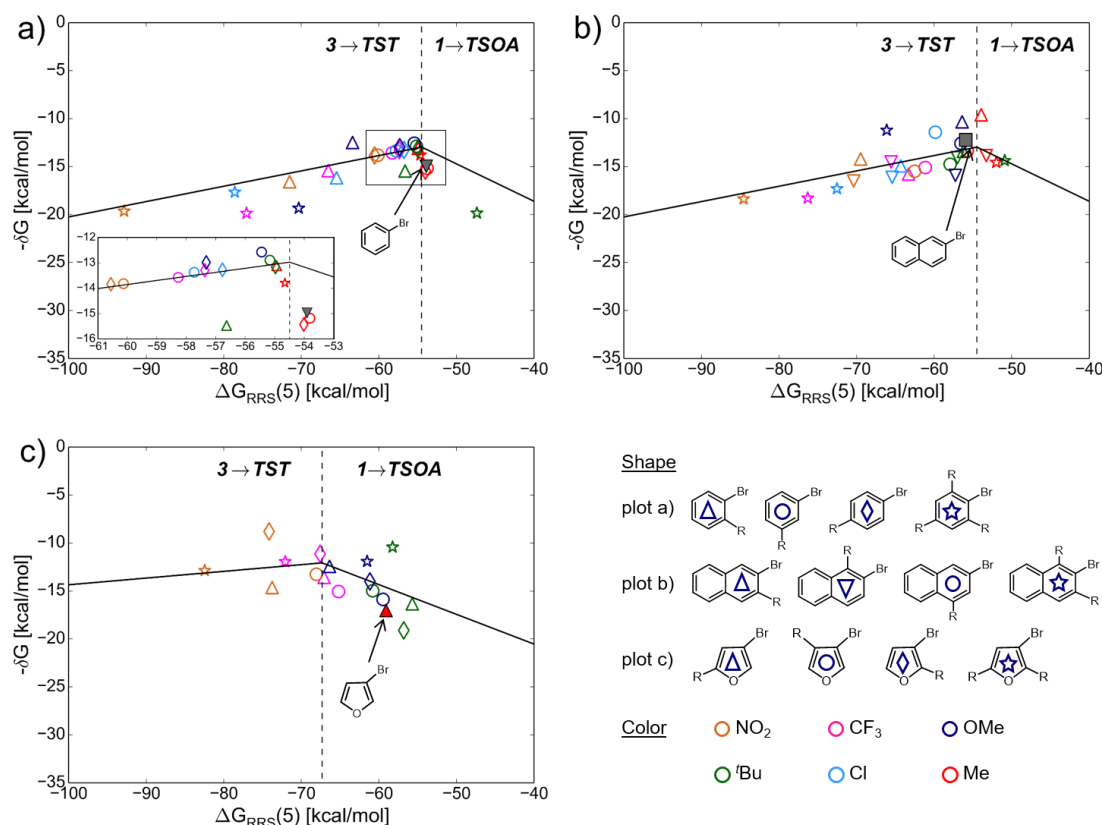


Figure 3.6 Volcano plots illustrating the influence of substituents on (a) six-membered rings with substituted bromobenzenes, (b) six-membered rings with substituted bromonaphthalenes, and (c) five-membered rings with substituted bromofurans.

Methyl- and *t*-butyl-substituted reactants deviate from the previously described trends, as these substrates are found near the volcano peaks or shifted along the right slope. Thus, electron-donating groups have a different impact than electron-withdrawing groups, regardless of the substitution position, as seen in the Figure 3.6c plot. Here, all methyl-substituted compounds are found to the right of the peak, where oxidative addition is rate-determining. Generally speaking, for oxidative addition to proceed, a substrate must be relatively electron-poor to accept two electrons from the metal center. Substituted bromobenzenes with electron-donating groups are electron-rich, making oxidative addition difficult. Furan is even more electron-rich than benzene and other aromatic six-membered rings, and thus is it unsurprising that nearly all substituted furan derivatives have difficulty undergoing oxidative addition. On the contrary, adding electron-



withdrawing groups to furan balances the intrinsic stereoelectronics and induces a leftward shift toward regions of higher overall activity. Altogether, these plots clearly capture that both the steric and the electronic properties of the substrate must be balanced to ensure a facile cross-coupling process.

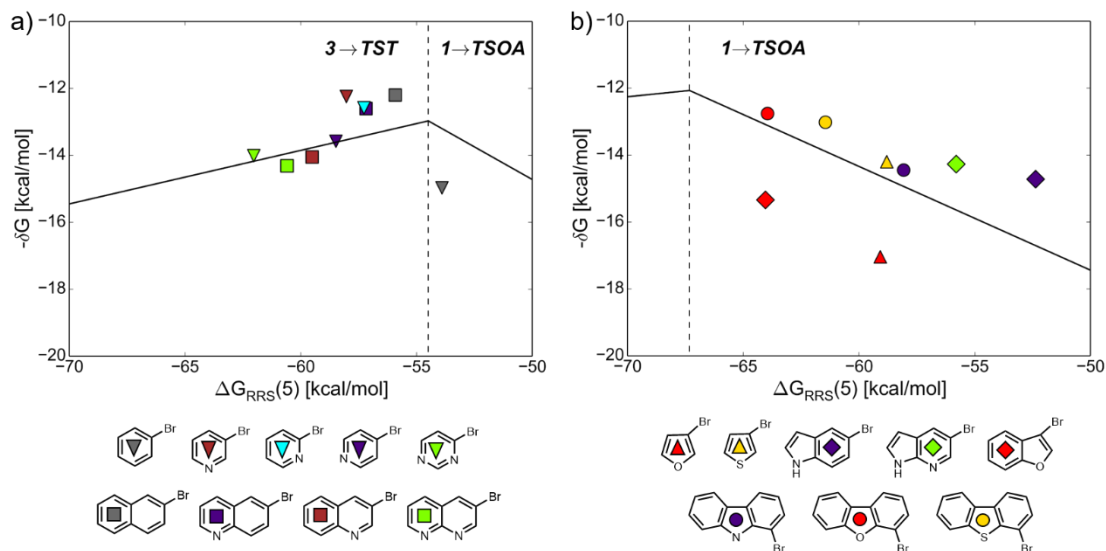


Figure 3.7 Volcano plots illustrating the influence of the substrate core on (a) six-membered rings and (b) five-membered rings.

Turning to the substrate core, Figure 3.7 shows that overall, the core influence is less pronounced than the influence of the substituents. Most 6MR cores are clustered relatively close to the volcano peak (Figure 3.7a), whereas the 5MR cores are shifted along the right slope (Figure 3.7b). For the 6MR, various bromopyridines and bromoquinolines can be used to unravel the influence of the heteroatom position. Within the same ring, having the heteroatom in the para position (*e.g.*, 4-bromopyridine, blue inverted triangle) retards the transmetalation process, resulting in a diminished experimental yield.<sup>72</sup> On the contrary, moving the heteroatom far from the reaction site into the outer ring (*e.g.*, from 3-bromoquinoline to 8-bromoquinoline) decreases the heteroatom effect, causing 8-bromoquinoline (blue square) to appear similar to naphthalene (gray square). Likewise, the presence of a second heteroatom increases the effect, resulting in a leftward shift away from the area of maximum activity (green points on the Figure 3.7a volcano).

Similar trends are observed for bromoindole (blue diamond) and bromopyrrolopyridine (green diamond) in the 5MR plot (Figure 3.7b). Here the influence of different heteroatoms can easily be examined. Considering tricyclic structures [bromocarbazole (blue circle), bromodibenzofuran (red circle), and bromodibenzothiophene (yellow circle)], the substrate containing a more electronegative oxygen atom is located closest to the volcano peak, whereas the less electronegative sulfur and nitrogen variants lie further down the right slope, indicating a diminished rate for oxidative addition. This results from the fact that an electronegative atom can reduce the electron density on the reaction site, making oxidative addition faster.

Figure 3.7b shows that the 5MR substrates predominantly lie to the right of the volcano peak (indicative of an overly weak catalyst/substrate interaction), potentially making cross-coupling using  $\text{Pd}(\text{PPh}_3)_2$  (the catalyst examined here) too energetically difficult. However, this does not

mean that these substrates are incapable of undergoing this reaction, as is well-known experimentally. Rather, the energetic profiles of these substrates can be improved by using stronger binding catalysts (*e.g.*, platinum and nickel complexes)<sup>5, 42</sup> that will result in a more balanced catalyst/substrate interaction and a corresponding improvement in the catalytic cycle free energy profile. Such a finding nicely illustrates how molecular volcanoes examining the substrate scope can be used in tandem with our other volcano plots that explore the influence of the catalyst on reaction energetics. This combination results in a series of pieces that can be mixed and matched to create more efficient catalytic processes.

### 3.4 Conclusions

In conclusion, we demonstrated how the substrate scope can be examined using volcano plots, which provides an additional tool within the molecular volcano family (and the broader set of tools for computational catalysis) that can be used to optimize homogeneous catalytic processes. These “substrate volcanoes” reveal a wealth of information regarding how the nature and location of substituents as well as core structures dictate catalytic cycle energetics. Specific to the cross-coupling reaction examined here, the presence of electronegative moieties (heteroatoms and/or electron-withdrawing groups) induces stronger catalyst/substrate binding interactions, causing a leftward shift on the volcano plot that either increases or decreases the reaction rate depending on the properties of the substrate core. Substrates with electropositive substituents, on the contrary, are often governed by the rate of oxidative addition. For substrates having transmetalation as the rate-determining step, steric effects play an important role, where bulky groups near the reaction site slow the reaction. Overall, our substrate volcanoes show that having balanced electronic parameters and not overly bulky substrates is essential for the cross-coupling reaction to proceed, thereby confirming what is known based on extensive experimental studies.

## 4 Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts

This chapter is based on following publication:

Meyer B., Sawatlon B., Heinen S., von Lilienfeld O. A., and Corminboeuf C., Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts. *Chem. Sci.*, **2018**, 9, 7069-7077.

Sawatlon B. did the molecular volcano plot, provided a machine learning database and did a catalyst prediction with an assistant of Meyer B. The machine learning part was conducted by Meyer B. and Heinen S.

### 4.1 Introduction

Chemists constantly pursue new molecular systems that provide increasingly higher yields and better control of selectivity. Rather than blindly searching for promising catalysts to meet their needs, numerous tools that aid in identifying the most appropriate species have been developed. These range from high-throughput screening<sup>99-100</sup> (including combinatorial methods<sup>101-102</sup>), which quickly evaluates reaction conditions and the structures of catalysts, to multidimensional modeling based on a design of experiments (DoE),<sup>2</sup> that relates steric and structural descriptors (*e.g.*, Charton values and Sterimol parameters) to enantioselectivity. Such methods have found broad application in asymmetric homogeneous catalysis.<sup>19, 21-22, 24, 103-107</sup> On the other hand, the tremendous increase in computer power accompanied by methodological advancements has also made computational studies of catalytic processes commonplace.<sup>108</sup> While virtually any catalytic system can be subjected to computational analysis, often the conclusions reached are not transferable and provide little insight into the best way to develop more active and selective catalysts. Thus, a tool that assesses the properties of untested catalysts based on a simple energetic or structural criterion would rapidly accelerate the discovery pace of new species. Indeed, similar concepts involving the mapping of a difficult to determine quantity onto an easily obtained variable have been a central pillar of catalysis and physical organic chemistry for more than 80 years, and are at the core of familiar concepts such as the Bell–Evans–Polanyi principle,<sup>30-31</sup> the Hammett equation,<sup>26-29</sup> or the Brønsted catalysis equation.<sup>32</sup> Today, volcano plots,<sup>45-46</sup> which relate easily accessible descriptor variables directly to catalytic performance, accomplish this objective and find regular use in the fields of heterogeneous catalysis<sup>14, 55, 75</sup> and electrocatalysis.<sup>13, 33-34, 47, 49, 109-110</sup>

Based on knowledge of a chosen descriptor variable, volcano plots function by discriminating catalytic performance using Sabatier's principle.<sup>36</sup> Sabatier conceived the notion of an ideal catalyst that should not bind a substrate too strongly or too weakly. The unique volcano shape facilitates rapid discrimination of catalytic activity. Species positioned highest on the plot (generally on or near the volcano plateau or peak) have the best profiles and fulfill Sabatier's principle. Species located along the left- and right-slopes have less ideal profiles and can be characterized as having either overly strong (left) or overly weak (right) catalyst/substrate interactions. While being commonly used in heterogeneous and electrocatalysis, and frequently invoked for homogeneous systems,<sup>60, 95, 111</sup> only recently have these appealing tools been concretely realized for molecular catalysts.<sup>5</sup> Corminboeuf *et al.* have pioneered the use of molecular volcano plots to study various aspects of prototypical C–C cross-coupling reactions in order to gauge the feasibility of using these tools to identify attractive homogeneous catalysts.<sup>5, 61-63</sup> Subsequent work has also focused on adapting volcano plots for applications in homogeneous catalysis *via* the inclusion of kinetics (as opposed to the typically used thermodynamic based criteria) of the catalytic cycle.<sup>64</sup>

The use of molecular volcano plots involves establishing linear scaling relationships that relate the quantitative value of a descriptor to the thermodynamic or kinetic performance of the catalyst. As such, this tool has clear utility in high-throughput screening applications that search for prospective catalysts by computing the value of this descriptor for any species desired. However, currently both the geometries and energies associated with multiple forms of each catalyst must be determined through a relatively slow process involving density functional theory computations. Clearly, increasing the speed at which the descriptor variable can be determined would result in an overall increase in the discovery pace of new catalysts because prospective species could be screened more rapidly. One route with the potential to provide virtually instantaneous access to the descriptor involves the application of quantum machine learning (ML) models, *i.e.*, ML models, which can be trained on, and used to predict quantum properties.<sup>112-114</sup> The application of ML models to estimate volcano plot energy descriptors offers increased speed for two reasons: first, the energy based value can be immediately accessed for any desired species, and second, the need to establish a precise geometry of the catalyst can be circumvented by also including this task into the ML model, as already demonstrated within the  $\Delta$ -ML approach.<sup>115</sup> As such, the ML model can predict an accurate descriptor value from an approximated 3D structure of a catalyst.

While, generally speaking, applications of machine learning methods in chemistry are still in their infancy, their use has begun to appear in the fields of materials science<sup>116-121</sup> and catalysis.<sup>122-129</sup> For example, a gradient-boosting regression method<sup>130</sup> has been used to predict the d-band center of mono- and bimetallic surfaces<sup>131</sup> and to estimate CO adsorption energies on Pt nanoparticles,<sup>132</sup> while a local similarity kernel could predict the catalytic activity of nanoparticles.<sup>133</sup> Moreover, applications of support vector machines (SVMs)<sup>134</sup> were able to anticipate CO<sub>2</sub> uptake in metal organic frameworks (MOFs)<sup>135</sup> by developing an atomic property-weighted radial distribution function (AP-RDF) based descriptor<sup>136</sup> that captures geometric and chemical features of periodic systems. Predictive structure–reactivity models have identified promising Pt-based electrocatalysts for the oxygen reduction reaction,<sup>137</sup> while artificial neural networks (ANNs) have recognized multimetallic alloys possessing high selectivity for

electrochemical CO<sub>2</sub> reduction to C<sub>2</sub> species.<sup>138-139</sup> Recently, Nørskov investigated various machine learning based approaches<sup>140</sup> to systematically search for the active sites of bimetallic (nickel gallium) nanoparticles,<sup>141</sup> to construct Pourbaix surface phase diagrams,<sup>142</sup> and to identify probable mechanisms of hydrocarbon–syngas reactions on rhodium(111).<sup>143</sup> Rappe *et al.* also exploited the regularized random forest machine learning algorithm,<sup>144</sup> and discovered the key role played by structure and charge descriptors (namely the Ni–Ni bond length and the Ni residual charge) in the hydrogen evolution reaction activity of Ni<sub>2</sub>P(0001).

Despite the considerable amount of progress in applying ML models to chemical problems, each of the aforementioned contributions tackled issues surrounding heterogeneous catalysis, while ML applications to homogeneous catalysis remain rare.<sup>77, 129, 145-150</sup> Significant advances with ML models to obtain fundamental molecular electronic properties (*e.g.*, atomization or total energies of molecules) have been made,<sup>151-158</sup> however, the prediction of catalytic cycle intermediates energies has never been attempted to the best of our knowledge. The purpose of this work is to demonstrate how ML models can be used to accelerate the screening of prospective homogeneous catalyst candidates, thereby enabling the computational discovery of novel catalytic materials. To this end, we selected the problem of finding catalysts for the Suzuki–Miyaura C–C cross-coupling reaction (Figure 4.1).<sup>66-67, 159</sup> Specifically, we trained and applied ML models using the reaction energy associated with oxidative addition (Equation 4.1), which has previously been shown to be a descriptor variable for analyzing the catalytic cycle thermodynamics using volcano plots.<sup>5</sup> Although kinetic profiles are obviously important for obtaining a full and accurate description of catalytic performance, here we rely on a simplified thermodynamic picture (Figure 4.1), which can be exploited to rapidly discriminate between catalysts with promising or inadequate energy profiles.<sup>5, 61-63</sup>



Using machine learning models of this quantity, along with previously constructed molecular volcano plots, it is possible to screen thousands of potential catalysts with controlled accuracy (by virtue of learning curves) and at a negligible computational overhead.

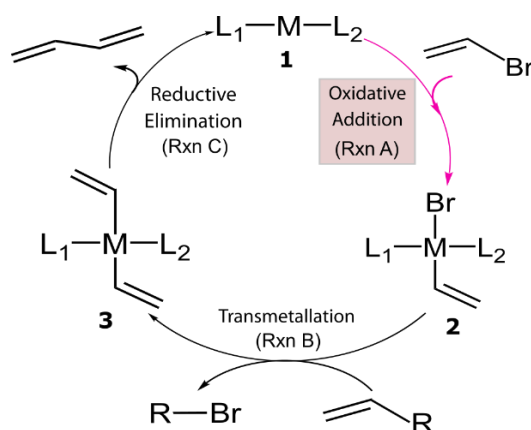


Figure 4.1 General catalytic cycle for C–C cross-coupling reactions. Coupling partners ( $R$ ) depend on specific cross-coupling reactions. Suzuki coupling undergoes a ligand exchange step replacing  $Br$  by an alkoxy group before transmetalation (Rxn B). The dissociated compound in Rxn B is alkoxy– $R$  instead of  $Br$ – $R$  and  $R$  is  $[B(OH)_2(O^tBu)]^-$  for the Suzuki reaction.<sup>5, 62</sup>

## 4.2 Computational Details

The initial set of Cartesian coordinates for each catalyst was obtained by converting Simplified Molecular Input Line Entry System (SMILES) formats (*i.e.*, a line notation for entering and representing molecules and reactions)<sup>160-161</sup> into three-dimensional structures with the 3D structure generator operation (*i.e.*, gen3d operation) of the OpenBabel software.<sup>162</sup> To generate target energy values for the training and test complexes, we proceeded as follows: computations involving geometry optimization and electronic energies were generated and executed *via* the AiiDA automated platform.<sup>163</sup> Gas phase geometry optimizations were computed at the B3LYP<sup>164-166</sup>-D3<sup>81, 167</sup> with 3-21G (for Ni, Pd, Cu and Ag complexes)<sup>168-171</sup> and a def2-SVP<sup>82</sup> basis set (for Pt and Au complexes) in Gaussian09.<sup>172</sup> Single point energies were computed at the B3LYP-D3/def2-TZVP level.<sup>82</sup> The oxidation states of the catalysts were adjusted to comply with the dominant  $14e^-/16e^-$  nature of the complexes in the Suzuki cross-coupling reaction. The reaction electronic energies (Equation 4.1) were computed and used as a descriptor (see a volcano plot in Figure 4.2) for training the machine learning models. The ML models were trained and applied using the Quantum Machine Learning toolkit QMLcode.<sup>173</sup>

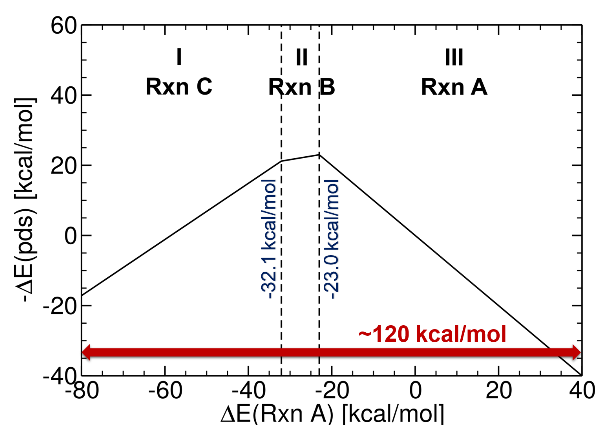


Figure 4.2 Reference volcano plot for the Suzuki cross-coupling reaction. Region (I) corresponds to reductive elimination, (II) to transmetalation, and (III) to oxidative addition. Acceptable catalysts should fall into the mid region (in between  $-32.1$  and  $-23.0$  kcal/mol).

The reference volcano plot associated with the catalytic cycle of Figure 4.1 was constructed according to the procedure outlined in our previous work<sup>5, 64</sup> using the same theoretical level as for the descriptors of the machine learning training set. Note that despite the relatively modest level of theory used herein (engendered by the large computational effort associated with generating the training set for the ML model), the geometries and key energetic properties are in line with those previously computed (Table 4.1).<sup>5, 62</sup> Similarly, we previously demonstrated that the same set of linear free energy scaling relationships capably describe variations in the number of coordinated ligands (*i.e.*, bis- vs. monoligated), as well as different oxidation or spin states of the catalyst.<sup>5, 61, 64</sup> Rather than predicting the entire volcano plot, the most essential property is the descriptor  $[\Delta E(\text{Rxn A})]$  (Equation 4.1), which can be machine learned, as well as knowledge about its target value, *i.e.*, the energy range corresponding to the ideal plateau region (extending from  $-32.1$  to  $-23.0$  kcal/mol, Figure 4.2).

Table 4.1 Comparison between electronic binding energies obtained from different level of theory (def2-TZVP//3-21G, def2-SVP and def2-TZVP) and the energies predicted by the machine learning model (BoB representation).

Metal	Ligand 1*	Ligand 2*	def2-TZVP//3-21G,def2-SVP	def2-TZVP	ML predictions
Pd	0	74	-26.52	-24.38	-26.56
Pd	15	74	-27.28	-26.82	-31.33
Pd	1	89	-22.64	-20.03	-25.46
Pd	4	72	-27.35	-29.35	-26.64
Pt	52	71	-29.01	-29.28	-28.50
Pt	71	84	-25.41	-25.68	-28.08
Pt	1	84	-27.14	-27.22	-29.02
Pt	17	23	-31.52	-31.67	-28.80

\*See the structures of ligands in **Appendix B**

## 4.3 Methods

### 4.3.1 Database

The training procedure relies upon constructing a large database of catalysts that are obtained through combining various ligands and metals. These species are then used for training and testing the ML models which, in turn, are used to predict descriptor values so rapidly and with such accuracy that large libraries can be scanned in order to identify acceptable catalyst candidates. Ninety-one ligands including CO, phosphines, N-heterocyclic carbenes and pyridines were combined with six transition metals (Ni, Pd, Pt, Cu, Ag, and Au) to form the database. All possible metal/ligand combinations (*i.e.*,  $L_1$  and  $L_2$ , where  $L_1ML_2$  is equivalent to  $L_2ML_1$ ) of catalytic cycle intermediates **1** and **2** (Figure 4.1) lead to a total library consisting of 25,116 species for each intermediate (see **Appendix B** for a complete list of ligands used). Rather than providing the optimized structures for each ligand to build the catalysts, the geometries of catalytic cycle intermediates **1** and **2** for each database entry were created by converting SMILES strings<sup>i</sup> (Figure 4.3)<sup>160-161</sup> to Cartesian coordinates using the OpenBabel implementation<sup>162</sup> of the Merck Molecular Force Field method (MMFF94).<sup>174-178</sup> This database was divided into two subsets: (1) the training/test set used within cross-validated learning curves (see details on the cross-validation procedure in section 4.3.2) for which the computed descriptor values [ $\Delta E(\text{Rxn A})$ ] were used as a reference and (2) the prediction set on which the model was applied to screen candidates based on their ML predicted descriptor values. Since collecting reference data for the training and test sets involves costly DFT geometry relaxations, we proceeded in two steps:<sup>ii</sup> first, an initial training set made of complexes involving a diverse set of ligands (72 in total) with Pd (2,595 complexes).<sup>iii</sup> Secondly, a small subset of illustrative ligands (12) with each of five other metals (Pt, Au, Ag, Cu, Ni) (390 complexes) was created. The final set consisted of

<sup>i</sup> The *trans* isomerism constraint was imposed using the general chiral specification syntax of the SMILES notation (*i.e.*, the @SP square-planar class symbol) as depicted (on the top right-hand corner) in Figure 4.3.

<sup>ii</sup> To refine the accuracy of the ML model in the targeted descriptor energy range, *i.e.*, the top of the volcano, we exploited the trained model to predict the binding energies on a subset of complexes combining the 5 metals (Pt, Au, Ag, Cu, and Ni) and 72 ligands (from no. 0 to 71) and selected the molecules for which the ML predicted reaction energy was in the selected range (as opposed to randomly selecting additional candidates to extend the training set).

<sup>iii</sup> Due to convergence problems, exactly 2,595 binding energies from Pd complexes were used in the training set.

a total of 7,054 reaction energy values corresponding to our descriptor. ML models were trained on this set (*vide infra*), and out-of-sample predictions were then made on the prediction set that consisted of all the other complexes (18,062 in total). Note that included in this set are 19 realistic ligands that have already been employed in experimental settings (*i.e.*, ligand no. 72–90<sup>iv</sup> in Figure 4.3).<sup>179-181</sup>

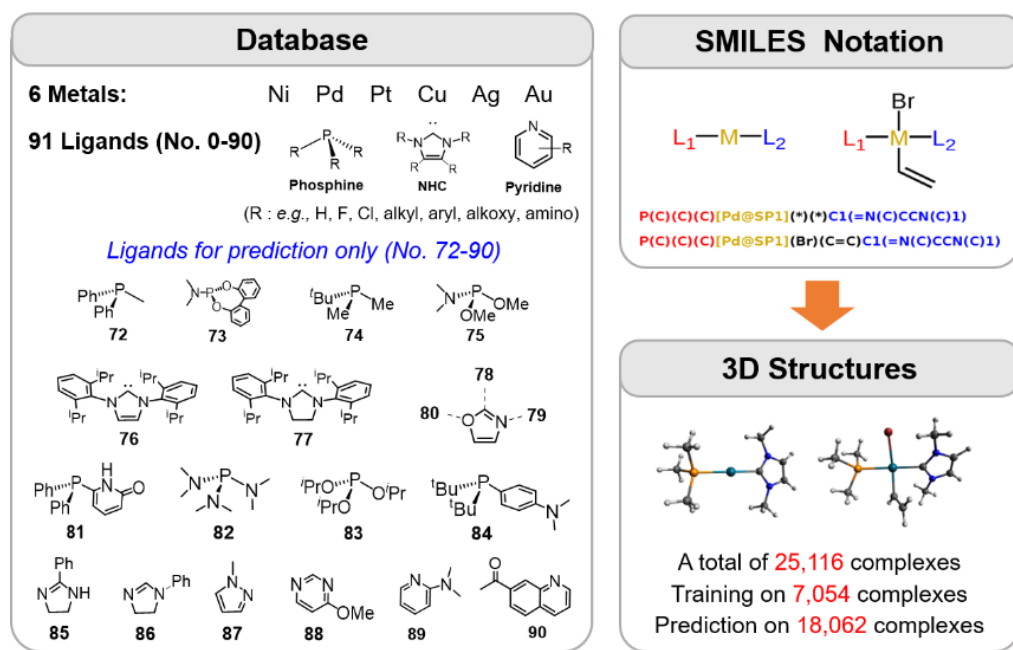


Figure 4.3 A database of 25,116 molecular transition metal catalyst candidates. Each complex consists of one out of six transition metals and a combination of two out of 91 ligands (left). Each ligand was written as the SMILES notation and all possible  $\text{L}_1\text{--M--L}_2$  combinations were constructed (top right-hand corner). SMILES strings were then converted into Cartesian coordinates through the 3D structure generator of the OpenBabel software (bottom right-hand corner). DFT reference results for training and testing of ML models were obtained for a sub-set of 7,054 candidates. Those structures were exploited for computing binding energies and for training the ML models.

### 4.3.2 Training

To begin the machine learning process, information intrinsically contained within each three-dimensional structure must be transformed into a suitable representation. The approach selected to represent a molecule has a crucial impact on the learning curve. It is of particular importance to construct a meaningful relationship between the representation and the catalyst candidate, that will be learned by the machine learning algorithm. For this reason, all the relevant variables for computing the target properties [in our case  $\Delta E(\text{Rxn A})$ ] should be represented in the chosen machine learning representation of the molecule. Over the last few years, increasingly improved representations<sup>112, 151, 155, 182-183</sup> that progressively encode increasing amounts of physical information have been proposed. Here, we focus on the sorted Coulomb Matrix (CM), the first representation introduced for ML models trained throughout chemical space and used to predict quantum properties,<sup>112</sup> a two-body bagged variant of the CM with superior performance, the Bag

<sup>iv</sup> Ligand number 78 was treated as anion.



of Bonds (BoB),<sup>155</sup> and the recently proposed Spectrum of London and Axilrod–Teller–Muto potential (SLATM).<sup>184</sup> The CM representation consists of a square atom by atom matrix, where the diagonal elements model the potential energies of free atoms while the off-diagonal elements correspond to the Coulomb nuclear repulsion between atom pairs. In the BoB representation, CM elements are bagged (*e.g.*, C–C, C–N, C–H, *etc.* are accounted for in separate bags.). SLATM is based on the dissociative limits of intermolecular dispersion contributions between unpolarized moieties. They account for interatomic two-body terms through London’s dispersion curve (rather than Coulomb), and for the three-body terms according to Axilrod–Teller–Muto.<sup>185–186</sup>

We stress that our principal objective is to describe the oxidative addition step directly from rough-coordinate estimates obtained from the SMILES structure (*i.e.*, without providing accurate geometry as an input). After conversion from SMILES to coordinates, we map our input representation onto the corresponding continuous label value [here  $\Delta E(\text{Rxn A})$ ] using kernel ridge regression (KRR),<sup>187</sup> which solves nonlinear problems by mapping data from the input space into a high-dimensional linear feature space (kernel trick). A Laplacian kernel function is used for the CM and BoB representations, and a Gaussian kernel for the SLATM representation. The quality of the models is evaluated by reporting test errors, which can be obtained by separating the dataset into training and test frames and calculating the average error [typically a mean absolute error (MAE)] for the predictions on the out-of-sample test set. This random sub-sampling cross-validation procedure<sup>114</sup> was used to shuffle the dataset randomly into different training sets. For every shuffling step the MAE for the model was calculated and the procedure repeated ten times for every training set size  $N$ . Afterwards, the errors for the different models were averaged into a single cross-validated error. Note that this error remains a random variable that is dependent on the initial splitting of the training/test datasets. When plotted on a log–log scale, successful learning is indicated by linearly decaying behavior for large training set sizes, as already suggested by Vapnik and others in the nineties.<sup>188–189</sup>

## 4.4 Results and Discussion

### 4.4.1 Machine Learning

In order to verify the performance and validity of our *ansatz*, we have trained and tested machine learning models for various training set sizes. The resulting learning curves, depicted in Figure 4.4, demonstrate the efficiency and accuracy of the learning process in terms of a near-linear decay of test error with training set size. While learning is observed for all representations, the learning curves illustrate the impact of the molecular representation on the off-set and slope. Overall, the performances of the ML models based on the SLATM and BoB are very similar (for the largest training set, the MAE is 2.61 kcal/mol and 2.73 kcal/mol respectively) and superior to CM (largest training set MAE = 3.05 kcal/mol). Despite these small variations, it is obvious that efficient learning is achieved by all three representations. This result contrasts with findings in ref. 117 where the CM was claimed to be of little use when constructing ML models for transition metal complexes. The poor performance of CM is more likely due to inappropriate choice of properties (electronic spin-states) than to the molecular systems themselves. It seems intuitive that any purely structure and composition based representation will struggle to account for various electronic states. When it comes to simple electronic ground state properties, such as the oxidative addition step studied here, Figure 4.4 clearly demonstrates that the CM is very

applicable to the machine learning modeling of properties of transition metal complexes. We also note that the BoB representation performs surprisingly well for this problem. We ascribe this behavior to the bagging which allows the model to place appropriate weights to bonds involving the transition metal.

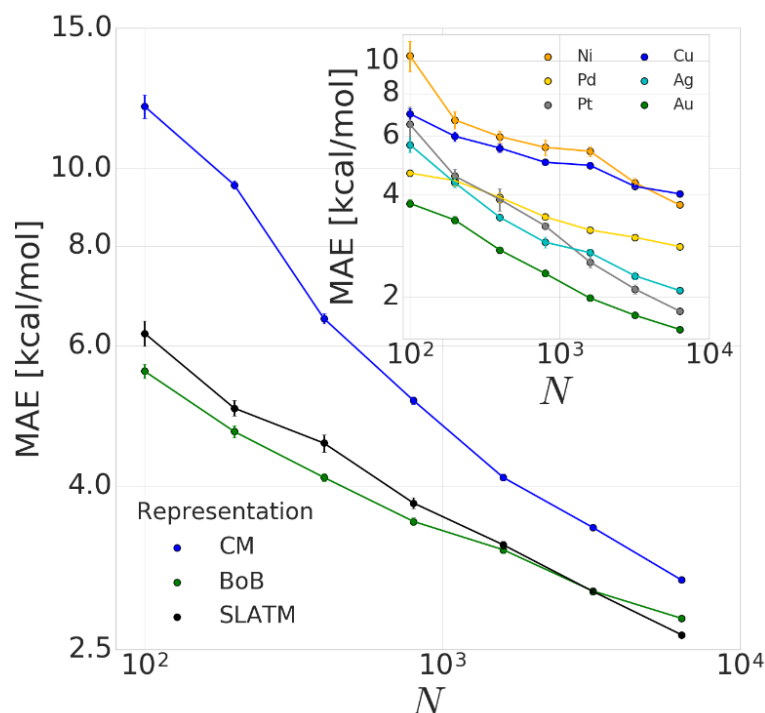


Figure 4.4 Learning curves [test error of catalytic descriptor values as a function of training set size ( $N$ )] for oxidative addition of vinyl bromide. Error bars correspond to standard deviation in cross validation. Inset: corresponding learning curves for individual metals for BoB.

The energy range for the descriptors of the training set (corresponding to the  $x$ -axis of the molecular volcano plot) is  $\sim 120$  kcal/mol (Figure 4.2). We therefore considered the ML model to be sufficiently well converged for the task of picking catalysts, once the learning curve dropped to less than 3 kcal/mol (*i.e.*, 2% of the descriptor range). The most efficient representations, SLATM and BoB, reached this threshold with a training set of 7,054 binding energies. The following discussion will thus be based on the less sophisticated representation, BoB. It is important to reiterate that while the machine learning models were trained on DFT reaction energies obtained for DFT optimized geometries, the molecular representations in the test set were constructed solely from the coordinates directly obtained from SMILES conversion.

The heterogeneity of the training set (*i.e.*, unequal representation of the six transition metals) has been looked into by evaluating the individual predictions of the BoB based machine learning model on each metal separately. The resulting learning curves depicted in the inset of Figure 4.4 demonstrate that learning is attained for all metals. For the largest training set size, the target MAE of 3 kcal/mol is achieved for Pd, Pt, Ag and Au, while the Ni and Cu metal complexes are less accurately described (best MAE = 3.74 and 4.04 kcal/mol, respectively). These larger errors certainly originate from the smaller sample of Ni complexes and from copper–ligand combinations featuring ligands that are less frequent in the rest of the training set. This leads to

a larger energy range in the descriptor variables which can be seen as a broader distribution/width [see the histograms (Figure B.1 and B.2) in **Appendix B**]. Overall, however, the ML performance for Ni and Cu-based complexes is still useful as it is not more than 5% of the descriptor's energy range (*i.e.*, inferior to 5 kcal/mol).

The application of the machine learning model studied here can accelerate the catalyst screening process using molecular volcano plots. However, it should be noted that the use of a crude geometric structure as an input might affect the accuracy of the ML model, where a large training set size is currently needed for the model to reach an acceptable MAE. In addition, as our ML model is used in tandem with molecular volcano plots, the limitations associated with the molecular volcano also indirectly influence the utility of the ML model. For example, complexes with coinage metals (group 11) were always treated as monocations because the oxidation states of each catalyst need to be adjusted to align with the well-known  $14e^-/16e^-$  nature of the coupling reactions. Thus, the ML models were trained with both neutral and cationic complexes (while the ligands used were mostly neutral). It is possible that the mixture of neutral and cationic complexes that compose the training set, in addition to the rough geometric structures used as inputs, could limit the accuracy and generality of the model.

#### 4.4.2 Catalyst Prediction

The trained ML models were subsequently exploited to predict the energy based descriptor of 18,062 potential out-of-sample catalysts with negligible computational cost (*vide supra*). At this point, it is worth noting that out-of-sample predictions that involve ligands not previously seen by the models should be considered with more care. Additionally, the predictive power of the model would be limited for catalysts that would suffer from a convergence problem in an actual computation. Because we are interested only in the catalysts predicted to have the best thermodynamic profile for the Suzuki–Miyaura reaction, emphasis was placed on a narrow range of descriptor energy values (from  $-32.1$  to  $-23.0$  kcal/mol) corresponding roughly to the plateau of the volcano. However, the same ML models would be relevant to the analysis of other cross-coupling reaction variants differing only by the width of the plateau region.<sup>62</sup> Using the BoB model, 557 catalysts were identified that fell into this region. A brief examination of the metal distribution (Figure 4.5) yields expected results, namely that catalysts incorporating group 10 metals (Ni, Pd, Pt) appear more frequently than their group 11 (Cu, Ag, Au) counterparts. This finding is in line with our earlier DFT-based molecular volcano plot analysis of the same reaction.<sup>5, 62-63</sup>

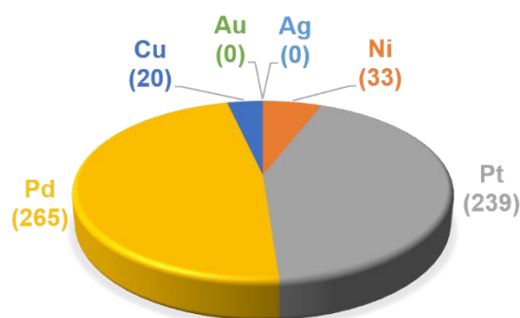


Figure 4.5 Occurrence of the six metal complexes in the selected range of  $-32.1/-23.0$  kcal/mol predicted by the machine learning model using the BoB representation.

A prevalent metal identified by the ML model is palladium, which has 265 species that appear on or near the volcano plateau (Figure 4.5). The large number of Pd catalysts attests to the accuracy of the ML models, as these species have a rich history in catalyzing cross-coupling reactions.<sup>190-192</sup> On the other hand, Pt catalysts are virtually experimentally unknown<sup>193</sup> and those that have been tested tend to show only moderate catalytic ability.<sup>194</sup> Nonetheless, their significant presence on the volcano plateau does align with our earlier DFT-based evaluations.<sup>5, 62-63</sup> Indeed, we previously postulated that the presence of Pt based catalysts on top of the volcano may indicate that the problem with these species is less thermodynamic and more kinetic in nature.<sup>62</sup> In addition, others have speculated that an enhanced M–R bond strength causes transmetalation in these species to be sluggish.<sup>195</sup> Despite being well-known cross-coupling catalysts,<sup>196</sup> only a handful of Ni based species are predicted by the ML model to appear near the volcano plateau. However, in its current state, the ML models consider only a single oxidation state, that for Ni corresponds to a Ni(0)/Ni(II) based catalytic cycle. Thus, the more catalytically active Ni(I) oxidation state, which is accessed *via* a one-electron redox process<sup>197</sup> and generally shifts Ni catalysts from the strong-binding side of the volcano onto the plateau,<sup>61</sup> is currently not assessed by the ML models (*vide supra*) but incorporation of alternative catalytic oxidation states represents an appealing future improvement of the current model. The volcano plot also reveals the influence of ligand type on the thermodynamics of the catalytic cycle. For example, Figure 4.6 clearly indicates that phosphine ligands generally outperform N-heterocyclic carbene and pyridine ligands when combined with group 10 metal (Ni, Pd, and Pt) complexes. More interesting is the presence of oxazole ligands for Pd metals. While the use of the monodentate variant (*e.g.*, ligands no. 78–80) appears elusive in the literature, the chemistry associated with the use of bidentate bis(oxazole) ligands for cross-coupling reactions is relatively well established.<sup>198</sup>

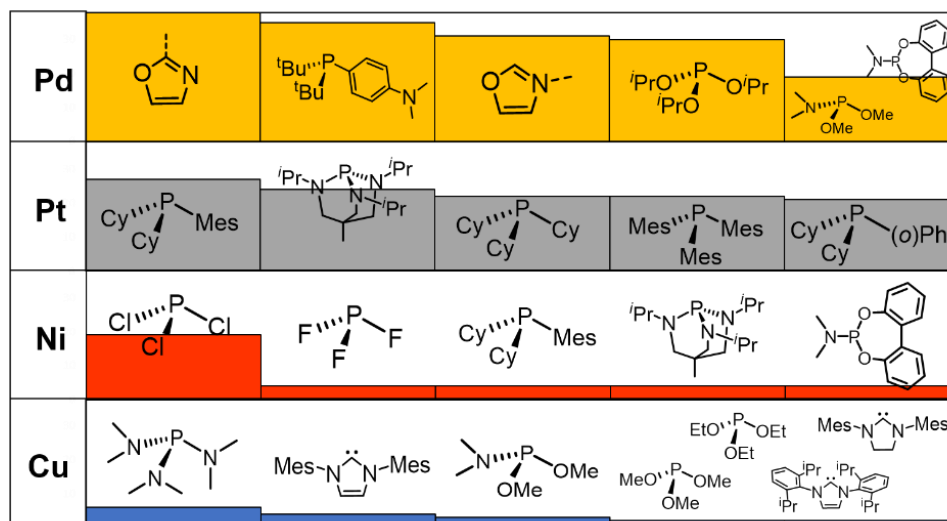


Figure 4.6 Histogram ranking of the five most identified ligands that appear on the volcano plateau (*i.e.*, with descriptor values between  $-32.1$  and  $-23.0$  kcal/mol) by metal type as predicted by the machine learning model using the BoB representation. The histogram is scaled relative to the Pd/oxazole ligand combination, which has the highest metal/ligand occurrence appearing 38 times on the volcano plateau.

By far, the vast majority of the coinage metal (group 11) catalysts have very weak binding energies and, correspondingly, lie on the right (weak-binding) slope of the volcano. Indeed, no Au or Ag based catalyst has sufficiently strong binding energy to appear on the volcano plateau (Figure 4.5). This finding directly agrees with experimental and computational studies that have found Ag and Au catalysts to have unfavorable free energies associated with oxidative addition.<sup>199</sup> On the other hand, a handful (20) of Cu based catalysts are found to have nearly ideal thermodynamic profiles. While instances of Cu-based Suzuki coupling have appeared in the literature,<sup>200-201</sup> these catalysts tend to employ bidentate acetylacetone (acac) or acetate/triflate ligands.<sup>202-203</sup> Thus, it is interesting to note that each of the thermodynamically most appealing Cu catalysts involves either a tris(dimethylamino)phosphine or bulky N-heterocyclic carbene (Figure 4.6). These findings represent a potentially interesting research direction that should be explored in more depth and that has been revealed solely through the application of ML models coupled with molecular volcano plots.

Finally, a more refined selection of catalysts was obtained based on their estimated price per mmol (Figure 4.7). Among the 557 catalysts with promising thermodynamic profiles, 37 complexes have an estimated price less than \$10 USD per mmol. These species include earth abundant metals (copper with tris(dimethylamino)phosphine) and a multitude of more standard palladium phosphine combinations.

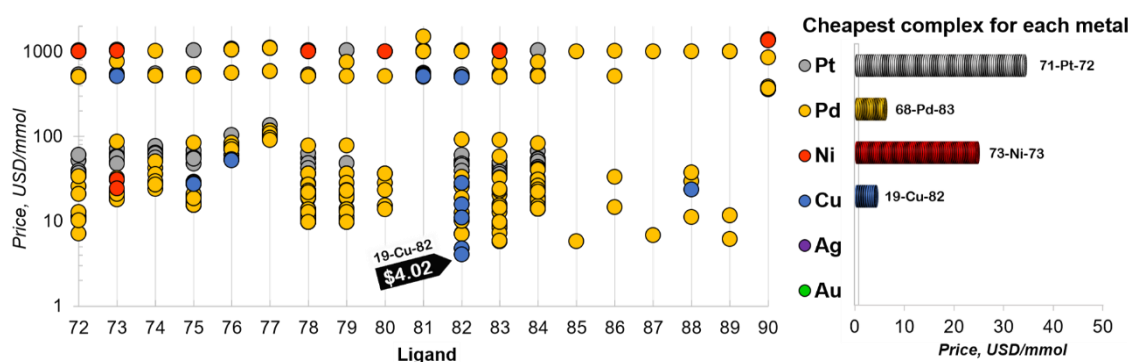


Figure 4.7 Estimated price (for one mmol in US dollars) of the catalysts in the selected range of  $-32.1$ – $-23.0$  kcal/mol (for ligand no. 72–90). The price is calculated as a summation of the commercial price of transition metal precursors (one mmol) and one mmol of each ligand. The cheapest complex for each metal is shown on the right.

## 4.5 Conclusions

We have trained and used machine learning models to dramatically accelerate the descriptor screening of 18,062 homogeneous catalysts for the Suzuki–Miyaura C–C cross-coupling reaction. The model was based on the capability of molecular volcano plots to identify thermodynamically attractive candidates with respect to a simple energy descriptor. Overall, we have identified 37 promising low-cost complexes featuring palladium and copper combined with both standard and less expected ligands. Our findings also indicate that machine learning can be used to screen thousands of catalysts, and that previously introduced machine learning representations can be used for property predictions of transition metal complexes. Exploitation of a  $\Delta$ -machine learning approach represents an appealing future improvement of the proposed ML models.<sup>115, 182</sup>



# 5 Data Mining the C–C Cross-Coupling Genome

This chapter is based on following publication:

Sawatlon B., Wodrich M. D., Meyer B., Fabrizio A., and Corminboeuf C., Data Mining the C–C Cross-Coupling Genome. *ChemCatChem* **2019**, 11, 4096-4107.

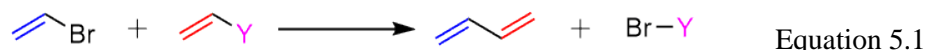
Sawatlon B. and Wodrich M. D. contributed equally to this work.

## 5.1 Introduction

Using machine learning (ML) to accelerate electronic structure methods has the potential to revolutionize the field of computational chemistry, having already demonstrated its promise by making significant headway in addressing several longstanding fundamental problems in chemistry and materials science over the past 15 years.<sup>112, 116, 119, 204-208</sup> In the field of computational catalysis, for example, recent work has tackled a diverse set of topics ranging from CO<sub>2</sub> uptake in metal organic frameworks<sup>135</sup> to identifying structure-activity relationships in Pt electrocatalysts for the oxygen reduction reaction<sup>137</sup> to reaction mechanism discovery on rhodium surfaces.<sup>143</sup> Such examples of ML only begin to scratch the surface.<sup>123, 209-211</sup> Applications to computational homogeneous catalysis do exist,<sup>77, 129, 145-150</sup> but occur with much less frequency than their heterogeneous counterparts. While ML techniques certainly can be used in a brute-force type manner to determine the properties (*e. g.*, bond energies,<sup>212</sup> spin-state splitting in inorganic complexes<sup>116</sup>) or reaction energetics<sup>39</sup> of a virtually unlimited number of prospective catalysts, the identification of new homogeneous species has historically placed significant emphasis on understanding reactivity through the establishment of heuristic devices such as structure-activity relationships.<sup>213-216</sup> This has led to, for example, multidimensional modeling based on design of experiments<sup>2</sup> that are widely used in asymmetric homogeneous catalysis.<sup>19, 21-22, 24, 104-105, 217-219</sup> Considerable effort has also been spent in determining appropriate descriptors (*e. g.*, bandgap, d-band center, adsorption energy) that can be coupled with machine-learning to estimate catalytic performance.<sup>131, 220-221</sup>

Along this line, our research group has done extensive work in crafting a computational toolkit built on volcano plots,<sup>46</sup> commonly used tools for identifying attractive species in heterogeneous/electro-catalysis,<sup>4, 13-14, 55, 109</sup> to study homogeneous reactions.<sup>5, 39, 43, 61-65</sup> Volcano plots are built upon Sabatier's principle,<sup>35-36</sup> which states that the interaction between a catalyst and a substrate should be neither too weak nor too strong. In addition to volcano plots, Sabatier's principle has also been used in other ways to understand and ultimately design better catalysts,<sup>6</sup> such as through

the use of Kozuch and Shaik’s energy span model.<sup>95, 97, 222–223</sup> The attractiveness of volcano plots and their utility for quickly screening catalysts arises from their distinctive shape, which reveals a considerable amount of information about each individual species. In essence, the most attractive species, those that obey Sabatier’s principle, sit on the volcano plateau or near the peak, while the sides represent either overly strong (left-side) or overly weak (right-side) catalyst/substrate interactions. A catalyst’s location on the plot is determined by the value of its descriptor variable (given on the *x*-axis), which is then related to either the thermodynamic or kinetic properties of the catalytic cycle (shown on the *y*-axis) through linear free energy scaling relationships (LFESRs).<sup>52, 55, 224</sup>



Our initial work on molecular volcano plots<sup>5</sup> began by examining a Suzuki C–C cross-coupling reaction (depicted in Equation 5.1). However, we also uncovered a type of cross-coupling genome that illustrates how the catalytic cycle thermodynamics are influenced by additional chemical factors such as the cross-coupling partner<sup>62</sup> [*i. e.*, Suzuki, Kumada, Negishi, Stille, and Hiyama coupling, (represented by “Y” in Figure 5.1a)]. Indeed, these different cross-coupling variants are each described by very similar volcano plots, where only the width of the plateau region changes (Figure 5.1b–f). Despite the considerable amount of information regarding the behavior of catalysts that can be extracted from these volcano plots, the screening for prospective new species requires determination of the descriptor value through a series of density functional theory computations. In an attempt to expedite the screening process, we recently turned to ML techniques in order to establish the value of the descriptor variable. This effort produced the desired effect and allowed us to considerably increase the number of species screened, up to 18,000 catalysts, for the aforementioned Suzuki C–C cross-coupling reaction.<sup>39</sup>

Since our previously established ML model was successful in obtaining the value of the descriptor variable for the aforementioned ~18,000 potential catalysts,<sup>i</sup> these same “descriptor” quantities can also be used to establish the thermodynamic viability for each species for the five different cross-coupling reactions illustrated in Figure 5.1. The purpose of this contribution is to use thermodynamic data accessed through the machine learning of quantum chemical properties to uncover chemical trends present in the cross-coupling genome by applying concepts from big-data analysis. This objective is accomplished with the assistance of interactive dimensionality-reducing data-clustering maps<sup>40–41, 225</sup> that group different catalysts based on their intrinsic properties. In turn, this information can be used to retrieve correlations between various chemical properties and catalytic cycle energetics. In essence, the amalgamation of big-data, dimensionality reduction algorithms, and volcano plots serve as effective tools for mining the cross-coupling genome and constitutes a promising protocol for guiding the development of novel homogeneous catalysts.

<sup>i</sup>For more details on training of the machine-learning model, including discussion on cross validation, see **Chapter 4**



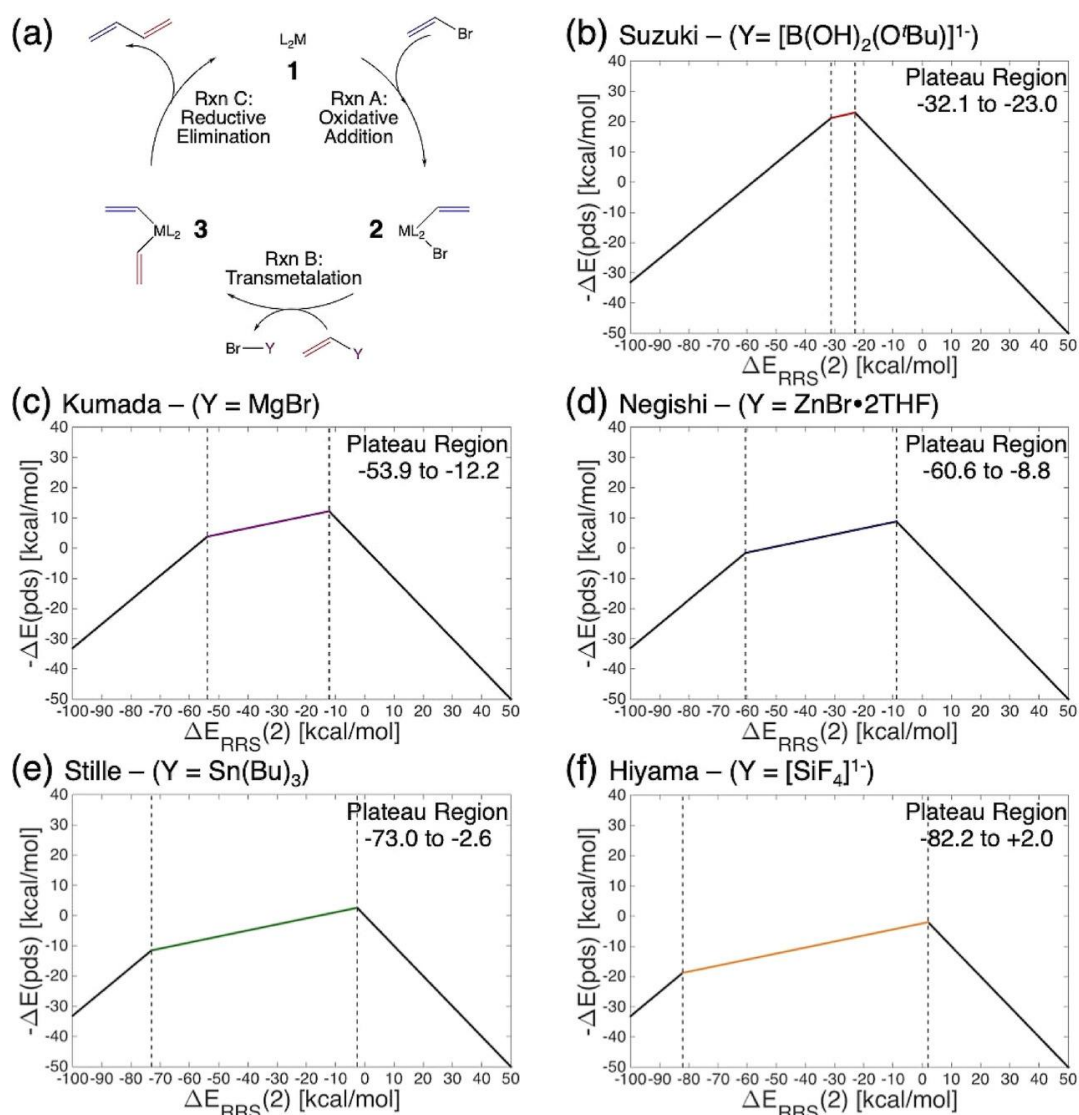


Figure 5.1 (a) The catalytic cycle for cross-coupling reaction of a vinylbromide (blue) and a vinyl metal complex (red). Note that changes in the cross-coupling partner,  $Y$  (purple), denote different cross-coupling name reactions, each of which are governed by different thermodynamics. Suzuki coupling involves a ligand exchange step prior transmetalation where  $Br$  is replaced by an alkoxy, resulting in  $Y$ -alkoxy rather than  $Y-Br$ . Both Suzuki and Hiyama coupling use activated coupling partners. Volcano plots and cross-coupling partners for (b) Suzuki, (c) Kumada, (d) Negishi, (e) Stille, and (f) Hiyama cross-coupling reactions. The energy descriptor displayed on the  $x$ -axis [ $\Delta E_{RRS}(2)$ ] is the energy of **2** relative to a reference state (**1**) while  $-\Delta E(pds)$ , displayed on the  $y$ -axis, is the energy associated with the potential determining step, or the largest thermodynamic barrier associated with moving between two connected intermediates in the catalytic cycle.

## 5.2 Computational Details

### 5.2.1 DFT Computations and Machine Learning

All intermediates (complex **1**, **2** and vinyl bromide substrate) were optimized in the gas phase at the B3LYP<sup>164-165</sup>-D3<sup>81, 167</sup> with the 3-21G (for Ni, Pd, Cu, and Ag complexes)<sup>168-171</sup> and the def2-SVP<sup>82</sup> basis sets for (Pt and Au complexes) in Gaussian 09.<sup>172</sup> Single point energies were

then computed on optimized structures at the B3LYP–D3/def2-TZVP<sup>82</sup> level. The oxidation states of the catalysts were adjusted to comply with the dominant 14e<sup>−</sup>/16e<sup>−</sup> nature of the complexes for the Suzuki cross-coupling reaction. This means that the overall charge for group 10 metals (Ni, Pd, Pt) is neutral while for group 11 metals (Cu, Ag, Au) carry a +1 charge. The machine-learning procedure was identical to that described in **Chapter 4**.

### 5.2.2 Sketch-Maps

The chemical and structural information of the molecules in the full dataset (~25,000 complexes from the ML database)<sup>39</sup> was vectorized using common cheminformatics descriptors.<sup>226</sup> In particular, each compound was characterized based on eight properties: molecular weight, shape index, molecular volume, total number of carbon atoms, number of phosphorus–metal bonds, number of carbon–metal bonds, number of nitrogen–metal bonds, and number of oxygen–metal bonds. The last four descriptors were weighted by the inverse of the bond lengths. The two-dimensional representation of the vectorized dataset was then constructed using Sketch-Maps.<sup>40–41, 225,227</sup> Sketch-Map is a dimensionality reduction algorithm aimed at mirroring the relationship between high dimensionality data in lower dimensional space. Specifically, the technique involves the transformation of distances in both dimensions with a sigmoid function (Equation 5.2):

$$F(r) = 1 - \left(1 + \left(2^{\frac{a}{b}} - 1\right)\left(\frac{r}{\sigma}\right)^a\right)^{-\frac{b}{a}} \quad \text{Equation 5.2}$$

where  $\sigma$ ,  $a$ , and  $b$  are three parameters that control the switching distance and the rate at which the function moves from zero to unity. Following the procedure detailed in ref. 40, parameters for the Sketch-Map transformation functions were set to  $\sigma = 1.0$ ,  $A = 4$ ,  $B = 4$ ,  $a = 1$ ,  $b = 2$  where  $\sigma$  is the threshold for the switching function, while  $A$ – $B$  and  $a$ – $b$  are, respectively, the high- and the low-dimensional exponents. In contrast to other dimensionality reduction algorithms such as principal component analysis (PCA), which are based on orthogonal linear transformation of data, the use of the sigmoid transformation of the distances confers to the Sketch-Maps enhanced flexibility: the algorithm can focus more on reproducing the pairwise connectivity of the data rather than trying to reproduce the exact distances between points. In practice, the projection of the data is performed through an iterative minimization of a non-linear fitness function, which makes the simultaneous fitting of all data points in the training set impractical. For this reason, the Sketch-Maps were constructed on a subset of 1,000 landmark points for each metal, chosen through farthest point sampling. The remainder of the dataset was projected into the Sketch-Maps using the two-dimensional embedding of the landmark points as a map. The interactive plots were then constructed with the interactive visualization library Bokeh<sup>228</sup> and a SQLite<sup>229</sup> database. The open-source HTML5 viewer JSmol<sup>230</sup> was used for visualization of the 3D structures.

## 5.3 Results and Discussion

The database for our cross-coupling machine-learning model consisted of approximately 25,000 complexes (*i. e.*, ~7,000 for training and ~18,000 for predicting)<sup>39</sup> created from combinations of six metals (Ni, Pd, Pt, Cu, Ag, and Au) and 91 ligands [phosphines, N-heterocyclic carbenes

(NHCs), pyridines, furan, acetone, and carbonyl (CO), see **Appendix B** for a complete list]. As shown in Figure 5.1a, each metal binds to two ligands (L), which can be either the same (*i. e.*,  $L_1=L_2$ ) or different (*i. e.*,  $L_1\neq L_2$ ), which means that complexes are characterized based on their specific combination of a metal and its two adjoined ligands,  $L_1-M-L_2$ . Using the chemical and structural information of each complex (see section 5.2.2 for more information on the collective variables), we created separate maps (*i. e.*, Sketch-Maps) for each metal using a dimensionality reduction algorithm (Figure 5.2). The dimensionality reduction reproduces the relationship between the data into two-dimensional space. Providing an interactive representation of the figures (available on the Materials Cloud at [www.materialsccloud.org/discover/ccg](http://www.materialsccloud.org/discover/ccg)) facilitates the exploration of the entirety of information revealed by the Sketch-Maps. Versatile graphics of this type offer a fast and intuitive way of visualizing the three-dimensional structures of prospective catalysts by selecting a metal and the appropriate range of the descriptor variable (*e. g.*, those corresponding the volcano plateaus presented in Figure 5.1).

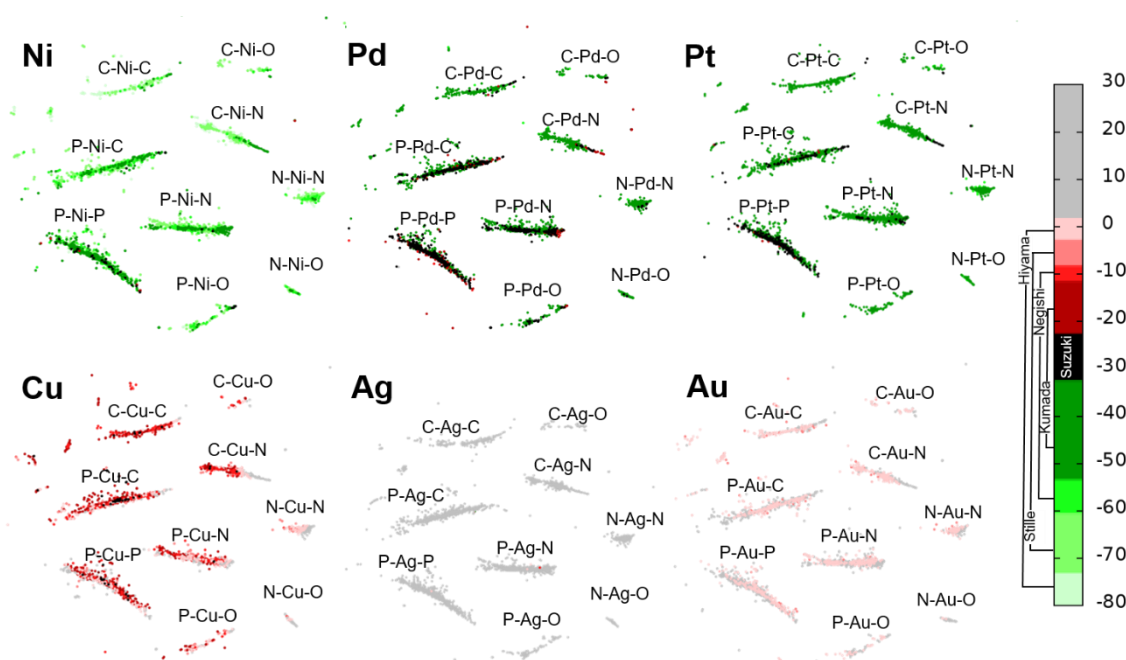


Figure 5.2 The Sketch-Map of each metal colored by the value of the energy-based descriptor variable [ $\Delta E_{RRS}(2)$ , in kcal/mol]. The color code corresponds to the volcano plateau of cross-coupling reactions, where black represent the volcano plateau for Suzuki coupling. Different shades of green and red (see color bar) indicate ideal thermodynamics (*i. e.*, these species fall on the plateau) for other the cross-coupling reactions represented by the other Figure 5.1 volcanoes. Catalysts colored gray are not ideal for any of the coupling reactions examined. Note that the structure (as and associated properties, *vide infra*) for each point can be visualized in an interactive manner on the Materials Cloud ([www.materialsccloud.org/discover/ccg](http://www.materialsccloud.org/discover/ccg)).

Clearly evident within the map for each metal are the nine clusters differentiated based on their chemical diversity, specifically the type of atoms bonded to the metal (*i. e.*, P for phosphines, C for N-heterocyclic carbenes (NHCs) and carbonyl (CO), N for pyridines, and O for furan and acetone). Each of these clusters is further separated based on ligand size, where bulky species are located further left and smaller species further right. For example, the CO ligand always appears on the far right of the clusters ( $L_1-M-L_2$ , wherever either  $L_1$  and/or  $L_2$  is C) owing to its very

To assure that our ML model was providing accurate descriptions for catalysts lying within each of the Figure 5.2 clusters, we examined deviations between the descriptor variable (*i. e.*,  $\Delta E_{\text{RRS}}(\mathbf{2})$  as defined by Equation 5.3) determined from DFT computations (lower circles, Figure 5.3) and those established by ML predictions (upper circles, Figure 5.3). Specifically, Figure 5.3 uses the average value of  $\Delta E_{\text{RRS}}(\mathbf{2})$  for each of clusters in the Figure 5.2 Sketch-Map, along with one additional group, X-M-CO, which represents the combination of one CO ligand with all other ligand types. Overall, our ML predictions for an additional set of  $\sim 18,000$  catalysts showed only minor differences from DFT computed values (generally less than 3 kcal/mol), which aligns well with the mean absolute error of 2.73 kcal/mol inherent to the ML representation.<sup>39</sup> This indicates that the ML derived data can be used to expand our set of catalysts (from  $\sim 7,000$  to  $\sim 25,000$  species), which provides more statistical validity to any chemical trends extracted when unraveling the chemical behavior of catalysts, which to the best of our knowledge, has not been done before. Note that the chemical implications of the Figure 5.3 will be discussed later.



Knowing which species are represented within each cluster, each point was then assigned a distinct color based on the value of its descriptor variable,  $\Delta E_{\text{RRS}}(\mathbf{2})$  (Equation 5.3), which is seen in Figure 5.2 and easily visualized interactively on the Materials Cloud. Using this quantity, the thermodynamic suitability for each of the five Figure 5.1 cross-coupling variants can be assessed based on whether a specific catalyst lies on the volcano plateau (see Figure 5.1b–f for the range

of descriptor values representing the volcano plateau). Catalysts depicted in black represent the best candidates for the Suzuki reaction, as they fall on the plateau of the Figure 5.1b volcano. Species that do not lie on the plateau for the Suzuki volcano (Figure 5.1b) are colored either green or red to indicate overly strong (green) or weak (red) catalyst/substrate binding interactions. While these green and red species are not thermodynamically ideal for Suzuki coupling, they may still be efficient catalysts for one or more of the other aforementioned cross-coupling variants. Indeed, as the volcano plateau is lowered and broadened when different cross-coupling partners are employed (see the Figure 5.1 volcanoes), new catalysts that were previously located on either the right or left slopes (*e. g.*, for Suzuki coupling) begin to appear on the volcano plateau for other cross-coupling variants. Thus, catalysts depicted in the darkest shades of green or red are thermodynamically ideal for all four cross-coupling variants examined here except Suzuki coupling, medium shaded catalysts are ideal for Negishi, Stille and Hiyama coupling, light shades are ideal for Stille and Hiyama coupling, and the palest shades are ideal for only Hiyama coupling. Those catalysts depicted in gray have exceptionally weak binding interactions and are ineffective for any of the variants of cross-coupling reactions.

As might be expected, the first trend that distinguishes itself in Figure 5.2 is that a majority of the P–M–P clusters have at least some black points, which corresponds to good thermodynamic profiles for Suzuki coupling. Taking a closer look, first at the Pd catalysts, it is seen that they are generally extremely robust for catalyzing all of the cross-coupling reactions studied.<sup>159, 191-192, 231-232</sup> Many catalysts found in the P–Pd–P cluster have descriptor values that place these species atop the Suzuki volcano plot. However, not just palladium phosphine catalysts are ideal, there are, at minima, a handful of species belonging to each ligand type combination that are also colored black (see Figure 5.4a). Thus, the Sketch-Maps illustrate that the intrinsic properties of Pd clearly place this metal atop the podium for catalyzing nearly any cross-coupling reactions, with nearly 24% (820 out of 4,185) of all the Pd catalysts assessed having ideal thermodynamics for Suzuki coupling. If we move to a cross-coupling variant with a wider plateau, such as Kumada coupling (Figure 5.4b), then the density of points increases considerably (from 820 to 4,081). Thus, ~98% of the Pd catalysts we tested have ideal thermodynamic profiles for Kumada coupling.

Ni catalysts are shifted the furthest toward strong binding of the group 10 metals. Correspondingly, only a handful of species appear atop the Suzuki volcano (only 63 out of 4,075, Figure 5.4c), with those being located predominately in the P–Ni–P cluster. However, if the plateau region is widened by moving to Kumada coupling, then many more catalysts with ideal thermodynamics appear (1,021 of 4,075). Figure 5.4c also shows that the clusters associated with all different ligand types reform when moving from Suzuki to Kumada coupling. This is true of species that traverse the catalytic cycle as Ni(0)/Ni(II) species (as was examined here) and does not consider the more reactive Ni(I)/Ni(III) oxidation states<sup>196</sup> where the thermodynamic profiles are greatly improved.<sup>61</sup> Platinum catalysts are located between Ni and Pd in terms of binding strength. Despite having a number of catalysts with good thermodynamic profiles, platinum species are virtually absent from the experimental literature.<sup>193-194</sup> It has been postulated that the principle drawbacks of using Pt catalysts are the presence of stronger M–R bonds (relative to Pd), which results in a slow transmetalation process<sup>195</sup> and, general difficulties in completing reductive elimination,<sup>233-234</sup> which slows the reaction rate.

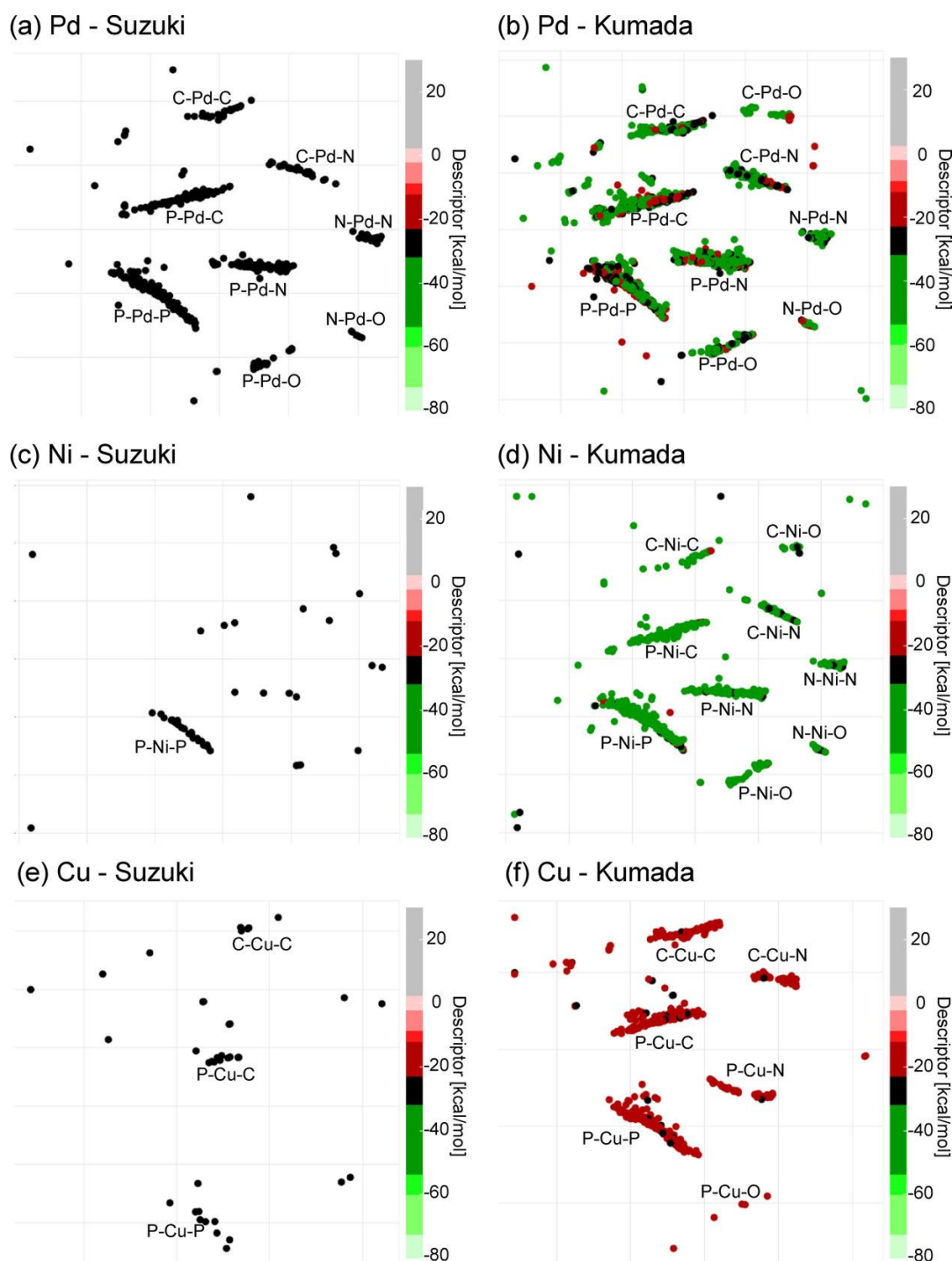


Figure 5.4 Snap-shots of the interactive Sketch-Maps located on the Materials Cloud ([www.materialscloud.org/discover/ccg](http://www.materialscloud.org/discover/ccg)) for selected metals depicting only catalysts with ideal thermodynamics for Suzuki coupling (a, c, e) and for both Suzuki and Kumada coupling (b, d, f).

Group 11 metal-based catalysts show remarkably different behavior than their group 10 counterparts, where the thermodynamic profiles are characterized by very weak binding energies (indicated by red, pink, and gray colored points in Figure 5.2). Cu catalysts represent the most promising group 11 species, with a handful (39 out of 4,184) even having ideal thermodynamics for Suzuki coupling (Figure 5.4e). Unlike Ni, which are often best paired with a specific ligand type (*e. g.*, two phosphine ligands) to induce better catalytic cycle thermodynamics, the 39 Suzuki

copper catalyst are distributed primarily in the C–Cu–C, P–Cu–C, and P–Cu–P ligand clusters. Indeed, numerous examples of Cu based cross-coupling catalysis using various ligand types are present in the literature.<sup>201, 235</sup> Moving from Suzuki to Kumada coupling results in the appearance of many new catalysts (739, Figure 5.4f), as well as several (but not all) different ligand combinations, including phosphine, carbene, and nitrogen coordinating ligands. On the other hand, the thermodynamics for Au, and particularly Ag, are worse than the Cu species and are characterized by binding energies that are far too weak. The origin of the contrasting situations between group 10 and group 11 metals lies in their ability to facilitate oxidative addition. Since group 11 catalysts traverse the catalytic cycle in their M(I)/M(III) oxidation states, there is an inherent problem with oxidative addition. Taking Cu, as an example, while these species can undergo oxidative addition in a similar manner to Pd,<sup>236</sup> the thermodynamics of the process are far worse, with reductive elimination to reform the starting materials<sup>236-237</sup> being a significant problem. The observation of this back reaction indicates that the thermodynamics of oxidative addition are more unfavorable in group 11 than in group 10 catalysts, which is reflected in the weak binding energies of Cu, Ag, and Au catalysts seen in Figure 5.2.

While the metal-center clearly plays a central role in dictating catalytic performance, the impact of the ancillary ligands is also critically important for fine-tuning a catalyst's affinity for facilitating oxidative addition and commencing the catalytic process. Indeed, the influence of the ligands in terms of stereoelectronics (*i. e.*,  $\sigma$ -donating and  $\pi$ -accepting ability), as well as ligand bulk, has long been of keen interest to the experimental organometallic community. Notably, work of Tolman concerning the influence of ligand sterics<sup>238</sup> and early observations that stereoelectronic effects were also important in influencing reactivity<sup>239-241</sup> laid the groundwork for a renaissance in examining the influence of ligands in cross-coupling reactions.<sup>8, 214, 242-247</sup> Figure 5.2 also highlights many of the anticipated trends that are firmly established by decades of experimental work. The energy of oxidative addition (the descriptor used in our volcano plots) is strongly influenced by the electronic properties of the ligands, where more electron donating species result in a more facile process.<sup>248-249</sup> For instance, N-heterocyclic carbenes are stronger  $\sigma$ -donors than even the most electron rich phosphines (based on measurements of CO stretching frequencies<sup>250-251</sup>). Turning again to the Pd map in Figure 5.2, the predominance of green colored catalysts indicate that the C–Pd–C catalysts (which includes the NHC ligands) have stronger binding energies than the P–Pd–P catalysts, which are predominantly colored black. Catalysts bearing one phosphine and one carbene ligand (*i. e.*, the P–Pd–C cluster) show the expected intermediary binding energies. Conversely, the presence of strong  $\pi$ -acceptors/weak  $\sigma$ -donors, such as CO, induce a strong shift toward weaker binding, and appear as red colored catalysts located at the right periphery in most of the X–Pd–C clusters.

In order to highlight that variations in descriptor value do arise from the  $\sigma$ -donating/ $\pi$ -accepting properties of the ligands, we computed the HOMO energy of intermediate **1** (see Figure 5.1a) for each catalyst. Through ligand field theory,<sup>252-253</sup> the HOMO energies should relate to the donor/acceptor properties of the ligand, which, in turn, should relate to the value of our descriptor variable. The Figure 5.5 maps for Pd catalysts shows the existence of this relationship. Strong  $\pi$ -acceptor ligands, such as CO, are characterized by very negative HOMO energies (*i. e.*, more red points in right map of Figure 5.5) while strong  $\sigma$ -donor, such as carbenes, are characterized by much higher HOMO energy levels (*i. e.*, more blue/green points in the right map of Figure 5.5). These values are roughly inversely correlated with the values of the energy descriptor



(Figure 5.5, left map) where, for instance, carbene species tend to have much stronger binding energies that are indicated by a considerable number of dark blue point in the C–Pd–C cluster.

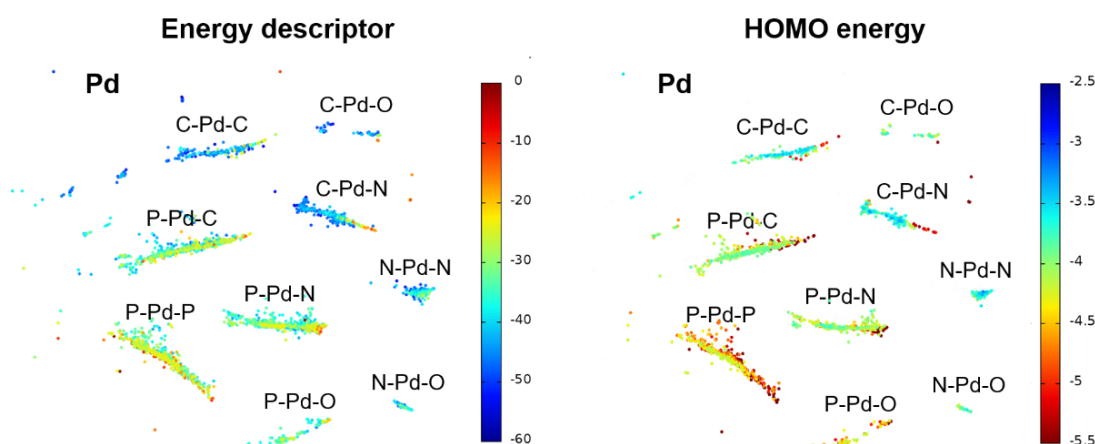


Figure 5.5 Sketch-Maps of palladium complexes colored by the energy descriptor (left, in kcal/mol) and the HOMO energy of intermediate **1** (right, in eV). The color bar of HOMO energy is set as the inverse of the energy descriptor for easier comparison. The Sketch-Maps for other metals can be found in the **Appendix B**.

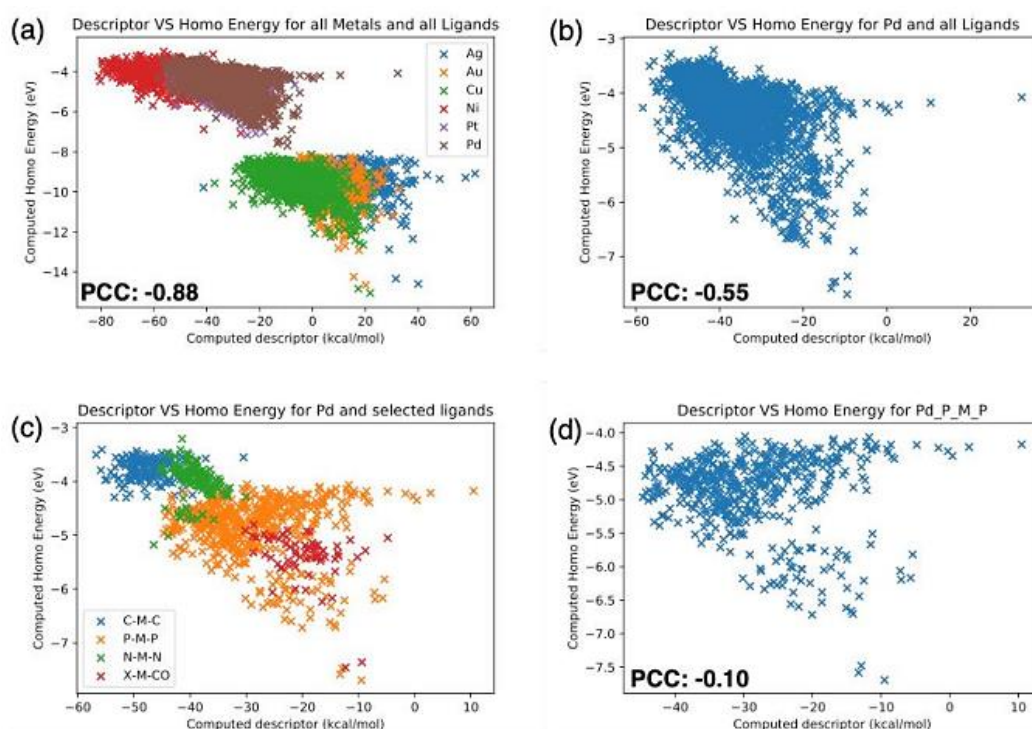


Figure 5.6 Correlations between the computed descriptor variable (in kcal/mol) and the HOMO energy (in eV) for (a) all catalysts, (b) all Pd catalysts, (c) selected Pd catalyst colored by ligand type, and (d) P–Pd–P catalysts. PCC values within the plots represent the Pearson correlation coefficient.

These correlations can be placed on more quantitative footing by examining statistical measurements of correlation. For example, determination of Pearson correlation coefficients (PCC) for all catalysts in the test set or just those containing a Pd metal center yield values of



−0.88 and −0.55 (−1 being a perfect negative correlation), respectively (Figure 5.6a and b). If the Pd catalysts are further broken down by different ligand combinations (such as those clustered together in the Figure 5.5 Sketch-Map) reveals that these different ligand classes are defined by a more compact grouping of descriptor variables and HOMO energies (Figure 5.6c). While the overall PCC can be relatively high, within each individual metal/ligand subgroup the same can be quite low. For instance, phosphine ligands show strong variations in sterics with minimal corresponding change in stereoelectronics (this appears as a horizontal line of catalysts with HOMO values between −4.0 and −4.5 eV, Figure 5.6d), which results in much lower correlations (*e. g.*, −0.10 for the P–Pd–P subgroup of catalysts). Other ligand classes show better correlations between steric and stereoelectronic properties (*e. g.*, N–M–N). Taken together, the separations in descriptor variable and HOMO energy allows us to extract meaningful chemical trends that exist between (but perhaps not within) different ligands classes.

The relationships between HOMO energy and descriptor value ( $\Delta E$ ) for each of the different metals are summarized more succinctly in Figure 5.7. Taking the average value of the P–M–O subgroup as an arbitrary reference (*i. e.*, setting the average  $\Delta E$  and HOMO values of all the catalysts in the P–M–O cluster to zero), each of the other clusters can then be plotted as a shift of the binding energy ( $\Delta\Delta E$ ) along the *x*-axis and  $\Delta HOMO$  along the *y*-axis relative to the P–M–O reference. The background colors of the plots correspond to the previously used color codes that represent the actual (as opposed to relative) value of the descriptor variable in the Figure 5.2 maps. Thus, a black background corresponds to being on the volcano plateau for Suzuki reaction, whereas green and red represent overly strong (green) or overly weak (red) binding interactions, and gray extremely weak interactions.

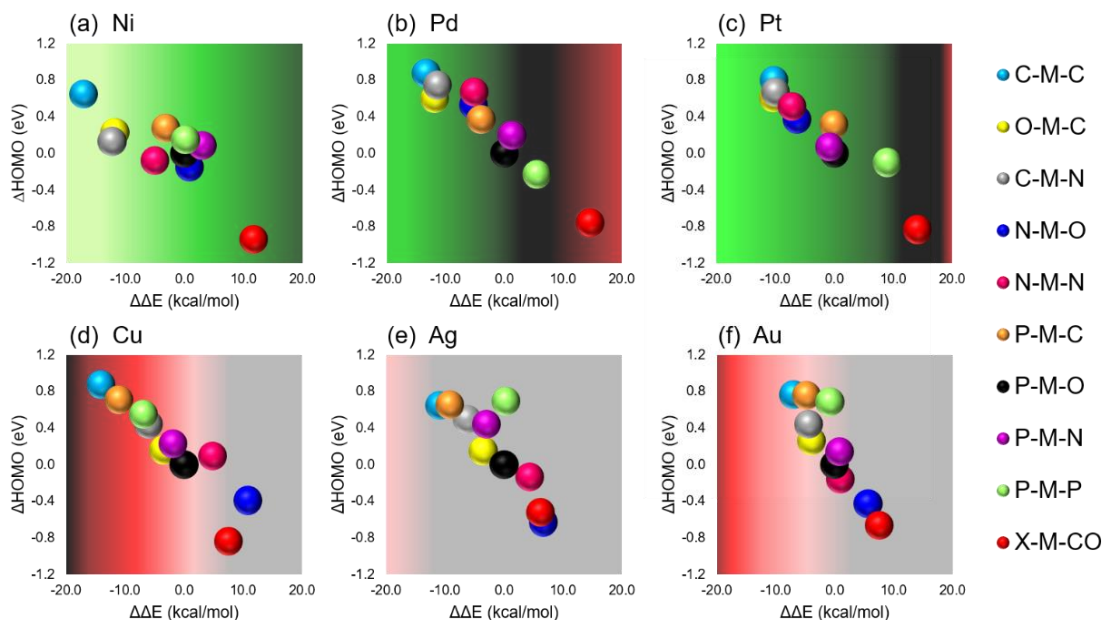


Figure 5.7 Relationships between the average energy descriptor and the HOMO energy of complex 1 ( $L_1$ –M– $L_2$ ) for each metal. All energies are relative to the values of P–M–O ligand type (black bubble). The background corresponds to the color code of the Sketch-Map in Figure 5.2, where the black region indicates the volcano plateau for Suzuki coupling, while green and red indicate overly strong and overly weak binding interaction, respectively. The Pearson correlation coefficients for each plot are: −0.51 (Ni), −0.96 (Pd), −0.92 (Pt), −0.90 (Cu), −0.85 (Ag), and −0.86 (Au).

Figure 5.7 illustrates that each class of ligands, with their varying degrees of  $\sigma$ -donating/ $\pi$ -accepting ability, influence catalysts belonging to the same metal group (*e. g.*, Ni, Pd, and Pt) in analogous ways. The aforementioned trends regarding shifts toward stronger binding brought about by strong  $\sigma$ -donors are seen in the Ni, Pd, and Pt plots, where all X–M–C (or C–M–X) species are located further to the left (indicative of stronger binding) and higher (indicative of less negative HOMO energies) than the reference P–M–O species. Catalysts with  $\pi$ -accepting ligands, such as those bearing two phosphine (P–M–P) or at least one CO (X–M–CO), show the opposite effect, and are shifted toward weaker binding while also illustrating the anticipated reduction in the HOMO energy (Figure 5.7a–c). Group 11 catalysts, curiously, show markedly different trends than the group 10 species, particularly in the influence brought about by phosphine ligands. While carbene (*e. g.*, C–M–C) and CO (X–M–CO) ligands still induce the largest shifts toward strong and weak binding, respectively, catalyst with two phosphine ligands (P–M–P) in tandem with a group 11 metal are shifted toward stronger, rather than weaker, binding. This non-intuitive behavior of phosphine ligands may explain experimental observations of Cu catalyzed Suzuki–Miyaura C–C cross-couplings with phosphine ligands.<sup>254</sup> Overall, the clearest route to improving the thermodynamic profiles of group 11 catalysts is including strong  $\sigma$ -donating ligands that result in stronger binding between the catalyst and the substrate. These ligands are capable of concentrating negative charge on the metal center (see correlation between the energy descriptor and Mulliken charges in **Appendix B**), which makes oxidative addition more thermodynamically favorable.

While the discussion above consisted of a more “traditional” chemical interpretation of metal/ligand effects on catalysis, we are also very interested in moving toward using big-data type analysis in computational catalysis research, with the objective of revealing hidden relationships that can only be uncovered by examining large numbers of species. Following this line of thinking, Figure 5.8 reveals a new way of investigating catalytic behavior by examining the distribution of energy descriptor values for various metal/ligand cluster combinations. One of the clearest, and perhaps least surprising, conclusions that is quickly drawn is that catalysts incorporating a palladium atom have nearly perfect thermodynamic profiles, regardless of the surrounding ligands. The “occurrence curves” depicted in Figure 5.8 for Pd catalysts that feature at least one phosphine ligand (*i. e.*, Figure 5.8a, b, c, f) are essentially centered within the yellow highlighted region, meaning they are thermodynamically ideal even for Suzuki coupling, whose volcano has the smallest plateau region. Pd catalysts with carbene ligands (Figure 5.8j) are centered toward stronger binding, but the distribution is bimodal (*i. e.*, there is a smaller second distribution lying to the right of the largest distribution), showing that a smaller subset of species exists that show ideal binding strengths. This same type of bimodal distribution is also observed in the N–M–O (Figure 5.8g), N–M–N (Figure 5.8h), and C–M–N ligand clusters (Figure 5.8i) not only for Pd, but also for each of the other metal types, which likely results from the distinct types (*i. e.*, amine and pyridines) of nitrogen bound ligands. Also very evident is that Ni and Ag consistently represent the extremes of strong and weak binding, respectively. Each of the four other metals are characterized by binding energies intermediary to these two metal extremes.

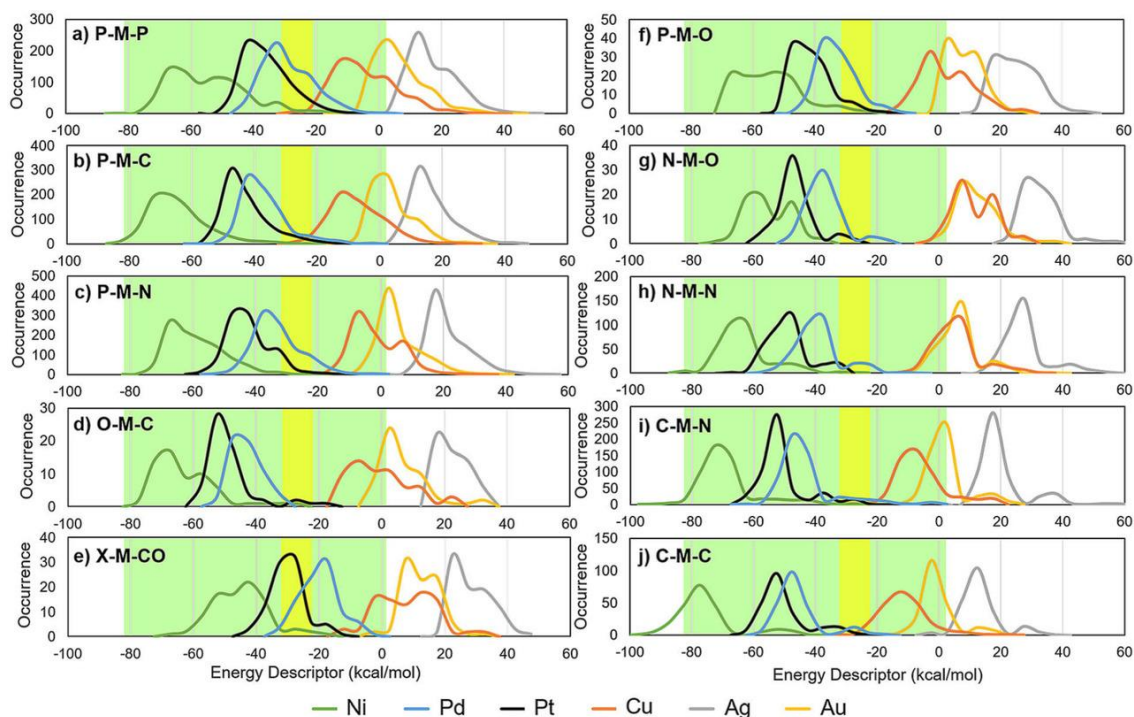


Figure 5.8 Distribution of metal complexes at different energy descriptor values according to ligand cluster type (a–j). The volcano plateau for the Suzuki coupling is highlighted in yellow and for Hiyama coupling in green.

Figure 5.8 also uncovers interesting facts regarding how, and to what extent, certain ligand groups are able to “polarize” the metal center. Ligands such as N–M–N (for the group 10 catalysts, Figure 5.8h) and C–M–N (for the group 11 catalysts, Figure 5.8i) induce markedly different binding energies for the different metals. Putting this into practical application, the distinct separation seen in, for example, the group 11 catalysts belonging to the C–M–N cluster show that this combination of ligands is particularly adept at shifting Cu catalysts away from their Ag and Au counterparts, toward stronger binding and better catalytic cycle thermodynamics. A collective examination of all of the Figure 5.8 plots shows that for the group 10 catalyst, Pd and Pt tend to move in tandem with one another. Nonetheless, certain ligand combinations, such as C–M–C, cause Pd and Pt to appear far different from Ni catalysts, while others (*e. g.*, P–M–O and P–M–P) cause each of the group 10 metals to look more similar to one another.

Similar to the ability of certain ligand subgroups to “polarize” the metals belonging to the same group, some ligands possess a noticeable ability to make metals belonging to different groups (*i. e.*, group 10 and 11) look more similar or dissimilar to one another. For example, Figure 5.8a shows that phosphine ligands cause group 10 and group 11 metals to appear similar to one another by shifting the group 10 catalysts toward weaker binding and shifting the group 11 catalysts toward stronger binding. The result is that the curves representing Pd (the weakest binding group 10 metal) and Cu (the strongest binding group 11 metal) significantly overlap one another (Figure 5.8a). The same effect is seen when at least one carbonyl ligand is present (Figure 5.8e). In contrast, ligands that interface with the metal center through a nitrogen atom (*e. g.*, N–M–O, N–M–N) induce the opposing situation where the group 10 and 11 catalysts shift away from one another (*i. e.*, the already strong binding group 10 catalysts move toward even stronger binding and the already weak binding group 11 catalysts toward even weaker binding).

Aside from the overarching inferences regarding how the different ligand clusters influence the behavior of various metal centers discussed above, Figure 5.8 also shows the exact metal/ligand recipe likely to produce good cross-coupling catalysts. While it is well known that certain metal/ligand combinations, most famously palladium-phosphine, nearly always constitute attractive catalysts, certain other combinations may be less intuitive and obvious. For example, the very broad (*i. e.*, spanning a large range of descriptor values) occurrence distributions seen for Cu based species indicates that some ligands sufficiently alter the metal through strong stereoelectronic influences such that their binding energy falls onto the Suzuki coupling volcano plateau (yellow highlighted area). As the plateau area is widened for other cross-coupling variants, a more diverse group of catalysts become thermodynamically viable. For example, Figure 5.8 shows that for Hiyama coupling (green highlighted area), all of the metals, with the notable exception of silver, have at least part of their occurrence curves overlapping the plateau region. This indicates that a host of “nontraditional” cross-coupling catalysts featuring Cu and Au metals should have ideal thermodynamic for catalyzing Hiyama coupling. For further discussion on synergy between thermodynamic volcano plots and experiment, the interested reader is referred to ref. 61.

Generally speaking, Figure 5.8 shows how the magnitude and direction of the shifts within and between disparate metal types induced by various ligand groups can be used as a tool to fine-tune a catalyst’s binding energy in order to obtain an ideal thermodynamic profile. Some particularly interesting classes of species that may be worthy of more detailed examination by computation or experiment include combinations of copper/gold with carbene or phosphine ligands. While we have focused on a class of reactions that is well understood through decades of both experimental and computational exploration, the true power of the type of approaches discussed here will be found in analyzing new reaction classes that are relevant to the pressing chemical problems of today.

## 5.4 Conclusions

In conclusion, we analyzed a combination of DFT-derived and machine-learned data for a series of more than 25,000 catalysts with the objective of employing a big-data type analysis to uncover hidden chemical trends in cross-coupling catalysts. The interactive use of data-driven representations aided in quickly revealing which metal/ligand combinations make effective catalysts, based on their position on a series of molecular volcano plots representing different “named” cross-coupling reactions. A large-scale data analysis revealed how certain ligand types are able to “polarize” different metals, which causes catalysts bearing different metal centers to appear more similar or dissimilar to one another. This information can, in turn, be used to systematically tune specific metal/ligand combinations to obtain a desired thermodynamic profile for the catalytic cycle. While we have studied a class of reactions with a long and storied history in chemistry, the real power of the type of approach described within will be in the exploration of new chemistries, where harnessing big-data type analyses will reveal the strengths and weaknesses of different catalytic processes.

## 6 Unraveling Metal/Pincer Ligand Effects in the Catalytic Hydrogenation of Carbon Dioxide to Formate

This chapter is based on following publication:

Sawatlon B., Wodrich M. D., and Corminboeuf C., Unraveling Metal/Pincer Ligand Effects in the Catalytic Hydrogenation of Carbon Dioxide to Formate. *Organometallics* **2018**, 37, 4568–4575.

### 6.1 Introduction

C1 compounds, *e.g.*, methanol, methane, and formic acid/formate salt,<sup>255-261</sup> are important platform chemicals that find use in applications ranging from fabric dyeing and printing processes to hydrogen storage.<sup>262-265</sup> As an example, the hydrogenation of carbon dioxide (CO<sub>2</sub>) to formic acid/formate salt provides a promising route using a readily available carbon feedstock<sup>266</sup> to develop liquid organic hydrogen-carriers (LOHCs) systems.<sup>267</sup> The principal shortcoming of the hydrogenation reaction (Figure 6.1a), however, is that CO<sub>2</sub> is an inherently inert and stable gas, meaning that transition metal catalysts are generally required to facilitate chemical transformations. Such catalytic processes typically operate under harsh conditions (high temperature and/or pressure of CO<sub>2</sub> and H<sub>2</sub>), thus it is hardly surprising that new systems that operate under milder conditions are continuously being developed.<sup>268-271</sup>

To date, the most promising homogeneous hydrogenation pincer-catalysts consist of transition metal complexes (*e.g.*, Ir, Rh, Ru, and Fe) incorporating PNP pincer ligands based on a 2,6-dimethylpyridine scaffold where the P and N atoms are directly bound to the metal center.<sup>272-275</sup> The connecting atoms of the pincer backbone can be swapped with either N and C, which leads to a host of different ligand families (*e.g.*, PNN, NNN, PCP),<sup>276-278</sup> as illustrated in Figure 6.1b. Perhaps the most well-known homogeneous catalytic system for the hydrogenation of CO<sub>2</sub> to formate used a Ir<sup>III</sup>–PNP catalyst (PNP = 2,6-bis(di-iso-propylphosphinomethyl)pyridine) and yielded a maximum turnover number of 3,500,000.<sup>279-280</sup> Subsequent experimental<sup>281-286</sup> and theoretical<sup>103, 287-294</sup> investigations have also revealed that this, as well as other metal–pincer complexes should be capable of catalyzing the hydrogenation reaction. Nevertheless, despite considerable experimental and computational work, a comprehensive picture on why certain metal/ligand combinations are superior to others remains lacking.

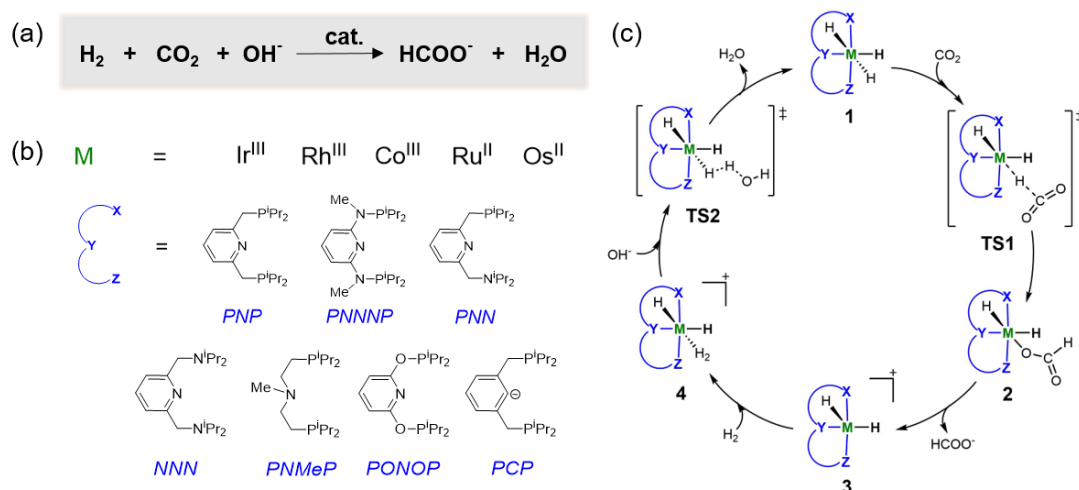


Figure 6.1 (a) The hydrogenation reaction of carbon dioxide to formate, (b) different metal–ligand combinations studied in this work, and (c) the proposed catalytic cycle.

The first step in uncovering such information is to firmly establish the details of the reaction mechanism, which can provide important evidence on how a catalyst might be modified to improve its activity. For example, the prototypical  $\text{Ir}^{\text{III}}$ –PNP catalyzed hydrogenation of  $\text{CO}_2$  (Figure 6.1c),<sup>290</sup> proceeds through a catalytic cycle that first involves complexation of carbon dioxide with the hydrido ( $\text{H}^-$ ) ligand of **1**, hydride extraction (**TS1**) then ultimately leads to a formate ion bound to the catalyst *via* its oxygen atom, **2**. Dissociation of the formate ion then produces cationic species, **3**, which is followed by the addition of molecular hydrogen ( $\text{H}_2$ ) to form **4**. The heterolytic cleavage of  $\text{H}_2$  and abstraction of the proton by a hydroxide base ( $\text{OH}^-$ ) (**TS2**) closes the catalytic cycle by producing a molecule of water and regenerating **1**. Previous computational results have identified the rate-determining step for the  $\text{Ir}^{\text{III}}$ –PNP catalyst as involving heterolytic  $\text{H}_2$  cleavage and proton transfer to the hydroxide anion (*i.e.*, **4**→**TS2**).<sup>290</sup> Nonetheless, the kinetic profiles that govern catalytic efficiency, or even the specific rate-determining step itself, may be different for other species. Obtaining this critically important information, however, requires computing the entire catalytic cycle for each individual catalyst, an arduous task that is both time consuming and computationally expensive.

Fortunately, we have recently shown that the kinetic profiles of catalysts can be estimated at greatly reduced computational cost using molecular volcano plots.<sup>5, 39, 61–65</sup> Volcano plots,<sup>45–46</sup> which are widely used in heterogeneous and electrocatalysis,<sup>4, 295–296</sup> and, more recently, have been invoked to rationalize<sup>60, 95, 223, 297</sup> and predict<sup>5, 39, 61–65</sup> the performance of homogeneous catalysts. By computing a single descriptor variable for each catalyst, such as the reaction energy of one step of the catalytic cycle, it is possible to assess the thermodynamics or kinetics of the entire catalytic cycle through linear free energy scaling relationships (LFESRs).<sup>52, 55, 224</sup> After constructing the volcano plot from the corresponding LFESRs, the properties of various catalysts can be quickly assessed from their position on the plot. Catalysts lying on the plateau or near the peak are considered good candidates that possess balanced catalyst/substrate interactions, in accordance with Sabatier’s principle.<sup>36</sup> On the other hand, catalysts located far from the peak (or plateau) on the left or right slopes have interaction energies that are either too strong (making release of the final products problematic) or too weak (making entry into the catalytic cycle difficult), respectively. Despite being based on thermodynamic concepts, these same principles

are valid for kinetic quantities such as transition state barrier heights.<sup>54, 64-65, 298-301</sup> The objective of this work is to unravel how various metals and ligands work in tandem to catalyze the hydrogenation of carbon dioxide to formate using molecular volcano plots. Using these tools, we reveal guiding principles that deliver strategies for improving the activity of prospective catalysts through manipulation of the catalytic cycle kinetics.

## 6.2 Computational Details

The geometries of all species were first optimized at the M06<sup>302-303</sup>/def2-SVPD<sup>82</sup> level using the “ultrafine” integration grid<sup>304</sup> along with the SMD<sup>83</sup> implicit solvation model (solvent = water) in Gaussian09.<sup>172</sup> An analysis of the resulting vibrational frequencies ensured that each structure was either a minima (zero imaginary frequencies) or a transition state (one imaginary frequency) on the potential energy surface. To obtain a complementary picture of the catalyst free energies profiles, single point electronic energies were determined on the M06 optimized geometries using the B3LYP<sup>164-165</sup> functional appended with a density-dependent dispersion correction –dDsC<sup>305-308</sup> (*i.e.*, B3LYP–dDsC) in tandem with the aug-TZ2P (TZ2P for the transition metal species) Slater-type basis sets as implemented in ADF.<sup>309-310</sup> Free energy corrections (from the M06 computations) were determined using the rigid-rotor harmonic oscillator model (as proposed by Grimme<sup>311</sup>) along with a correction for translational entropy in solution (as proposed by Whitesides<sup>312</sup>) within the GoodVibes program<sup>313</sup> developed by Paton and Funes–Ardoiz. Final solvation correction values (in water) were then determined at the B3LYP–dDsC/aug-TZ2P level using the COSMO–RS solvation model,<sup>314</sup> also as implemented in ADF. For all intermediates and transition states, wave function stability was tested to confirm that the singlet state is the ground state.

## 6.3 Results and Discussion

### 6.3.1 Linear Free Energy Scaling Relationships

Based on the reaction mechanism depicted in Figure 6.1c, we computed free energies for each of the catalytic cycle intermediates and transition states for a total of 35 pincer catalysts (combinations of five metals and seven ligands shown in Figure 6.1b). Analysis of the Figure 6.1c reaction mechanism, as well as an alternative ligand cooperative pathway (see **Appendix C** for details), confirmed the favorability of the non-cooperative mechanism.<sup>279, 290-291</sup> Linear free energy scaling relationships (LFESRs) were then established by calculating the stabilities of each intermediate and transition state relative to a reference state ( $\Delta G_{\text{RRS}}$ ), which we selected as the entry point into the catalytic cycle, **1**. An assessment of different potential descriptor variables identified  $\Delta G_{\text{RRS}}(\mathbf{4})$  (as defined in Equation 6.1) as being the most suitable, as strong LFESRs exist between this variable and the stabilities of other catalytic cycle intermediates and transition states, as illustrated in Figure 6.2.





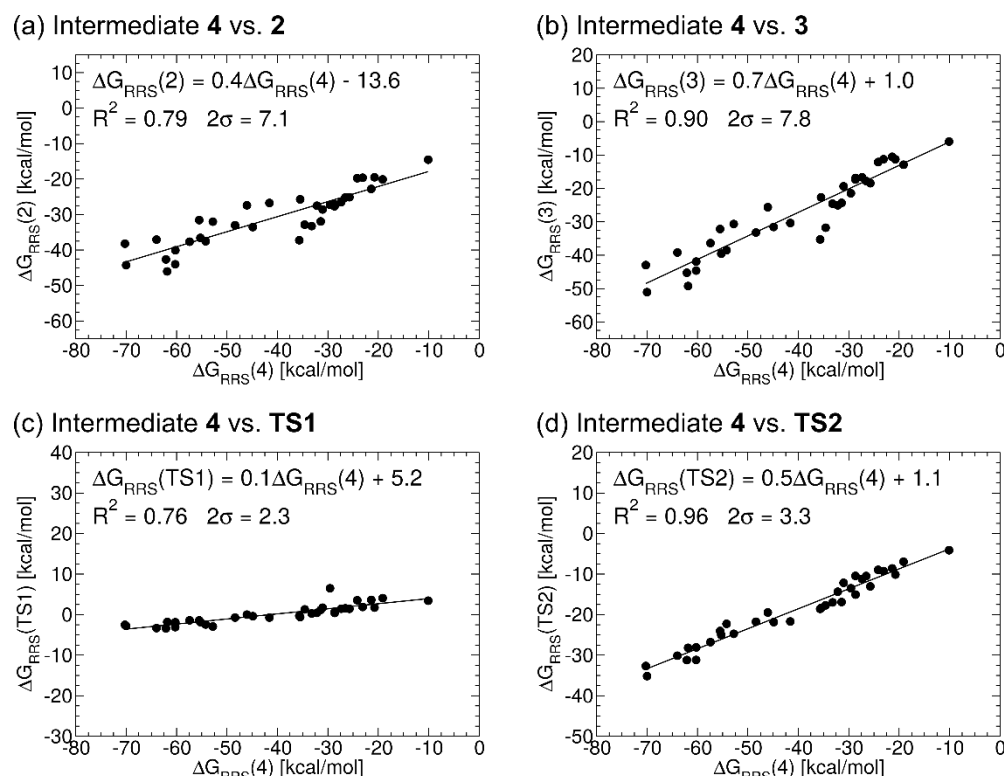


Figure 6.2 Linear free energy scaling relationships between the descriptor variable,  $\Delta G_{\text{RRS}}(4)$  and intermediates [(a)  $\Delta G_{\text{RRS}}(2)$  and (b)  $\Delta G_{\text{RRS}}(3)$ ] as well as transition states [(c)  $\Delta G_{\text{RRS}}(\text{TS1})$  and (d)  $\Delta G_{\text{RRS}}(\text{TS2})$ ] for the hydrogenation of carbon dioxide to formate.

### 6.3.2 Molecular Volcano Plots

Having established LFESRs along with a suitable descriptor variable, the Figure 1 reaction can be cast into a simulated reaction profile (Figure 6.3a). This profile provides an estimate of the free energy associated with moving between any two linked intermediates and transition states of the catalytic cycle directly by knowing the value of the descriptor [*e.g.*,  $\Delta G_{\text{RRS}}(4)$ ] for the reaction. The overall nature of the steps largely agrees with previous work. For instance, formation of the formate ion through hydride abstraction from the catalyst by  $\text{CO}_2$  is shown to be a very energetically facile process (*i.e.*, the black line indicates most values are roughly thermoneutral), in agreement with the recent work of Heimann *et al.*<sup>315</sup> Moreover, in many cases proton abstraction by the hydroxide basis in anticipated to be the rate determining step (*i.e.*, the magenta line predicts very endergonic reaction values for species falling in the left section of the plot), as is the case for the aforementioned  $\text{Ir}^{\text{III}}$ –PNP catalyst.<sup>290</sup>

The final volcano plot (Figure 6.3b) can be obtained directly from the simulated reaction profile (Figure 6.3a) by taking only the most energetically costly reaction step [*i.e.*, the kinetic determining step,  $-\Delta G(\text{kds})$ , plotted along the y-axis] for each descriptor (*x*-axis) value. For the hydrogenation of  $\text{CO}_2$  to formate, the volcano plot is divided into two sections, each of which is governed by a different reaction step that is the most energetically costly. The two steps correspond to the dissociation of the oxygen bound formate ion from the catalyst ( $2 \rightarrow 3$ ) as well as heterolytic  $\text{H}_2$  cleavage and proton abstraction by the hydroxide base ( $4 \rightarrow \text{TS2}$ ). Curiously, in contrast to the more typically seen molecular volcano picture in which the slopes can be



directly interpreted as paralleling substrate binding (*i.e.*, oxidative addition) and product release (*i.e.*, reductive elimination), the Figure 6.3 plot is a bit more complicated. While our previous work used the binding energy of a substrate directly with the metal center as a descriptor variable,<sup>5, 61-65</sup> here the chosen descriptor represents the energy difference between the binding of a hydrido ligand (**1**) and molecular H<sub>2</sub> (**4**). As such, the typical strong-binding/weak-binding nature of the left/right slopes becomes harder to distinguish, yet remains present. In essence, species with negative *x*-axis values have stronger interaction with molecular hydrogen than with a hydride ion. Thus, catalysts falling on the left slope bind H<sub>2</sub> in an overly strong fashion, making the transition from H<sub>2</sub> to H<sup>-</sup> (*i.e.*, **4**→**TS2**) difficult and energetically costly. Moving rightward along the *x*-axis, species will have an increasingly strong interaction for H<sup>-</sup> and an accompanying reduction in H<sub>2</sub> binding ability, which manifests itself in a gradual reduction in the **4**→**TS2** barrier height. Ultimately, the ability of the catalysts to bind anionic species (such as H<sup>-</sup> or formate ion) becomes quite strong, which causes the release of oxygen bound formate species (**2**→**3**) to become the most energetically costly reaction step.<sup>i</sup> Taken as a whole, the Figure 6.3b volcano plot indicates that an ideal hydrogenation catalyst must have a finely tuned ability to bind both neutral (*e.g.*, molecular hydrogen) and anionic (*e.g.*, hydride and formate), each of which are present in the catalytic cycle, neither too weakly nor too strongly.

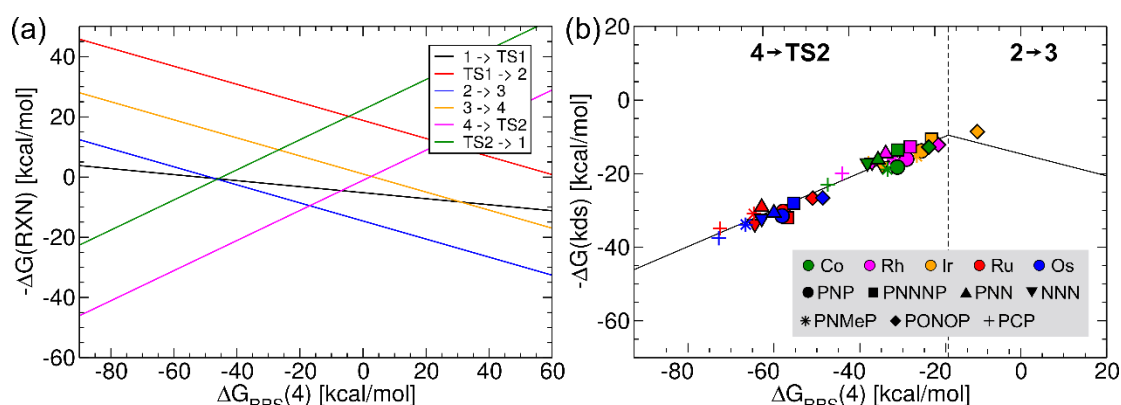


Figure 6.3 (a) Simulated reaction profile depicting the anticipated energy associated with each step [ $\Delta G(\text{RXN})$ ] of the catalytic cycle for a specific value of the descriptor variable (*x*-axis). (b) A volcano plot that illustrates anticipated activity for the hydrogenation of CO<sub>2</sub> to formate by predicting the value of the most energetically difficult reaction step to complete [*e.g.*, the kinetic determining step (kds)]. Note that Ir(PNP) (orange circle) is an experimentally known catalyst that facilitates the hydrogenation reaction.

Individual catalysts are placed onto the volcano according to their energy-based descriptor (*x*-axis) and the free energy associated with the appropriate kinetic determining step (kds, *y*-axis). For the 35 catalysts tested here, a vast majority fall on the left side of the volcano, meaning that the most energetically costly reaction step is the heterolytic H<sub>2</sub> cleavage and proton transfer (**4**→**TS2**). As mentioned earlier, this matches the findings of a previous computational study of Ir(PNP) pincer.<sup>290</sup> The sole exception is Ir(PONOP) (orange diamond, Figure 6.3b), which lies on the opposing side of the volcano as the other catalysts where the most costly step is dissociation of the formate ion (**2**→**3**).

<sup>i</sup> Note that a scan of the potential energy surface revealed no TS for the dissociation of the formate ion from the catalyst. This is because this dissociation process is, overall, thermodynamically unfavorable.

A closer examination of Figure 6.3b shows that the catalysts are roughly grouped into two regions based on their metal center. Catalysts bearing group 9 metals (Co, Rh, Ir) lie closer to the volcano peak, indicating that they are anticipated to have more pronounced ability to catalyze the reaction in question than group 8 metals (Ru, Os), which fall further from the volcano peak along the strong-binding (left) slope. As expected, the experimentally characterized Ir(PNP) complex<sup>279</sup> (orange circle, Figure 6.3b) is amongst a handful of catalysts that the volcano plot predicts will have nearly maximal activity. Other species having roughly comparable activities to Ir(PNP) include: Ir(PONOP), Rh(PONOP), and Co(PONOP). The latter should be of particular interest given the cost associated with using earth-abundant metals to catalyze reactions. On the other hand, the Ru and Os catalysts studied here show significant stabilization of the catalytic cycle intermediates, particularly **4**, relative to the desired reaction product. As such, these species are anticipated to exhibit little or no activity for the hydrogenation of CO<sub>2</sub>.

Figure 6.4 provides a free energy diagram that depicts the catalytic energetics for these four species. As expected, the highest energy barriers of these species are identical to those predicted by the volcano plot, indicating that the free energy predictions made by the linear free energy scaling relationships closely match actual values obtained directly from DFT computations. For Ir(PNP), Rh(PONOP), and Co(PONOP), heterolytic H<sub>2</sub> cleavage and proton transfer (**4**→**TS2**) is most costly and requires 13.8, 12.2, and 12.8 kcal/mol, respectively. In contrast, Ir(PONOP) needs only 6.0 kcal/mol to complete this reaction step. For Ir(PONOP), the most costly reaction step is dissociation of the formate ion (**2**→**3**), which requires 8.6 kcal/mol. Overall, Figure 6.4 shows that all of these catalysts are experimentally viable, with Ir(PONOP) anticipated to have particularly good activity.

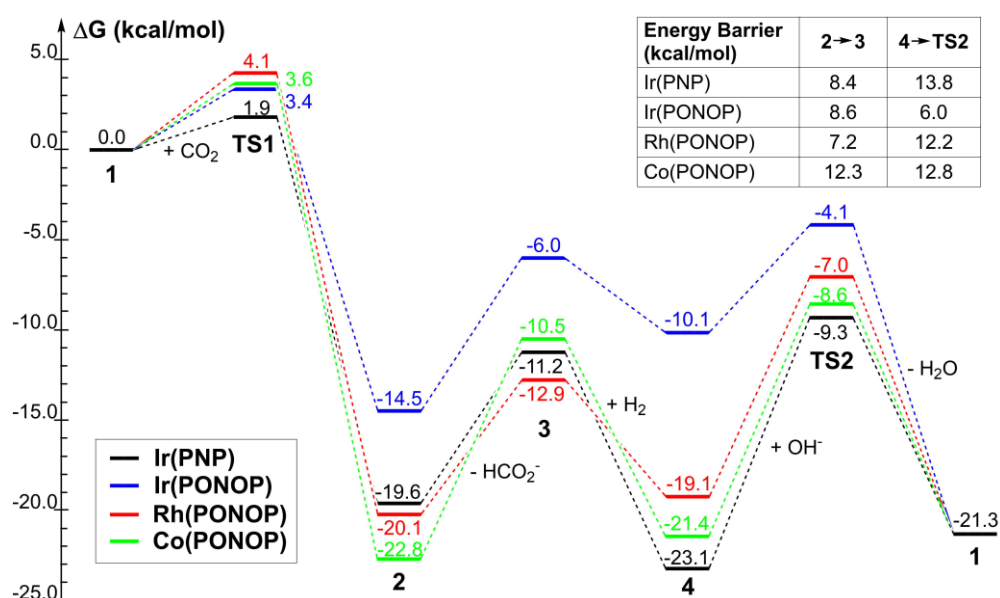


Figure 6.4 Computed free energy diagrams of the catalytic cycles of selected species.

Despite the fact that all catalysts appear to be clustered according to their metal center, the nature of the ligand does have an influence on the efficiency of the catalytic process. These changes become most evident if the relative positions of ligands are considered. Clear trends emerge if we arbitrarily set the PNP-ligand complexes for each metal type as the “zero point” and plot the

corresponding shifts to weaker or stronger binding of the other ligands. This reveals, for instance, that the PONOP ligand produces a considerable shift of a catalyst toward weaker binding (*i.e.*, right along the  $x$ -axis, Figure 6.5). Since the Rh(PNP) and Co(PNP) catalysts already lie on the strong binding (left) side of the volcano peak, the rightward shift brought about by replacing the PNP with the PONOP ligand leads to enhanced activity, as indicated by Rh(PONOP) and Ir(PONOP) being closer to the volcano peak than Rh(PNP) and Ir(PNP). A smaller, yet similar effect is seen for the PNNNP ligand (Figure 6.5). In contrast, the PCP ligand provides an opposing effect, producing a significant shift toward stronger binding (*i.e.*, leftward along the  $x$ -axis, Figure 6.5). For the reaction studied here, replacing the PNP ligand with a PCP ligand reduces catalytic activity by shifting all catalyst leftward away from the volcano peak (Figure 6.3b). The use of NNN, PNMeP and PNN ligands also produce similar, yet less dramatic, leftward shifts along the volcano plot.

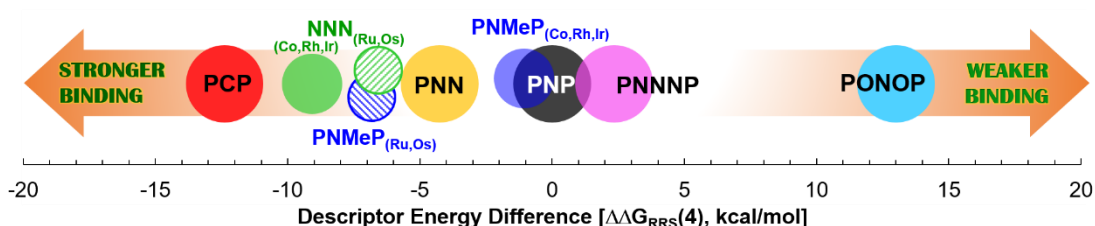


Figure 6.5 The average deviation of different pincer ligands on the volcano plots from the PNP ligand, which is used as a reference and arbitrarily set to zero. The influence of the ligand on the descriptor value [ $\Delta G_{RRS}(4)$ ] is then measured by subtracting each ligand value with the PNP complex of the same metal.

Further in depth examination also uncovers the influence of the pincer ligand's connecting atoms (X, Y, and Z symbols in Figure 6.1). In essence, when the flanking X and Z atoms (*i.e.*, as in the PNP, PNN, and NNN ligands), as well as the central pincer atom ( $Y = N_{\text{pyridine}}$ ,  $N_{\text{amine}}$ , and  $C_{\text{phenyl}}$  in PNP, PNMeP, and PCP respectively), are strong  $\pi$ -acceptor/more  $\pi$ -acidic moieties, such as phosphorus, the position of the catalyst tends to fall further to the right on the volcano plot than other pincer ligands with the same metal type (Figure 6.5), indicative of a weaker catalyst/substrate binding interaction.<sup>316-317</sup> Collectively, these “ligand backbone effects” are known to significantly impact overall catalytic performance.<sup>318</sup> Moreover, changing the number or type of connecting atoms located between the X and Y or the Y and Z atoms also influences activity, as is clearly evident by examining difference between the PONOP, PNNNP, and PNP complexes. The presence of more electronegative oxygen atoms in the PONOP ligand produces a larger shift of the catalysts toward weak binding (rightward along the  $x$ -axis) than the less electronegative N and C atoms found in PNNNP and PNP. As such, the clearest path toward improving catalytic performance is utilizing  $\pi$ -acidic ligands with group 9 metals (Ir, Rh, Co). Doing so will reduce the energy associated with heterolytic  $H_2$  cleavage and proton transfer that represents the most energetically costly step of the catalytic cycle.

## 6.4 Conclusions

In conclusion, an examination of the hydrogenation of  $CO_2$  to formate using molecular volcano plots shows that myriad metal and pincer ligand combinations have the ability to catalyze this

reaction. Of a set of 35 catalysts tested, iridium, rhodium and cobalt complexes with  $\pi$ -acidic ligands stand out as having particularly attractive catalytic cycle energetic profiles. For nearly all of the tested catalysts, the heterolytic  $H_2$  cleavage and the proton transfer to hydroxide base represent the most energetically costly step of the catalytic cycle. This energetic cost, however, can be minimized and the overall kinetics of the reaction improved by selecting catalysts bearing a strong  $\pi$ -accepting pincer ligand. Overall, this work demonstrates how molecular volcano plots can be used not only for estimating the performance of and designing new catalytic species, but also for uncovering the underlying trends that make functional catalysts tick.

# 7 Activity-Based Screening of Homogeneous Catalysts through the Rapid Assessment of Theoretically Derived Turnover Frequencies

This chapter is based on following publication:

Wodrich M. D., Sawatlon B., Solel E., Kozuch S., and Corminboeuf C., Activity-Based Screening of Homogeneous Catalysts through the Rapid Assessment of Theoretically Derived Turnover Frequencies. *ACS Catal.* **2019**, 9, 5716–5725.

## 7.1 Introduction

In experiment, measurements of catalytic activity are often reported as a turnover frequency (TOF). This single numeric provides a clear and unambiguous means of directly comparing the activities of two or more catalysts. However, when chemists turn to computation, comparisons become more difficult since, frequently, the output consists only of a reaction free energy profile containing a series of peaks and valleys representing the energies of the transition states and intermediates found in the catalytic cycle. To a large degree, the shortcomings of free energy profiles can be overcome and the computed activity of catalysts directly compared with one another if a TOF can be directly computed. In homogeneous catalysis, this task is often accomplished using Kozuch and Shaik's energy (energetic) span model (ESM)<sup>95-97, 222-223, 297</sup> or through microkinetic modeling.<sup>i,319</sup> Based on a term coined by Amatore and Jutand,<sup>320</sup> the ESM uses steady-state equations<sup>321</sup> and Eyring's transition state theory<sup>322</sup> to draw connections between the Gibbs free energies of the catalytic cycle and the TOF. According to the ESM the basic equation for the TOF is given by<sup>95, 223</sup>

$$TOF = \frac{k_B T}{h} \frac{1 - e^{\frac{\Delta G_r}{RT}}}{\sum_{i,j=1}^N e^{\frac{T_i - I_j + \delta G_{ij}}{RT}}}$$
$$\delta G_{ij} = \begin{cases} 0 & \text{if } T_i \text{ after } I_j \\ \Delta G_r & \text{if } T_i \text{ before } I_j \end{cases}$$

Equation 7.1

<sup>i</sup> Examination of TOFs using the energy span model and microkinetic modeling yielded very similar results for Nozaki's Ir(PNP) complex. At 273 K, the TOF using the energy span model was 83 s<sup>-1</sup> and that using microkinetic modeling was 82 s<sup>-1</sup>.

where  $T_i$  and  $I_j$  are the Gibbs free energies of the reaction's transition states and intermediates, respectively, and  $\Delta G_r$  is the total change in the reaction's Gibbs free energy. The  $T_i - I_j$  energy differences represent the obstacles that a catalyst must "climb over".  $\delta G_{ij}$  is a correction arising from the cyclic nature of catalytic reactions that ensures that the difference between an intermediate and its subsequent transition state (even if found in a second cycle) is always considered. This equation may seem complicated; however, often the denominator contains only a single dominant term, which results in a simplification to

$$TOF \approx \frac{k_B T}{h} e^{-\frac{\delta E}{RT}}$$

$$\delta E = \max_{i,j}(T_i - I_j + \delta G_{ij}) \quad \text{Equation 7.2}$$

In this form, the TOF depends only on the highest obstacle encountered by the catalyst, *i.e.* the energy span ( $\delta E$ ). The two states forming the energy span are known as the TOF-determining intermediate (TDI) and the TOF-determining transition state (TDTS), and they do not necessarily belong to the same step. The use of Equation 7.2 facilitates the easy determination of the theoretical TOF of any catalytic reaction directly from its computed potential energy surface.

From the above it is clear that an essential concept underpinning the ESM is that no single rate-determining step exists, but rather there are two or more TOF-determining states.<sup>95</sup> This idea is not limited to homogeneous catalysis, as it also applies to enzyme kinetics,<sup>323-324</sup> organic reactions,<sup>325</sup> and heterogeneous catalysis.<sup>326-327</sup> Taken as a whole, it is clear that the ESM represents an ideal tool for screening catalytic activity in a straightforward and unambiguous manner. The principle drawback, however, is that the model requires accurate free energies for all intermediates and transition states present in a catalytic cycle to obtain a TOF, which can quickly become very computationally expensive if one desires to screen hundreds or even thousands of catalysts.

To overcome this type of computational bottleneck, numerous methods have been developed aimed at discovering new catalysts more quickly either by reducing (*e.g.*, through the establishment of structure–activity type relationships),<sup>2, 4, 328</sup> by completely automating the underlying computations,<sup>329-340</sup> or through the use of machine-learning.<sup>122-129, 341</sup> Volcano plots<sup>45-46</sup> are another tool borrowed from heterogeneous and electrocatalysis<sup>6</sup> that facilitate the rapid screening of prospective catalysts by relating an easily computable descriptor variable with the thermodynamics or kinetics of the catalytic cycle.<sup>4, 75, 109</sup> Based on Sabatier's principle that an ideal catalyst should not bind intermediates too weakly or too strongly,<sup>35-36</sup> volcano plots facilitate the quick, visual discrimination of good and bad catalysts based on their location on the volcano, with the best species lying on the volcano plateau or near the peak. While originally applied to heterogeneous catalysts, Sabatier's concept is also equally applicable for describing homogeneous systems.<sup>60, 111</sup> For instance, if a catalyst binds a reactant or intermediate too strongly, then low TOFs will be observed because the accompanying TS will be very high. Conversely, an overly weak binding of intermediates causes binding of the reactant with the catalyst to be thermodynamically unfavorable.<sup>111</sup>

In 2015, we established the viability of molecular volcano plots by showing that Sabatier’s principle also applies to homogeneous catalytic reactions.<sup>5</sup> This proof-of-principle example reproduced known experimental trends and distinguished thermodynamically “good” from “bad” catalysts based on the computation of a single binding free energy. Subsequent advancements have aimed at improving the robustness<sup>61</sup> and the speed (*via* machine-learning)<sup>39</sup> at which these plots can be constructed and utilized. Today, an increasingly large toolkit based on molecular volcanoes exists that is capable of describing the kinetics<sup>43, 64</sup> and selectivity<sup>65</sup> of homogeneous catalysts, as well as providing comprehensive pictures of families of reactions.<sup>62-63</sup>

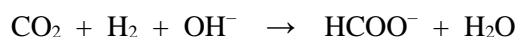
Despite being very useful for identifying new catalysts as well as better understanding catalytic behavior, one shortcoming of the previously reported molecular volcano plots is that the suitability of catalysts is evaluated solely from the most energetically costly step of the catalytic cycle [either the highest transition state (for kinetic volcanoes) or the largest free energy difference between linked intermediates (for thermodynamic volcanoes)]. Clearly, this type of picture is fundamentally at odds with the idea of a series of rate-determining states (as opposed to a single rate-determining state) that dictate the TOF, as is central to the ESM. The objective of this work is to recast molecular volcano plots within the more accurate ESM model with the aim of providing an unambiguous description of catalytic activity by directly determining the theoretical turnover frequencies for a set of 35 catalysts we previously studied.<sup>43</sup> We then harness the speed-up provided by volcano plots to screen for prospective new catalysts based on their TOFs, which can be rapidly estimated using free energy profiles obtained from linear free energy scaling relationships (LFESRs). To demonstrate the power of utility of TOF volcanoes, we chose to examine the homogeneously catalyzed conversion of carbon dioxide to formate facilitated by a transition metal and pincer ligand.

## 7.2 Computational Details

The molecular geometries of all species were obtained by optimization at the M06<sup>302-303</sup>/def2-SVP(D)<sup>82</sup> level in implicit water solvent (SMD solvation model<sup>83</sup>) using the “ultrafine” integration grid<sup>304</sup> in Gaussian09.<sup>172</sup> Analysis of the vibrational frequencies characterized the nature of each structure as either a minima (no imaginary frequencies) or a transition state (one imaginary frequency). Single point energies were then obtained on the M06 geometries using the B3LYP<sup>164-165</sup> functional appended with a density dependent dispersion correction,  $-dDsC$ <sup>305-308</sup> in tandem with the aug-TZ2P (TZ2P for the transition metals) basis set (*i.e.*, B3LYP- $dDsC$ /aug-TZ2P) as implemented in ADF.<sup>309-310</sup> Free energy corrections were determined using the rigid-rotor harmonic oscillator model<sup>311</sup> and a correction for translational entropy in solution<sup>312</sup> (water) using the Goodvibes program.<sup>313</sup> Final solvation corrections were established using the COSMO-RS<sup>314</sup> solvation model (in water). Reported free energies include electronic energies (computed at the B3LYP- $dDsC$ /aug-TZ2P/M06/def2-SPVD level), free energy corrections (computed at the M06/def2-SVPD level), and COSMO-RS solvation energies (computed at the B3LYP- $dDsC$ /aug-TZ2P level). Theoretical turnover frequencies were obtained by inputting the corresponding reaction profile into the AUTOF program of Uhe, Kozuch, and Shaik.<sup>222-223, 297</sup>

## 7.3 Results and Discussion

The hydrogenation of CO<sub>2</sub> to formate (Equation 7.3) represents an appealing pathway for the creation of C1 compounds from a readily available carbon feedstock.<sup>266</sup> The current state-of-the-art homogeneous catalytic systems rely on a tandem of transition metals and tridentate pincer ligands.<sup>272-273, 275</sup> Of these, notable examples include the following: Nozaki's Ir(PNP) catalyst [PNP = 2,6-bis(diisopropylphosphinomethyl)pyridine] with a TOF of 73,000 h<sup>-1</sup> and an impressive TON of 3,500,000,<sup>279</sup> Sanford's Ru(PNN) catalyst [PNN = 6-(di-*tert*-butylphosphino-methylene)-2-(*N,N*-diethylaminomethyl)-1,6-dihydropyridine] with a TOF of 2,200 h<sup>-1</sup> and a maximum of 23,000 turnovers under ideal conditions,<sup>286</sup> and Milstein's Fe(*t*Bu-PNP) complex with a TOF of 156 h<sup>-1</sup> and maximum TON of 788.<sup>285</sup> A host of other metal/pincer-ligand combinations are also known to catalyze this reaction and, as a result, have been the subject of intensive experimental<sup>1281-284, 288, 342-343</sup> and theoretical<sup>103, 287, 289-294</sup> examination.



Equation 7.3

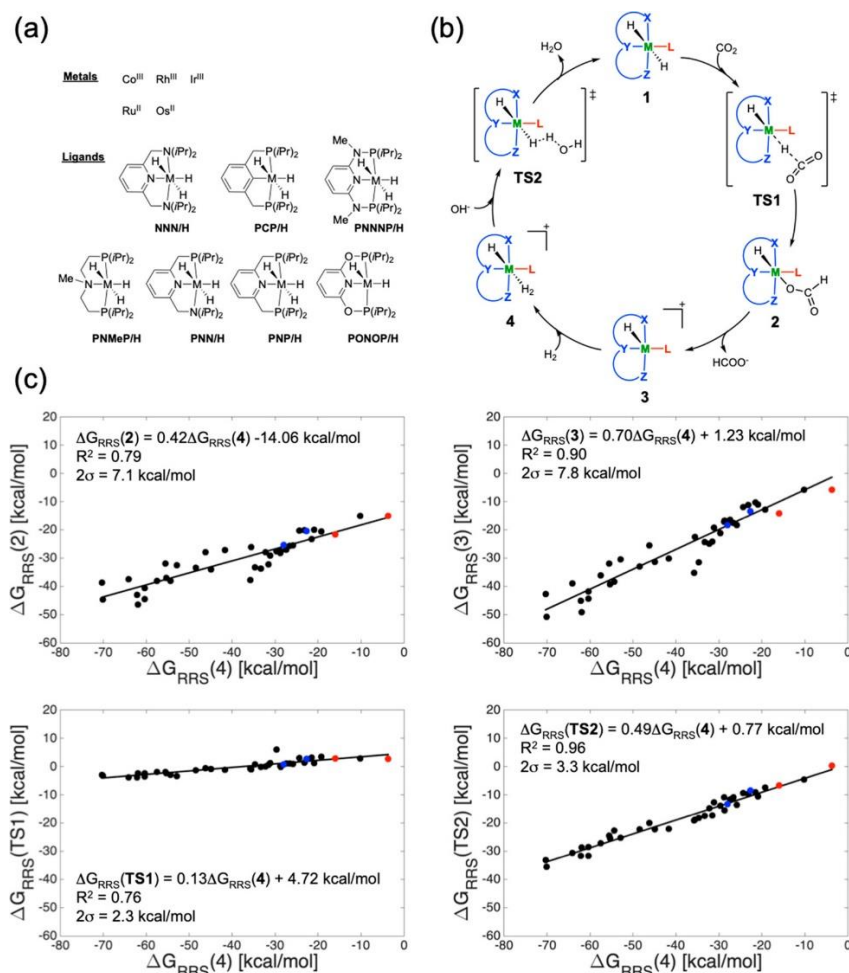


Figure 7.1 (a) Combinations of metals and pincer ligands used to establish linear free energy scaling relationships (LFESRs). (b) Catalytic cycle for the conversion of CO<sub>2</sub> to formate. (c) Linear free energy scaling relationships for the (b) catalytic cycle. Black points represent combinations of the metals and ligand depicted in (a), while red and blue points were used to confirm that newly tested catalyst follow the same linear scaling relationships as earlier tested species (*vide infra*).



To uncover the inner workings of why certain metal/pincer-ligand combinations are such effective catalysts, we recently used molecular volcano plots to examine the energetics of a series of 35 catalysts (Figure 7.1a) for the conversion of carbon dioxide to formate (Figure 7.1b).<sup>43</sup> Creating the volcanoes involved establishing linear free energy scaling relationships (LFESRs, Figure 7.1c), which show that the relative free energy of each intermediate and transition state (depicted on the y-axis) of the catalytic cycle can be estimated with a high degree of accuracy by knowing the value of a single reaction energy for each catalyst (depicted on the x-axis), as defined in Equation 7.4. Through a series of simple mathematical operations (see **Appendix C** for full details), the LFESRs can be converted into a plot like Figure 7.2a, which estimates the free energy associated with moving between any two connected intermediates or transition states. A volcano plot is then typically created from a profile of this type by taking the most energetically costly reaction step associated with each value of the descriptor variable. For the Figure 7.2b volcano plot, the catalytic cycle kinetics are controlled by one of two steps ( $4 \rightarrow \text{TS2}$  or  $2 \rightarrow 3$ ), which represent the largest free energy barrier that must be overcome in the catalytic cycle. This plot can then be used to quickly estimate the suitability of prospective catalysts by computing the value of the descriptor variable, which gives an approximate free energy for the largest energy barrier encountered in the catalytic cycle.

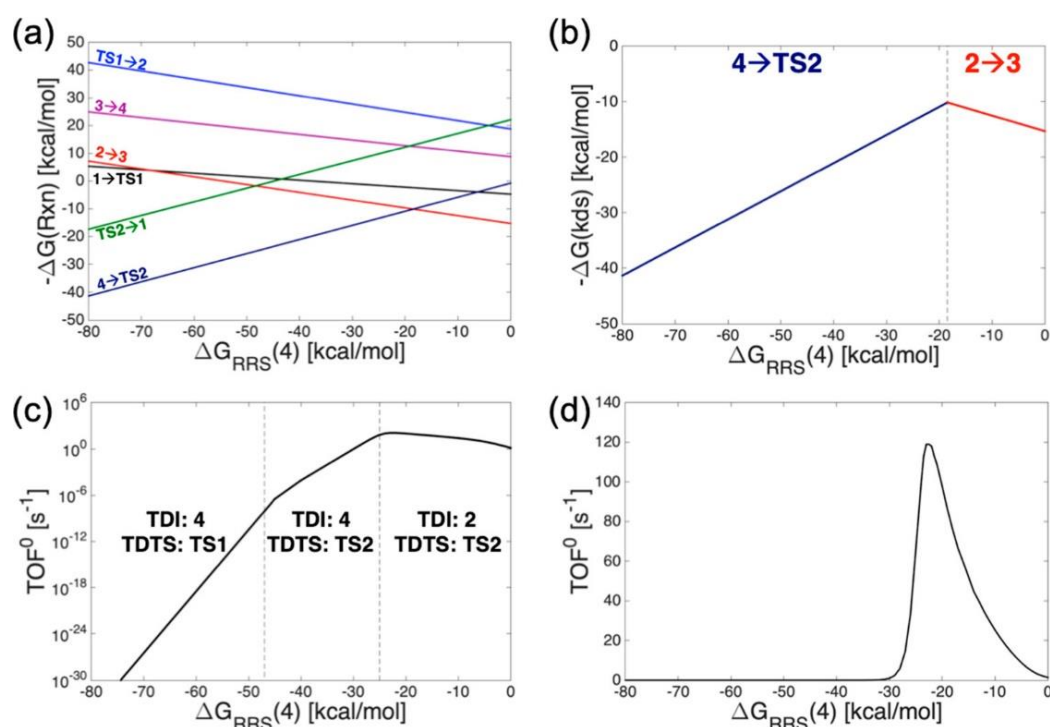


Figure 7.2 (a) Estimations of the free energy associated with each reaction step in the Figure 7.1b catalytic cycle derived from linear free energy scaling relationships. (b) Volcano plot obtained by taking the most energetically costly reaction step of the catalytic cycle for each descriptor ( $x$ -axis) value. (c) TOF volcano plot for the Figure 7.1b reaction ( $y$ -axis given in log scale). (d) The same TOF volcano ( $y$ -axis given in linear scale). The TOF volcanoes are obtained by inputting the theoretical free energy profile obtained from the Figure 7.1c LFESRs for each  $x$ -axis value and determining the associated TOF using the AUTOF program.

While seemingly simple to interpret, the Figure 7.2b volcano plot does possess problems, the principal shortcoming being that the kinetics of the catalytic cycle are always defined by a single reaction energy, which ignores situations where multiple transition states control the kinetics of the catalytic cycle. Moreover, in kinetic volcanoes it may also be difficult to determine when a catalyst becomes close enough to the volcano peak to possess a “good” kinetic profile. Nonetheless, the ability of volcano plots to rapidly screen prospective species based on computing only one reaction energy is a highly desirable feature and has led to their widespread use. The Figure 7.2a profile, however, already reveals that the free energy of each step of the catalytic cycle can be estimated, provided we know the descriptor variable of a catalyst. Given this fact, a more comprehensive treatment of the catalytic cycle, one that yields an unambiguous TOF, is possible through use of the energy span model.

To create the TOF volcano shown in Figure 7.2c, we computed hypothetical reaction profiles for a series of descriptor values ranging between  $-80$  and  $0$  kcal/mol ( $x$ -axis). The resulting energetic profiles were then input into the AUTOF program and the TOFs obtained plotted on the  $y$ -axis. To treat all reactions on equal footing, we choose to use the “standard” TOF (*i.e.*,  $\text{TOF}^0$ ),<sup>344</sup> which is measured at a concentration of  $1$  M and at  $273.15$  K. When plotted with a logarithmic  $y$ -axis, the shape of the new TOF volcano (Figure 7.2c) closely resembles the shape seen in standard volcano plots (*e.g.*, Figure 7.2b). Much like standard energetic volcano plots, the TOF volcano in Figure 7.2c can also be divided into three parts based on the identity of the determining states. For  $\Delta G_{\text{RRS}}(\mathbf{4}) < -47$  the TDI is **4**, while the TDTS is **TS1**. For descriptor values ranging from  $-47$  and  $-25$ , **4** is still the TDI but the TDTS changes to **TS2**, and for  $\Delta G_{\text{RRS}}(\mathbf{4}) > -25$  the TDI changes to **2** while the TDTS remains **TS2**. The changes in the TDI or TDTS are seen as distinct slope changes in the logarithmic TOF volcano, which arise from differences in the  $E_{\text{TDTS}} - E_{\text{TDI}}$  linear equation (as a function of the descriptor value). If the TOF volcano is viewed with a standard linear scale for the  $y$ -axis (Figure 7.2d), it becomes very clear that only catalysts lying within a rather narrow range of binding energies (*i.e.*, between  $-30$  and  $0$  kcal/mol) will show any appreciable activity. This information is much more difficult to gather from the more frequently seen energy-based volcano plot, such as that depicted in Figure 7.2b.

To more clearly illustrate the advantages of TOF volcanoes, we plotted the 35 catalysts shown in Figure 7.1a on both a standard energy-based volcano (Figure 7.3a) and a TOF volcano (Figure 7.3b). An examination of these two different volcano plots quickly reveals that the same species appear atop each, which effectively serves as proof of concept that both the TOF (Figure 7.3b volcano) and energy based criterion (Figure 7.3a volcano) yield similar chemical conclusions, as well as the fact that both can be accurately estimated from linear scaling relationships (shown by the solid lines in the two plots). However, using the TOF volcano is more intuitive and has the added benefit of being directly connecting to a well-characterized experimental observable. For the hydrogenation of carbon dioxide to formate, each of these “most active” catalysts possesses a group 9 metal center and a more  $\pi$ -acidic ligand, such as PONOP, PNNNP, PNP, and PNMeP (see Figure 7.1a for structures). Generally speaking, activity decreases as the periodic table is ascended (*i.e.*, first row transition metals show diminished activity relative to second and third row species). The coupling of highly  $\pi$ -acidic ligands with an Ir metal center yields catalysts having the highest activities, with some species possessing theoretical  $\text{TOF}^0$ s exceeding  $10,000 \text{ s}^{-1}$ . On the other hand, Figure 7.3b also shows that Ru and

Os based catalysts show virtually no activity, with the “most active” species tested here [Os(PONOP)] having a TOF of  $5.4 \times 10^{-9} \text{ s}^{-1}$ , or roughly one turnover every six years. As seen in Figure 7.3b, most of the computed catalysts lie in the two leftmost regions, where **4** is the TDI. Both the LFESRs of **TS1** and **TS2** have slopes that are positive and smaller than one, so that destabilizing **4** decreases the energy span (in both regions where **4** is the TDI). Destabilization of **4** can be brought on by making the catalyst more electron deficient, which explains why group 9 metals with  $\pi$ -acidic ligands have the highest computed TOFs.

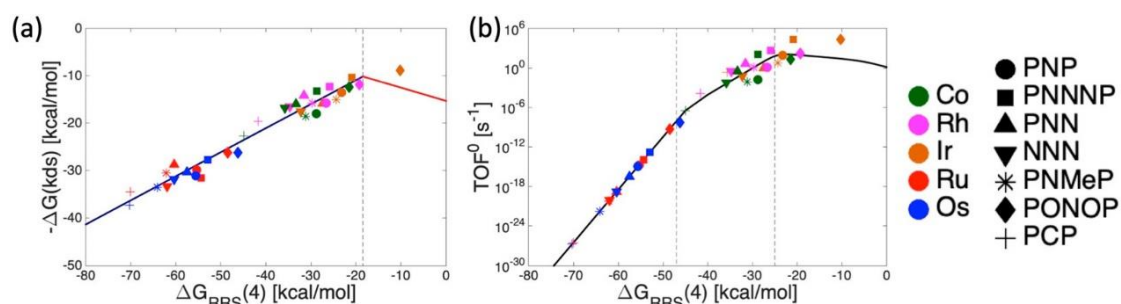


Figure 7.3 (a) Standard energy-based volcano plot and (b) the  $\text{TOF}^0$  volcano plot. Note that the vertical line in (a) represents a change in the nature of the most energetically costly step of the catalytic cycle while those in (b) represent changes in the nature of the turnover determining transition state and turnover determining intermediate within the context of the energy span model.

### 7.3.1 Screening with TOF Volcanoes

Having shown the benefits of TOF volcano plots, we were eager to utilize the TOF volcano shown in Figure 7.3b to screen for prospective new catalysts by computing only the value of the descriptor variable, which can then be converted into an estimated free energy profile and finally a TOF *via* the previously established linear scaling relationships. The benefit of proceeding in this manner is the considerable reduction in computational time needed to estimate the TOF, which, for the conversion of  $\text{CO}_2$  to formate, is only  $\sim 20\text{--}25\%$  of the time necessary to compute the full catalytic cycle. In particular, we were interested in broadening the number of group 9 catalysts capable of catalyzing this reaction, as well as altering the stereoelectronics of the group 8 species with the aim of dramatically improving their activity. Rather than change the metals or the pincer-ligands themselves, we instead examined the influence of hydrido ( $\text{H}^-$ ) ligand (red “L” in the Figure 7.1b catalytic cycle) that lies in the equatorial plane of the pincer ligand. To accomplish this, we tested two new possibilities, replacing the hydrido either with a more electronegative chloro ligand or with a neutral carbonyl ligand, the latter of which would change the overall charge of the complex. To ensure that the catalytic cycles of species bearing these new ligands followed the previously established scaling relationships, we computed the entire catalytic cycle for two of these catalyst bearing the new ligands and then compared them to the existing LFESRs. The blue [CO ligands, *i.e.*, Ru(PNMeP/–CO) and Ru(PONOP/–CO)] and red [Cl ligands, *i.e.*, Rh(PONOP/–Cl) and Ir(PONOP/–Cl)] points in the Figure 7.1c LFESRs show that the new species we desire to screen are governed by the same sets of scaling relationships as the 35 previously tested catalysts, and thus the use of the Figure 7.3 TOF volcano is justified.

Recall that the screening of catalyst involves directly determining the descriptor energy of each prospective catalyst *via* DFT computations. This value is then used to establish a theoretical energetic profile (through the previously established LFESRs) and, ultimately, to arrive at a theoretical TOF. Because these energy profiles are directly determined from the LFESRs, each catalyst falls directly on the volcano. Figure 7.4 shows the results for the 35 metal/pincer-ligand combinations given in Figure 7.1a, but now with a chloro replacing the hydrido ligand. Generally speaking, many of the same metal/pincer-ligand combinations remain atop the volcano. However, there is a discernable shift toward weaker-binding interactions (rightward along the  $x$ -axis) when the chloro ligand is present. As a result, the activities of a number of Co and Rh species are improved, while a reduction in activity is seen for the previously most active Ir catalysts. With the exception of the Co(PCP/-Cl) and Rh(PCP/-Cl), all of the Co, Rh, and Ir based complexes now show a theoretical  $\text{TOF}^0$  of at least  $1 \text{ s}^{-1}$ . Noteworthy among these are Co(PNNNP/-Cl), Co(PNP/-Cl), and Co(PNMeP/-Cl), which incorporate a desirable earth-abundant metal and have predicted  $\text{TOF}^0$ s of  $\sim 100 \text{ s}^{-1}$ . Several Rh species [Rh(PNNNP/-Cl), Rh(PNN/-Cl), Rh(PNP/-Cl)], as well as Ir(PNMeP/-Cl) also have predicted values exceeding  $100 \text{ s}^{-1}$ .

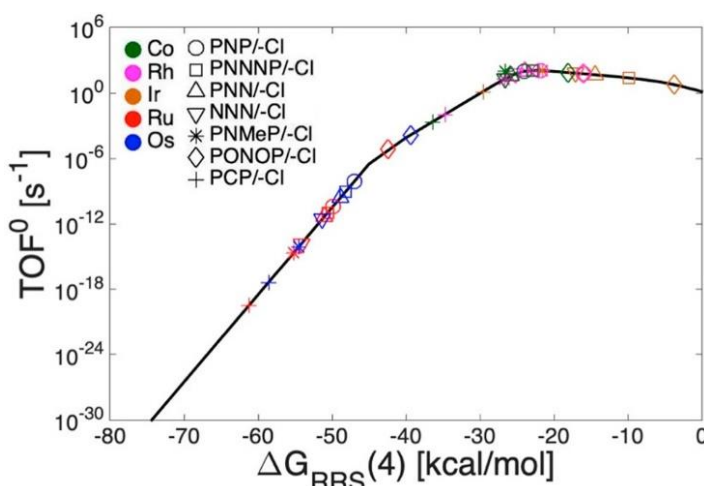


Figure 7.4  $\text{TOF}^0$  volcano plot used to screen catalysts bearing a  $-\text{Cl}$  ligand.

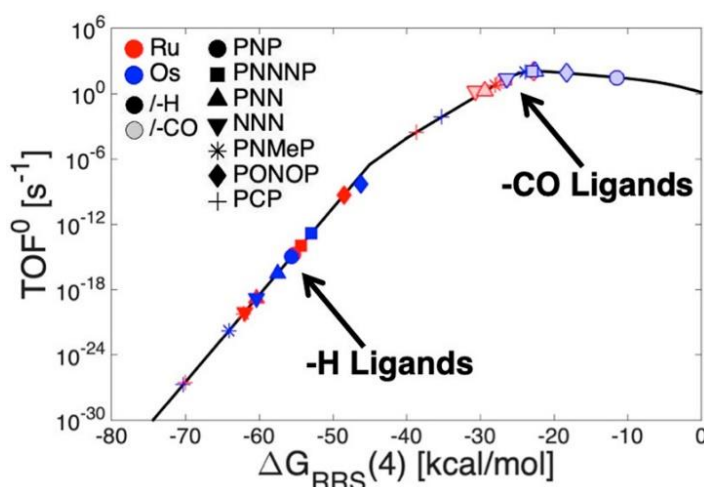


Figure 7.5  $\text{TOF}^0$  volcano plot used to screen catalysts bearing a  $-\text{CO}$  ligand.

Also evident in Figure 7.4 is that replacing the hydrido by a chloro ligand does not appreciably improve the activities of Ru or Os based catalysts, which still fall on the strong-binding (left) side of the volcano. Because each of the Ru and Os complexes studied above carries an overall negative charge, we were interested in forming neutral complexes by replacing the negatively charged hydrido/chloro ligand with a neutral carbonyl ligand. Figure 7.5 illustrates the resulting dramatic shifts of these species toward weaker binding and the corresponding significant improvements in the  $\text{TOF}^0$  values. Indeed, there are now three Os [Os(PNMeP/-CO), Os(PNNNP/-CO), and Os(PNN/-CO)] and one Ru catalyst [Ru(PONOP/-CO)] with predicted  $\text{TOF}^0$ s greater than  $100 \text{ s}^{-1}$ . It should be noted that catalysts incorporating neutral carbonyl ligands closely resemble, for instance, Sanford's Ru pincer catalyst<sup>286</sup> that has shown the ability to catalyze the same reaction.

### 7.3.2 Influence of Temperature on TOF Volcanoes

Having demonstrated that  $\text{TOF}^0$  values can be quickly estimated by computing a single descriptor and using established LFESRs, we were interested in probing the influence of temperature on the TOF volcanoes. Elevating the temperature often increases the reaction rate, with the conversion of carbon dioxide to formate being no exception. For instance, Nozaki's seminal work<sup>279</sup> was conducted at a temperature of  $\sim 393 \text{ K}$ . Figure 7.6 presents three energy span (6a) and TOF (6b) volcano plots at  $273 \text{ K}$  (equivalent to  $\text{TOF}^0$ , *vide supra*),  $298 \text{ K}$ , and  $393 \text{ K}$ . As expected, an increase in temperature significantly accelerates the TOFs, with a change from  $273$  to  $393 \text{ K}$  resulting in an increase of  $\sim 3$  orders of magnitude. Because the y-axes of these plots are given in a logarithmic scale, the energy span can be expressed as a linear function of the descriptor variable, leading to a TOF as follows:

$$\begin{aligned} \log \text{TOF} &= \frac{1}{2.3} \ln \text{TOF} \approx \frac{1}{2.3} \left[ \ln \left( \frac{k_B T}{h} \right) - \frac{\delta E}{RT} \right] \\ &= \frac{1}{2.3} \left[ \ln \left( \frac{k_B T}{h} \right) - \frac{\alpha \cdot \Delta G_{\text{RRS}}(4) + \beta}{RT} \right] \end{aligned} \quad \text{Equation 7.5}$$

Here, the first term explains the observed increase in the TOF with increasing temperatures while the second term is the reason the slopes observed in Figure 7.6b become closer to zero at higher temperatures.

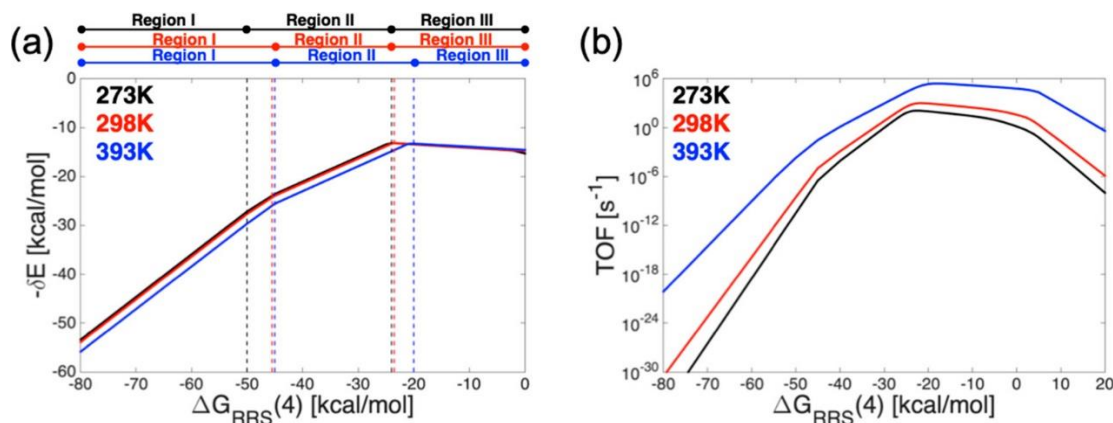


Figure 7.6 (a) Energy span and (b) TOF volcanoes depicting the influence of temperature.

One noteworthy feature of the TOF volcanoes at different temperatures is that the plateau region of the 393 K volcano (blue, Figure 7.6b) is shifted toward weaker-binding (rightward along the  $x$ -axis) relative to the lower temperature volcanoes, meaning that altering the temperature may lead to different catalysts being the most active. The temperature-induced shift in the maximum of the TOF essentially arises from two factors: first, the entropy difference between the states that define the energy span for each region (*e.g.*, where large entropy differences in the leftmost regions I and II of Figure 7.6a lead to larger temperature effects) and, second, the presence of more than one dominant energy difference near the volcano peak (see denominator of Equation 7.1) and especially the relative weighting of these dominant contributions at different temperatures that must be considered for accurately determining the point of maximum TOF.<sup>ii</sup> Note that such factors are not considered in standard, energy-based, volcano plots.

Since a commonly used reaction temperature for the hydrogenation of CO<sub>2</sub> to formate is 393.15 K, we recomputed the binding energies and determined the resulting TOF values for the earlier screened –Cl and –CO catalysts. The resulting TOF volcano shown in Figure 7.7 indicates that a plethora of pincer ligand catalysts have quite high TOF values, with Os(PONOP/–CO) (shaded blue diamond), Co(PONOP/–Cl) (empty green diamond), Ir(NNN/–Cl) (empty orange inverted triangle), and Rh(PONOP/–Cl) (empty fuchsia diamond) predicted to have the highest activities.

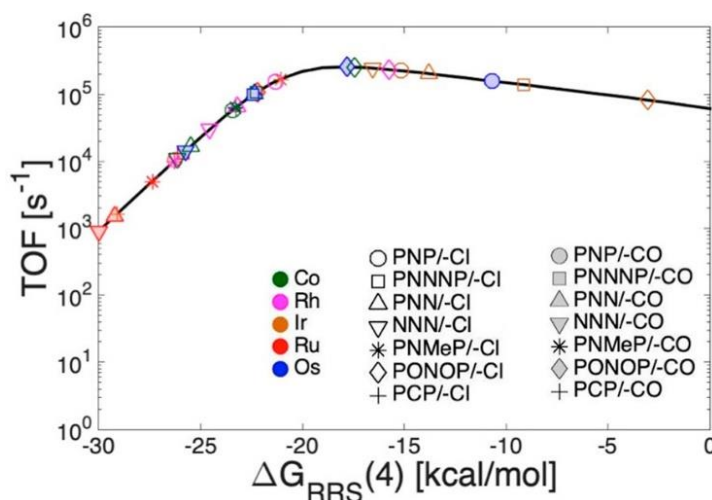


Figure 7.7 TOF volcano plots for the hydrogenation of CO<sub>2</sub> to formate determined at 393.15 K.

### 7.3.3 Strengths and Weaknesses of the Model

Having established a method that rapidly determines the TOFs of prospective catalysts, we were interested in comparing the TOFs obtained from free energy profiles derived from linear scaling relationships (as is done in the “screening” procedure) *versus* those established directly from DFT computations. The simplest assessment involves directly comparing the TOFs of the 35 catalysts depicted in Figure 7.1a. Generally speaking, there is strong agreement between “predicted” (from DFT computations) and “estimated” (from LFESRs) TOF values (Figure 7.8). Thus, the TOF volcano plots provide a quite accurate picture, where catalysts with “high” TOFs

<sup>ii</sup> A more detailed explanation and mathematical derivation of this phenomenon is given in **Appendix C**.



can quickly and unambiguously be distinguished from those with low TOFs. However, defining a single most active catalyst among the “best species” is more difficult. Of course, the quality of the estimated TOF values depends entirely on the quality of the linear scaling relationships, where better LFESRs will yield more accurate predictions of the catalytic cycle free energy profile, which will give rise to a better agreement between computed and estimated values. Despite some quantitative uncertainty associated with our predictions, we reiterate that, broadly speaking, volcano plots (of all types) are best used to identify a handful of best performing species. The best candidates should then be subjected to a more detailed analysis to identify, for instance, competing mechanistic and deactivation pathways that may play a significant role in dictating the experimentally observed TOF.

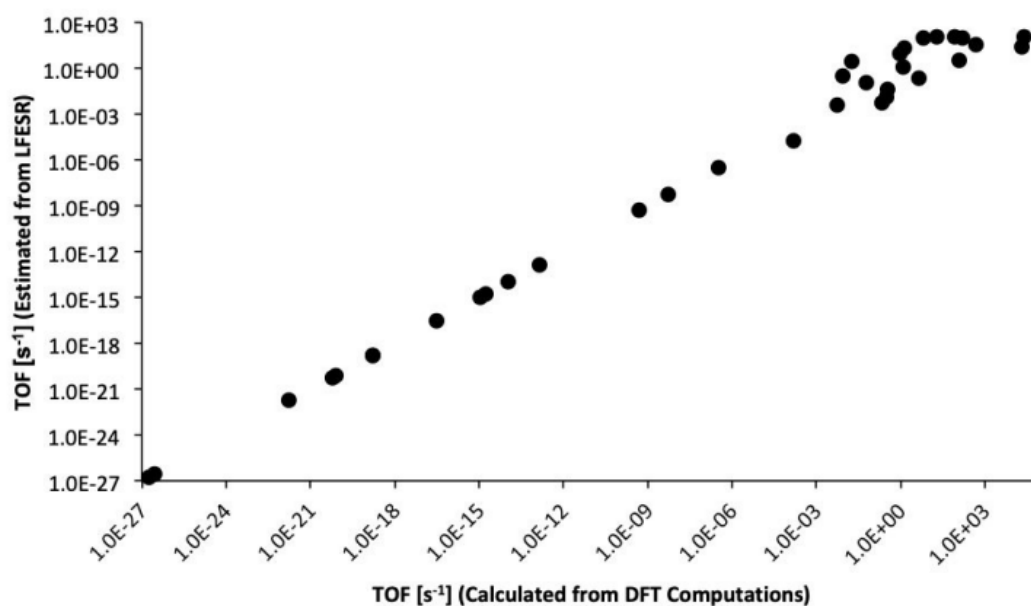


Figure 7.8 Comparison of 273 K TOF values determined directly through DFT computations and those predicted from the estimated free energy profiles obtained from linear free energy scaling relationships.

The direct comparison of theoretically computed TOFs with experimental values is a difficult business. Perhaps the most logical choice for comparison for the conversion of CO<sub>2</sub> to formate is with Nozaki’s Ir(PNP) catalyst, with a reported TOF of  $730 \times 10^2 \text{ h}^{-1}$  at 393 K.<sup>279</sup> This experimental value is far less than our theoretical prediction (obtained directly from DFT computations, not estimated from the TOF volcano) of  $4.5 \times 10^4 \text{ s}^{-1}$ . Note that our free energy profile aligns well with previous theoretical assessments.<sup>290, 292-293</sup> Indeed, the free energy values from Yang’s seminal work<sup>290</sup> also yield an overestimated TOF of  $2.3 \times 10^2 \text{ s}^{-1}$ ; thus, the theory/experiment disagreement does not seem to arise from the quality of the DFT computations. It is important to note, however, that our computations do not consider any side or deactivation pathways or issues of catalyst solubility (which, as Nozaki noted,<sup>342</sup> can greatly influence the measured TOF). Overall, we can conclude that theoretically estimated TOFs are (not surprisingly) extremely sensitive to the height of the TOF determining transition state(s) and that the origin of the disagreement between theory and computation likely arises from experimental factors not considered by our model. While the TOF obtained from experiment is overestimated here, the use of computed standard TOFs (*i.e.*, TOF<sup>0</sup>, *vide supra*), nonetheless,

allows for an unbiased comparison of theoretical determined turnover frequencies that serves as a basis for the screening of prospective new catalysts.

## **7.4 Conclusions**

In conclusion, we developed a method that rapidly screens prospective catalysts based on their activity, as defined by a theoretical turnover frequency. Specifically, establishing linear free energy scaling relationships allows a connection to be drawn between an easily computed descriptor variable and the free energy associated with each step of a catalytic cycle. This information, in turn, can be used to establish a turnover frequency through the use of the energy span model and the AUTOF program. Plotting these TOF values over a range of descriptors yields a TOF volcano plot that can be used to rapidly screen catalysts based on their anticipated activity. To demonstrate this, we aimed to identify new metal/pincer-ligand catalysts for the hydrogenation of CO<sub>2</sub> to formate. Our findings revealed that both altering the connecting atoms of the pincer ligand and replacing a hydrido ligand lying in the pincer equatorial plane with either a chloro or carbonyl ligand serves as an effective means of manipulating the energetics of the catalytic cycle. At experimental temperatures, our screening procedure identifies a handful of catalysts [Os(PONOP/–CO), Co(PONOP/–Cl), Ir(NNN–Cl), and Rh(PONOP/–Cl)] expected to exhibit high catalytic activity.



## 8 General Conclusions and Outlook

Rational catalyst design is central to identifying new well-defined catalysts that can lower reaction times while also operating under milder conditions. The creation and development of tools aimed at uncovering various aspects of catalytic performance are key for finding an ideal catalyst for a reaction of interest. In this sense, volcano plots, a tool with an established history of estimating catalyst performance in heterogeneous catalysis and electrocatalysis, have also shown great promise in estimating the performance of homogeneous catalysis. This thesis emphasizes the exploration and the application of molecular volcano plots to further predict the performance of catalysts and to better understand their effects toward reactivity. Overall, molecular volcanoes have the capability to serve as rational design tools for both catalysts and other components (*e.g.*, substrates) of catalytic reactions. A brief summary of the work presented in this thesis can be found below, following the three objectives stated in the introduction.

First, the underlying volcano concept and functionality is explored to broaden screening scope of volcano plots. Specifically, the newly revealed feature of using volcanoes to examine substrate scope has been presented. Taking Suzuki cross-coupling reaction as an example, substrate volcanoes reveal a wealth of information regarding how the nature and location of substituents as well as core structures dictate catalytic cycle energetics. It also mirrors known experimental trends confirming that the computational-based volcano plots provide useful input for experimentalists. Altogether, the realization that single volcanoes can treat multiple catalytic states and the development of substrate volcanoes provide additional features within the molecular volcano family that can be used to optimize homogeneous catalytic processes.

Second, the use of machine-learning models to reduce computational cost represents an ideal path to greatly expand the number of catalysts able to be screened. In this thesis, we have shown that machine-learning models can be trained and used to rapidly screen the descriptor variables of over eighteen thousand catalysts for Suzuki–Miyaura C–C cross-coupling reaction. Through the use of a pre-constructed volcano plot, a number of attractive candidates were identified having ideal reaction profiles and cost less than \$10 USD per mmol. Furthermore, this study not only sped up the screening process, but also provided new insight into the underlying chemistry through the use of big-data type analysis. Specifically, we used Sketch-Map to classify the catalysts into different subgroups based on their intrinsic chemical properties that uncovered various chemical trends for C–C cross-coupling reactions. For example, some ligand classes are able to polarize the complexes bearing different metal centers, which causes the energetic profiles

to appear more similar. This information, in turn, can be used to systematically tune specific metal/ligand combinations to obtain a desired thermodynamic profile for the catalytic cycle.

Finally, the ability of molecular volcano plots to provide useful chemical information for a specific chemical reaction is exemplified by our study on the hydrogenation of carbon dioxide to formate. Various metal and pincer ligand combinations showed the ability to catalyze this reaction, with iridium, rhodium and cobalt complexes bearing  $\pi$ -acidic ligands standing out as having particularly attractive catalytic cycle energetics. By adjusting the  $\pi$ -acidic strength of the ligand, an ideal catalyst that balances the strength of the catalyst/substrate interaction can be developed. Using this reaction, we also developed a method that rapidly screens prospective catalysts based on an experimental measurement of overall activity, the turnover frequency. A theoretical turnover frequency (TOF) was established from the corresponding free energy profile of each catalyst. The same linear scaling relationships used in conventional volcanoes were then used to create theoretical profiles corresponding to different values of the descriptor variable, which yielded the corresponding TOF volcano shape. The shape of this volcano slope indicates a large drop in anticipated activity when moving away from the volcano peak.

Although in this thesis we have focused on using volcanoes to study C–C cross-coupling reactions and the hydrogenation of carbon dioxide, it is important to remember that the underlying tools developed can be applied to any desired chemical reactions. As an example, our group have also worked on hydroformylation and reductive cleavage of aryl ethers, where we were also able to identify feasible catalysts and better understand the underlying chemistry of these processes.

Molecular volcano plots are undoubtedly valuable tools for estimating the performance of and designing new catalytic species, while also being convenient for uncovering underlying trends that make functional catalysts tick. However, we believe that the capability of volcano plots has not yet been fully exhausted. There are still unexplored areas, some of which are listed below:

- *Combined Volcano Plots for Catalysts and Substrates*

We have shown that volcano plots can probe substrates as similar as catalyst cases. The chemical information obtained from both plots could be integrated for the creation of two (or three) dimensional maps depicting the influence of both factors and the appropriate catalyst/substrate combinations while the axes are substrate and catalyst energetic descriptor variables. In **Chapter 3**, we realized that, for Suzuki cross-coupling, electron-rich substrates have a difficulty to surpass the oxidative addition and tend to bind  $\text{Pd}(\text{PPh}_3)_2$  catalysts “too weakly”. However, if coupling these substrates with “too strongly” binding catalysts such as nickel or platinum complexes, both excessive interaction sides can act as a push-pull character to balance the binding interaction, which will lower the activation barriers of the reaction. Supremely, combined substrate and catalyst volcanoes could be a catalog storing the reactivity and performance of a substrate pool and a catalyst collection in a particular reaction.

---

- *Volcano Plots for Reactions with Different Mechanistic Pathways*

One limitation of volcano plots is that their construction assumes that all catalysts transform a substrate to a product *via* the same mechanistic pathway. Even though we choose catalyst set from the combination of metal in a particular group and similar type of ligand, there is no guarantee that those catalysts will act identical to one another in reality. Finding ways to create volcanoes that can describe more than one possible reaction pathway represents an important “next-step” in their further generalization. Such tools could also be used to investigate what causes the behavior of catalysts with diverging pathways.

- *Volcano Plots for Reversible Reactions*

In some instances, a reversible process can be considered as a competing reaction pathway when the activation barriers that separate forward and reverse reactions are similar. In this situation, the reverse reaction can become dominant and reform the substrate from the product until equilibrium is reached. The addition of a chemical stimulus, *e.g.*, additive reagents or energies, is often needed to drive the reaction in the desired direction. Exploring catalytic systems where we can develop full control of the reaction direction (and formation of the products) would be profoundly important for contributing to solutions to important world issues, such as hydrogen storage and drug delivery. Finding proper descriptors and creating volcano plots for studying reactions of this type will unravel the key criteria that can be manipulated to influence which reaction direction is dominant.

- *Alternative Chemical Descriptors for Volcano Plots*

We have shown in this thesis that the volcano plot’s *y*-axis could be converted from the computed highest energy barrier to an experimentally-friendly quantity like the turnover frequency. In principle, the descriptor variables plotted along the *x*-axis could be also be modified to reflect other properties of the catalyst. The descriptor variable used for linear free energy scaling relationships and molecular volcano plots is generally taken as the relative energy of one catalytic cycle intermediate (sometimes called the binding energy), which is easily accessible *via* computation. However, this same quantity is not as easily discerned for an experimental chemist. The idea of altering the descriptor variable to reflect other chemical properties (*e.g.*, acidity  $pK_a$ , molar mass, density, and melting or boiling point) could give additional flexibility in constructing molecular volcanoes. Moreover, by using these familiar and measurable descriptor values, it allows us to quantitatively compare volcanoes created from experimental versus computational data.

- *An Alternative to the Energy-Based Volcano Plots*

Since the beginning of our studies on molecular volcanoes, we have created plots using the energetic characteristics of the catalysts. While the broad family of “energy-

based volcanoes” clearly is quite capable of predicting the performance and understanding the effects of catalysts, discovering alternative approaches for constructing volcano plots could be an additional route to build complementary catalyst design tools. For example, plotting the experimentally percent yield of the product against  $\text{pK}_{\text{a}}$  value or other chemical properties in the form of volcano plot could lead to new ways of assessing the performance of catalysts.

In closing, we believe that molecular volcano plots are highly precise and provide qualitatively accurate predictions of catalytic performance that can lead to the development of design principles. The chemical trends gathered from volcanoes are rational and practical, and can be instantly adapted and applied in experimental settings. Overall, this work demonstrates the uses of volcano plots for treating chemical reactions catalyzed by homogeneous transition metal catalysts.

# A Supplementary for Probing Substrate Scope with Molecular Volcanoes

This appendix is based on the supporting information of following publication:

- Sawatlon B., Wodrich M. D., and Corminboeuf C. Probing Substrate Scope with Molecular Volcanoes, *Org. Lett.* **2020**, 22, 7936–7941.

## A.1 Construction of Substrate Volcano Plots

Similar to the methods of constructing other types of volcano plots (see SI of *Chem. Sci.* **2015**, 6, 6754 and *Chem. Sci.* **2016**, 7, 5723 for detailed descriptions), we created substrate volcanoes based on the linear scaling relationship equations shown in Figure 3.4 where the relative energy of each intermediate/transition state found in the catalytic cycle is cast in terms of the energy descriptor variable [ $\Delta G_{\text{RRS}}(\mathbf{5})$ ]. However, unlike in our earlier molecular volcanoes, changing the substrate also changes the total reaction energy [ $\Delta G(\mathbf{RXN})$ ]. To aid the reader below we describe the complete process of constructing substrate volcanoes.

As the substrate volcanoes employ are aimed at finding the largest energetic barriers that must be overcome in the catalytic cycle, the two rate-determining states (*i.e.*, the two intermediates or intermediate/transition state that lead to the largest energy difference,  $\delta G$ ) must first be identified. Within the context of the Kozuch and Shaik's energy span model,<sup>95</sup> these states are called the turnover determining intermediate (TDI, representing the catalyst resting state) and the turnover determining transition state (TDTS, representing the highest energy transition state). The TDI and TDTS can be any species in the catalytic cycle and they need not be directly connected to one another (*e.g.*, the TDI can be the first intermediate and the TDTS can be the last transition state in a hypothetical catalytic cycle). In order to find these rate-determining states (TDI and TDTS), we calculated the energy difference ( $\delta G$ ) of each possible intermediate/transition state pair. Note that, if the intermediate (TDI) is after the transition state (TDTS), the total reaction energy [ $\Delta G(\mathbf{RXN})$ ] need to be added for the calculation of the energy difference (Equation A.1).

$$\delta G = T_i - I_j + \delta E_{ij}$$
$$\delta E_{ij} = \begin{cases} 0 & \text{if } T_i \text{ after } I_j \\ \Delta G(\mathbf{RXN}) & \text{if } T_i \text{ before } I_j \end{cases}$$

Equation A.1

To illustrate how the volcanoes are constructed, below are step-by-step instruction on how to build the volcano plot for 6-membered ring substrates. We first need the mathematical equations that describe the relative free energy of the intermediates and transition states [ $\Delta G_{\text{RRS}}(\text{X})$ ] in terms of the descriptor variable [*i.e.*,  $\Delta G_{\text{RRS}}(\mathbf{5})$ ], which are obtained from linear scaling relationships shown in Figure 3.4. The mathematical equations for 6-membered ring group are:

$$\begin{aligned}\Delta G_{\text{RRS}}(\mathbf{TSOA}) &= 0.39 \Delta G_{\text{RRS}}(\mathbf{5}) + 34.22 \text{ kcal/mol} \\ \Delta G_{\text{RRS}}(\mathbf{2}) &= 0.85 \Delta G_{\text{RRS}}(\mathbf{5}) + 22.64 \text{ kcal/mol} \\ \Delta G_{\text{RRS}}(\mathbf{3}) &= 0.90 \Delta G_{\text{RRS}}(\mathbf{5}) + 9.14 \text{ kcal/mol} \\ \Delta G_{\text{RRS}}(\mathbf{4}) &= 0.73 \Delta G_{\text{RRS}}(\mathbf{5}) + 4.77 \text{ kcal/mol} \\ \Delta G_{\text{RRS}}(\mathbf{TST}) &= 0.74 \Delta G_{\text{RRS}}(\mathbf{5}) + 13.39 \text{ kcal/mol} \\ \Delta G_{\text{RRS}}(\mathbf{TSRE}) &= 0.88 \Delta G_{\text{RRS}}(\mathbf{5}) + 3.23 \text{ kcal/mol} \\ \Delta G(\mathbf{RXN}) &= 0.23 \Delta G_{\text{RRS}}(\mathbf{5}) - 72.01 \text{ kcal/mol}\end{aligned}$$

Note that, as we set the intermediate **1** as the reference state,  $\Delta G_{\text{RRS}}(\mathbf{1})$  is equal to zero.

As stated above, the energy difference ( $\delta G$ ) of every possible pair of states must be determined (*i.e.*, **1**→**TSOA**, **1**→**2**, **1**→**3**, **1**→**4**, **1**→**TST**, **1**→**5**, **1**→**TSRE**, **2**→**1**, **2**→**TSOA**, **2**→**3**, **2**→**4**, **2**→**TST**, **2**→**5**, **2**→**TSRE**, **3**→**1**, **3**→**2**, ...) where the intermediate before the arrow indicates the TDI and the intermediate/transition state after the arrow corresponds to TDTS. An example of how to calculate the equations of the energy difference are shown below:

For **1**→**TSOA**:

$$\begin{aligned}-[\Delta G(\mathbf{1} \rightarrow \mathbf{TSOA})] &= -[\Delta G_{\text{RRS}}(\mathbf{TSOA}) - \Delta G_{\text{RRS}}(\mathbf{1})] \\ &= -[\Delta G_{\text{RRS}}(\mathbf{TSOA}) - 0] \\ -[\Delta G(\mathbf{1} \rightarrow \mathbf{TSOA})] &= -[0.39 \Delta G_{\text{RRS}}(\mathbf{5}) + 34.22 \text{ kcal/mol} - 0] \\ &= -0.39 \Delta G_{\text{RRS}}(\mathbf{5}) - 34.22 \text{ kcal/mol}\end{aligned}$$

Note that as in the next step we have to plot the negative of the energy difference; therefore, we directly calculated  $-[\Delta G(\mathbf{1} \rightarrow \mathbf{TSOA})]$  instead of  $[\Delta G(\mathbf{1} \rightarrow \mathbf{TSOA})]$ .

For **2**→**1**:

$$\begin{aligned}-[\Delta G(\mathbf{2} \rightarrow \mathbf{1})] &= -[\Delta G_{\text{RRS}}(\mathbf{1}) - \Delta G_{\text{RRS}}(\mathbf{2}) + \Delta G(\mathbf{RXN})]^{\dagger} \\ &= -[0 - \Delta G_{\text{RRS}}(\mathbf{2}) + \Delta G(\mathbf{RXN})] \\ &= -\{0 - [0.85 \Delta G_{\text{RRS}}(\mathbf{5}) + 22.64 \text{ kcal/mol}] + \\ &\quad [0.23 \Delta G_{\text{RRS}}(\mathbf{5}) - 72.01 \text{ kcal/mol}]\} \\ &= -[-0.62 \Delta G_{\text{RRS}}(\mathbf{5}) - 94.65 \text{ kcal/mol}] \\ -[\Delta G(\mathbf{2} \rightarrow \mathbf{1})] &= 0.62 \Delta G_{\text{RRS}}(\mathbf{5}) + 94.65 \text{ kcal/mol}\end{aligned}$$

Repeating these same procedure multiple times then yields the energy difference equations for 6-membered ring group seen below.

<sup>†</sup> According to the energy span model, if the TDTS appears before the TDI (*i.e.*, in this case, intermediate **1** is before intermediate **2**), we need to add the total reaction energy into the equation as shown in Equation A.1

Intermediate **1** as the TDI:

$$\begin{aligned}
-[\Delta G(1 \rightarrow \text{TSOA})] &= -0.39 \Delta G_{\text{RRS}}(\mathbf{5}) - 34.22 \text{ kcal/mol} \\
-[\Delta G(1 \rightarrow \mathbf{2})] &= -0.85 \Delta G_{\text{RRS}}(\mathbf{5}) - 22.64 \text{ kcal/mol} \\
-[\Delta G(1 \rightarrow \mathbf{3})] &= -0.90 \Delta G_{\text{RRS}}(\mathbf{5}) - 9.14 \text{ kcal/mol} \\
-[\Delta G(1 \rightarrow \mathbf{4})] &= -0.73 \Delta G_{\text{RRS}}(\mathbf{5}) - 4.77 \text{ kcal/mol} \\
-[\Delta G(1 \rightarrow \text{TST})] &= -0.74 \Delta G_{\text{RRS}}(\mathbf{5}) - 13.39 \text{ kcal/mol} \\
-[\Delta G(1 \rightarrow \mathbf{5})] &= -\Delta G_{\text{RRS}}(\mathbf{5}) \\
-[\Delta G(1 \rightarrow \text{TSRE})] &= -0.23 \Delta G_{\text{RRS}}(\mathbf{5}) + 72.01 \text{ kcal/mol}
\end{aligned}$$

Intermediate **2** as the TDI:

$$\begin{aligned}
-[\Delta G(2 \rightarrow \mathbf{1})] &= 0.62 \Delta G_{\text{RRS}}(\mathbf{5}) + 94.65 \text{ kcal/mol} \\
-[\Delta G(2 \rightarrow \text{TSOA})] &= 0.23 \Delta G_{\text{RRS}}(\mathbf{5}) + 60.43 \text{ kcal/mol} \\
-[\Delta G(2 \rightarrow \mathbf{3})] &= -0.05 \Delta G_{\text{RRS}}(\mathbf{5}) + 13.50 \text{ kcal/mol} \\
-[\Delta G(2 \rightarrow \mathbf{4})] &= 0.12 \Delta G_{\text{RRS}}(\mathbf{5}) + 17.87 \text{ kcal/mol} \\
-[\Delta G(2 \rightarrow \text{TST})] &= 0.11 \Delta G_{\text{RRS}}(\mathbf{5}) + 9.25 \text{ kcal/mol} \\
-[\Delta G(2 \rightarrow \mathbf{5})] &= -0.15 \Delta G_{\text{RRS}}(\mathbf{5}) + 22.64 \text{ kcal/mol} \\
-[\Delta G(2 \rightarrow \text{TSRE})] &= -0.03 \Delta G_{\text{RRS}}(\mathbf{5}) + 19.41 \text{ kcal/mol}
\end{aligned}$$

Intermediate **3** as the TDI:

$$\begin{aligned}
-[\Delta G(3 \rightarrow \mathbf{1})] &= 0.67 \Delta G_{\text{RRS}}(\mathbf{5}) + 81.15 \text{ kcal/mol} \\
-[\Delta G(3 \rightarrow \text{TSOA})] &= 0.28 \Delta G_{\text{RRS}}(\mathbf{5}) + 46.93 \text{ kcal/mol} \\
-[\Delta G(3 \rightarrow \mathbf{2})] &= -0.18 \Delta G_{\text{RRS}}(\mathbf{5}) + 58.51 \text{ kcal/mol} \\
-[\Delta G(3 \rightarrow \mathbf{4})] &= 0.17 \Delta G_{\text{RRS}}(\mathbf{5}) + 4.37 \text{ kcal/mol} \\
-[\Delta G(3 \rightarrow \text{TST})] &= 0.16 \Delta G_{\text{RRS}}(\mathbf{5}) - 4.25 \text{ kcal/mol} \\
-[\Delta G(3 \rightarrow \mathbf{5})] &= -0.10 \Delta G_{\text{RRS}}(\mathbf{5}) + 9.14 \text{ kcal/mol} \\
-[\Delta G(3 \rightarrow \text{TSRE})] &= 0.02 \Delta G_{\text{RRS}}(\mathbf{5}) + 5.91 \text{ kcal/mol}
\end{aligned}$$

Intermediate **4** as the TDI:

$$\begin{aligned}
-[\Delta G(4 \rightarrow \mathbf{1})] &= 0.50 \Delta G_{\text{RRS}}(\mathbf{5}) + 76.78 \text{ kcal/mol} \\
-[\Delta G(4 \rightarrow \text{TSOA})] &= 0.11 \Delta G_{\text{RRS}}(\mathbf{5}) + 42.56 \text{ kcal/mol} \\
-[\Delta G(4 \rightarrow \mathbf{2})] &= -0.35 \Delta G_{\text{RRS}}(\mathbf{5}) + 54.14 \text{ kcal/mol} \\
-[\Delta G(4 \rightarrow \mathbf{3})] &= -0.40 \Delta G_{\text{RRS}}(\mathbf{5}) + 67.64 \text{ kcal/mol} \\
-[\Delta G(4 \rightarrow \text{TST})] &= -0.01 \Delta G_{\text{RRS}}(\mathbf{5}) - 8.62 \text{ kcal/mol} \\
-[\Delta G(4 \rightarrow \mathbf{5})] &= -0.27 \Delta G_{\text{RRS}}(\mathbf{5}) + 4.77 \text{ kcal/mol} \\
-[\Delta G(4 \rightarrow \text{TSRE})] &= -0.15 \Delta G_{\text{RRS}}(\mathbf{5}) + 1.54 \text{ kcal/mol}
\end{aligned}$$

Intermediate **5** as the TDI:

$$\begin{aligned}
-[\Delta G(5 \rightarrow \mathbf{1})] &= 0.77 \Delta G_{\text{RRS}}(\mathbf{5}) + 72.01 \text{ kcal/mol} \\
-[\Delta G(5 \rightarrow \text{TSOA})] &= 0.38 \Delta G_{\text{RRS}}(\mathbf{5}) + 37.79 \text{ kcal/mol} \\
-[\Delta G(5 \rightarrow \mathbf{2})] &= -0.08 \Delta G_{\text{RRS}}(\mathbf{5}) + 49.37 \text{ kcal/mol} \\
-[\Delta G(5 \rightarrow \mathbf{3})] &= -0.13 \Delta G_{\text{RRS}}(\mathbf{5}) + 62.87 \text{ kcal/mol} \\
-[\Delta G(5 \rightarrow \mathbf{4})] &= 0.04 \Delta G_{\text{RRS}}(\mathbf{5}) + 67.24 \text{ kcal/mol} \\
-[\Delta G(5 \rightarrow \text{TST})] &= 0.03 \Delta G_{\text{RRS}}(\mathbf{5}) + 58.62 \text{ kcal/mol} \\
-[\Delta G(5 \rightarrow \text{TSRE})] &= 0.12 \Delta G_{\text{RRS}}(\mathbf{5}) - 3.23 \text{ kcal/mol}
\end{aligned}$$

We then plot these mathematical equations as a function of  $\Delta G_{\text{RRS}}(\mathbf{5})$  (Figure A.1). When reading this graph recall that it depicts the negative of the energy difference, meaning that the most energetically costly steps fall on the bottom of the plot. As such, the lowest line for any value of the descriptor variable corresponds to the states that define the energy span for substrates lying in the region [*i.e.*, TDI: **5**/TDTS: **1** for  $\Delta G_{\text{RRS}}(\mathbf{5})$  values less than around  $-125$  kcal/mol, TDI: **3**/TDTS: **TST** for  $\Delta G_{\text{RRS}}(\mathbf{5})$  values between  $-125$  and  $-55$  kcal/mol, and TDI: **1**/TDTS: **TSOA** for  $\Delta G_{\text{RRS}}(\mathbf{5})$  values greater than  $-55$  kcal/mol]. These three scaling relationships form the shape of the volcano plot seen in Figure 3.5.

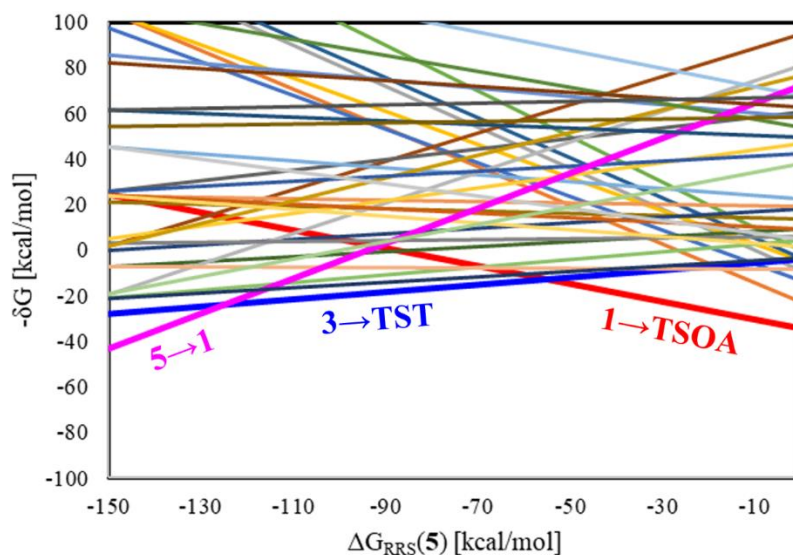


Figure A.1 Plot of lines corresponding to all possible TDI/TDTS combinations for the reaction of the 6-membered ring substrates.



## A.2 Correlation between the Charge on Pd, Substrate Sterics and the Energy of Transmetalation

Figure A.2a and c indicates that the charge on palladium center of intermediate **3** stay relatively constant (particularly for the bromobenzenes) even as the energy barrier of transmetalation increases (moving from right to left). However, Figure A.2b and d illustrate that steric factors of the substrate (as demonstrated by the percent buried volume, %  $V_{\text{bur}}$ ) largely influence the energy associated with transmetalation with more steric bulk (larger %  $V_{\text{bur}}$ ) having higher energy barriers. Overall, this result confirms that the electronic properties of palladium have less effect on transmetalation rate than the steric factor. Note that charges on Pd were calculated using Hirshfeld population analysis<sup>345</sup> in Gaussian16.<sup>84</sup> The %  $V_{\text{bur}}$  of substituted aryl rings on substrates were computed using the SambVca web application.<sup>346</sup>

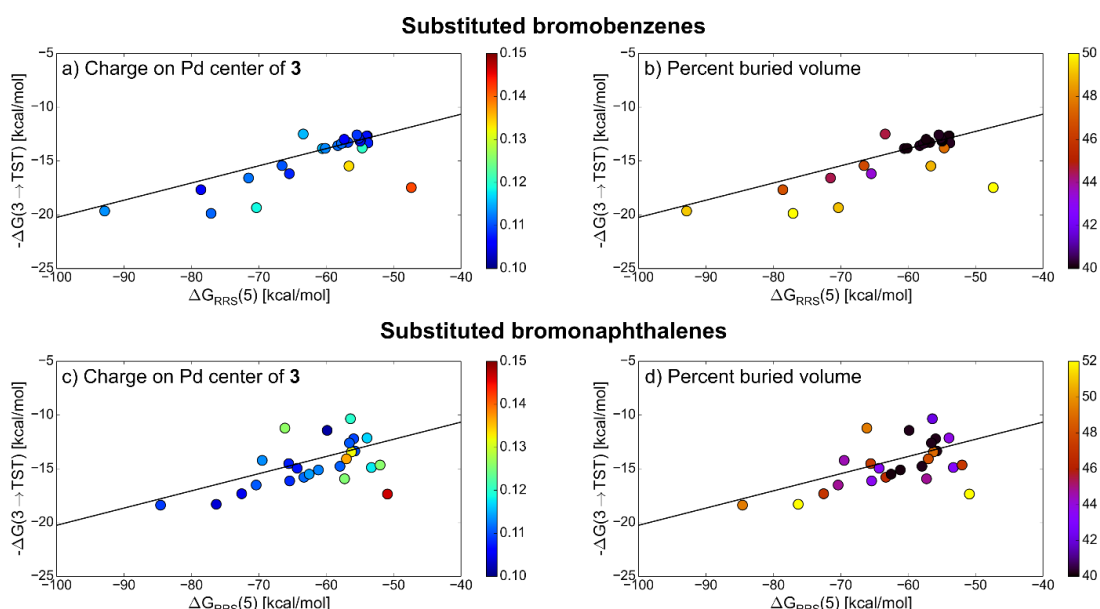


Figure A.2 Plots of the descriptor variable  $[\Delta G_{\text{RRS}}(5)]$  and the energy difference in transmetalation step  $[-\Delta G(3 \rightarrow \text{TST})]$  of substituted bromobenzenes (top) and substituted bromonaphthalenes (bottom). Each substrate in the plot is colored by the charge on Pd of intermediate **3** (a and c) or a measure of steric bulk (percent buried volume, %  $V_{\text{bur}}$ , where larger values indicate more sterics) (b and d). The solid line corresponds to the estimated outline of the  $3 \rightarrow \text{TST}$  from volcano plot.



# B Supplementary for Machine Learning Meets Volcano Plots and Data Mining of the C–C Cross-Coupling

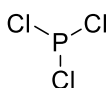
This appendix is based on the supporting information of following publications:

- Meyer B., Sawatlon B., Heinen S., von Lilienfeld O. A., and Corminboeuf C., Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts. *Chem. Sci.*, **2018**, 9, 7069-7077.
- Sawatlon B., Wodrich M. D., Meyer B., Fabrizio A., and Corminboeuf C., Data Mining the C–C Cross-Coupling Genome. *ChemCatChem* **2019**, 11, 4096-4107.

## B.1 Ligand Database



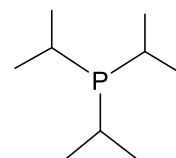
0



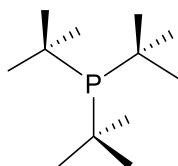
1



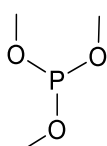
2



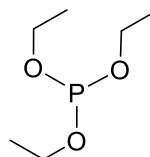
3



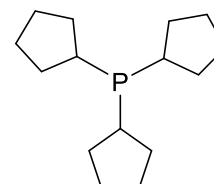
4



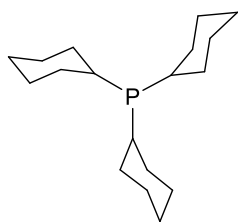
5



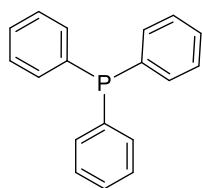
6



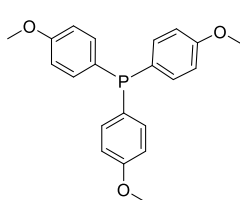
7



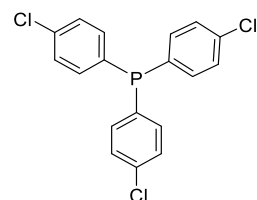
8



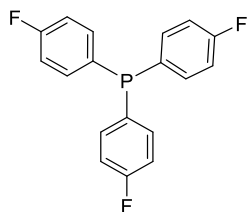
9



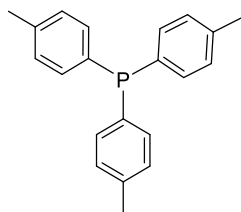
10



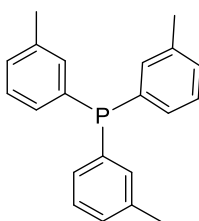
11



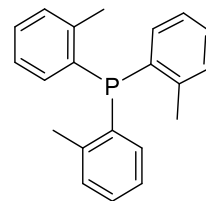
12



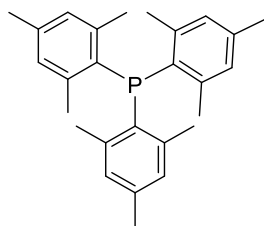
13



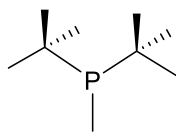
14



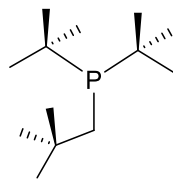
15



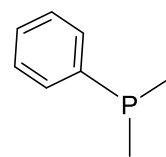
16



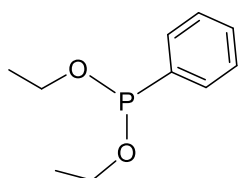
17



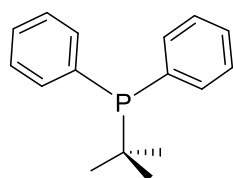
18



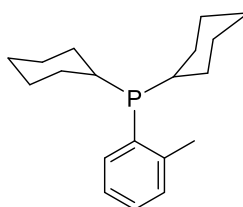
19



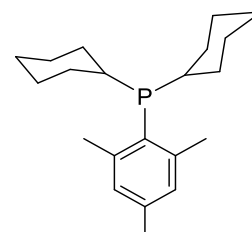
20



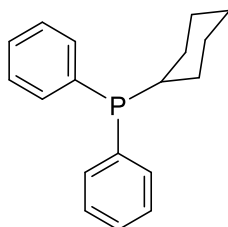
21



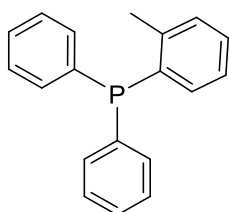
22



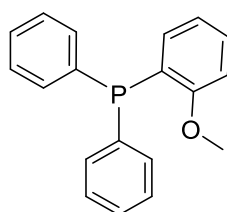
23



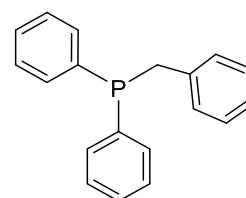
24



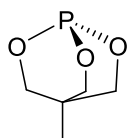
25



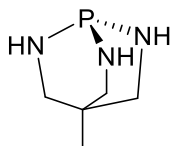
26



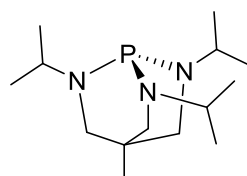
27



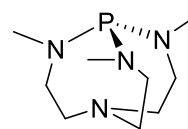
28



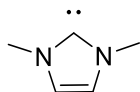
29



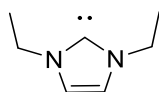
30



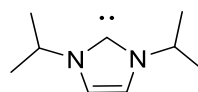
31



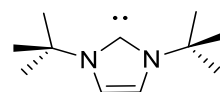
32



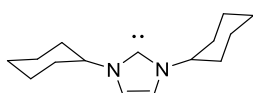
33



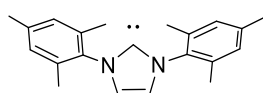
34



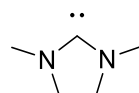
35



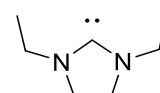
36



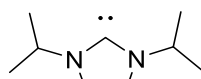
37



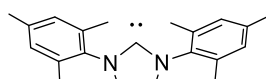
38



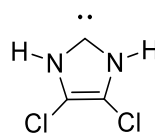
39



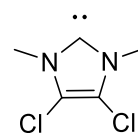
40



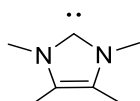
41



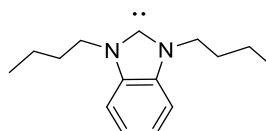
42



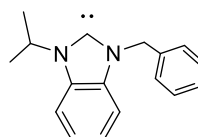
43



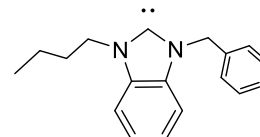
44



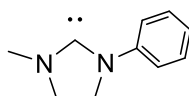
45



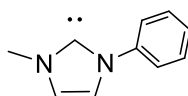
46



47



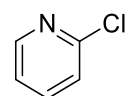
48



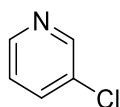
49



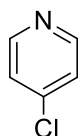
50



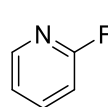
51



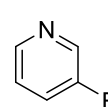
52



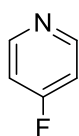
53



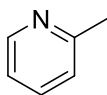
54



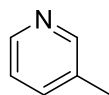
55



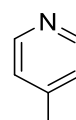
56



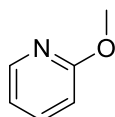
57



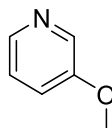
58



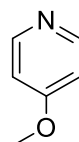
59



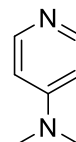
60



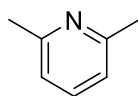
61



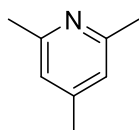
62



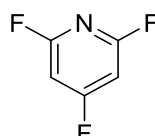
63



64



65



66



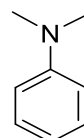
67



68



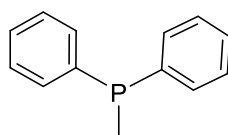
69



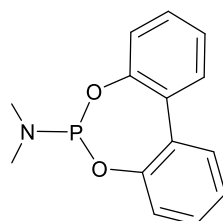
70



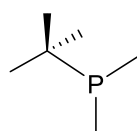
71



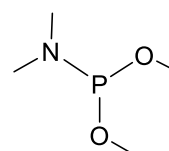
72



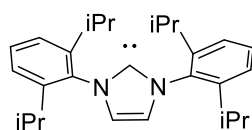
73



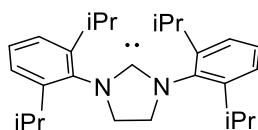
74



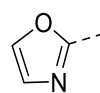
75



76

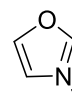


77



78

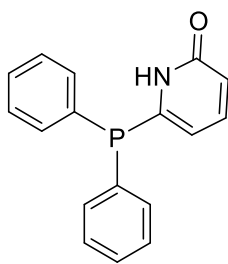
(charge: -1)



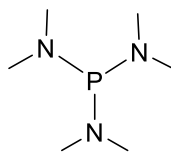
79



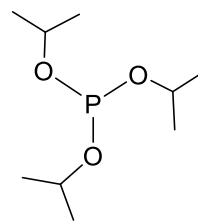
80



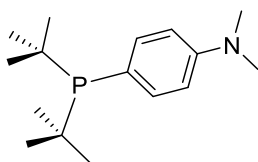
81



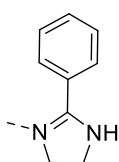
82



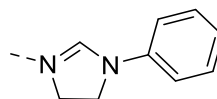
83



84



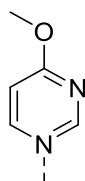
85



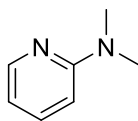
86



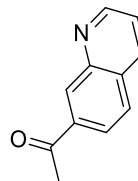
87



88



89



90

## B.2 Distribution of Binding Energies in the Training Set

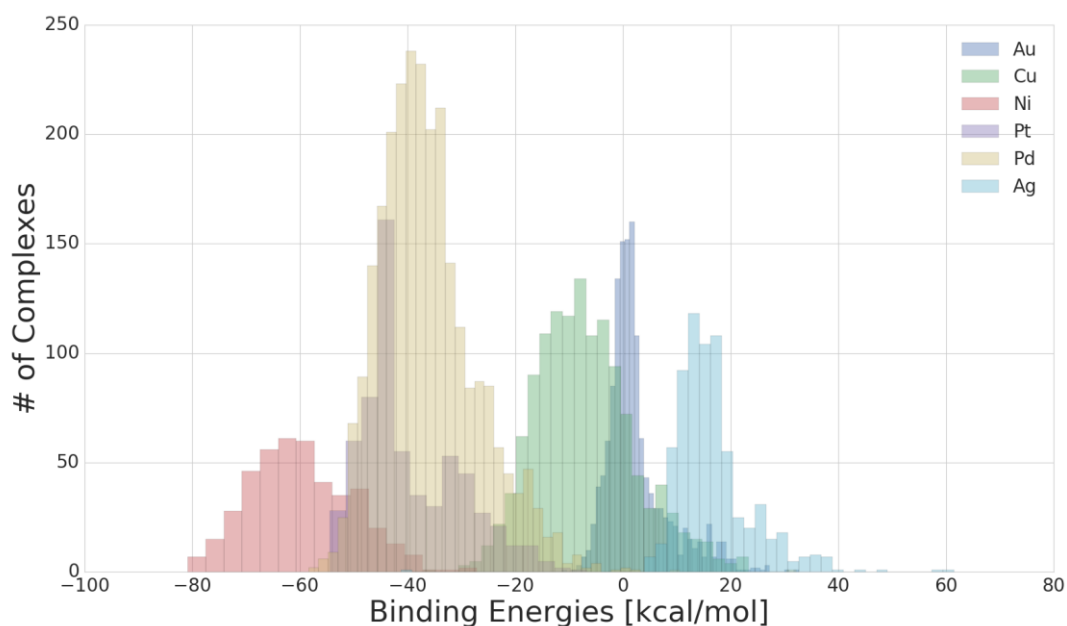


Figure B.1 Histogram representing the occurrence of each metal complexes in the training set. The size of the beans is selected following the Freedman–Diaconis rule.

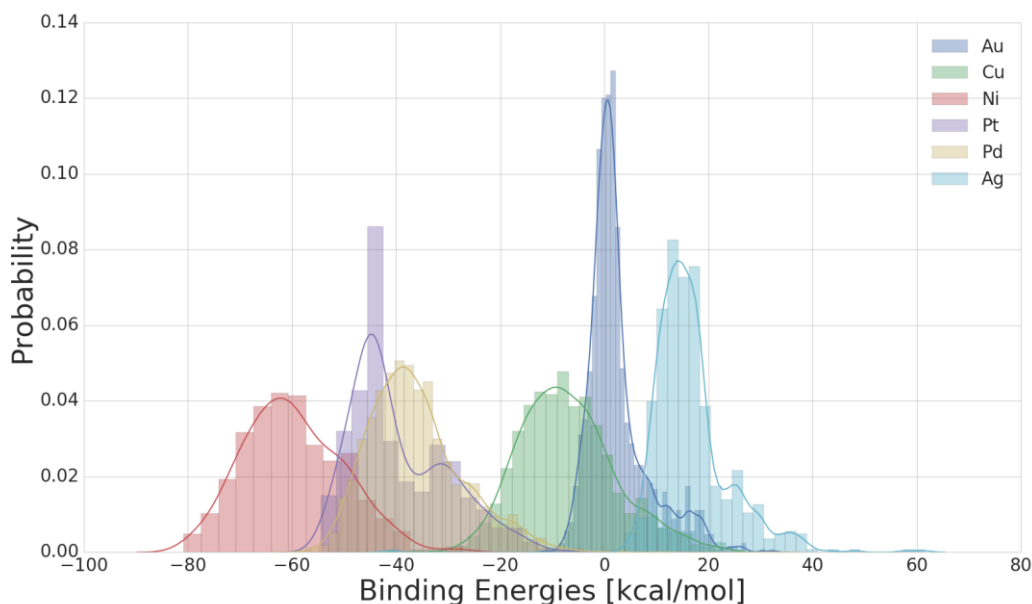


Figure B.2 Histogram representing the probability of each metal complexes in the training set.



## B.3 Sketch-Maps (Energy Descriptors vs HOMO Energies) of Complex 1

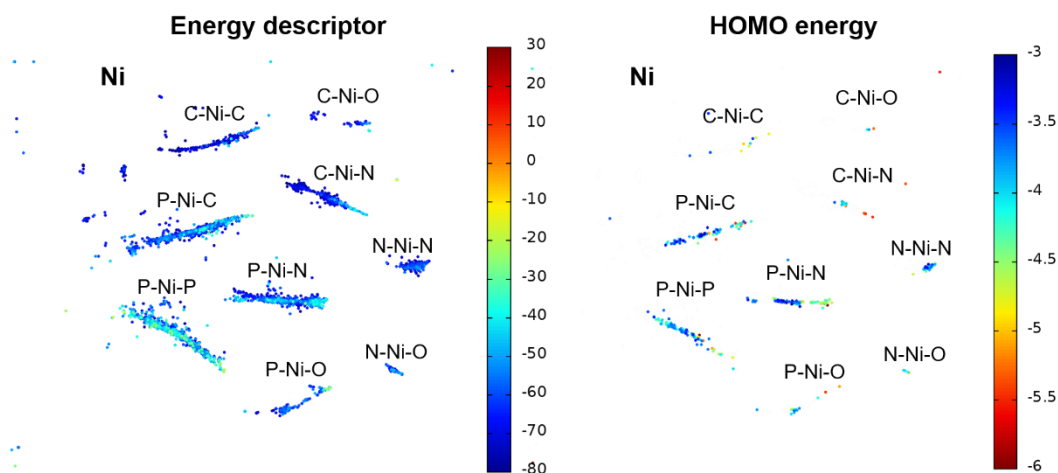


Figure B.3 The Sketch-Maps of nickel complexes colored by the energy descriptor (left, in kcal/mol) and the HOMO energy of complex **1** (right, in eV). Note that the HOMO value was not computed for species appearing in white in the HOMO energy plot since the energy descriptor value was taken directly from machine-learning.

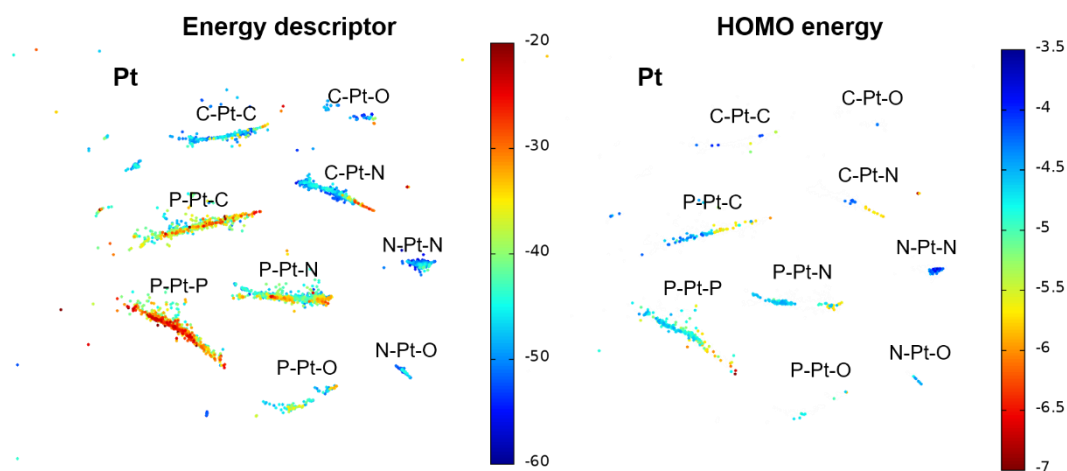


Figure B.4 The Sketch-Maps of platinum complexes colored by the energy descriptor (left, in kcal/mol) and the HOMO energy of complex **1** (right, in eV).

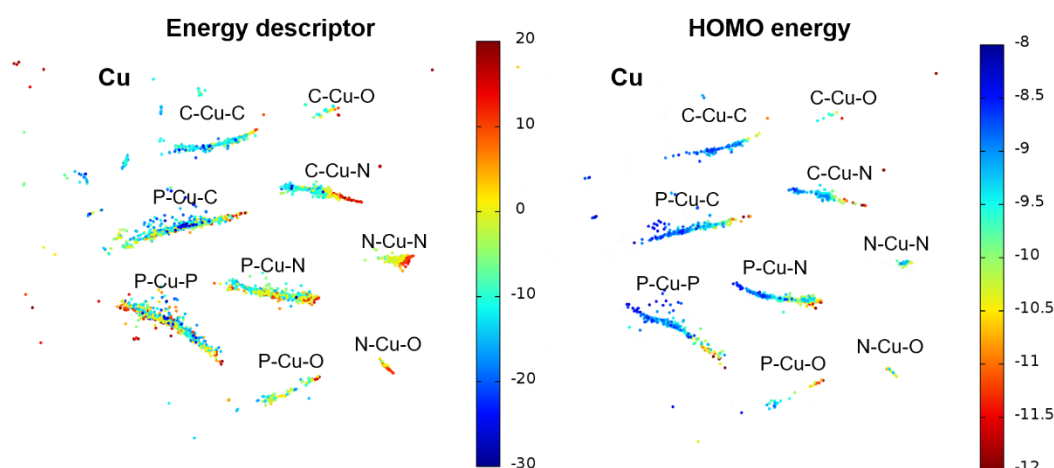


Figure B.5 The Sketch-Maps of copper complexes colored by the energy descriptor (left, in kcal/mol) and the HOMO energy of complex 1 (right, in eV).

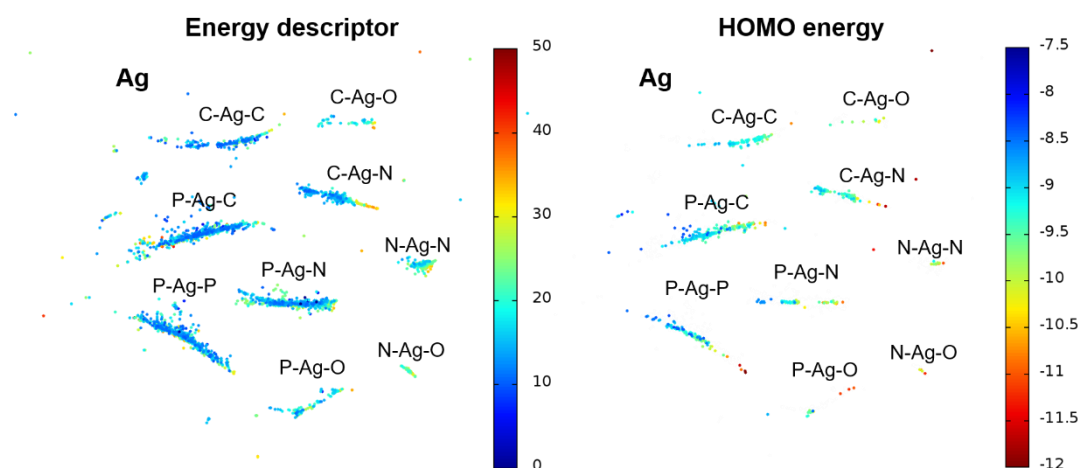


Figure B.6 The Sketch-Maps of silver complexes colored by the energy descriptor (left, in kcal/mol) and the HOMO energy of complex 1 (right, in eV).

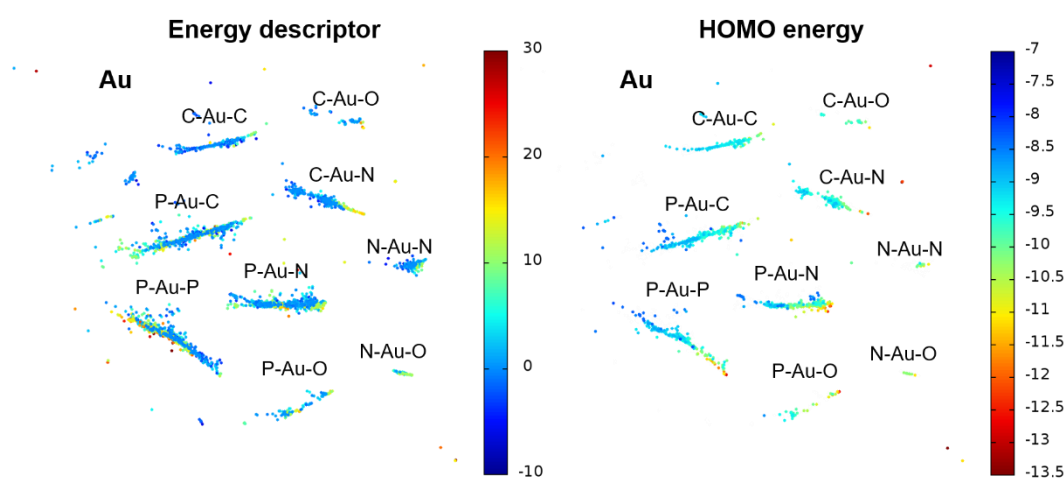


Figure B.7 The Sketch-Maps of gold complexes colored by the energy descriptor (left, in kcal/mol) and the HOMO energy of complex 1 (right, in eV).

## B.4 The Shifts of Ligands from P–M–O References

Table B.1 The absolute values of the average energy descriptor, HOMO energy of complex **1** and the Mulliken charge on metal of complex **1** for nickel complexes.

Ni complexes	$\Delta E$ (kcal/mol)	HOMO (eV)	Mulliken charge
C–M–C	–76.030	–3.543	–0.102
O–M–C	–70.732	–3.963	0.011
C–M–N	–71.228	–4.058	0.022
N–M–N	–63.992	–4.267	–0.046
P–M–C	–62.123	–3.913	–0.292
P–M–O	–58.932	–4.188	–0.159
P–M–P	–58.762	–4.037	–0.470
N–M–O	–58.061	–4.341	0.067
P–M–N	–55.953	–4.099	–0.288
X–M–CO	–47.166	–5.126	0.072

Table B.2 The absolute values of the average energy descriptor, HOMO energy of complex **1** and the Mulliken charge on metal of complex **1** for palladium complexes.

Pd complexes	$\Delta E$ (kcal/mol)	HOMO (eV)	Mulliken charge
C–M–C	–47.444	–3.773	–0.504
O–M–C	–45.982	–4.043	–0.320
C–M–N	–45.502	–3.901	–0.393
N–M–O	–39.367	–4.115	–0.207
N–M–N	–39.233	–3.974	–0.326
P–M–C	–38.119	–4.275	–0.444
P–M–O	–34.150	–4.645	–0.286
P–M–N	–32.840	–4.449	–0.361
P–M–P	–28.573	–4.873	–0.395
X–M–CO	–19.534	–5.406	–0.193

Table B.3 The absolute values of the average energy descriptor, HOMO energy of complex **1** and the Mulliken charge on metal of complex **1** for platinum complexes.

Pt complexes	$\Delta E$ (kcal/mol)	HOMO (eV)	Mulliken charge
O–M–C	–52.405	–4.312	–0.471
C–M–C	–52.288	–4.123	–0.574
C–M–N	–52.109	–4.252	–0.555
N–M–N	–49.248	–4.409	–0.501
N–M–O	–48.371	–4.547	–0.422
P–M–N	–42.909	–4.851	–0.464
P–M–C	–42.187	–4.602	–0.515
P–M–O	–42.090	–4.919	–0.374
P–M–P	–33.173	–5.019	–0.493
X–M–CO	–27.970	–5.753	–0.294

## Appendix B. Supplementary for Machine Learning Meets Volcano Plots and Data Mining of the C–C Cross-Coupling

Table B.4 The absolute values of the average energy descriptor, HOMO energy of complex **1** and the Mulliken charge on metal of complex **1** for copper complexes.

Cu complexes	$\Delta E$ (kcal/mol)	HOMO (eV)	Mulliken charge
C–M–C	–15.654	–9.062	–0.578
P–M–C	–12.676	–9.226	–0.487
P–M–P	–8.436	–9.383	–0.516
C–M–N	–7.619	–9.494	–0.316
O–M–C	–5.088	–9.776	–0.079
P–M–N	–3.420	–9.710	–0.279
P–M–O	–1.572	–9.936	–0.085
N–M–N	3.288	–9.844	–0.071
X–M–CO	5.941	–10.775	0.038
N–M–O	9.226	–10.328	0.183

Table B.5 The absolute values of the average energy descriptor, HOMO energy of complex **1** and the Mulliken charge on metal of complex **1** for silver complexes.

Ag complexes	$\Delta E$ (kcal/mol)	HOMO (eV)	Mulliken charge
C–M–C	11.796	–9.299	–0.054
P–M–C	13.505	–9.291	–0.023
C–M–N	16.437	–9.448	0.103
O–M–C	19.197	–9.797	0.184
P–M–N	19.779	–9.504	0.121
P–M–O	22.733	–9.942	0.186
P–M–P	23.090	–9.250	0.043
N–M–N	27.072	–10.075	0.285
C–M–O	28.934	–10.462	0.178
N–M–O	29.420	–10.570	0.384

Table B.6 The absolute values of the average energy descriptor, HOMO energy of complex **1** and the Mulliken charge on metal of complex **1** for gold complexes.

Au complexes	$\Delta E$ (kcal/mol)	HOMO (eV)	Mulliken charge
C–M–C	–2.015	–9.348	0.053
P–M–C	0.216	–9.351	0.037
C–M–N	0.533	–9.670	0.118
O–M–C	1.027	–9.853	0.236
P–M–P	4.175	–9.421	–0.004
P–M–O	4.992	–10.108	0.205
P–M–N	5.820	–9.967	0.078
N–M–N	5.912	–10.268	0.162
N–M–O	10.657	–10.533	0.280
C–M–O	12.505	–10.774	0.269

# C Supplementary for Volcano Plots of the Hydrogenation of Carbon Dioxide to Formate

This appendix is based on the supporting information of following publications:

- Sawatlon B., Wodrich M. D., and Corminboeuf C., Unraveling Metal/Pincer Ligand Effects in the Catalytic Hydrogenation of Carbon Dioxide to Formate. *Organometallics* **2018**, 37, 4568–4575.
- Wodrich M. D., Sawatlon B., Solel E., Kozuch S., and Corminboeuf C., Activity-Based Screening of Homogeneous Catalysts through the Rapid Assessment of Theoretically Derived Turnover Frequencies. *ACS Catal.* **2019**, 9, 5716–5725.

## C.1 Comparison of Reaction Mechanism

According to previous studies on the mechanistic pathways of the hydrogenation of carbon dioxide,<sup>279, 290-291</sup> two possible pathways are considered here. Both mechanisms start the reaction by the insertion of CO<sub>2</sub> to the hydrido ligand of catalyst *via* **TS1**. Then, for the direct pathway, the addition of H<sub>2</sub> to the catalyst **3** results in the molecular H<sub>2</sub> complex **4**, which can regenerate the starting complex **1** by proton abstraction of OH<sup>−</sup> base and dissociation of water (Figure C.1, black line). In contrast, for the ligand cooperative pathway, the formation of water starts by the proton abstraction of OH<sup>−</sup> base at the backbone of pincer ligand (**TS3**) while the formate adduct still binds with the catalyst (Figure C.1, red line). The formate then dissociates to open a vacant site at the metal center where a molecular hydrogen can bind to and forms the complex **7**. The starting complex **1** can be regenerated by the heterolytic cleavage of H<sub>2</sub> in which proton moves to carbon at the pincer backbone to regenerate the aromatic pyridine ring *via* **TS4**. This pathway is sometimes called as aromatic/dearomatic mechanism.

We computed both catalytic cycles of IrPNP and RhPNP to evaluate the most feasible pathway and found that the direct one has the lowest energy barriers for both catalysts [13.8 and 16.0 kcal/mol for Ir(PNP) and Rh(PNP), respectively] where the rate-determining step was found at the heterolytic H<sub>2</sub> cleavage and proton transfer. The cooperative pathway, on the other hand, has the higher barriers than the direct pathway in two steps, which are the proton abstraction at pincer backbone (**2**→**TS3**) and the regeneration of the starting complexes **1** (**8**→**TS4**). Both Ir(PNP) and Rh(PNP) require 21.6 kcal/mol to overcome the highest barrier. Therefore, we computed the energies and constructed the linear scaling relationships and the molecular volcano plot based on

## Appendix C. Supplementary for Volcano Plots of the Hydrogenation of Carbon Dioxide to Formate

the direct pathway. Note that, to moderate the complexity of the mechanism, we discounted some intermediates that do not affect the energy barriers, *i.e.*, the species with H-bound formate (**2\_H**), the complex with solvent (water) coordinated (**3\_H2O**), and the weak interaction between proton of water and hydrido ligand species (**5**).

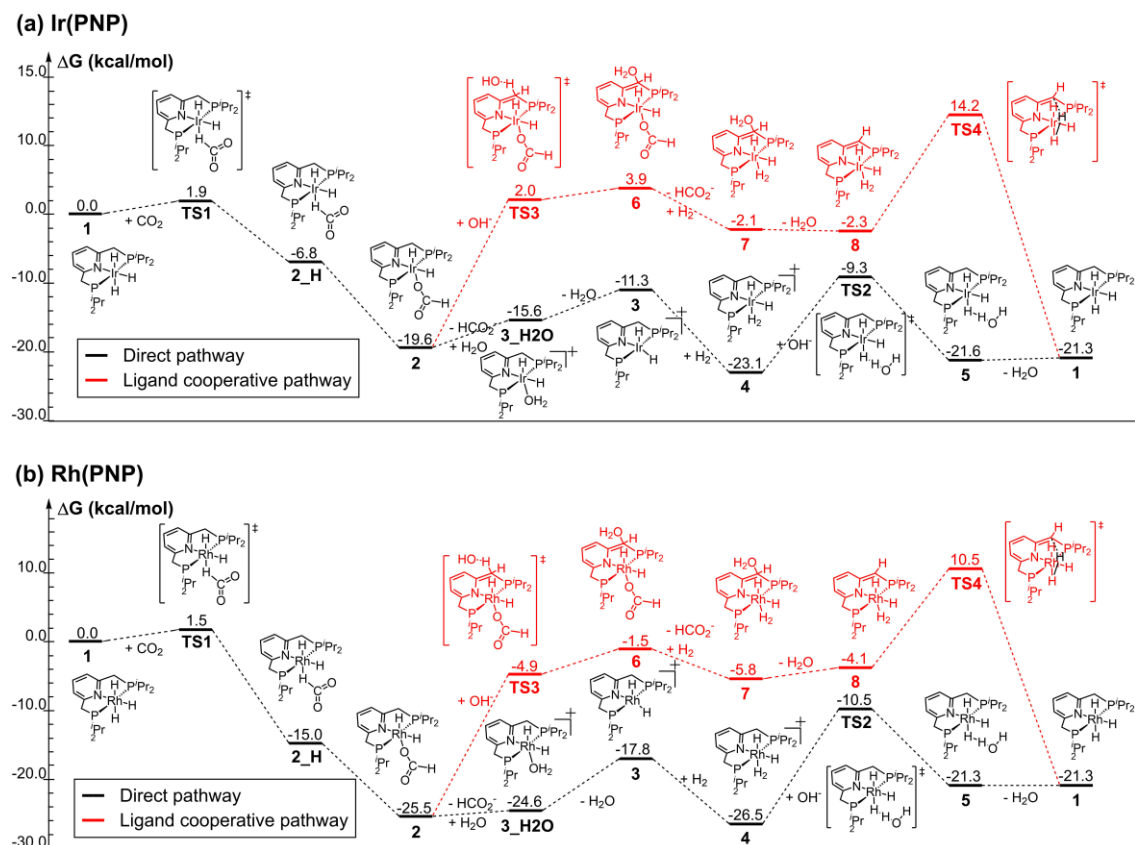


Figure C.1 Energy diagrams of (a) IrPNP and (b) RhPNP comparing two possible pathways: the direct pathway without ligand involvement (black line) and the non-innocent ligand pathway (red line).

## C.2 Construction of the Molecular Volcano Plot

The method for constructing molecular volcano plots has previously been summarized in detail elsewhere.<sup>5, 62, 64</sup> Both reactants and products are used as reference states. Here, the free energy associated with the formation of species **4** and formate was used as a descriptor variable (Equation 7.1).

The free energies of the different reaction steps [*i.e.*, (**1**→**TS1**), (**TS1**→**2**), (**2**→**3**), (**3**→**4**), (**4**→**TS2**), and (**TS2**→**1**)] are estimated from the linear free energy scaling relationships (Figure 6.2):

$$\begin{aligned}
 \Delta G_{\text{RRS}}(\mathbf{2}) &= 0.4\Delta G_{\text{RRS}}(\mathbf{4}) - 13.6 \text{ kcal/mol}, \\
 \Delta G_{\text{RRS}}(\mathbf{3}) &= 0.7\Delta G_{\text{RRS}}(\mathbf{4}) + 1.0 \text{ kcal/mol}, \\
 \Delta G_{\text{RRS}}(\mathbf{TS1}) &= 0.1\Delta G_{\text{RRS}}(\mathbf{4}) + 5.2 \text{ kcal/mol}, \\
 \Delta G_{\text{RRS}}(\mathbf{TS2}) &= 0.5\Delta G_{\text{RRS}}(\mathbf{4}) + 1.1 \text{ kcal/mol}
 \end{aligned}$$

using the procedure outlined below.

The reactants and products are used to define the reference states:

$$G(\mathbf{1}) = G(\text{CO}_2) = G(\text{HCOO}^-) = G(\text{H}_2) = G(\text{OH}^-) = 0$$

$$G(\text{H}_2\text{O}) = -21.28 \text{ kcal/mol}$$

**1→TS:**

$$-\Delta G_{\text{theo}}(\mathbf{1} \rightarrow \mathbf{TS1}) = - [\Delta G(\mathbf{TS1}) - (G(\mathbf{1}) + G(\text{CO}_2))]$$

Since  $G(\mathbf{1})$  and  $G(\text{CO}_2)$  are zero by definition and substituting in the linear free energy scaling relationship for  $\Delta G(\mathbf{TS1})$  gives:

$$-\Delta G_{\text{theo}}(\mathbf{1} \rightarrow \mathbf{TS1}) = -0.1\Delta G_{\text{RRS}}(\mathbf{4}) - 5.2 \text{ kcal/mol}$$

**TS1→2:**

$$-\Delta G_{\text{theo}}(\mathbf{TS1} \rightarrow \mathbf{2}) = - [\Delta G(\mathbf{2}) - \Delta G(\mathbf{TS1})]$$

Substituting in the linear free energy scaling relationship for  $\Delta G(\mathbf{TS1})$  and  $\Delta G(\mathbf{2})$ :

$$-\Delta G_{\text{theo}}(\mathbf{TS1} \rightarrow \mathbf{2}) = - [(0.4\Delta G_{\text{RRS}}(\mathbf{4}) - 13.6) - (0.1\Delta G_{\text{RRS}}(\mathbf{4}) + 5.2)]$$

$$-\Delta G_{\text{theo}}(\mathbf{TS1} \rightarrow \mathbf{2}) = -0.3\Delta G_{\text{RRS}}(\mathbf{4}) + 18.8 \text{ kcal/mol}$$

**2→3:**

$$-\Delta G_{\text{theo}}(\mathbf{2} \rightarrow \mathbf{3}) = - [G(\mathbf{3}) - (G(\mathbf{2}) + G(\text{HCOO}^-))]$$

$G(\text{HCOO}^-)$  are zero by definition and substituting the linear free energy scaling relationship for  $\Delta G(\mathbf{2})$ :

$$-\Delta G_{\text{theo}}(\mathbf{2} \rightarrow \mathbf{3}) = - [(0.7\Delta G_{\text{RRS}}(\mathbf{4}) + 1.0) - (0.4\Delta G_{\text{RRS}}(\mathbf{4}) - 13.6)]$$

$$-\Delta G_{\text{theo}}(\mathbf{2} \rightarrow \mathbf{3}) = -0.3\Delta G_{\text{RRS}}(\mathbf{4}) - 14.6 \text{ kcal/mol}$$

**3→4:**

$$-\Delta G_{\text{theo}}(\mathbf{3} \rightarrow \mathbf{4}) = - [G(\mathbf{4}) - (G(\mathbf{3}) + G(\text{H}_2))]$$

$G(\text{H}_2)$  is zero by definition and substituting the linear free energy scaling relationship for  $\Delta G(\mathbf{4})$ :

$$-\Delta G_{\text{theo}}(\mathbf{3} \rightarrow \mathbf{4}) = - [\Delta G_{\text{RRS}}(\mathbf{4}) - (0.7\Delta G_{\text{RRS}}(\mathbf{4}) + 1.0)]$$

$$-\Delta G_{\text{theo}}(\mathbf{3} \rightarrow \mathbf{4}) = -0.3\Delta G_{\text{RRS}}(\mathbf{4}) - 1.0 \text{ kcal/mol}$$

**4→TS2:**

$$-\Delta G_{\text{theo}}(\mathbf{4} \rightarrow \mathbf{TS2}) = -[(G(\mathbf{TS2}) + G(\text{OH}^-) - G(\mathbf{4}))]$$

$G(\text{OH}^-)$  is zero by definition and substituting in the linear free energy scaling relationship for  $\Delta G(\mathbf{TS2})$ :

$$-\Delta G_{\text{theo}}(\mathbf{4} \rightarrow \mathbf{TS2}) = -[(0.5\Delta G_{\text{RRS}}(\mathbf{4}) - 1.1) - \Delta G_{\text{RRS}}(\mathbf{4})]$$

$$-\Delta G_{\text{theo}}(\mathbf{4} \rightarrow \mathbf{TS2}) = (0.5\Delta G_{\text{RRS}}(\mathbf{4}) - 1.1 \text{ kcal/mol})$$

**TS2→1:**

$$-\Delta G_{\text{theo}}(\mathbf{TS2} \rightarrow \mathbf{1}) = -[(G(\mathbf{1}) + G(\text{H}_2\text{O}) - G(\mathbf{TS2}))]$$

$G(\mathbf{1})$  are zero by definition, substituting in the linear free energy scaling relationship for  $\Delta G(\mathbf{TS2})$  and substituting  $G(\text{H}_2\text{O})$  with the overall reaction free energy:

$$-\Delta G_{\text{theo}}(\mathbf{TS2} \rightarrow \mathbf{1}) = -[-21.28 - (0.5\Delta G_{\text{RRS}}(\mathbf{4}) + 1.1)]$$

$$-\Delta G_{\text{theo}}(\mathbf{TS2} \rightarrow \mathbf{1}) = 0.5\Delta G_{\text{RRS}}(\mathbf{4}) + 22.4 \text{ kcal/mol}$$

Plotting the lines for each of these reaction energies gives a simulated reaction profile (Figure 6.3a).

The final volcano plot is then obtained by taking the most energetically costly reaction step (as defined by equation below) for each descriptor value ( $x$ -axis), which is then plotted on the  $y$ -axis.

$$\Delta G(kds) = \max \left[ \begin{matrix} \Delta G(1 \rightarrow \mathbf{TS1}), \Delta G(\mathbf{TS1} \rightarrow 2), \Delta G(2 \rightarrow 3), \\ \Delta G(3 \rightarrow 4), \Delta G(4 \rightarrow \mathbf{TS2}), \Delta G(\mathbf{TS2} \rightarrow 1) \end{matrix} \right]$$

### C.3 Linear Free Energy Scaling Relationships on Different Temperatures

Linear free energy scaling relationships (LFESRs) were developed to estimate the relative free energies of catalytic cycle intermediates and transition states based on the free energy value of the descriptor variable. We previously found  $\Delta G_{\text{RRS}}(\mathbf{4})$  to be the most suitable descriptor, which relates to the other intermediates and transition states as follows:

**LFESRs at 273.15K:**

$\Delta G_{\text{RRS}}(\mathbf{TS1})$	=	$0.125 \Delta G_{\text{RRS}}(\mathbf{4}) + 4.72 \text{ kcal/mol}$	$R^2 = 0.76$
$\Delta G_{\text{RRS}}(\mathbf{2})$	=	$0.424 \Delta G_{\text{RRS}}(\mathbf{4}) - 14.06 \text{ kcal/mol}$	$R^2 = 0.79$
$\Delta G_{\text{RRS}}(\mathbf{3})$	=	$0.704 \Delta G_{\text{RRS}}(\mathbf{4}) + 1.23 \text{ kcal/mol}$	$R^2 = 0.90$



$$\Delta G_{\text{RRS}}(\text{TS2}) = 0.493 \Delta G_{\text{RRS}}(\mathbf{4}) + 0.77 \text{ kcal/mol} \quad R^2 = 0.96$$

**LFESRs at 298.15K:**

$$\Delta G_{\text{RRS}}(\text{TS1}) = 0.125 \Delta G_{\text{RRS}}(\mathbf{4}) + 5.20 \text{ kcal/mol} \quad R^2 = 0.76$$

$$\Delta G_{\text{RRS}}(\mathbf{2}) = 0.424 \Delta G_{\text{RRS}}(\mathbf{4}) - 13.61 \text{ kcal/mol} \quad R^2 = 0.79$$

$$\Delta G_{\text{RRS}}(\mathbf{3}) = 0.704 \Delta G_{\text{RRS}}(\mathbf{4}) + 0.99 \text{ kcal/mol} \quad R^2 = 0.90$$

$$\Delta G_{\text{RRS}}(\text{TS2}) = 0.493 \Delta G_{\text{RRS}}(\mathbf{4}) + 1.14 \text{ kcal/mol} \quad R^2 = 0.96$$

**LFESRs at 393.15K:**

$$\Delta G_{\text{RRS}}(\text{TS1}) = 0.125 \Delta G_{\text{RRS}}(\mathbf{4}) + 7.04 \text{ kcal/mol} \quad R^2 = 0.76$$

$$\Delta G_{\text{RRS}}(\mathbf{2}) = 0.424 \Delta G_{\text{RRS}}(\mathbf{4}) - 11.95 \text{ kcal/mol} \quad R^2 = 0.79$$

$$\Delta G_{\text{RRS}}(\mathbf{3}) = 0.704 \Delta G_{\text{RRS}}(\mathbf{4}) + 0.06 \text{ kcal/mol} \quad R^2 = 0.90$$

$$\Delta G_{\text{RRS}}(\text{TS2}) = 0.493 \Delta G_{\text{RRS}}(\mathbf{4}) + 2.61 \text{ kcal/mol} \quad R^2 = 0.96$$

These LFESRs can then be used to estimate the free energy associated with moving between two connected points of the catalytic cycle and to create the Figure 7.2b and 8.6a volcano plot.

**$\Delta G(\text{RXN})$  at 273.15K:**

$$-\Delta G(\mathbf{1} \rightarrow \text{TS1}) = -0.125 \Delta G_{\text{RRS}}(\mathbf{4}) - 4.72 \text{ kcal/mol}$$

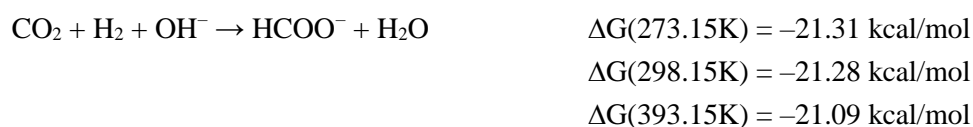
$$-\Delta G(\text{TS1} \rightarrow \mathbf{2}) = -0.299 \Delta G_{\text{RRS}}(\mathbf{4}) + 18.78 \text{ kcal/mol}$$

$$-\Delta G(\mathbf{2} \rightarrow \mathbf{3}) = -0.280 \Delta G_{\text{RRS}}(\mathbf{4}) - 15.29 \text{ kcal/mol}$$

$$-\Delta G(\mathbf{3} \rightarrow \mathbf{4}) = -0.296 \Delta G_{\text{RRS}}(\mathbf{4}) + 1.23 \text{ kcal/mol}$$

$$-\Delta G(\mathbf{4} \rightarrow \text{TS2}) = 0.507 \Delta G_{\text{RRS}}(\mathbf{4}) - 0.77 \text{ kcal/mol}$$

$$-\Delta G(\text{TS2} \rightarrow \mathbf{1}) = 0.493 \Delta G_{\text{RRS}}(\mathbf{4}) + 22.08 \text{ kcal/mol}$$



## C.4 Temperature Influence on the TOF Volcano

If we compute the TOF according to Equation 7.2 the position of the maximum TOF is expected to correspond with the minimum energy span. However, at the crossing points between the straight lines of the different energy span regions, there are small areas in which the two different energy differences are very close in magnitude. In these areas both energy differences should be considered in the denominator of Equation 7.1, and so the simplified Equation 7.2 is not a good

## Appendix C. Supplementary for Volcano Plots of the Hydrogenation of Carbon Dioxide to Formate

---

approximation for the TOF. Around these crossing points the TOF calculated from Equation 7.1 will be smaller than the one calculated from Equation 7.2. This will lower the TOF volcano's tip and move its maximum. Near the crossing points, we can assume that Equation 7.1 has only two dominant energy differences,  $T_i - I_j + \delta G_{ij}$  and  $T_m - I_n + \delta G_{nm}$ . Each of the two differences can be expressed as a linear function of the descriptor value ( $x$ ). In order to form a maximum in the volcano plot, one of these differences should have a positive slope and the other a negative slope:

$$TOF = \frac{k_B T}{h} \frac{1 - e^{\frac{\Delta G_r}{RT}}}{e^{\frac{T_i - I_j + \delta G_{ij}}{RT}} + e^{\frac{T_m - I_n + \delta G_{nm}}{RT}}} \approx \frac{k_B T}{h} \frac{1 - e^{\frac{\Delta G_r}{RT}}}{e^{\frac{-\alpha x + \beta}{RT}} + e^{\frac{\gamma x + \delta}{RT}}}$$

$\alpha$  and  $\gamma$  here are both positive values. In exothermic reactions we can neglect the second term of the numerator, and so we get:

$$TOF \approx \frac{k_B T}{h} \frac{1}{e^{\frac{-\alpha x + \beta}{RT}} + e^{\frac{\gamma x + \delta}{RT}}}$$

In order to find the maximum of the graph we should solve for  $x$  values in which the derivative of this equation is zero:

$$\frac{dTOF}{dx} \approx \frac{d\left(\frac{k_B T}{h} \frac{1}{e^{\frac{-\alpha x + \beta}{RT}} + e^{\frac{\gamma x + \delta}{RT}}}\right)}{dx} = 0 \therefore x = \frac{RT \log\left(\frac{\alpha}{\gamma}\right) + \beta - \delta}{\alpha + \beta}$$

and we can see that the  $x$  value for the maximum is dependent on the temperature and on the relative slopes of the two energy differences. Increasing the temperature will cause a shift of the maximum, while the direction of the shift is towards the straight line with the lower absolute slope.

# Bibliography

- [1] Recognizing the Best in Innovation: Breakthrough Catalyst. *R&D Magazine* 2005, p 20.
- [2] Sigman, M. S.; Harper, K. C.; Bess, E. N.; Milo, A., The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond. *Acc. Chem. Res.* **2016**, *49*, 1292-1301.
- [3] Phillips, K. A.; Wambaugh, J. F.; Grulke, C. M.; Dionisio, K. L.; Isaacs, K. K., High-Throughput Screening of Chemicals as Functional Substitutes using Structure-based Classification Models. *Green Chem.* **2017**, *19*, 1063-1074.
- [4] Nørskov, J. K.; Bligaard, T.; Rossmeisl, J.; Christensen, C. H., Towards the Computational Design of Solid Catalysts. *Nat. Chem.* **2009**, *1*, 37-46.
- [5] Busch, M.; Wodrich, M. D.; Corminboeuf, C., Linear Scaling Relationships and Volcano Plots in Homogeneous Catalysis - Revisiting the Suzuki Reaction. *Chem. Sci.* **2015**, *6*, 6754-6761.
- [6] Falivene, L.; Kozlov, S. M.; Cavallo, L., Constructing Bridges between Computational Tools in Heterogeneous and Homogeneous Catalysis. *ACS Catal.* **2018**, *8*, 5637-5656.
- [7] Anand, M.; Rohr, B.; Statt, M. J.; Nørskov, J. K., Scaling Relationships and Volcano Plots in Homogeneous Catalysis. *J. Phys. Chem. Lett.* **2020**, 8518-8526.
- [8] Durand, D. J.; Fey, N., Computational Ligand Descriptors for Catalyst Design. *Chem. Rev.* **2019**, *119*, 6561-6594.
- [9] Hammes-Schiffer, S., Catalysts by Design: The Power of Theory. *Acc. Chem. Res.* **2017**, *50*, 561-566.
- [10] Cheng, G.-J.; Zhang, X.; Chung, L. W.; Xu, L.; Wu, Y.-D., Computational Organic Chemistry: Bridging Theory and Experiment in Establishing the Mechanisms of Chemical Reactions. *J. Am. Chem. Soc.* **2015**, *137*, 1706-1725.
- [11] Poree, C.; Schoenebeck, F., A Holy Grail in Chemistry: Computational Catalyst Design: Feasible or Fiction? *Acc. Chem. Res.* **2017**, *50*, 605-608.
- [12] Muster, T. H.; Trinchì, A.; Markley, T. A.; Lau, D.; Martin, P.; Bradbury, A.; Bendavid, A.; Dligatch, S., A Review of High Throughput and Combinatorial Electrochemistry. *Electrochim. Acta* **2011**, *56*, 9679-9699.

- [13] Seh, Z. W.; Kibsgaard, J.; Dickens, C. F.; Chorkendorff, I.; Nørskov, J. K.; Jaramillo, T. F., Combining Theory and Experiment in Electrocatalysis: Insights into Materials Design. *Science* **2017**, 355, eaad4998.
- [14] Calle-Vallejo, F.; Loffreda, D.; Koper, M. T. M.; Sautet, P., Introducing Structural Sensitivity into Adsorption–Energy Scaling Relations by Means of Coordination Numbers. *Nat. Chem.* **2015**, 7, 403–410.
- [15] Wang, Z.; Wang, H.-F.; Hu, P., Possibility of Designing Catalysts Beyond the Traditional Volcano Curve: A Theoretical Framework for Multi-Phase Surfaces. *Chem. Sci.* **2015**, 6, 5703-5711.
- [16] Wang, Y.; Montoya, J. H.; Tsai, C.; Ahlquist, M. S. G.; Nørskov, J. K.; Studt, F., Scaling Relationships for Binding Energies of Transition Metal Complexes. *Catal. Lett.* **2016**, 146, 304-308.
- [17] Audran, G.; Brémond, P.; Marque, S. R. A.; Siri, D.; Santelli, M., Calculated Linear Free Energy Relationships in the Course of the Suzuki–Miyaura Coupling Reaction. *Tetrahedron* **2014**, 70, 2272-2279.
- [18] Gustafson, J. L.; Sigman, M. S.; Miller, S. J., Linear Free-Energy Relationship Analysis of a Catalytic Desymmetrization Reaction of a Diarylmethane-bis(phenol). *Org. Lett.* **2010**, 12, 2794-2797.
- [19] Harper, K. C.; Sigman, M. S., Three-Dimensional Correlation of Steric and Electronic Free Energy Relationships Guides Asymmetric Propargylation. *Science* **2011**, 333, 1875.
- [20] Hickey, D. P.; Schiedler, D. A.; Matanovic, I.; Doan, P. V.; Atanassov, P.; Minter, S. D.; Sigman, M. S., Predicting Electrocatalytic Properties: Modeling Structure–Activity Relationships of Nitroxyl Radicals. *J. Am. Chem. Soc.* **2015**, 137, 16179-16186.
- [21] Chen, Z.-M.; Hilton, M. J.; Sigman, M. S., Palladium-Catalyzed Enantioselective Redox-Relay Heck Arylation of 1,1-Disubstituted Homoallylic Alcohols. *J. Am. Chem. Soc.* **2016**, 138, 11461-11464.
- [22] Niemeyer, Z. L.; Milo, A.; Hickey, D. P.; Sigman, M. S., Parameterization of Phosphine Ligands Reveals Mechanistic Pathways and Predicts Reaction Outcomes. *Nat. Chem.* **2016**, 8, 610-617.
- [23] Neel, A. J.; Milo, A.; Sigman, M. S.; Toste, F. D., Enantiodivergent Fluorination of Allylic Alcohols: Data Set Design Reveals Structural Interplay between Achiral Directing Group and Chiral Anion. *J. Am. Chem. Soc.* **2016**, 138, 3863-3875.
- [24] Guo, J.-Y.; Minko, Y.; Santiago, C. B.; Sigman, M. S., Developing Comprehensive Computational Parameter Sets To Describe the Performance of Pyridine-Oxazoline and Related Ligands. *ACS Catal.* **2017**, 7, 4144-4151.
- [25] Park, Y.; Niemeyer, Z. L.; Yu, J.-Q.; Sigman, M. S., Quantifying Structural Effects of Amino Acid Ligands in Pd(II)-Catalyzed Enantioselective C–H Functionalization Reactions. *Organometallics* **2018**, 37, 203-210.
- [26] Hammett, L. P., Some Relations between Reaction Rates and Equilibrium Constants. *Chem. Rev.* **1935**, 17, 125-136.

- 
- [27] Hammett, L. P., The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96-103.
- [28] Hammett, L. P., Linear Free Energy Relationships in Rate and Equilibrium Phenomena. *Trans. Faraday Society* **1938**, *34*, 156-165.
- [29] Santiago, C. B.; Milo, A.; Sigman, M. S., Developing a Modern Approach To Account for Steric Effects in Hammett-Type Correlations. *J. Am. Chem. Soc.* **2016**, *138*, 13424-13430.
- [30] Evans, M. G.; Polanyi, M., Inertia and Driving Force of Chemical Reactions. *Trans. Faraday Society* **1938**, *34*, 11-24.
- [31] Bell, R. P.; Hinshelwood, C. N., The Theory of Reactions involving Proton Transfers. *Proc. R. Soc. London, Ser. A* **1936**, *154*, 414-429.
- [32] Brønsted, J.; Pedersen, K., Die Katalytische Zersetzung des Nitramids und Ihre Physikalisch-Chemische Bedeutung. *Z. Phys. Chem.* **1924**, *108*, 185-235.
- [33] Vorotnikov, V.; Vlachos, D. G., Group Additivity and Modified Linear Scaling Relations for Estimating Surface Thermochemistry on Transition Metal Surfaces: Application to Furanics. *J. Phys. Chem. C* **2015**, *119*, 10417-10426.
- [34] Kiss, I. Z.; Kazsu, Z.; Gaspar, V., Scaling Relationship for Oscillating Electrochemical Systems: Dependence of Phase Diagram on Electrode Size and Rotation Rate. *Phys. Chem. Chem. Phys.* **2009**, *11*, 7669-7677.
- [35] Sabatier, P., Hydrogénations et Déshydrogénations par Catalyse. *Berichte der deutschen chemischen Gesellschaft* **1911**, *44*, 1984-2001.
- [36] Sabatier, P., *La Catalyse en Chimie Organique*. Librairie polytechnique: 1913.
- [37] Wodrich, M. D.; Sawatlon, B.; Busch, M.; Corminboeuf, C., The Genesis of Molecular Volcano Plots. *Acc. Chem. Res.* **2021**, *54*, 1107-1117.
- [38] Sawatlon, B.; Wodrich, M. D.; Corminboeuf, C., Probing Substrate Scope with Molecular Volcanoes. *Org. Lett.* **2020**, *22*, 7936-7941.
- [39] Meyer, B.; Sawatlon, B.; Heinen, S. N.; von Lilienfeld, A.; Corminboeuf, C., Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts. *Chem. Sci.* **2018**, *9*, 7069-7077.
- [40] Ceriotti, M.; Tribello, G. A.; Parrinello, M., Simplifying the Representation of Complex Free-Energy Landscapes using Sketch-Map. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13023.
- [41] Ceriotti, M.; Tribello, G. A.; Parrinello, M., Demonstrating the Transferability and the Descriptive Power of Sketch-Map. *J. Chem. Theory Comput.* **2013**, *9*, 1521-1532.
- [42] Sawatlon, B.; Wodrich, M. D.; Meyer, B.; Fabrizio, A.; Corminboeuf, C., Data Mining the C-C Cross-Coupling Genome. *ChemCatChem* **2019**, *11*, 4096-4107.
- [43] Sawatlon, B.; Wodrich, M. D.; Corminboeuf, C., Unraveling Metal/Pincer Ligand Effects in the Catalytic Hydrogenation of Carbon Dioxide to Formate. *Organometallics* **2018**, *37*, 4568-4575.

- [44] Wodrich, M. D.; Sawatlon, B.; Solel, E.; Kozuch, S.; Corminboeuf, C., Activity-Based Screening of Homogeneous Catalysts through the Rapid Assessment of Theoretically Derived Turnover Frequencies. *ACS Catal.* **2019**, *9*, 5716-5725.
- [45] Gerischer, H., Mechanism of Electrolytic Discharge of Hydrogen and Adsorption Energy of Atomic Hydrogen. *Bull. Soc. Chim. Belg.* **1958**, *67*, 506.
- [46] Parsons, R., The Rate of Electrolytic Hydrogen Evolution and the Heat of Adsorption of Hydrogen. *Trans. Faraday Society* **1958**, *54*, 1053-1063.
- [47] Trasatti, S., Electrocatalysis in the Anodic Evolution of Oxygen and Chlorine. *Electrochim. Acta* **1984**, *29*, 1503-1512.
- [48] Trasatti, S., Work Function, Electronegativity, and Electrochemical Behaviour of Metals: III. Electrolytic Hydrogen Evolution in Acid Solutions. *J. Electroanal. Chem. Interfacial Electrochem.* **1972**, *39*, 163-184.
- [49] Nørskov, J. K.; Bligaard, T.; Logadottir, A.; Kitchin, J. R.; Chen, J. G.; Pandelov, S.; Nørskov, J. K., Trends in the Exchange Current for Hydrogen Evolution. *J. Electrochem. Soc.* **2005**, *152*, J23-J26.
- [50] Nørskov, J. K.; Rossmeisl, J.; Logadottir, A.; Lindqvist, L.; Kitchin, J. R.; Bligaard, T.; Jónsson, H., Origin of the Overpotential for Oxygen Reduction at a Fuel-Cell Cathode. *J. Phys. Chem. B* **2004**, *108*, 17886-17892.
- [51] Bligaard, T.; Nørskov, J. K.; Dahl, S.; Matthiesen, J.; Christensen, C. H.; Sehested, J., The Brønsted–Evans–Polanyi Relation and the Volcano Curve in Heterogeneous Catalysis. *J. Catal.* **2004**, *224*, 206-217.
- [52] Calle-Vallejo, F.; Martínez, J. I.; García-Lastra, J. M.; Rossmeisl, J.; Koper, M. T. M., Physical and Chemical Nature of the Scaling Relations between Adsorption Energies of Atoms on Metal Surfaces. *Phys. Rev. Lett.* **2012**, *108*, 116103.
- [53] Ferrin, P.; Simonetti, D.; Kandoi, S.; Kunkes, E.; Dumesic, J. A.; Nørskov, J. K.; Mavrikakis, M., Modeling Ethanol Decomposition on Transition Metals: A Combined Application of Scaling and Brønsted–Evans–Polanyi Relations. *J. Am. Chem. Soc.* **2009**, *131*, 5809-5815.
- [54] Nørskov, J. K.; Bligaard, T.; Logadottir, A.; Bahn, S.; Hansen, L. B.; Bollinger, M.; Bengaard, H.; Hammer, B.; Sljivancanin, Z.; Mavrikakis, M.; Xu, Y.; Dahl, S.; Jacobsen, C. J. H., Universality in Heterogeneous Catalysis. *J. Catal.* **2002**, *209*, 275-278.
- [55] Man, I. C.; Su, H.-Y.; Calle-Vallejo, F.; Hansen, H. A.; Martínez, J. I.; Inoglu, N. G.; Kitchin, J.; Jaramillo, T. F.; Nørskov, J. K.; Rossmeisl, J., Universality in Oxygen Evolution Electrocatalysis on Oxide Surfaces. *ChemCatChem* **2011**, *3*, 1159-1165.
- [56] Rossmeisl, J.; Qu, Z. W.; Zhu, H.; Kroes, G. J.; Nørskov, J. K., Electrolysis of Water on Oxide Surfaces. *J. Electroanal. Chem.* **2007**, *607*, 83-89.
- [57] Exner, K. S., Recent Advancements Towards Closing the Gap between Electrocatalysis and Battery Science Communities: The Computational Lithium Electrode and Activity–Stability Volcano Plots. *ChemSusChem* **2019**, *12*, 2330-2344.

- 
- [58] Busch, M.; Halck, N. B.; Kramm, U. I.; Siahrostami, S.; Krttil, P.; Rossmeisl, J., Beyond the Top of the Volcano? – A Unified Approach to Electrocatalytic Oxygen Reduction and Oxygen Evolution. *Nano Energy* **2016**, *29*, 126-135.
- [59] Busch, M., Water Oxidation: From Mechanisms to Limitations. *Curr. Opin. Electrochem.* **2018**, *9*, 278-284.
- [60] Swiegers, G., *Mechanical Catalysis: Methods of Enzymatic, Homogeneous, and Heterogeneous Catalysis*. Wiley: 2008.
- [61] Wodrich, M. D.; Sawatlon, B.; Busch, M.; Corminboeuf, C., On the Generality of Molecular Volcano Plots. *ChemCatChem* **2018**, *10*, 1586-1591.
- [62] Busch, M.; Wodrich, M. D.; Corminboeuf, C., A Generalized Picture of C–C Cross-Coupling. *ACS Catal.* **2017**, *7*, 5643-5653.
- [63] Busch, M.; Wodrich, M. D.; Corminboeuf, C., Improving the Thermodynamic Profiles of Prospective Suzuki–Miyaura Cross-Coupling Catalysts by Altering the Electrophilic Coupling Component. *ChemCatChem* **2018**, *10*, 1592-1597.
- [64] Wodrich, M. D.; Busch, M.; Corminboeuf, C., Accessing and Predicting the Kinetic Profiles of Homogeneous Catalysts from Volcano Plots. *Chem. Sci.* **2016**, *7*, 5723-5735.
- [65] Wodrich, M. D.; Busch, M.; Corminboeuf, C., Expedited Screening of Active and Regioselective Catalysts for the Hydroformylation Reaction. *Helv. Chim. Acta* **2018**, *101*, e1800107.
- [66] Miyaura, N.; Yamada, K.; Suzuki, A., A New Stereospecific Cross-Coupling by the Palladium-Catalyzed Reaction of 1-Alkenylboranes with 1-Alkenyl or 1-Alkynyl Halides. *Tetrahedron Lett.* **1979**, *20*, 3437-3440.
- [67] Miyaura, N.; Suzuki, A., Palladium-Catalyzed Cross-Coupling Reactions of Organoboron Compounds. *Chem. Rev.* **1995**, *95*, 2457-2483.
- [68] Hooshmand, S. E.; Heidari, B.; Sedghi, R.; Varma, R. S., Recent Advances in the Suzuki–Miyaura Cross-Coupling Reaction using Efficient Catalysts in Eco-Friendly Media. *Green Chem.* **2019**, *21*, 381-405.
- [69] Das, P.; Sarmah, C.; Tairai, A.; Bora, U., Highly Efficient Amine-Based Catalytic System for Room Temperature Suzuki–Miyaura Reactions of Aryl Halides with Arylboronic Acids. *Appl. Organomet. Chem.* **2011**, *25*, 283-288.
- [70] Ramaotsoa, G. V.; Strydom, I.; Panayides, J.-L.; Riley, D., Immobilized Tetrakis(triphenylphosphine)palladium(0) for Suzuki–Miyaura Coupling Reactions under Flow Conditions. *React. Chem. Eng.* **2019**, *4*, 372-382.
- [71] Tang, R.-J.; Milcent, T.; Crousse, B., Regioselective Halogenation of Arenes and Heterocycles in Hexafluoroisopropanol. *J. Org. Chem.* **2018**, *83*, 930-938.
- [72] Tagata, T.; Nishida, M., Palladium Charcoal-Catalyzed Suzuki–Miyaura Coupling to Obtain Arylpyridines and Arylquinolines. *J. Org. Chem.* **2003**, *68*, 9412-9415.
- [73] Maetani, S.; Fukuyama, T.; Ryu, I., Rhodium-Catalyzed Decarbonylative C–H Arylation of 2-Aryloxybenzoic Acids Leading to Dibenzofuran Derivatives. *Org. Lett.* **2013**, *15*, 2754-2757.

- [74] Šiaučiulis, M.; Ahlsten, N.; Pulis, A. P.; Procter, D. J., Transition-Metal-Free Cross-Coupling of Benzothiophenes and Styrenes in a Stereoselective Synthesis of Substituted (E,Z)-1,3-Dienes. *Angew. Chem. Int. Ed.* **2019**, *58*, 8779-8783.
- [75] Dau, H.; Limberg, C.; Reier, T.; Risch, M.; Roggan, S.; Strasser, P., The Mechanism of Water Oxidation: From Electrolysis via Homogeneous to Biological Catalysis. *ChemCatChem* **2010**, *2*, 724-761.
- [76] Greeley, J., Theoretical Heterogeneous Catalysis: Scaling Relationships and Computational Catalyst Design. *Annu. Rev. Chem. Biomol. Eng.* **2016**, *7*, 605-635.
- [77] Cordova, M.; Wodrich, M. D.; Meyer, B.; Sawatlon, B.; Corminboeuf, C., Data-Driven Advancement of Homogeneous Nickel Catalyst Activity for Aryl Ether Cleavage. *ACS Catal.* **2020**, 7021-7031.
- [78] Adamo, C.; Barone, V., Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158-6170.
- [79] Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.
- [80] Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **1997**, *78*, 1396-1396.
- [81] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- [82] Weigend, F.; Ahlrichs, R., Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305.
- [83] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378-6396.
- [84] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams, D. J.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J., Gaussian 16, Revision A.03. Wallingford, CT, 2016.
- [85] Bootsma, A. N.; Wheeler, S., Popular Integration Grids Can Result in Large Errors in DFT-Computed Free Energies. DOI: 10.26434/chemrxiv.8864204.v5 **2019**.
- [86] Miyaura, N., Cross-Coupling Reaction of Organoboron Compounds via Base-Assisted Transmetalation to Palladium(II) Complexes. *J. Organomet. Chem.* **2002**, *653*, 54-57.



- 
- [87] Braga, A. A. C.; Morgon, N. H.; Ujaque, G.; Lledós, A.; Maseras, F., Computational Study of the Transmetalation Process in the Suzuki–Miyaura Cross-Coupling of Aryls. *J. Organomet. Chem.* **2006**, *691*, 4459–4466.
- [88] Carrow, B. P.; Hartwig, J. F., Distinguishing Between Pathways for Transmetalation in Suzuki–Miyaura Reactions. *J. Am. Chem. Soc.* **2011**, *133*, 2116–2119.
- [89] Lennox, A. J. J.; Lloyd-Jones, G. C., Transmetalation in the Suzuki–Miyaura Coupling: The Fork in the Trail. *Angew. Chem. Int. Ed.* **2013**, *52*, 7362–7370.
- [90] Ortuño, M. A.; Lledós, A.; Maseras, F.; Ujaque, G., The Transmetalation Process in Suzuki–Miyaura Reactions: Calculations Indicate Lower Barrier via Boronate Intermediate. *ChemCatChem* **2014**, *6*, 3132–3138.
- [91] Payard, P.-A.; Perego, L. A.; Ciofini, I.; Grimaud, L., Taming Nickel-Catalyzed Suzuki–Miyaura Coupling: A Mechanistic Focus on Boron-to-Nickel Transmetalation. *ACS Catal.* **2018**, *8*, 4812–4823.
- [92] Yunker, L. P. E.; Ahmadi, Z.; Logan, J. R.; Wu, W.; Li, T.; Martindale, A.; Oliver, A. G.; McIndoe, J. S., Real-Time Mass Spectrometric Investigations into the Mechanism of the Suzuki–Miyaura Reaction. *Organometallics* **2018**, *37*, 4297–4308.
- [93] Thomas, A. A.; Denmark, S. E., Pre-Transmetalation Intermediates in the Suzuki–Miyaura Reaction Revealed: The Missing Link. *Science* **2016**, *352*, 329.
- [94] Thomas, A. A.; Wang, H.; Zahrt, A. F.; Denmark, S. E., Structural, Kinetic, and Computational Characterization of the Elusive Arylpalladium(II)boronate Complexes in the Suzuki–Miyaura Reaction. *J. Am. Chem. Soc.* **2017**, *139*, 3805–3821.
- [95] Kozuch, S.; Shaik, S., How to Conceptualize Catalytic Cycles? The Energetic Span Model. *Acc. Chem. Res.* **2011**, *44*, 101–110.
- [96] Kozuch, S., A Refinement of Everyday Thinking: The Energetic Span Model for Kinetic Assessment of Catalytic Cycles. *WIREs Comput. Mol. Sci.* **2012**, *2*, 795–815.
- [97] Solel, E.; Tarannam, N.; Kozuch, S., Catalysis: Energy is the Measure of All Things. *Chem. Commun.* **2019**, *55*, 5306–5322.
- [98] Jover, J.; Fey, N.; Purdie, M.; Lloyd-Jones, G. C.; Harvey, J. N., A Computational Study of Phosphine Ligand Effects in Suzuki–Miyaura Coupling. *J. Mol. Catal. A: Chem.* **2010**, *324*, 39–47.
- [99] Collins, K. D.; Gensch, T.; Glorius, F., Contemporary Screening Approaches to Reaction Discovery and Development. *Nat. Chem.* **2014**, *6*, 859–871.
- [100] Jäkel, C.; Paciello, R., High-Throughput and Parallel Screening Methods in Asymmetric Hydrogenation. *Chem. Rev.* **2006**, *106*, 2912–2942.
- [101] Senkan, S., Combinatorial Heterogeneous Catalysis—A New Path in an Old Field. *Angew. Chem. Int. Ed.* **2001**, *40*, 312–329.
- [102] Reetz, M. T., Combinatorial and Evolution-Based Methods in the Creation of Enantioselective Catalysts. *Angew. Chem. Int. Ed.* **2001**, *40*, 284–310.

- [103] Filonenko, G. A.; Smykowski, D.; Szyja, B. M.; Li, G.; Szczygieł, J.; Hensen, E. J. M.; Pidko, E. A., Catalytic Hydrogenation of CO<sub>2</sub> to Formates by a Lutidine-Derived Ru–CNC Pincer Complex: Theoretical Insight into the Unrealized Potential. *ACS Catal.* **2015**, *5*, 1145-1154.
- [104] Friedfeld, M. R.; Shevlin, M.; Hoyt, J. M.; Krska, S. W.; Tudge, M. T.; Chirik, P. J., Cobalt Precursors for High-Throughput Discovery of Base Metal Asymmetric Alkene Hydrogenation Catalysts. *Science* **2013**, *342*, 1076.
- [105] Robbins, D. W.; Hartwig, J. F., A Simple, Multidimensional Approach to High-Throughput Discovery of Catalytic Reactions. *Science* **2011**, *333*, 1423.
- [106] Reetz, M. T., New Methods for the High-Throughput Screening of Enantioselective Catalysts and Biocatalysts. *Angew. Chem. Int. Ed.* **2002**, *41*, 1335-1338.
- [107] Sigman, M. S.; Jacobsen, E. N., Schiff Base Catalysts for the Asymmetric Strecker Reaction Identified and Optimized from Parallel Synthetic Libraries. *J. Am. Chem. Soc.* **1998**, *120*, 4901-4902.
- [108] Sperger, T.; Sanhueza, I. A.; Kalvet, I.; Schoenebeck, F., Computational Studies of Synthetically Relevant Homogeneous Organometallic Catalysis Involving Ni, Pd, Ir, and Rh: An Overview of Commonly Employed DFT Methods and Mechanistic Insights. *Chem. Rev.* **2015**, *115*, 9532-9586.
- [109] Greeley, J.; Markovic, N. M., The Road from Animal Electricity to Green Energy: Combining Experiment and Theory in Electrocatalysis. *Energy Environ. Sci.* **2012**, *5*, 9246-9256.
- [110] Bockris, J. O.; Otagawa, T., The Electrocatalysis of Oxygen Evolution on Perovskites. *J. Electrochem. Soc.* **1984**, *131*, 290-302.
- [111] Ananikov, V. P., *Understanding Organometallic Reaction Mechanisms and Catalysis: Computational and Experimental Tools*. John Wiley & Sons: 2014.
- [112] Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A., Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
- [113] Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Müller, K.-R.; Anatole von Lilienfeld, O., Machine Learning of Molecular Electronic Properties in Chemical Compound Space. *New J. Phys.* **2013**, *15*, 095003.
- [114] Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R., Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404-3419.
- [115] Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A., Big Data Meets Quantum Chemistry Approximations: The  $\Delta$ -Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087-2096.
- [116] Janet, J. P.; Chan, L.; Kulik, H. J., Accelerating Chemical Discovery with Machine Learning: Simulated Evolution of Spin Crossover Complexes with an Artificial Neural Network. *J. Phys. Chem. Lett.* **2018**, *9*, 1064-1071.
- [117] Janet, J. P.; Kulik, H. J., Predicting Electronic Structure Properties of Transition Metal Complexes with Neural Networks. *Chem. Sci.* **2017**, *8*, 5137-5152.

- 
- [118] Janet, J. P.; Kulik, H. J., Resolving Transition Metal Chemical Space: Feature Selection for Machine Learning and Structure–Property Relationships. *J. Phys. Chem. A* **2017**, *121*, 8939-8954.
- [119] Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J., Machine-Learning-Assisted Materials Discovery using Failed Experiments. *Nature* **2016**, *533*, 73-76.
- [120] Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A., Quantitative Structure–Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* **2012**, *112*, 2889-2919.
- [121] von Lilienfeld, O. A., Quantum Machine Learning in Chemical Compound Space. *Angew. Chem. Int. Ed.* **2018**, *57*, 4164-4169.
- [122] Noh, J.; Back, S.; Kim, J.; Jung, Y., Active Learning with Non-Ab Initio Input Features Toward Efficient CO<sub>2</sub> Reduction Catalysts. *Chem. Sci.* **2018**, *9*, 5152-5159.
- [123] Kitchin, J. R., Machine Learning in Catalysis. *Nat. Catal.* **2018**, *1*, 230-232.
- [124] Vignola, E.; Steinmann, S. N.; Vandegehuchte, B. D.; Curulla, D.; Stamatakis, M.; Sautet, P., A Machine Learning Approach to Graph-Theoretical Cluster Expansions of the Energy of Adsorbate Layers. *J. Chem. Phys.* **2017**, *147*, 054106.
- [125] Timoshenko, J.; Lu, D.; Lin, Y.; Frenkel, A. I., Supervised Machine-Learning-Based Determination of Three-Dimensional Structure of Metallic Nanoparticles. *J. Phys. Chem. Lett.* **2017**, *8*, 5091-5098.
- [126] Madaan, N.; Shiju, N. R.; Rothenberg, G., Predicting the Performance of Oxidation Catalysts using Descriptor Models. *Catal. Sci. Technol.* **2016**, *6*, 125-133.
- [127] Ras, E.-J.; Rothenberg, G., Heterogeneous Catalyst Discovery using 21st Century Tools: A Tutorial. *RSC Adv.* **2014**, *4*, 5963-5974.
- [128] Ras, E.-J.; Louwerse, M. J.; Rothenberg, G., New Tricks by Very Old Dogs: Predicting the Catalytic Hydrogenation of HMF Derivatives using Slater-Type Orbitals. *Catal. Sci. Technol.* **2012**, *2*, 2456-2464.
- [129] Maldonado, A. G.; Rothenberg, G., Predictive Modeling in Homogeneous Catalysis: A Tutorial. *Chem. Soc. Rev.* **2010**, *39*, 1891-1902.
- [130] Friedman, J. H., Stochastic Gradient Boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367-378.
- [131] Takigawa, I.; Shimizu, K.-i.; Tsuda, K.; Takakusagi, S., Machine-Learning Prediction of the d-Band Center for Metals and Bimetals. *RSC Adv.* **2016**, *6*, 52587-52595.
- [132] Gasper, R.; Shi, H.; Ramasubramaniam, A., Adsorption of CO on Low-Energy, Low-Symmetry Pt Nanoparticles: Energy Decomposition Analysis and Prediction via Machine-Learning Models. *J. Phys. Chem. C* **2017**, *121*, 5612-5619.
- [133] Jinnouchi, R.; Asahi, R., Predicting Catalytic Activity of Nanoparticles by a DFT-Aided Machine-Learning Algorithm. *J. Phys. Chem. Lett.* **2017**, *8*, 4279-4283.
- [134] Cortes, C.; Vapnik, V., Support-Vector Networks. *Machine Learning* **1995**, *20*, 273-297.

- [135] Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z.; Woo, T. K., Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for CO<sub>2</sub> Capture. *J. Phys. Chem. Lett.* **2014**, *5*, 3056-3060.
- [136] Fernandez, M.; Trefiak, N. R.; Woo, T. K., Atomic Property Weighted Radial Distribution Functions Descriptors of Metal–Organic Frameworks for the Prediction of Gas Uptake Capacity. *J. Phys. Chem. C* **2013**, *117*, 14095-14105.
- [137] Xin, H.; Holewinski, A.; Linic, S., Predictive Structure–Reactivity Models for Rapid Screening of Pt-Based Multimetallic Electrocatalysts for the Oxygen Reduction Reaction. *ACS Catal.* **2012**, *2*, 12-16.
- [138] Ma, X.; Li, Z.; Achenie, L. E. K.; Xin, H., Machine-Learning-Augmented Chemisorption Model for CO<sub>2</sub> Electroreduction Catalyst Screening. *J. Phys. Chem. Lett.* **2015**, *6*, 3528-3533.
- [139] Li, Z.; Wang, S.; Chin, W. S.; Achenie, L. E.; Xin, H., High-Throughput Screening of Bimetallic Catalysts Enabled by Machine Learning. *J. Mater. Chem. A* **2017**, *5*, 24131-24138.
- [140] Rasmussen, C. E.; Williams, C. K. I., *Gaussian Processes for Machine Learning*. MIT Press: 2006.
- [141] Ulissi, Z. W.; Tang, M. T.; Xiao, J.; Liu, X.; Torelli, D. A.; Karamad, M.; Cummins, K.; Hahn, C.; Lewis, N. S.; Jaramillo, T. F.; Chan, K.; Nørskov, J. K., Machine-Learning Methods Enable Exhaustive Searches for Active Bimetallic Facets and Reveal Active Site Motifs for CO<sub>2</sub> Reduction. *ACS Catal.* **2017**, *7*, 6600-6608.
- [142] Ulissi, Z. W.; Singh, A. R.; Tsai, C.; Nørskov, J. K., Automated Discovery and Construction of Surface Phase Diagrams Using Machine Learning. *J. Phys. Chem. Lett.* **2016**, *7*, 3931-3935.
- [143] Ulissi, Z. W.; Medford, A. J.; Bligaard, T.; Nørskov, J. K., To Address Surface Reaction Network Complexity using Scaling Relations Machine Learning and DFT Calculations. *Nat. Commun.* **2017**, *8*, 14621.
- [144] Wexler, R. B.; Martirez, J. M. P.; Rappe, A. M., Chemical Pressure-Driven Enhancement of the Hydrogen Evolving Activity of Ni<sub>2</sub>P from Nonmetal Surface Doping Interpreted via Machine Learning. *J. Am. Chem. Soc.* **2018**, *140*, 4678-4683.
- [145] Landrum, G. A.; Penzotti, J. E.; Putta, S., Machine-Learning Models for Combinatorial Catalyst Discovery. *Meas. Sci. Technol.* **2004**, *16*, 270-277.
- [146] dos Passos Gomes, G.; Pollice, R.; Aspuru-Guzik, A., Navigating through the Maze of Homogeneous Catalyst Design with Machine Learning. *Trends in Chemistry* **2021**, *3*, 96-110.
- [147] Yang, W.; Fidelis, T. T.; Sun, W.-H., Machine Learning in Catalysis, From Proposal to Practicing. *ACS Omega* **2020**, *5*, 83-88.
- [148] Wodrich, M. D.; Fabrizio, A.; Meyer, B.; Corminboeuf, C., Data-Powered Augmented Volcano Plots for Homogeneous Catalysis. *Chem. Sci.* **2020**, *11*, 12070-12080.
- [149] Tkatchenko, A., Machine Learning for Chemical Discovery. *Nat. Commun.* **2020**, *11*, 4125.
- [150] Singh, S.; Pareek, M.; Changotra, A.; Banerjee, S.; Bhaskararao, B.; Balamurugan, P.; Sunoj, R. B., A Unified Machine-Learning Protocol for Asymmetric Catalysis as a Proof of

---

Concept Demonstration Using Asymmetric Hydrogenation. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 1339.

[151] Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A., Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *J. Chem. Theory Comput.* **2017**, *13*, 5255-5264.

[152] Browning, N. J.; Ramakrishnan, R.; von Lilienfeld, O. A.; Roethlisberger, U., Genetic Optimization of Training Sets for Improved Machine Learning Models of Molecular Properties. *J. Phys. Chem. Lett.* **2017**, *8*, 1351-1359.

[153] Huang, B.; Lilienfeld, O. A. v., Communication: Understanding Molecular Representations in Machine Learning: The Role of Uniqueness and Target Similarity. *J. Chem. Phys.* **2016**, *145*, 161102.

[154] Rupp, M.; Ramakrishnan, R.; von Lilienfeld, O. A., Machine Learning for Quantum Mechanical Properties of Atoms in Molecules. *J. Phys. Chem. Lett.* **2015**, *6*, 3309-3313.

[155] Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A., Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326-2331.

[156] Faber, F.; Lindmaa, A.; von Lilienfeld, O. A.; Armiento, R., Crystal Structure Representations for Machine Learning Models of Formation Energies. *Int. J. Quantum Chem* **2015**, *115*, 1094-1101.

[157] Bereau, T.; Andrienko, D.; von Lilienfeld, O. A., Transferable Atomic Multipole Machine Learning Models for Small Organic Molecules. *J. Chem. Theory Comput.* **2015**, *11*, 3225-3233.

[158] Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A., Quantum Chemistry Structures and Properties of 134 Kilo Molecules. *Sci. Data* **2014**, *1*, 140022.

[159] Suzuki, A., Cross-Coupling Reactions Of Organoboranes: An Easy Way To Construct C-C Bonds (Nobel Lecture). *Angew. Chem. Int. Ed.* **2011**, *50*, 6722-6737.

[160] Weininger, D., SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31-36.

[161] Weininger, D.; Weininger, A.; Weininger, J. L., SMILES. 2. Algorithm for Generation of Unique SMILES Notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97-101.

[162] O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An Open Chemical Toolbox. *J. Cheminformatics* **2011**, *3*, 33.

[163] Pizzi, G.; Cepellotti, A.; Sabatini, R.; Marzari, N.; Kozinsky, B., AiiDA: Automated Interactive Infrastructure and Database for Computational Science. *Comput. Mater. Sci.* **2016**, *111*, 218-230.

[164] Becke, A. D., Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

[165] Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785-789.

- [166] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623-11627.
- [167] Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456-1465.
- [168] Ditchfield, R.; Hehre, W. J.; Pople, J. A., Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724-728.
- [169] Binkley, J. S.; Pople, J. A.; Hehre, W. J., Self-Consistent Molecular Orbital Methods. 21. Small Split-Valence Basis Sets for First-Row Elements. *J. Am. Chem. Soc.* **1980**, *102*, 939-947.
- [170] Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J., Self-Consistent Molecular-Orbital Methods. 22. Small Split-Valence Basis Sets for Second-Row Elements. *J. Am. Chem. Soc.* **1982**, *104*, 2797-2803.
- [171] Pietro, W. J.; Francl, M. M.; Hehre, W. J.; DeFrees, D. J.; Pople, J. A.; Binkley, J. S., Self-Consistent Molecular Orbital Methods. 24. Supplemented Small Split-Valence Basis Sets for Second-Row Elements. *J. Am. Chem. Soc.* **1982**, *104*, 5039-5048.
- [172] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09, Revision D.01. Wallingford CT, 2009.
- [173] Christensen, A. S. F., Felix A.; Huang, Bing; Bratholm, Lars Andersen; Tkatchenko, Alexandre; Müller, Klaus-Robert; von Lilienfeld, O. Anatole. *qmlcode/qml: Release v0.3.1*, 0.3.1; Zenodo: 2017.
- [174] Halgren, T. A., Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490-519.
- [175] Halgren, T. A., Merck Molecular Force Field. II. MMFF94 Van der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comput. Chem.* **1996**, *17*, 520-552.
- [176] Halgren, T. A., Merck Molecular Force Field. III. Molecular Geometries and Vibrational Frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553-586.
- [177] Halgren, T. A.; Nachbar, R. B., Merck Molecular Force Field. IV. Conformational Energies and Geometries for MMFF94. *J. Comput. Chem.* **1996**, *17*, 587-615.
- [178] Halgren, T. A., Merck molecular Force Field. V. Extension of MMFF94 using Experimental Data, Additional Computational Data, and Empirical Rules. *J. Comput. Chem.* **1996**, *17*, 616-641.

- 
- [179] Lei, P.; Meng, G.; Ling, Y.; An, J.; Szostak, M., Pd-PEPPSI: Pd-NHC Precatalyst for Suzuki–Miyaura Cross-Coupling Reactions of Amides. *J. Org. Chem.* **2017**, *82*, 6638–6646.
- [180] Martin, R.; Buchwald, S. L., Palladium-Catalyzed Suzuki–Miyaura Cross-Coupling Reactions Employing Dialkylbiaryl Phosphine Ligands. *Acc. Chem. Res.* **2008**, *41*, 1461–1473.
- [181] Surry, D. S.; Buchwald, S. L., Biaryl Phosphane Ligands in Palladium-Catalyzed Amination. *Angew. Chem. Int. Ed.* **2008**, *47*, 6338–6361.
- [182] Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M., Machine Learning Unifies the Modeling of Materials and Molecules. *Sci. Adv.* **2017**, *3*, e1701816.
- [183] Faber, F. A.; Christensen, A. S.; Huang, B.; von Lilienfeld, O. A., Alchemical and Structural Distribution based Representation for Universal Quantum Machine Learning. *J. Chem. Phys.* **2018**, *148*, 241717.
- [184] Huang, B. v. L., O. Anatole., Efficient Accurate Scalable and Transferable Quantum Machine Learning with Amons. *arXiv e-print* **2017**, 1707.04146.
- [185] Axilrod, B. M.; Teller, E., Interaction of the van der Waals Type Between Three Atoms. *J. Chem. Phys.* **1943**, *11*, 299–300.
- [186] Muto, Y., Force between Nonpolar Molecules. *J. Phys. Math. Soc. Jpn* **1943**, *17*, 629–631.
- [187] Hastie, T.; Tibshirani, R.; Friedman, J. H., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer: 2001.
- [188] Vapnik, V. N., *The Nature of Statistical Learning Theory*. Springer New York: 2013.
- [189] Cortes, C.; Jackel, L. D.; Chiang, W.-P. In *Limits on Learning Machine Accuracy Imposed by Data Quality*, Advances in Neural Information Processing Systems, 1995; pp 239–246.
- [190] De Meijere, A.; Bräse, S.; Oestreich, M., *Metal Catalyzed Cross-Coupling Reactions and More*. John Wiley & Sons: 2013.
- [191] Colacot, T. J.; Nolan, S.; Stradiotto, M.; Lautens, M.; Lloyd-Jones, G.; Wu, X.-F.; Koide, K.; Magano, J.; Tudge, M.; Maiti, D., *New Trends in Cross-Coupling: Theory and Applications*. Royal Society of Chemistry: 2014.
- [192] Nishihara, Y., *Applied Cross-Coupling Reactions*. Springer: 2013; Vol. 80.
- [193] Bedford, R. B.; Hazelwood, S. L.; Albisson, D. A., Platinum Catalysts for Suzuki Biaryl Coupling Reactions. *Organometallics* **2002**, *21*, 2599–2600.
- [194] Mateo, C.; Fernández-Rivas, C.; Cárdenas, D. J.; Echavarren, A. M., Intramolecular Transmetalation of Arylpalladium(II) and Arylplatinum(II) Complexes with Silanes and Stannanes. *Organometallics* **1998**, *17*, 3661–3669.
- [195] Ananikov, V. P.; Musaev, D. G.; Morokuma, K., Theoretical Insight into the C–C Coupling Reactions of the Vinyl, Phenyl, Ethynyl, and Methyl Complexes of Palladium and Platinum. *Organometallics* **2005**, *24*, 715–723.

- [196] Han, F.-S., Transition-Metal-Catalyzed Suzuki–Miyaura Cross-Coupling Reactions: A Remarkable Advance from Palladium to Nickel Catalysts. *Chem. Soc. Rev.* **2013**, *42*, 5270-5298.
- [197] Tasker, S. Z.; Standley, E. A.; Jamison, T. F., Recent Advances in Homogeneous Nickel Catalysis. *Nature* **2014**, *509*, 299-309.
- [198] Zhang, D.; Wang, Q., Palladium Catalyzed Asymmetric Suzuki–Miyaura Coupling Reactions to Axially Chiral Biaryl Compounds: Chiral Ligands and Recent Advances. *Coord. Chem. Rev.* **2015**, *286*, 1-16.
- [199] Livendahl, M.; Goehry, C.; Maseras, F.; Echavarren, A. M., Rationale for the Sluggish Oxidative Addition of Aryl Halides to Au(I). *Chem. Commun.* **2014**, *50*, 1533-1536.
- [200] Maaliki, C.; Thiery, E.; Thibonnet, J., Emergence of Copper-Mediated Formation of C–C Bonds. *Eur. J. Org. Chem.* **2017**, *2017*, 209-228.
- [201] Thapa, S.; Shrestha, B.; Gurung, S. K.; Giri, R., Copper-Catalysed Cross-Coupling: an Untapped Potential. *Org. Biomol. Chem.* **2015**, *13*, 4816-4827.
- [202] Rao, H. S. P.; Rao, A. V. B., Copper-Catalyzed C(sp<sup>3</sup>)–OH Cleavage with Concomitant C–C Coupling: Synthesis of 3-Substituted Isoindolinones. *J. Org. Chem.* **2015**, *80*, 1506-1516.
- [203] Hoshi, M.; Kawamura, N.; Shirakawa, K., Construction of Terminal Conjugated Enynes: Cu-Mediated Cross-Coupling Reaction of Alkenyldialkylborane with (Trimethylsilyl)ethynyl Bromide. *Synthesis* **2006**, *2006*, 1961-1970.
- [204] Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G., Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.
- [205] Behler, J.; Parrinello, M., Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- [206] Duan, C.; Janet, J. P.; Liu, F.; Nandy, A.; Kulik, H. J., Learning from Failure: Predicting Electronic Structure Calculation Outcomes with Machine Learning Models. *J. Chem. Theory Comput.* **2019**, *15*, 2331-2345.
- [207] Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A., Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8*, 13890.
- [208] Pyzer-Knapp, E. O.; Li, K.; Aspuru-Guzik, A., Learning from the Harvard Clean Energy Project: The Use of Neural Networks to Accelerate Materials Discovery. *Adv. Funct. Mater.* **2015**, *25*, 6495-6502.
- [209] Rothenberg, G., Data Mining in Catalysis: Separating Knowledge from Garbage. *Catal. Today* **2008**, *137*, 2-10.
- [210] Medford, A. J.; Kunz, M. R.; Ewing, S. M.; Borders, T.; Fushimi, R., Extracting Knowledge from Data through Catalysis Informatics. *ACS Catal.* **2018**, *8*, 7403-7429.
- [211] Goldsmith, B. R.; Esterhuizen, J.; Liu, J.-X.; Bartel, C. J.; Sutton, C., Machine Learning for Heterogeneous Catalyst Design and Discovery. *AIChE J.* **2018**, *64*, 2311-2323.
- [212] Yao, K.; Herr, J. E.; Brown, S. N.; Parkhill, J., Intrinsic Bond Energies from a Bonds-in-Molecules Neural Network. *J. Phys. Chem. Lett.* **2017**, *8*, 2689-2694.



- 
- [213] Fey, N.; Tsipis, A. C.; Harris, S. E.; Harvey, J. N.; Orpen, A. G.; Mansson, R. A., Development of a Ligand Knowledge Base, Part 1: Computational Descriptors for Phosphorus Donor Ligands. *Chem. Eur. J.* **2005**, *12*, 291-302.
- [214] Fey, N.; Orpen, A. G.; Harvey, J. N., Building Ligand Knowledge Bases for Organometallic Chemistry: Computational Description of Phosphorus(III)-Donor Ligands and the Metal–Phosphorus Bond. *Coord. Chem. Rev.* **2009**, *253*, 704-722.
- [215] Jover, J.; Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Orpen, A. G.; Owen-Smith, G. J. J.; Murray, P.; Hose, D. R. J.; Osborne, R.; Purdie, M., Expansion of the Ligand Knowledge Base for Monodentate P-Donor Ligands (LKB-P). *Organometallics* **2010**, *29*, 6245-6258.
- [216] Jover, J.; Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Orpen, A. G.; Owen-Smith, G. J. J.; Murray, P.; Hose, D. R. J.; Osborne, R.; Purdie, M., Expansion of the Ligand Knowledge Base for Chelating P,P-Donor Ligands (LKB-PP). *Organometallics* **2012**, *31*, 5302-5306.
- [217] Piou, T.; Romanov-Michailidis, F.; Ashley, M. A.; Romanova-Michaelides, M.; Rovis, T., Stereodivergent Rhodium(III)-Catalyzed cis-Cyclopropanation Enabled by Multivariate Optimization. *J. Am. Chem. Soc.* **2018**, *140*, 9587-9593.
- [218] Harper, K. C.; Bess, E. N.; Sigman, M. S., Multidimensional Steric Parameters in the Analysis of Asymmetric Catalytic Reactions. *Nat. Chem.* **2012**, *4*, 366.
- [219] Harper, K. C.; Sigman, M. S., Predicting and Optimizing Asymmetric Catalyst Performance using the Principles of Experimental Design and Steric Parameters. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2179.
- [220] Toyao, T.; Suzuki, K.; Kikuchi, S.; Takakusagi, S.; Shimizu, K.-i.; Takigawa, I., Toward Effective Utilization of Methane: Machine Learning Prediction of Adsorption Energies on Metal Alloys. *J. Phys. Chem. C* **2018**, *122*, 8315-8326.
- [221] Pilania, G.; Mannodi-Kanakkithodi, A.; Uberuaga, B. P.; Ramprasad, R.; Gubernatis, J. E.; Lookman, T., Machine Learning Bandgaps of Double Perovskites. *Sci. Rep.* **2016**, *6*, 19375.
- [222] Uhe, A.; Kozuch, S.; Shaik, S., Automatic Analysis of Computed Catalytic Cycles. *J. Comput. Chem.* **2010**, *32*, 978-985.
- [223] Kozuch, S.; Shaik, S., A Combined Kinetic–Quantum Mechanical Model for Assessment of Catalytic Cycles: Application to Cross-Coupling and Heck Reactions. *J. Am. Chem. Soc.* **2006**, *128*, 3355-3365.
- [224] Abild-Pedersen, F.; Greeley, J.; Studt, F.; Rossmeisl, J.; Munter, T. R.; Moses, P. G.; Skúlason, E.; Bligaard, T.; Nørskov, J. K., Scaling Properties of Adsorption Energies for Hydrogen-Containing Molecules on Transition-Metal Surfaces. *Phys. Rev. Lett.* **2007**, *99*, 016105.
- [225] Tribello, G. A.; Ceriotti, M.; Parrinello, M., Using Sketch-Map Coordinates to Analyze and Bias Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 5196.
- [226] Todeschini, R.; Consonni, V.; Mannhold, R.; Kubinyi, H.; Timmerman, H., *Handbook of Molecular Descriptors*. Wiley: 2008.
- [227] <http://sketchmap.org>

- [228] Bokeh Development Team, Bokeh: Python Library for Interactive Visualization. 2018, <http://www.bokeh.pydata.org>
- [229] SQLite: A C-Language Library that Implements a SQL Database Engine. <http://www.sqlite.org/index.html>
- [230] JSmol: An Open-Source HTML5 Viewer for Chemical Structures in 3D. <http://wiki.jmol.org/index.php/JSmol>
- [231] Negishi, E.-i., Magical Power of Transition Metals: Past, Present, and Future (Nobel Lecture). *Angew. Chem. Int. Ed.* **2011**, *50*, 6738-6764.
- [232] Johansson Seechurn, C. C. C.; Kitching, M. O.; Colacot, T. J.; Snieckus, V., Palladium-Catalyzed Cross-Coupling: A Historical Contextual Perspective to the 2010 Nobel Prize. *Angew. Chem. Int. Ed.* **2012**, *51*, 5062-5085.
- [233] Low, J. J.; Goddard, W. A., Theoretical Studies of Oxidative Addition and Reductive Elimination. 3. Carbon-Hydrogen and Carbon-Carbon Reductive Coupling from Palladium and Platinum Bis(phosphine) Complexes. *J. Am. Chem. Soc.* **1986**, *108*, 6115-6128.
- [234] Low, J. J.; Goddard, W. A., Theoretical Studies of Oxidative Addition and Reductive Elimination. 2. Reductive Coupling of Hydrogen-Hydrogen, Hydrogen-Carbon, and Carbon-Carbon Bonds from Palladium and Platinum Complexes. *Organometallics* **1986**, *5*, 609-622.
- [235] Bhunia, S.; Pawar, G. G.; Kumar, S. V.; Jiang, Y.; Ma, D., Selected Copper-Based Reactions for C–N, C–O, C–S, and C–C Bond Formation. *Angew. Chem. Int. Ed.* **2017**, *56*, 16136-16179.
- [236] Casitas, A.; King, A. E.; Parella, T.; Costas, M.; Stahl, S. S.; Ribas, X., Direct Observation of CuI/CuIII Redox Steps Relevant to Ullmann-Type Coupling Reactions. *Chem. Sci.* **2010**, *1*, 326-330.
- [237] Cohen, T.; Cristea, I., Kinetics and Mechanism of the Copper(I)-Induced Homogeneous Ullmann Coupling of o-Bromonitrobenzene. *J. Am. Chem. Soc.* **1976**, *98*, 748-753.
- [238] Tolman, C. A., Steric Effects of Phosphorus Ligands in Organometallic Chemistry and Homogeneous Catalysis. *Chem. Rev.* **1977**, *77*, 313-348.
- [239] Ben-David, Y.; Portnoy, M.; Gozin, M.; Milstein, D., Palladium-Catalyzed Vinylation of Aryl Chlorides. Chelate Effect in Catalysis. *Organometallics* **1992**, *11*, 1995-1996.
- [240] Nishiyama, M.; Yamamoto, T.; Koie, Y., Synthesis of N-Arylpiperazines from Aryl Halides and Piperazine under a Palladium Tri-tert-butylphosphine Catalyst. *Tetrahedron Lett.* **1998**, *39*, 617-620.
- [241] Shen, W., Palladium Catalyzed Coupling of Aryl Chlorides with Arylboronic Acids. *Tetrahedron Lett.* **1997**, *38*, 5575-5578.
- [242] Fu, G. C., The Development of Versatile Methods for Palladium-Catalyzed Coupling Reactions of Aryl Electrophiles through the Use of P(t-Bu)<sub>3</sub> and PCy<sub>3</sub> as Ligands. *Acc. Chem. Res.* **2008**, *41*, 1555-1564.
- [243] Marion, N.; Nolan, S. P., Well-Defined N-Heterocyclic Carbenes–Palladium(II) Precatalysts for Cross-Coupling Reactions. *Acc. Chem. Res.* **2008**, *41*, 1440-1449.

- 
- [244] Fleckenstein, C. A.; Plenio, H., Sterically Demanding Trialkylphosphines for Palladium-Catalyzed Cross Coupling Reactions—Alternatives to PtBu<sub>3</sub>. *Chem. Soc. Rev.* **2010**, *39*, 694-711.
- [245] Dröge, T.; Glorius, F., The Measure of All Rings—N-Heterocyclic Carbenes. *Angew. Chem. Int. Ed.* **2010**, *49*, 6940-6952.
- [246] Lundgren, R. J.; Stradiotto, M., Addressing Challenges in Palladium-Catalyzed Cross-Coupling Reactions Through Ligand Design. *Chem. Eur. J.* **2012**, *18*, 9758-9769.
- [247] Couzijn, E. P. A.; Lai, Y.-Y.; Limacher, A.; Chen, P., Intuitive Quantifiers of Charge Flows in Coordinate Bonding. *Organometallics* **2017**, *36*, 3205-3214.
- [248] McMullin, C. L.; Fey, N.; Harvey, J. N., Computed Ligand Effects on the Oxidative Addition of Phenyl Halides to Phosphine Supported Palladium(0) Catalysts. *Dalton Trans.* **2014**, *43*, 13545-13556.
- [249] Corbet, J.-P.; Mignani, G., Selected Patented Cross-Coupling Reaction Technologies. *Chem. Rev.* **2006**, *106*, 2651-2710.
- [250] Chianese, A. R.; Li, X.; Janzen, M. C.; Faller, J. W.; Crabtree, R. H., Rhodium and Iridium Complexes of N-Heterocyclic Carbenes via Transmetalation: Structure and Dynamics. *Organometallics* **2003**, *22*, 1663-1667.
- [251] Dorta, R.; Stevens, E. D.; Scott, N. M.; Costabile, C.; Cavallo, L.; Hoff, C. D.; Nolan, S. P., Steric and Electronic Properties of N-Heterocyclic Carbenes (NHC): A Detailed Study on Their Interaction with Ni(CO)<sub>4</sub>. *J. Am. Chem. Soc.* **2005**, *127*, 2485-2495.
- [252] Griffith, J. S.; Orgel, L. E., Ligand-Field Theory. *Quarterly Reviews, Chemical Society* **1957**, *11*, 381-393.
- [253] Griffith, J. S., *The Theory of Transition-Metal Ions*. Cambridge University Press: 1964.
- [254] Zhou, Y.; You, W.; Smith, K. B.; Brown, M. K., Copper-Catalyzed Cross-Coupling of Boronic Esters with Aryl Iodides and Application to the Carboboration of Alkynes and Allenes. *Angew. Chem. Int. Ed.* **2014**, *53*, 3475-3479.
- [255] Álvarez, A.; Bansode, A.; Urakawa, A.; Bavykina, A. V.; Wezendonk, T. A.; Makkee, M.; Gascon, J.; Kapteijn, F., Challenges in the Greener Production of Formates/Formic Acid, Methanol, and DME by Heterogeneously Catalyzed CO<sub>2</sub> Hydrogenation Processes. *Chem. Rev.* **2017**, *117*, 9804-9838.
- [256] Gunasekar, G. H.; Park, K.; Jung, K.-D.; Yoon, S., Recent Developments in the Catalytic Hydrogenation of CO<sub>2</sub> to Formic Acid/Formate using Heterogeneous Catalysts. *Inorg. Chem. Front.* **2016**, *3*, 882-895.
- [257] Upadhyay, P.; Srivastava, V., Carbon Sequestration: Hydrogenation of CO<sub>2</sub> to Formic Acid. In *Present Environment and Sustainable Development*, 2016; Vol. 10, p 13.
- [258] Ashley, A. E.; Thompson, A. L.; O'Hare, D., Non-Metal-Mediated Homogeneous Hydrogenation of CO<sub>2</sub> to CH<sub>3</sub>OH. *Angew. Chem. Int. Ed.* **2009**, *48*, 9839-9843.
- [259] Ménard, G.; Stephan, D. W., Room Temperature Reduction of CO<sub>2</sub> to Methanol by Al-Based Frustrated Lewis Pairs and Ammonia Borane. *J. Am. Chem. Soc.* **2010**, *132*, 1796-1797.

- [260] Berkefeld, A.; Piers, W. E.; Parvez, M., Tandem Frustrated Lewis Pair/Tris(pentafluorophenyl)borane-Catalyzed Deoxygenative Hydrosilylation of Carbon Dioxide. *J. Am. Chem. Soc.* **2010**, *132*, 10660-10661.
- [261] Wang, W.-H.; Himeda, Y.; Muckerman, J. T.; Manbeck, G. F.; Fujita, E., CO<sub>2</sub> Hydrogenation to Formate and Methanol as an Alternative to Photo- and Electrochemical CO<sub>2</sub> Reduction. *Chem. Rev.* **2015**, *115*, 12936-12973.
- [262] Johnson, W., Jr.; Heldreth, B.; Bergfeld, W. F.; Belsito, D. V.; Hill, R. A.; Klaassen, C. D.; Liebler, D. C.; Marks, J. G., Jr.; Shank, R. C.; Slaga, T. J.; Snyder, P. W.; Andersen, F. A., Safety Assessment of Formic Acid and Sodium Formate as Used in Cosmetics. *Int. J. Toxicol.* **2016**, *35*, 41S-54S.
- [263] He, T.; Pachfule, P.; Wu, H.; Xu, Q.; Chen, P., Hydrogen Carriers. *Nat. Rev. Mater.* **2016**, *1*, 16059.
- [264] Zhong, H.; Iguchi, M.; Chatterjee, M.; Himeda, Y.; Xu, Q.; Kawanami, H., Formic Acid-Based Liquid Organic Hydrogen Carrier System with Heterogeneous Catalysts. *Adv. Sustainable Syst.* **2018**, *2*, 1700161.
- [265] Formic Acid. In *Ullmann's Encyclopedia of Industrial Chemistry*.
- [266] Sakakura, T.; Choi, J.-C.; Yasuda, H., Transformation of Carbon Dioxide. *Chem. Rev.* **2007**, *107*, 2365-2387.
- [267] Sordakis, K.; Tang, C.; Vogt, L. K.; Junge, H.; Dyson, P. J.; Beller, M.; Laurenczy, G., Homogeneous Catalysis for Sustainable Hydrogen Storage in Formic Acid and Alcohols. *Chem. Rev.* **2018**, *118*, 372-433.
- [268] Ohnishi, Y.-y.; Nakao, Y.; Sato, H.; Sakaki, S., Ruthenium(II)-Catalyzed Hydrogenation of Carbon Dioxide to Formic Acid. Theoretical Study of Significant Acceleration by Water Molecules. *Organometallics* **2006**, *25*, 3352-3363.
- [269] Rohmann, K.; Kothe, J.; Haenel, M. W.; Englert, U.; Hölscher, M.; Leitner, W., Hydrogenation of CO(2) to Formic Acid with a Highly Active Ruthenium Acridophos Complex in DMSO and DMSO/Water. *Angew. Chem., Int. Ed. Engl.* **2016**, *55*, 8966-8969.
- [270] Ge, H.; Chen, X.; Yang, X., Hydrogenation of Carbon Dioxide to Methanol Catalyzed by Iron, Cobalt, and Manganese Cyclopentadienone Complexes: Mechanistic Insights and Computational Design. *Chem. Eur. J.* **2017**, *23*, 8850-8856.
- [271] Schneidewind, J.; Adam, R.; Baumann, W.; Jackstell, R.; Beller, M., Low-Temperature Hydrogenation of Carbon Dioxide to Methanol with a Homogeneous Cobalt Catalyst. *Angew. Chem. Int. Ed.* **2017**, *56*, 1890-1893.
- [272] Chakraborty, S.; Lagaditis, P. O.; Förster, M.; Bielinski, E. A.; Hazari, N.; Holthausen, M. C.; Jones, W. D.; Schneider, S., Well-Defined Iron Catalysts for the Acceptorless Reversible Dehydrogenation-Hydrogenation of Alcohols and Ketones. *ACS Catal.* **2014**, *4*, 3994-4003.
- [273] Bernskoetter, W. H.; Hazari, N., Reversible Hydrogenation of Carbon Dioxide to Formic Acid and Methanol: Lewis Acid Enhancement of Base Metal Catalysts. *Acc. Chem. Res.* **2017**, *50*, 1049-1058.

- 
- [274] Zeng, G.; Maeda, S.; Taketsugu, T.; Sakaki, S., Catalytic Hydrogenation of Carbon Dioxide with Ammonia–Borane by Pincer-Type Phosphorus Compounds: Theoretical Prediction. *J. Am. Chem. Soc.* **2016**, *138*, 13481-13484.
- [275] Werkmeister, S.; Neumann, J.; Junge, K.; Beller, M., Pincer-Type Complexes for Catalytic (De)Hydrogenation and Transfer (De)Hydrogenation Reactions: Recent Progress. *Chem. Eur. J.* **2015**, *21*, 12226-12250.
- [276] Arduengo, A. J.; Stewart, C. A.; Davidson, F.; Dixon, D. A.; Becker, J. Y.; Culley, S. A.; Mizen, M. B., The Synthesis, Structure, and Chemistry of 10-Pn-3 Systems: Tricoordinate Hypervalent Pnictogen Compounds. *J. Am. Chem. Soc.* **1987**, *109*, 627-647.
- [277] Culley, S. A.; Arduengo, A. J., Synthesis and Structure of the First 10-P-3 Species. *J. Am. Chem. Soc.* **1984**, *106*, 1164-1165.
- [278] Lawrence, M. A. W.; Green, K.-A.; Nelson, P. N.; Lorraine, S. C., Review: Pincer Ligands—Tunable, Versatile and Applicable. *Polyhedron* **2018**, *143*, 11-27.
- [279] Tanaka, R.; Yamashita, M.; Nozaki, K., Catalytic Hydrogenation of Carbon Dioxide Using Ir(III)–Pincer Complexes. *J. Am. Chem. Soc.* **2009**, *131*, 14168-14169.
- [280] Osadchuk, I.; Tamm, T.; Ahlquist, M. S. G., Theoretical Investigation of a Parallel Catalytic Cycle in CO<sub>2</sub> Hydrogenation by (PNP)IrH<sub>3</sub>. *Organometallics* **2015**, *34*, 4932-4940.
- [281] Filonenko, G. A.; van Putten, R.; Schulpen, E. N.; Hensen, E. J. M.; Pidko, E. A., Highly Efficient Reversible Hydrogenation of Carbon Dioxide to Formates Using a Ruthenium PNP-Pincer Catalyst. *ChemCatChem* **2014**, *6*, 1526-1530.
- [282] Filonenko, G. A.; Conley, M. P.; Copéret, C.; Lutz, M.; Hensen, E. J. M.; Pidko, E. A., The Impact of Metal–Ligand Cooperation in Hydrogenation of Carbon Dioxide Catalyzed by Ruthenium PNP Pincer. *ACS Catal.* **2013**, *3*, 2522-2526.
- [283] Vogt, M.; Gargir, M.; Iron, M. A.; Diskin-Posner, Y.; Ben-David, Y.; Milstein, D., A New Mode of Activation of CO<sub>2</sub> by Metal–Ligand Cooperation with Reversible C–C and M–O Bond Formation at Ambient Temperature. *Chem. Eur. J.* **2012**, *18*, 9194-9197.
- [284] Schmeier, T. J.; Dobereiner, G. E.; Crabtree, R. H.; Hazari, N., Secondary Coordination Sphere Interactions Facilitate the Insertion Step in an Iridium(III) CO<sub>2</sub> Reduction Catalyst. *J. Am. Chem. Soc.* **2011**, *133*, 9274-9277.
- [285] Langer, R.; Diskin-Posner, Y.; Leitus, G.; Shimon, L. J. W.; Ben-David, Y.; Milstein, D., Low-Pressure Hydrogenation of Carbon Dioxide Catalyzed by an Iron Pincer Complex Exhibiting Noble Metal Activity. *Angew. Chem. Int. Ed.* **2011**, *50*, 9948-9952.
- [286] Huff, C. A.; Sanford, M. S., Catalytic CO<sub>2</sub> Hydrogenation to Formate by a Ruthenium Pincer Complex. *ACS Catal.* **2013**, *3*, 2412-2416.
- [287] Rawat, K. S.; Pathak, B., Aliphatic Mn-PNP Complexes for the CO<sub>2</sub> Hydrogenation Reaction: a Base Free Mechanism. *Catal. Sci. Technol.* **2017**, *7*, 3234-3242.
- [288] Filonenko, G. A.; Cosimi, E.; Lefort, L.; Conley, M. P.; Copéret, C.; Lutz, M.; Hensen, E. J. M.; Pidko, E. A., Lutidine-Derived Ru-CNC Hydrogenation Pincer Catalysts with Versatile Coordination Properties. *ACS Catal.* **2014**, *4*, 2667-2671.

- [289] Bernskoetter, W. H.; Hazari, N., A Computational Investigation of the Insertion of Carbon Dioxide into Four- and Five-Coordinate Iridium Hydrides. *Eur. J. Inorg. Chem.* **2013**, *2013*, 4032-4041.
- [290] Yang, X., Hydrogenation of Carbon Dioxide Catalyzed by PNP Pincer Iridium, Iron, and Cobalt Complexes: A Computational Design of Base Metal Catalysts. *ACS Catal.* **2011**, *1*, 849-854.
- [291] Ahlquist, M. S. G., Iridium Catalyzed Hydrogenation of CO<sub>2</sub> under Basic Conditions—Mechanistic Insight from Theory. *J. Mol. Catal. A: Chem.* **2010**, *324*, 3-8.
- [292] Praveen, C. S.; Comas-Vives, A.; Copéret, C.; VandeVondele, J., Role of Water, CO<sub>2</sub>, and Noninnocent Ligands in the CO<sub>2</sub> Hydrogenation to Formate by an Ir(III) PNP Pincer Catalyst Evaluated by Static-DFT and ab Initio Molecular Dynamics under Reaction Conditions. *Organometallics* **2017**, *36*, 4908-4919.
- [293] Tanaka, R.; Yamashita, M.; Chung, L. W.; Morokuma, K.; Nozaki, K., Mechanistic Studies on the Reversible Hydrogenation of Carbon Dioxide Catalyzed by an Ir-PNP Complex. *Organometallics* **2011**, *30*, 6742-6750.
- [294] Li, J.; Yoshizawa, K., Catalytic Hydrogenation of Carbon Dioxide with a Highly Active Hydride on Ir(III)–Pincer Complex: Mechanism for CO<sub>2</sub> Insertion and Nature of Metal–Hydride Bond. *Bull. Chem. Soc. Jpn.* **2011**, *84*, 1039-1048.
- [295] Seh, Z. W.; Kibsgaard, J.; Dickens, C. F.; Chorkendorff, I.; Nørskov, J. K.; Jaramillo, T. F., Combining Theory and Experiment in Electrocatalysis: Insights into Materials Design. *Science* **2017**, 355.
- [296] Medford, A. J.; Vojvodic, A.; Hummelshøj, J. S.; Voss, J.; Abild-Pedersen, F.; Studt, F.; Bligaard, T.; Nilsson, A.; Nørskov, J. K., From the Sabatier Principle to a Predictive Theory of Transition-Metal Heterogeneous Catalysis. *J. Catal.* **2015**, *328*, 36-42.
- [297] Kozuch, S.; Shaik, S., Kinetic-Quantum Chemical Model for Catalytic Cycles: The Haber–Bosch Process and the Effect of Reagent Concentration. *J. Phys. Chem. A* **2008**, *112*, 6032-6041.
- [298] Zaffran, J.; Michel, C.; Delbecq, F.; Sautet, P., Trade-Off between Accuracy and Universality in Linear Energy Relations for Alcohol Dehydrogenation on Transition Metals. *J. Phys. Chem. C* **2015**, *119*, 12988-12998.
- [299] Sutton, J. E.; Vlachos, D. G., A Theoretical and Computational Analysis of Linear Free Energy Relations for the Estimation of Activation Energies. *ACS Catal.* **2012**, *2*, 1624-1634.
- [300] Santen, R. A. v.; Neurock, M.; Shetty, S. G., Reactivity Theory of Transition-Metal Surfaces: A Brønsted–Evans–Polanyi Linear Activation Energy–Free-Energy Analysis. *Chem. Rev.* **2010**, *110*, 2005-2048.
- [301] Michaelides, A.; Liu, Z. P.; Zhang, C. J.; Alavi, A.; King, D. A.; Hu, P., Identification of General Linear Relationships between Activation Energies and Enthalpy Changes for Dissociation Reactions at Surfaces. *J. Am. Chem. Soc.* **2003**, *125*, 3704-3705.
- [302] Zhao, Y.; Truhlar, D. G., The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

- 
- [303] Zhao, Y.; Truhlar, D. G., Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res.* **2008**, *41*, 157-167.
- [304] Wheeler, S. E.; Houk, K. N., Integration Grid Errors for Meta-GGA-Predicted Reaction Energies: Origin of Grid Errors for the M06 Suite of Functionals. *J. Chem. Theory Comput.* **2010**, *6*, 395-404.
- [305] Steinmann, S. N.; Corminboeuf, C., A System-Dependent Density-Based Dispersion Correction. *J. Chem. Theory Comput.* **2010**, *6*, 1990-2001.
- [306] Steinmann, S. N.; Corminboeuf, C., A Density Dependent Dispersion Correction. *Chimia* **2011**, *65*, 240-244.
- [307] Steinmann, S. N.; Corminboeuf, C., A Generalized-Gradient Approximation Exchange Hole Model for Dispersion Coefficients. *J. Chem. Phys.* **2011**, *134*, 044117.
- [308] Steinmann, S. N.; Corminboeuf, C., Comprehensive Benchmarking of a Density-Dependent Dispersion Correction. *J. Chem. Theory Comput.* **2011**, *7*, 3567-3577.
- [309] te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T., Chemistry with ADF. *J. Comput. Chem.* **2001**, *22*, 931-967.
- [310] Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J., Towards an Order-N DFT Method. *Theor. Chem. Acc.* **1998**, *99*, 391-403.
- [311] Grimme, S., Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem. Eur. J.* **2012**, *18*, 9955-9964.
- [312] Mammen, M.; Shakhnovich, E. I.; Deutch, J. M.; Whitesides, G. M., Estimating the Entropic Cost of Self-Assembly of Multiparticle Hydrogen-Bonded Aggregates Based on the Cyanuric Acid-Melamine Lattice. *J. Org. Chem.* **1998**, *63*, 3821-3830.
- [313] Funes-Ardoiz, I.; Paton, R. S., GoodVibes: version 2.0.3. **2018**, <https://doi.org/10.5281/zenodo.1435820>.
- [314] Klamt, A., The COSMO and COSMO-RS Solvation Models. *WIREs: Comput. Mol. Sci.* **2011**, *1*, 699-709.
- [315] Heimann, J. E.; Bernskoetter, W. H.; Hazari, N.; Mayer, James M., Acceleration of CO<sub>2</sub> Insertion into Metal Hydrides: Ligand, Lewis Acid, and Solvent Effects on Reaction Kinetics. *Chem. Sci.* **2018**, *9*, 6629-6638.
- [316] Carroll, M. P.; Guiry, P. J., P,N Ligands in Asymmetric Catalysis. *Chem. Soc. Rev.* **2014**, *43*, 819-833.
- [317] Mitoraj, M. P.; Michalak, A.,  $\sigma$ -Donor and  $\pi$ -Acceptor Properties of Phosphorus Ligands: An Insight from the Natural Orbitals for Chemical Valence. *Inorg. Chem.* **2010**, *49*, 578-582.
- [318] Leis, W.; Mayer, H. A.; Kaska, W. C., Cycloheptatrienyl, Alkyl and Aryl PCP-Pincer Complexes: Ligand Backbone Effects and Metal Reactivity. *Coord. Chem. Rev.* **2008**, *252*, 1787-1797.
- [319] Besora, M.; Maseras, F., Microkinetic Modeling in Homogeneous Catalysis. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1372.

- [320] Amatore, C.; Jutand, A., Mechanistic and Kinetic Studies of Palladium Catalytic Systems. *J. Organomet. Chem.* **1999**, 576, 254-278.
- [321] Christiansen, J. A., The Elucidation of Reaction Mechanisms by the Method of Intermediates in Quasi-Stationary Concentrations. *Advances in Catalysis* **1953**, 5, 311-353.
- [322] Eyring, H., The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, 3, 107-115.
- [323] Siegbahn, P. E. M.; Tye, J. W.; Hall, M. B., Computational Studies of [NiFe] and [FeFe] Hydrogenases. *Chem. Rev.* **2007**, 107, 4414-4435.
- [324] Northrop, D. B., Minimal Kinetic Mechanism and General Equation for Deuterium Isotope Effects on Enzymic Reactions: Uncertainty of Detecting a Rate-Limiting Step. *Biochemistry* **1981**, 20, 4056-4061.
- [325] Curtin, D. Y., Stereochemical Control of Organic Reactions Differences in Behaviour of Diastereoisomers. *Rec. Chem. Prog* **1954**, 15, 110-128.
- [326] Avanesian, T.; Christopher, P., Scaled Degree of Rate Control: Identifying Elementary Steps That Control Differences in Performance of Transition-Metal Catalysts. *ACS Catal.* **2016**, 6, 5268-5272.
- [327] Stegelmann, C.; Andreasen, A.; Campbell, C. T., Degree of Rate Control: How Much the Energies of Intermediates and Transition States Control Rates. *J. Am. Chem. Soc.* **2009**, 131, 8077-8082.
- [328] Fey, N.; Tsipis, A. C.; Harris, S. E.; Harvey, J. N.; Orpen, A. G.; Mansson, R. A., Development of a Ligand Knowledge Base, Part 1: Computational Descriptors for Phosphorus Donor Ligands. *Chem. Eur. J.* **2006**, 12, 291-302.
- [329] Rooks, B. J.; Haas, M. R.; Sepúlveda, D.; Lu, T.; Wheeler, S. E., Prospects for the Computational Design of Bipyridine N,N'-Dioxide Catalysts for Asymmetric Propargylation Reactions. *ACS Catal.* **2015**, 5, 272-280.
- [330] Doney, A. C.; Rooks, B. J.; Lu, T.; Wheeler, S. E., Design of Organocatalysts for Asymmetric Propargylations through Computational Screening. *ACS Catal.* **2016**, 6, 7948-7955.
- [331] Guan, Y.; Wheeler, S. E., Automated Quantum Mechanical Predictions of Enantioselectivity in a Rhodium-Catalyzed Asymmetric Hydrogenation. *Angew. Chem. Int. Ed.* **2017**, 56, 9101-9105.
- [332] Guan, Y.; Ingman, V. M.; Rooks, B. J.; Wheeler, S. E., AARON: An Automated Reaction Optimizer for New Catalysts. *J. Chem. Theory Comput.* **2018**, 14, 5249-5261.
- [333] Habershon, S., Sampling Reactive Pathways with Random Walks in Chemical Space: Applications to Molecular Dissociation and Catalysis. *J. Chem. Phys.* **2015**, 143, 094106.
- [334] Habershon, S., Automated Prediction of Catalytic Mechanism and Rate Law Using Graph-Based Reaction Path Sampling. *J. Chem. Theory Comput.* **2016**, 12, 1786-1798.
- [335] Martínez-Núñez, E., An Automated Method to Find Transition States using Chemical Dynamics Simulations. *J. Comput. Chem.* **2015**, 36, 222-234.



- 
- [336] Varela, J. A.; Vázquez, S. A.; Martínez-Núñez, E., An Automated Method to Find Reaction Mechanisms and Solve the Kinetics in Organometallic Catalysis. *Chem. Sci.* **2017**, *8*, 3843-3851.
- [337] Maeda, S.; Morokuma, K., Communications: A Systematic Method for Locating Transition Structures of A+B→X Type Reactions. *J. Chem. Phys.* **2010**, *132*, 241102.
- [338] Maeda, S.; Ohno, K.; Morokuma, K., Systematic Exploration of the Mechanism of Chemical Reactions: the Global Reaction Route Mapping (GRRM) Strategy using the ADDF and AFIR Methods. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3683-3701.
- [339] Maeda, S.; Taketsugu, T.; Morokuma, K., Exploring Transition State Structures for Intramolecular Pathways by the Artificial Force Induced Reaction Method. *J. Comput. Chem.* **2014**, *35*, 166-173.
- [340] Yoshimura, T.; Maeda, S.; Taketsugu, T.; Sawamura, M.; Morokuma, K.; Mori, S., Exploring the Full Catalytic Cycle of Rhodium(I)-BINAP-Catalysed Isomerisation of Allylic Amines: a Graph Theory Approach for Path Optimisation. *Chem. Sci.* **2017**, *8*, 4475-4488.
- [341] Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G., Predicting Reaction Performance in C–N Cross-Coupling using Machine Learning. *Science* **2018**, *360*, 186.
- [342] Takaoka, S.; Eizawa, A.; Kusumoto, S.; Nakajima, K.; Nishibayashi, Y.; Nozaki, K., Hydrogenation of Carbon Dioxide with Organic Base by PCIP-Ir Catalysts. *Organometallics* **2018**, *37*, 3001-3009.
- [343] Aoki, W.; Wattanavinin, N.; Kusumoto, S.; Nozaki, K., Development of Highly Active Ir–PNP Catalysts for Hydrogenation of Carbon Dioxide with Organic Bases. *Bull. Chem. Soc. Jpn.* **2016**, *89*, 113-124.
- [344] Kozuch, S.; Martin, J. M. L., “Turning Over” Definitions in Catalytic Cycles. *ACS Catal.* **2012**, *2*, 2787-2794.
- [345] Hirshfeld, F. L., Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theor. Chim. Acta* **1977**, *44*, 129-138.
- [346] Falivene, L.; Cao, Z.; Petta, A.; Serra, L.; Poater, A.; Oliva, R.; Scarano, V.; Cavallo, L., Towards the Online Computer-Aided Design of Catalytic Pockets. *Nat. Chem.* **2019**, *11*, 872-879.



# Curriculum Vitae

## Boodsarin Sawatlon



*EPFL SB ISIC LCMD*

*CH-1015 Lausanne*

*Boodsarin.sawatlon@epfl.ch*

*Date of birth: July 3, 1991, Nationality: Thai*

*Languages: Thai (mother tongue), English (fluent), French (basic)*

---

### Education

- 2017–2021     **Ph.D. student in Chemistry**, EPFL, Lausanne, Switzerland  
Doctoral Thesis: Molecular Volcano Plots: Tools for Rationalizing and Predicting Catalytic Performance  
Advisor: Prof. Clemence Corminboeuf
- 2013–2016     **M.Sc. in Chemistry**, Mahidol University, Bangkok, Thailand  
Master Thesis: Mechanistic Study of Dehydrogenation and Hydrogenation of N-Heterocycles using PNP-Pincer-Supported Iron-Catalyst: a Density Functional Study  
Advisor: Assoc. Prof. Panida Surawatanawong
- 2009–2012     **B.Sc. in Chemistry (1<sup>st</sup> class honors)**, Mahidol University, Bangkok, Thailand  
Bachelor Thesis: Mechanistic Study of Nickel-Catalyzed C–O Bond Activation in Hydrogenolysis by Density Functional Theory  
Advisor: Assoc. Prof. Panida Surawatanawong

---

### Awards & Grants

- 2019             SCNAT/SCS Chemistry travel awards from the Swiss Academy of Sciences and the Swiss Chemical Society
- 2009–2015     Science Achievement Scholarship of Thailand (SAST) by Office of the Higher Education Commission, Ministry of Education, Thailand

---

### Contributed Talks

- 2020      **Effect of Substrates on Suzuki Reaction: a Molecular Volcano Plot Investigation** in the Pure and Applied Chemistry International Conference (PACCON 2020), Bangkok, Thailand
- 2018      **Machine Learning Models for Homogeneous Catalysis** in Challenges & Opportunities in Energy Research (NRG2018) winter school, Crans-Montana, Switzerland
- 2017      **Machine Learning Models for Homogeneous Catalysis** in National Center of Competence in Research (NCCR) MARVEL – Agility Plus symposium, Lausanne, Switzerland

---

### Poster Presentations

- 2019      **Data Mining the C–C Cross-Coupling** in the Swiss Chemical Society Fall Meeting, Lausanne, Switzerland
- 2019      **Data Mining the C–C Cross-Coupling** in the 2019 Gordon Research Conference (GRC) on physical organic chemistry: the intersection of organic structure, reactivity and mechanism, Holderness, New Hampshire, United states
- 2018      **Machine Learning Models for Homogeneous Catalysis** in the International Workshop on Computational Design and Discovery of Novel Materials (COMDI) 2018, Lausanne, Switzerland
- 2018      **Unravelling Metal–Ligand Combination Effects for the Hydrogenation of Carbon Dioxide using Molecular Volcano Plots** in the Swiss Chemical Society Fall Meeting, Lausanne, Switzerland
- 2018      **Machine Learning Models for Homogeneous Catalysis** in the Swiss Chemical Society Fall Meeting, Lausanne, Switzerland
- 2018      **Machine Learning Models for Homogeneous Catalysis** in MARVEL site visit 2018, Lausanne, Switzerland
- 2017      **Unravelling Metal–Ligand Combination Effects for the Hydrogenation of Carbon Dioxide using Molecular Volcano Plots** in the 11th Triennial Congress of the World Association of Theoretical and Computational Chemists WATOC 2017, Munich, Germany
- 2017      **Machine Learning Models for Homogeneous Catalysis** in MARVEL site visit 2017, Lausanne, Switzerland
- 2015      **Mechanistic Study of Dehydrogenation and Hydrogenation of N-heterocycles using PNP-Pincer-Supported Iron-Catalyst: a Density Functional Study** in the 15th International Congress of Quantum Chemistry (ICQC 2015), Beijing, China

- 
- 2015      **Mechanistic Study of Dehydrogenation and Hydrogenation of N-heterocycles using PNP-Pincer-Supported Iron-Catalyst: a Density Functional Study** in the Pure and Applied Chemistry International Conference (PACCON 2015), Bangkok, Thailand
- 2013      **Mechanistic Study of Nickel-Catalyzed C–O Bond Activation in Hydrogenolysis by Density Functional Theory** in the 8th Conference on Science and Technology for Youths, Bangkok, Thailand

---

## Publications

1. Wodrich M. D., Sawatlon B., Busch M., Corminboeuf C., The Genesis of Molecular Volcano Plots. *Acc. Chem. Res.*, **2021**, 54, 1107–1117.
2. Sawatlon B., Wodrich M. D., Corminboeuf C., Probing Substrate Scope with Molecular Volcano Plots. *Org. Lett.* **2020**, 22, 20, 7936–7941.
3. Naweephattana P., Sawatlon B., Surawatanawong P., Insights into the Regioselectivity of Hydroheteroarylation of Allylbenzene with Pyridine Catalyzed by Ni/AlMe<sub>3</sub> with N-Heterocyclic Carbene: the Concerted Hydrogen Transfer Mechanism. *J. Org. Chem.* **2020**, 85, 11340–11349.
4. Cordova M., Wodrich M. D., Meyer B., Sawatlon B., Corminboeuf C., Data-Driven Advancement of Homogeneous Nickel Catalyst Activity for Aryl-Ether Cleavage. *ACS Catal.* **2020**, 10, 7021–7031.
5. Wodrich M. D., Sawatlon B., Solel E., Kozuch S., Corminboeuf C., Activity-Based Screening of Homogeneous Catalysts Through the Rapid Assessment of Theoretically Derived Turnover Frequencies. *ACS Catal.* **2019**, 9, 5716–5725.
6. Sawatlon B., Wodrich M. D., Meyer B., Fabrizio A., Corminboeuf C., Data Mining the C–C Cross-Coupling Genome. *ChemCatChem*, **2019**, 11, 4096–4107.
7. Sawatlon B., Wodrich M. D., Corminboeuf C., Unraveling Metal/Pincer Ligand Effects in the Catalytic Hydrogenation of Carbon Dioxide to Formate. *Organometallics*, **2018**, 37, 4568–4575.
8. Meyer B., Sawatlon B., Heinen S. N., von Lilienfeld A., Corminboeuf C., Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts. *Chem. Sci.*, **2018**, 9, 7069–7077.
9. Wodrich M. D., Sawatlon B., Busch M., Corminboeuf C., On the Generality of Molecular Volcano Plots. *ChemCatChem*, **2018**, 10, 1586–1591.
10. Sawatlon B., Surawatanawong P., Mechanisms for Dehydrogenation and Hydrogenation of N-heterocycles using PNP-Pincer-Supported Iron Catalysts: a Density Functional Study. *Dalton Trans.*, **2016**, 45, 14965–14978.

11. Sawatlon B., Wititsuwannakul T., Tantirungrotechai Y., Surawatanawong P., Mechanisms of Ni N-heterocyclic Carbene Catalysts for C-O Bond Hydrogenolysis of Diphenyl Ether: a Density Functional Study. *Dalton Trans.*, **2014**, 43, 18123–18133.