DISSERTATION

# COMBINING MECHANISTIC AND STATISTICAL MODELS FOR PREDICTING REACTION OUTCOMES IN ORGANIC SYNTHESIS

Submitted by:

Liliana Cabrera Gallegos

Department of Chemistry

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2023

Doctoral Committee

Advisor: Robert S. Paton

Andrew McNally
Anthony Rappé
Ann Hess

ABSTRACT

COMBINING MECHANISTIC AND STATISTICAL MODELS FOR PREDICTING REACTION OUTCOMES IN ORGANIC SYNTHESIS

Computational modeling and machine learning tools have assisted in the fundamental challenge of predicting the "over-the-arrow" optimal reaction conditions to maximize the output (e.g., yield and selectivity). The work presented here explores multiple challenging synthetic reactions for reaction optimization ranging from: (i) precise photocatalytic transformations in chemical biology, (ii) new reactivity using organobismuth(V) reagents, (iii) challenging reversible nucleophilic alcohol addition reactions influence at equilibrium, and (iv) a late-stage key reaction step in a total synthesis project. Overall, this dissertation aims to assist in predicting optimal reaction outcomes by understanding and formulating reaction mechanisms from quantum mechanics and statistical methods while using open-source automated workflows to improve transparency and reproducibility within data-chemistry fields.

Chapter 1 provides the necessary background to introduce the methods behind computational and statistical models that assist in addressing the challenges faced within the optimization process and the limitations of each strategy. First, there will be a brief overview of the computational protocols to generate and understand reaction mechanisms using quantum mechanical methods. Then, a summary of the data-driven approach introduces the statistical methods and metrics that build relationships to chemical reactivity using computer-readable mechanistically derived molecular descriptions.

Chapter 2 tackles the challenge of studying the chemical reactivity in large biological systems (e.g., peptides and proteins) with quantum mechanical methods. First, the precise photocatalytic functionalization at selenocysteine reaction developed by the Payne lab is simulated using a simplified model substrate followed by a more realistic model that generates the final energy profile. Based on the resulting computational analysis, the utility of this late-stage functionalization reaction is later demonstrated on large polypeptide chains.

Chapters 3 and 4 embark on a journey into new bismuth chemistry developed by the Ball group. The bismuth arylation reaction published in *Nature* transformed the following collaborative work discussed here, ranging from the computational protocols implemented in selectivity problems to the versatile chemical reactivity originating from bismuth(V) reagents. From the previously reported but otherwise unexplored DFT integration grid effects, the computed free energies on organobismuth reactions explored here would have led to significant errors and incorrectly predicting selectivities. With the optimal computational protocols, new reactivity using organobismuth reagents is investigated in Chapter 3 to propose a reaction mechanism for the selective arylation of 2- and 4-pyridiones. Chapter 4 describes the mechanistic investigation of the developed palladium-catalyzed cross-coupling reaction to achieve challenging C-C couplings in mild reaction conditions with the amino-bridged bismacycle reagent. A statistical modeling approach using automated workflows discussed in Chapter 7 is applied here to predict an optimal reaction design and capture the origin of the reactivity for various coupling substrates and modified organobismuth(V)-reagents for the developed Bisma-Stille cross-coupling reaction.

Chapter 5 describes a mechanistic investigation to optimize a challenging key reaction in the total synthesis of the natural product of allopupukeanane developed by the Sarpong group. The reaction success in late-stage synthetic plans becomes detrimental as the availability of reactants

in a multiple-step natural product synthesis becomes limiting. The elementary step influencing the reactivity is identified in the palladium-mediated cascade reaction. Then, a data-driven approach is implemented to screen various ligands and collect mechanistically derived molecular DFT features to incorporate into a Bayesian optimization tool developed by the Doyle lab. Automated workflows discussed in Chapter 7 were utilized to collect the features. This approach successfully identified more suitable and efficient reaction conditions for racemic mixture, byproduct formation, and catalyst decomposition challenges. The overall synthesis plan to access multiple natural products *via* the bridged bicycle scaffold highlighted in this chapter is an ongoing project by the Sarpong group.

Chapter 6 pivots into data-driven approaches to formulate statistical relationships sampled over small and large datasets. First, the collaborative research in section 6.2 dives into building a multivariate linear regression model with a small dataset to explain the reaction performance in various solvents on the challenging reversible nucleophilic alcohol addition reaction developed by the Bandar group. The statistical conclusions provide the bases for modeling the solvent effects *via* DFT methods. Next, in section 6.3, a machine learning model is trained on a large diverse molecule dataset to predict NMR chemical shifts with high accuracy to DFT-derived NMR values at only a fraction of the cost of DFT methods. Here are two examples where a successful prediction is evaluated based on the research goal to obtain model accuracy or interpretability.

Chapter 7 focuses on facilitating the transparency and reproducibility for collecting and generating meaningful statistical models for the data chemist in low- and high-throughput studies. The open-source, automated workflows, DISCO and REGGAE, allowed for the execution of projects mentioned in Chapters 4 to 6 at different stages of the research process (e.g., chemical data collection, feature selection, and then statistical modeling).

# ACKNOWLEDGMENTS

As I approach the end of this remarkable and transformative journey, my heart is brimming with gratitude. The support I have received along the way has been instrumental in reaching this significant milestone. With immense appreciation, I take this moment to express my deepest gratitude to those who have played a pivotal role in shaping and nurturing my Ph.D. journey.

First, I would like to sincerely thank my advisor, Prof. Robert S. Paton. Your exceptional guidance, patience, and continuous encouragement have given me the courage to believe in myself, especially when applying for awards. Your insightful feedback has been instrumental in shaping my research ideas and enhancing the quality of my work. I especially want to thank my first Ph.D. advisor, Prof. Jeremy May. I am genuinely grateful to you for reminding me of my love for organic chemistry and for your mentorship and experience under your supervision. I would not have continued my Ph.D. program if it weren't for your support.

Second, I would like to thank my thesis committee: Professors Andy McNally, Anthony Rappé, and Ann Hess. Additionally, Dr. Ben Sharp for sitting in on my candidacy exam and guiding me through the statistical course. I am thankful to all for your time, effort, and commitment to evaluating my work and providing valuable suggestions and recommendations.

Third, I would like to thank my current and former colleagues in the Paton and Kim group. The bond we have forged within our group is a treasure I will forever hold dear. The memories we have created, both within and outside the lab, have made my Ph.D. journey unforgettable. Specifically, many thanks are given to Guilian Luchini, Heidi Klem, Shree Sowndarya S V, Louis

de Lescure, Brandon S. Portela, Alex Platt, Abhijeet Singh Bhadauria, Niket Manoj, Turki Alturaifi, Lukas Sigmund, Dr. Juanvi Alegre Requena, Dr. Yingzi Li, Dr. Sreenithya Avadakkam, Dr. Santeri Aikonen, Dr. Yanfei Guan, and Dr. Mihai Popescu.

Mom and Dad, *gracias* for teaching me how to work hard, be kind, and value the most important things in life, and *gracias* for the love and support all these years. I am forever grateful, and this success is for you. *Los quiero mucho mama y papa*. To my siblings, Andres, Jaime, Monica, and the rest of my family, thank you for your love and support from miles away. To my dear *abuelitos*, *gracias*, for listening to my dreams; you have imparted me with wisdom well beyond my years.

Carlos, we began this journey with dreams of traveling in a van, and here we are, visiting new horizons while getting caught in dangerous scenarios with Supernova. Thank you, *amor*, for taking that hike with me to begin our life together. Thank you for the endless laughs and your kind support through this challenging journey. I look forward to a lifetime of moments filled with laughter, warm hugs, and great music sessions. I hope to support you in achieving everything that you want, and I will spend the rest of my life returning, in kind, the love and support that you have given me. I wouldn't have it any other way, *amor*.

TABLE OF CONTENTS

LIST OF KEYWORDS

| | |
|---|---|
| Ala | Alanine (A) |
| ANI-1 | Accurate neural network engine for molecular energies |
| ANOVA | Analysis of variance |
| AQME | Automated quantum mechanical environments |
| AVE | Average steric occupancy |
| Bi | Bismuth |
| $B_{max}$ | Largest Sterimol width parameter of a molecule |
| $B_{min}$ | Smallest Sterimol width parameter of a molecule |
| B3PW91 | Range-separated dispersion-corrected functional |
| BSSE | Basis set superposition errors |
| BV | Buried volume |
| CASCADE | Chemical shift calculation with deep learning |
| CHESHIRE | Chemical shift repository for computed NMR scaling factors |
| COMFA | Comparative molecular field analysis |
| COSMO-RS | Quantum chemical continuum solvation model COSMO |
| CREST | Conformer-rotamer ensemble sampling tool |
| CSV | Common-separated values file format |
| CV | Cross validation in statistical analysis or Cyclic voltammetry in electrochemistry |
| Cys | Cysteine (C) |
| D3/D4 | London dispersion correction functions |
| DBSTEP | Python program to compute DFT-based Steric Parameters |
| def2-SVP | Split valence with polarization functions on heavy atoms |
| def2-TZVP | Triple-zeta split-valence polarized basis set |
| def2-QZVPP | Quadruple-zeta split-valence polarized basis set with additional diffusion function |
| DCM | Dichloromethane |
| DFT | Density functional theory |
| DISCO | Distributing computed outputs workflow |
| $E_{disp}$ | London dispersion interaction energy |

| | |
|---|---|
| $E_{elec}$ | Electrostatic energy |
| $E_{ind}$ | Orbital interaction energy |
| $E_{Pauli}$ | Pauli repulsion energy |
| EAS | Electrophilic aromatic substitutions |
| ECP | Effective core potentials |
| EDA | Energy decomposition analysis |
| EDBO | Experimental design *via* bayesian optimization |
| eq. | Equivalent(s) |
| GIAO | Gauge-independent atomic orbital |
| GNN | Graph neural networks |
| GOODVIBES | Python program to compute thermochemical data |
| HOMO | Highest occupied molecular orbital |
| hr | Hour(s) |
| HPLC | High performance liquid chromatography |
| IQR | Interquartile range |
| INT | Ground-state intermediate |
| IRC | Intrinsic reaction coordinate |
| $K_{eq}$ | Reaction equilibrium constant |
| LANL2DZ | Los Alamos National Laboratory 2 Double-Zeta |
| LED | Light emitting diode |
| LUMO | Lowest unoccupied molecular orbital |
| Lys | Lysine (K) |
| MAE | Mean absolute error |
| MeCN | acetonitrile |
| ML | Machine learning |
| MMFF | Merck molecular force field |
| MORFEUS | Python package for calculating molecular features |
| MVLR | Multivariate linear regression |
| NBO | Natural bonding orbital |
| NCI | Non-covalent interaction |
| NCIPLOT | Program for plotting non-covalent interactions |

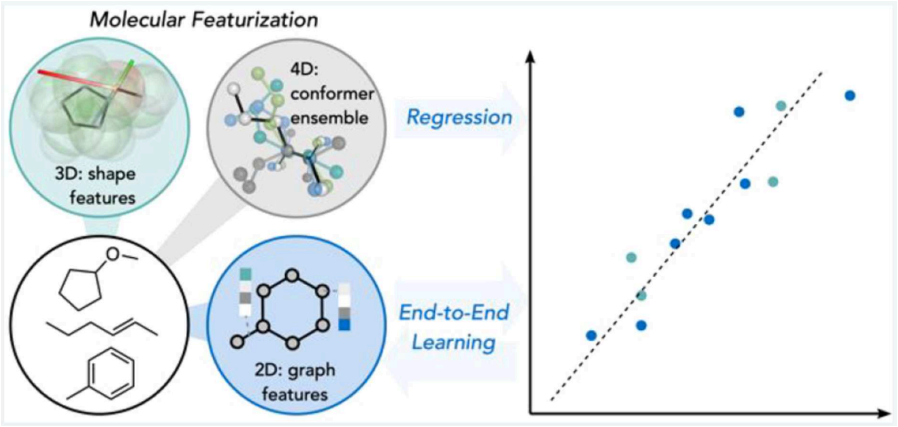
| | |
|---|---|
| NMR | Nuclear magnetic resonance |
| NPA | Natural population analysis |
| OA | Oxidative addition elementary step |
| PCA | Principal component analysis |
| Pd | Palladium |
| PDC | Photocatalytic diselenide contraction |
| PES | Potential energy surface |
| PL | Photoluminescence |
| PNG | Portable network graphics file format |
| PTA | 1,3,5-triaza-7-phos-phaadamantane |
| p-value | Probability of finding the observed result due to chance (e.g., null hypothesis) |
| PyMol | Software by Schrödinger for molecular visualization |
| Pyr | Pyridine |
| $Q^2$ | Predictive squared correlation coefficient |
| QM | Quantum mechanics |
| QSAR | Quantitative structure-activity relationships |
| QSPR | Quantitative structure-property relationships |
| R | Programming language for statistical computing and graphics |
| $R^2$ | Squared correlation coefficient |
| RDTS | Rate-determining Transition State |
| RE | Reductive elimination elementary step |
| RF | Random Forest |
| REGGAE | Regression generator and analyzer workflow |
| RMSE | Root mean squared error |
| rt | Room temperature |
| R-Studio | Computational program for running R |
| Se | Selenium |
| Sec | Selenocysteine (U) |
| $S_H2$ | Homolytic substitution reaction |
| $S_N2$ | Nucleophilic substitution reaction |
| $S_NAr$ | Nucleophilic aromatic substitution reaction |

| | |
|---|---|
| SIE | Self-interaction errors |
| GIAO | gauge-invariant atomic orbital |
| RDKit | Software for cheminformatics, computational chemistry, and predictive modeling |
| SET | Single electron transfer |
| SMD | Solvation model based on density |
| SMILES | Simplified molecular input line entry system |
| T1 | Lowest triplet excited state |
| TDTS | Turnover-frequency determining transition atate |
| THF | tetrahydrofuran |
| TM | Transmetallation elementary step |
| TMS | Tetramethyl-silane |
| TOF | Turnover-frequency |
| TS | Transition state |
| TST | Transition state theory |
| UTS | Universal training set |
| ωB97X-D | Long-range corrected functional including empirical dispersion correction |
| xTB | Semiempirical extended tight-binding program package (also GFNn-xTB) |

# 1 INTRODUCTION

## *1.1 Chapter Overview*

This chapter is intended to provide the necessary background to introduce the methods behind computational and statistical modeling. This overview aims to appreciate the tools that have assisted in the fundamental challenge of predicting the "over-the-arrow" optimal reaction conditions to maximize the desired reaction outcome. We begin with an introduction of the computational workflow to generate and understand reaction mechanisms for interesting reactions using quantum mechanical methods. Then, we describe the data-driven approach with computer-readable molecular descriptions and introduce the methods and metrics that build relationships to the chemical reactivity. We briefly mention (i) the limitations within each strategy, and (ii) the automated workflows developed to facilitate the transparency and reproducibility for collecting and analyzing the data presented herein. Molecular feature descriptions from one to four dimensions, the steric, and electronic features discussion in section 1.3 includes published material with co-authorship with Guilian Luchini in the *Accounts of Chemical Research*[1] and material currently under preparation to be submitted to *ACS Catalysis*.



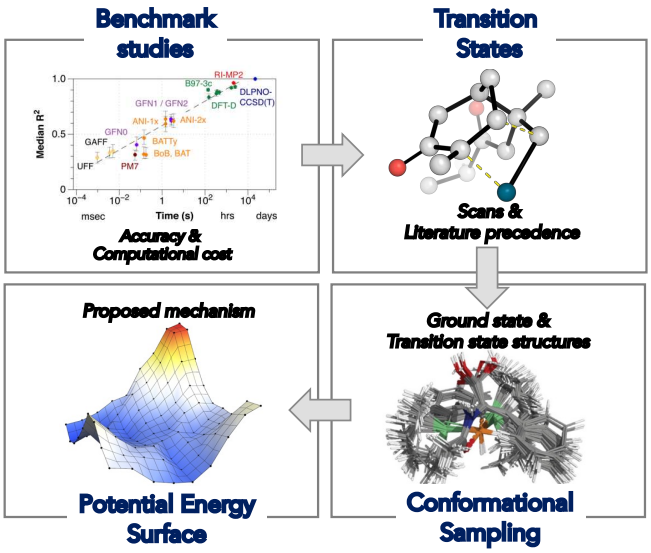

**Scheme 1.1.** General overview discussed in this chapter of the machine-readable molecular structure representations (with emphases on mechanistically derived features) that harness statistical relationships for the prediction of reactivities, selectivities, and chemical properties.

### *1.2 Mechanistically driven Methods in Reaction Optimization*

Reaction optimization often follows from a detailed mechanistic understanding obtained from experimental and/or computational means. Knowledge of the most likely reaction pathways, which steps control selectivity and/or turnover, and the existence of competing side reactions can all be used to logically inform steps taken in the laboratory. Inherent challenges are related (but not limited) to the intrinsic reactivity, which controls whether product formation occurs on a reasonable timescale, and the selectivity, which relates to the formation of unwanted side-products. Both challenges can be explored through computer simulations to explain the experimentally observed chemical reactivity that occurs in laboratory settings and in biological systems, using techniques ranging from kinetic and thermodynamic calculations, electronic structure analysis, and molecular property (e.g., NBO and NMR) calculations, amongst many others. Application of computational chemistry in these chapters consists of solving challenging chemical reactivity and selectivity problems from calculations and resulting mechanistic insights.

#### *1.2.1 Overview for Applying Quantum Mechanics*

Quantum mechanical (QM) methods aim to understand the electronic structure of a chemical system in which include (a) *ab initio* methods based entirely on the first principles of quantum mechanics, without any experimental data input while (b) semi-empirical methods include additional empirical parameters to improve speed and accuracy targeted for relatively larger systems. The typical steps taken by a computational chemist to gain mechanistic insight into a reaction of interest using quantum mechanical (QM) methods are demonstrated in Scheme 1.2.




**Scheme 1.2.** Simplified computational modeling workflow for obtaining a mechanistic understanding of a reaction of interest.

Four stages are exemplified here: (i) calibration of a computational protocol based on considerations of time and quantitative precision; (ii) the computation of transition structures/transition states (e.g., species that control the rates/selectivities of elementary chemical reactions); (iii) consideration of possible molecular conformations; (iv) characterization of the potential energy surface (PES) that includes desirable and undesirable reaction pathways.

### *1.2.2 Computational Protocols for Mechanistic Studies*

With DFT, we can express the energy and molecular properties as a functional of the electron density. However, calibration for the ideal computational protocol with DFT based on quantitative precision is best achieved by performing benchmarking studies or scavenging the literature for comprehensive benchmarking studies suitable for the specific system of interest. In this section, the computational protocols utilized in our mechanistic studies are defined here to establish its reasoning behind its implementation. In particular, we consider the following: (i) self-interaction errors, (ii) dispersion corrections, (iii) basis set superposition errors (BSSE), (iv) proper description of the valence electrons for heavier elements, (v) diffuse functions, (vi) polarization functions,

(vii) solvation effects, and (viii) integration grid effects to balance between accuracy and efficiency in these methods.[2]

Self-interaction errors (SIE), which result from an artificial residual interaction of an electron with itself, can be reduced using range-separated functionals. London dispersion interactions (e.g., D3/D4 corrections) are required to account for inter- and intramolecular non-covalent interactions, thereby accurately calculating the thermodynamics and kinetics.[3] Basis set superposition errors result from an artificial energy stabilization (e.g., energy lowering) with compact structures as spatially close atoms begin to access each other's basis function; however, this error can be reduced by avoiding small basis sets. For systems involving heavier elements (e.g., elements in the third row and above, also more commonly applied for transition metals), effective core potentials (ECP) or pseudopotentials, with a larger number of basis functions (e.g., Los Alamos National Laboratory (LANL) ECPs) are implemented to capture the inert core electrons. Diffuse functions are applied in anionic systems to accurately describe the loose electrons away from the nucleus indicated by "+" on atoms other than hydrogens and "++" for all atoms. Polarization functions are utilized to represent atomic orbitals with higher angular momentums (e.g., during bond formation) which are denoted as "(d,p)" or "**" indicating d-type functions added to heavy atoms, p-type functions added to hydrogens, and f-type functions added to transition-metals. In general, implicit solvation models are applied within computation studies as it represents a solvated molecule within a continuous solvent medium without explicitly including solvent molecules. The continuum solvation model based on molecular electron density called SMD from Truhlar and co-workers was applied within most of these studies;[4] COSMO-RS, a conductor-like screening model for real solvents, was applied in the solvent analysis discussed in section 6.2 to obtain solvent features for a more descriptive representation.[5] Orientational dependence of Gibbs energy differences has been

reported by Wheeler *et al* as a result to the sensitivity of vibrational frequencies (e.g., low-frequency vibrational modes).[6] The integration grid size to reduce the magnitude of grid errors (e.g., especially for calculating selectivities) is recommended to at least the default "ultrafine" pruned (75,302) grid or higher. The following density functionals were applied within our mechanistic studies.

**B3PW91**. A range-separated dispersion-corrected density functional.[7-8] This functional includes Becke's three-parameter exact exchange functional (B3) and Perdew and Wang's 1991 (PW91) gradient-corrected exchange and correlation functions with 20% Hartree−Fock exchange.

**ωB97X-D**. A range-separated hybrid functional where the inter-electronic interactions are separated to short-range and long-range interactions by the ω parameter. Additionally, this functional includes Grimme's D2 empirical dispersion correction to capture weak non-covalent interactions (NCIs) (e.g., the van der Waals attractive interactions).[9-10]

*1.2.3 Generating and Understanding Reaction Mechanisms with QM*

The generated potential energy surface (PES) is the electronic energy landscape of the system over the nuclear coordinates in which connects reactant, intermediate and product molecules *via* the intervening transition structures (TSs). TSs are saddle points on the PES characterized by negative curvature along a single direction (e.g., that chemists often refer to as an imaginary normal mode). Motion along this direction is mapped out by an intrinsic reaction coordinate (IRC) calculation that illustrates changes in chemical bonding that occur in this step.

Nevertheless, finding a hypothesized transition state and then using this knowledge to develop a conceptual model can be a challenging task. Depending on the novelty of the reaction or literature precedence, transition states can be found through constrained scans of the bond distances and

angles from the ground-state intermediates. It is critical to compute the stabilities of transition structures and products when computing the selectivity of a reaction, since there are two different regimes under which this can take place. Enantioselectivity of an irreversible reaction (e.g., typically at low temperatures) is a function of the Gibbs energy difference between the competing transition structures (e.g., difference between the activation barriers, $\Delta\Delta G^{\ddagger}$) and is said to be under kinetic control. In contrast, a reaction under thermodynamic control (e.g., a reversible reaction) only requires the Gibbs energy difference (i.e., $\Delta G$) of products and reactants (Figure 1.1).

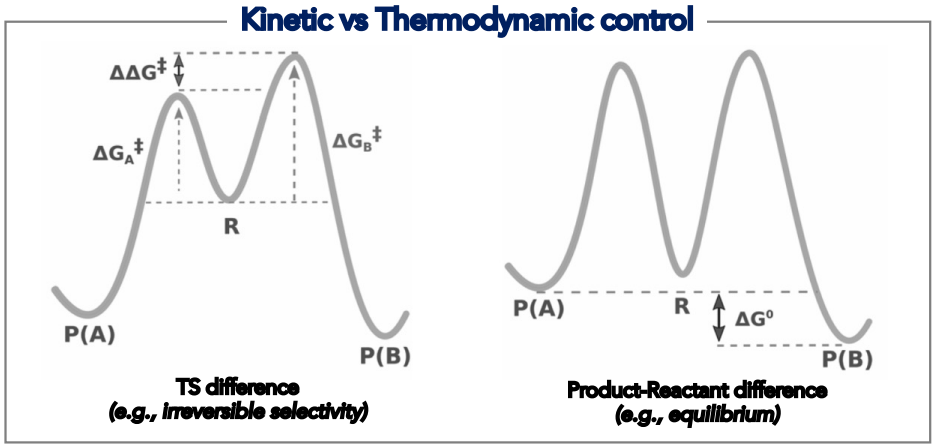



**Figure 1.1.** Gibbs free-energy profiles under kinetic control (left) and thermodynamic control (right).

Obtaining an accurate energy for the system described is rendered when performing an effective conformation sampling. Molecular conformation influences steric and electronic properties. The experimentally observable reactivity is a result from an accumulation of contributions from individual molecular conformations. Therefore, describing an accurate range of conformers appropriate for the system will help in accurately modeling a chemical process when comparing to experiment. In Chapters 3-4 and section 6.2, conformations of starting materials, intermediates, and TS structures were surveyed manually in terms of rotations about single bonds. However, this technique can become tedious with large flexible compounds. Several low-cost tools and algorithms exist to search for a molecule's conformers by systematically scanning dihedrals

of all rotatable bonds and applying force fields to identify the most stable compounds with tools like RDKit.[11] This approach was applied to systems mentioned in section 6.3. CREST is a program from Grimme used to sample the conformational space of molecules using semiempirical methods.[12] In Chapters 2 and 5, we apply CREST *via* a comprehensive workflow called AQME[13], to generate these conformers with considerably less user input and manipulation than manual optimization.

*1.2.4 Advantages and Limitations in Computational Modeling*

Computational modeling can lead to predictions of selectivity by applying transition state theory to the computed energies of competing transition states. However, the margin of error of typical computations is often on the order of ±1 kcal/mol.[14] Therefore, prediction of the favored product is not always achieved. Subtle energy differences can be approximated by subjecting the intermediates and transition structures to conformational sampling. The multiple steps to investigate the reaction's activity as shown in Scheme 1.2 makes computational screening of various substrates, particularly substrates with flexible groups (e.g., generates multiple conformers from the number of rotatable bonds), and reaction conditions relatively "low-throughput." Additionally, while some aspects of chemical reaction modeling are relatively mature, other aspects such as solvent mixtures, interfacial processes (e.g., changes in phase between solid/liquid/gas), integration grid effects, and large and flexible systems are still in their infancy from a computational perspective.

Quantum mechanical studies have provided tools to predict reactivity trends and insights into transition states to build successful generalized models of reactivity.[15] However, such studies typically take months or years to complete and often involve multiple rounds of experimental and computational investigations. These restrictions have led chemists to seek alternative approaches

that require shorter timescales and which utilize more empirical information rather than constructing computational models of reaction mechanism from first principles.

### *1.3 Data-driven Methods in Reaction Optimization*

Statistical modeling in chemistry aims to find a quantitative relationship between an observable outcome (e.g., reaction yield, selectivity, reaction conditions, *etc.*) and the chemical inputs involved.[16] So-called quantitative structure-activity relationships (QSARs) were first used by computational biologists to relate the biological activity of small molecules with their chemical structures.[17] Subsequently, chemical applications have adopted the same philosophy, *albeit* with different endpoints (e.g., reaction yield or selectivity). A key area of active research is the numerical encoding of chemical structures called cheminformatics for various chemical reactivity descriptions.

#### *1.3.1 Molecular Representations: From One to Four Dimensions*

The quantitative description of molecules and their physical and chemical properties is an essential first step towards building statistical relationships for reaction outcomes (Scheme 1.1). A large number of molecular "descriptors", alternatively referred to as "parameters", "features", and "feature vectors" have been developed to date. Feature selection is a core concept of statistical modeling and machine learning. Physics-based molecular features originate from either macroscopic bulk measurements (e.g. molecular volume, polarizability, boiling point, *etc*) or atomistic/molecular descriptions generated computationally have shown to improve model accuracy.[18]

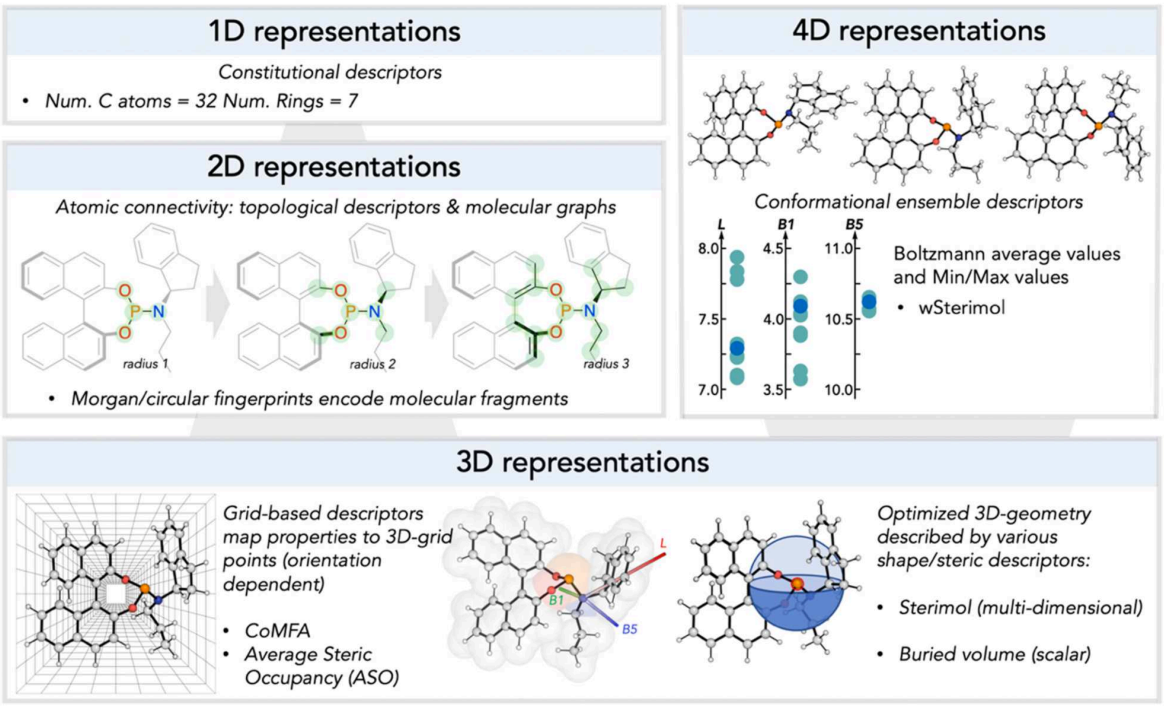



**Figure 1.2.** Hierarchy of molecular representations used to encode organic and organometallic structures.

Traditional cheminformatics representations largely focus on 2D or topological molecular representations that define the connectivity and bonding types of atoms in a molecule. 2D molecular descriptors, such as topological fingerprints, are simple to define and can be obtained without geometry optimization. However, a number of features of mechanistic relevance to reactivity and selectivity depend upon a molecule's 3D structure and conformation, including electronic properties such as atomic and molecular charge distributions, as well as other structure-dependent features that capture a molecule's steric influence, chirality, volume or surface area. Quantum mechanically (QM) optimized molecular coordinates (e.g., using density functional theory, DFT) can be used to obtain such descriptors, in addition to other QM-computed properties such as thermochemical values, molecular orbital energies, vibrational frequencies, and noncovalent interaction energies. In QSAR vernacular, 3D-descriptors generally refer to those that map molecular interactions to a pre-aligned grid of points, as in Comparative Molecular Field

Analysis (COMFA)[19] and the more recently developed Average Steric Occupancy (ASO)[20] from Denmark. More generally, properties dependent upon 3D-structure, and especially spatial/steric occupancies, have been described by scalar parameters such as buried volume and higher dimensional objects such as topological maps or multidimensional Sterimol parameters, approaches pioneered by the Cavallo[21] and Sigman[22] groups, respectively. This hierarchy of molecular representations used across organic chemistry is shown in Figure 1.2.

The dependence of 3D or DFT-derived descriptors upon the molecular geometry means that unlike 2D representations, conformational dynamics are important to consider. Indeed, the presence of multiple conformations may itself be an important descriptor relating to catalytic proficiency.[23] Thus, in addition to featurization of the most stable conformer, consideration of the full conformational ensemble may be necessary to quantitatively encode macroscopic behavior. Ensemble approaches have been referred to as 4D-QSAR.[24] Spatial parameters may be sensitive to conformational behavior and to the level of theory employed to generate an ensemble, which led us to develop a software tool, wSterimol, to automate conformer-sampling and featurization.[25] With this, we have been able to include estimates of parameter uncertainty into regression models of stereoselectivity. Broadly speaking as data-chemists, we can draw a distinction between electronic and steric (i.e., 3D-spatial) features to build statistical relationships to explain chemical reactivity although there is not always complete separation between these groupings.

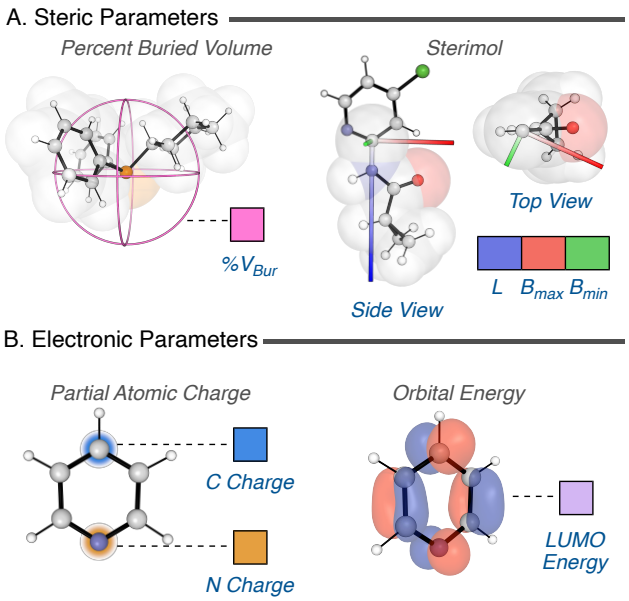



**Figure 1.3.** Visualizations (generated by Guilian Luchini) of (A) steric parameters percent buried volume ($\%V_{bur}$) and Sterimol parameters and (B) electronic parameters including partial atomic charge and orbital energies.

Steric effects can be described as nonbonding interactions between molecules occupying the same space that leads to repulsion interactions. Structure-dependent features (e.g., Sterimol and buried volume) are often used towards capturing the steric effects of molecules. The Sterimol descriptors, also shown in Figure 1.3A, are an anisotropic class of three descriptors describing the length (L), minimum width ($B_{min}$) and maximum width ($B_{max}$) of a molecule. Originally, these descriptors were designed showing their usefulness in predicting reactivity of pesticide molecules, and in the years following quickly found use in insecticide, herbicide, and pharmaceutical studies.[26] These parameters have since been revisited and have found use in asymmetric catalysis, proving useful to describe how the steric bulk of ligand molecules and substituents influence reaction outcome.[22, 27]

Electronic descriptors capture either the whole molecule, single atom, or group of atoms from computed molecular wavefunctions.[28] Formulations for partial atomic charges involve partitioning the charge of a molecule into individual atoms and can be derived from a molecule's electrostatic

potential, atomic orbitals, or electron density.[29] The Natural Population Analysis (NPA) is a method to compute atomic partial charges from molecular orbital populations, providing quantitative values for atomic charges, as well as orbital energies.[30] Qualitative methods of visualizing molecular orbitals, like the LUMO orbital shown in Figure 1.3B, obtained through these analyses have also been used to provide mechanistic insight into a reaction. Recently, Modak and others utilized NPA to aid in understanding regioselectivity differences in substituents for homologation reactions of aryl halides.[31] Using key $\sigma^*$ molecular orbital energies obtained with NPA, authors constructed a classification model identifying an energetic threshold separating highly regioselective substrates against lower selectivities. Workflows (e.g., DISCO discussed in section 7.2), were developed to facilitate collecting these various electronic molecular representations from different DFT calculations.[32]

Solvent atomistic descriptors and molecular properties are also considered in predicting reaction selectivity.[33]

Although quantum mechanical (QM) methods have proven to capture the relationship to the reaction outcome, the choice of method and size of the basis set can exponentially scale up the computational demand. As a result, efforts to limit the cost of resources while maintaining chemical accuracy have exploded with newly developed machine learning algorithms, integrated automated chemistry research workflows, and the availability of data and high-performing computing resources.[34-36] Machine-learned molecular descriptors derived from graph neural networks (GNN) like NMR chemical shifts is discussed in section 6.3.

Advances in computational methods such as dispersion corrections in long-range correlations has improved the accuracy of reaction barriers by capturing dispersion interactions between non-bonded atoms, or non-covalent interactions (NCI).[9] NCI analysis of computed structures assist in

identifying "beyond the molecule" descriptors to capture van der Waals interactions, steric clashes, and hydrogen bonds influencing the selectivity of the reaction. Additionally, DFT-computed interaction energies between molecules can be investigated using an energy decomposition analysis (EDAs) approach. This method breakdowns the binding energetic components into distortion energies, and various interaction energy terms: electrostatic ($E_{elec}$), Pauli repulsion ($E_{Pauli}$), London dispersion interactions ($E_{disp}$), and orbital interactions (e.g., $E_{ind}$, including induction effects from polarization and charge transfer). A variety of computational techniques and programs (e.g., PSI4[37], ORCA, Gaussian, and QChem) are used to visualize these interactions to explain the reaction outcomes. In Chapters 3-4, we utilize NCIPLOT[38-39] to visualize the intermolecular interactions influencing selectivity and reactivity in bismuth systems.

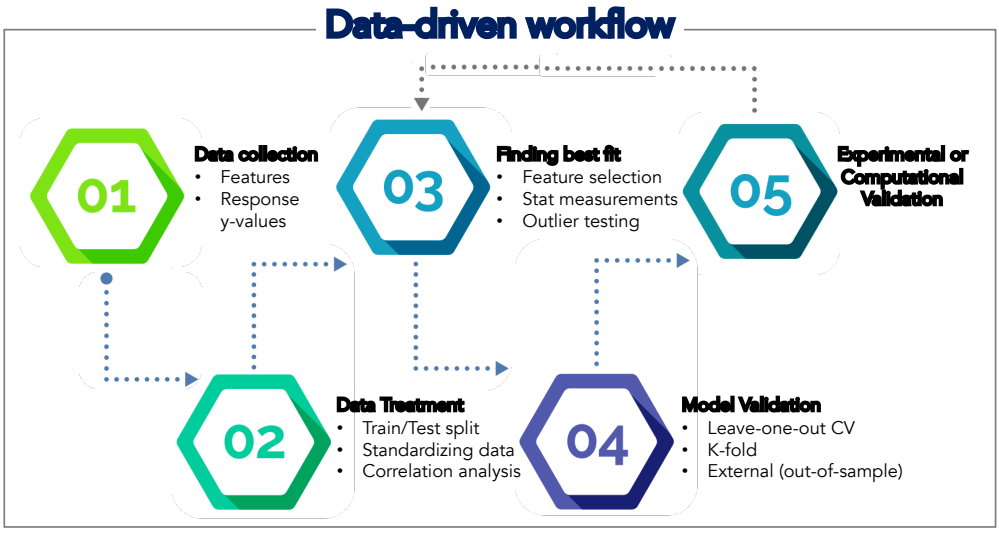
*1.3.2 Building a Linear Predictive Model*

The two classifications for statistical modeling either fall under supervised (e.g., classification, regression) or unsupervised learning (e.g., clustering, principal component analysis). Under unsupervised learning, unlabeled data is trained and clustered to discover hidden patterns while supervised learning deals with labeled input and output data to measure model accuracy and learning over time. Types of supervised learning models frequently used by chemists for reactivity or property predictions are regression[40], random forest[41], and neural networks[42]; other machine learning methods such as the Markov decision process in reinforcement learning models have recently emerged for maximizing molecules' drug-likeness[43].

Random forest, (graph) neural networks, and Bayesian optimization are black-box machine learning algorithms used in regression and classification models with labeled datasets (e.g., observations or y-response column) to train the algorithm to the collected features (e.g., x-variables). The random forest model is composed of multiple decision trees where observations

are split at each decision node in the tree to follow the branch that best fits the labeled target. Neural networks or artificial neural networks are comprised of connected nodes where data is passed along the input layer, multiple hidden layers with an associated weight, and finally out the output layer of the network. The Bayesian optimization algorithm consists of a Gaussian process surrogate model fitted around labeled data to build a response surface with broad areas corresponding to higher uncertainty and tighter areas containing more sampled and explored observations. On the other hand, multivariate linear regression (MVLR) utilizes multiple independent variables (e.g., molecular representations/features, x-variables) to predict the observations while estimating the coefficients that builds the linear relationship between the variables.

Although, random forest, (graph) neural networks, and Bayesian optimization models[44-45] are less interpretable, the quantitative accuracy in chemical applications has yielded interest and priority for a successful predictive model.[46] In this thesis, we focus primarily on multivariate linear regression (MVLR) models based on the ability to interpret the results from the variable's coefficients and the limited data size within the projects discussed in Chapter 4 and Section 6.2, such that more complex models tend to overfit.




**Scheme 1.3.** Statistical analysis workflow to generate a predictive model.

Statistical analysis workflows (e.g., REGGAE in section 7.3) were developed with the steps listed in Scheme 1.3 to facilitate the application of multivariate linear regression models. This statistical workflow encompasses generalized "best practice" guidelines descried in the cheminformatics literature for generating and validating predictive models within applications in synthetic chemistry.[47-49]

A model's statistical performance is validated using data held-back from the training process. Chemists continue to debate optimal strategies for dividing reaction data sets into training and testing sub-sets. For example, while a random split (e.g., 70% Train and 30% Test split or 80% Train and 20% Test split) is common, Denmark has put forward the idea of a "universal" training set (UTS).[50] In part, this is because chemical data sets tend to be limited in size. The UTS is automatically selected using a dimensionality reduction and clustering algorithm with the intention of sampling uniformly across each variable dimension. Pairwise correlation analyses assist in identifying highly correlated variables dependent on each other (e.g., $R^2$ greater than 0.6 is a general rule of thumb). By removing these highly correlated variables, we increase the model's

robustness while reducing any collinearity amongst the variables. Common feature selection functions for finding a best linear model fit include (i) forward-stepwise feature selection or (ii) utilizing a dredge function in R-studio (e.g., function for comparison of all possible linear models). Tropsha (a leader in the field of cheminformatics) has emphasized that a particular feature's statistical contribution to the model is deemed a satisfactory model fit if the following criteria are met: $R^2 > 0.6$, variables with p-values $< 0.05$ (e.g., statistically significant), standardized residuals $< \pm 2.5$ (e.g., to determine possible outliers), amongst others. The model's predictive power is determined by validation methods such as leave-one-out and K-fold cross-validation with resulting $Q^2$ values $> 0.5$.[48]

*1.3.3 Advantages and Limitations in Statistical Modeling*

The data-driven approach aims to eliminate unnecessary experiments and/or expensive calculations by effectively learning from historical data. However, the predictive performance of a given statistical model generally deteriorates in areas of chemical space that lie outside of the so-called domain of applicability. In this sense, interpolation between areas of existing data tends to be more successful than extrapolation (e.g., areas outside the trained model) to new molecules or reactions, which has led to a focus in the creation of "universal" training sets of molecules, with as broad coverage as possible of chemical space. Furthermore, MVLR models require a linear correlation to the descriptors in which is not always the case (e.g., when different types of substrates have a change in mechanism). Additional strategies, such as Bayesian (algorithm applied to project discussed in Chapter 5) or reinforcement learning methods[51] have been successfully used as optimization methods that allows a flexible model to fit around the observations and approximate outcome predictions through exploration and exploitation of the chemical space.[44, 52] The exploration and exploitation can be seen as the tension between collecting

new data to boost the model's performance versus selecting the greedy strategy to optimize the outcome. Attempts at these modern machine learning methods for reaction optimization demonstrate promising results at tackling the challenges of limited applicability across multiple reaction mechanisms.

### *1.4 Conclusions*

This chapter provided an overview of the importance of computational protocols tailored towards studying each specific system and predicting reaction outcomes based on DFT methods. These methods are applied to Chapters 2-5 to propose reaction mechanisms and explain the experimentally observed reactivity. In Chapter 6, we progress into applying statistical modeling approaches towards predicting reaction outcomes. Specifically, in section 6.2, we apply DFT methods to derive mechanistic informed descriptors towards predicting reaction outcomes *via* statistical modeling; in section 6.3, we utilize DFT NMR features towards training and building a predictive model for ML-derived NMR chemical shift descriptors. In Chapter 7, we discuss the automated workflows developed to facilitate the processes in data-driven approaches in reaction optimization projects.

## REFERENCES


1. Gallegos, L. C.; Luchini, G.; St. John, P. C.; Kim, S.; Paton, R. S., Importance of Engineered and Learned Molecular Representations in Predicting Organic Reactivity, Selectivity, and Chemical Properties. *Acc. Chem. Res.* **2021,** *54* (4), 827-836.
2. Bursch, M.; Mewes, J.-M.; Hansen, A.; Grimme, S., Best-Practice DFT Protocols for Basic Molecular Computational Chemistry**. *Angew. Chem., Int. Ed.* **2022,** *61* (42), e202205735.
3. Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C., Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016,** *116* (9), 5105-5154.
4. 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* (18), 6378-6396.
5. Klamt, A., Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995,** *99* (7), 2224-2235.
6. Bootsma, A. N.; Wheeler, S., Popular Integration Grids Can Result in Large Errors in DFT-Computed Free Energies. *ChemRxiv* **2019**.
7. Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C., Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* **1992,** *46* (11), 6671-6687.
8. Adamo, C.; Barone, V., Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J. Chem. Phys.* **1998,** *108* (2), 664-675.
9. Chai, J.-D.; Head-Gordon, M., Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008,** *10* (44), 6615-6620.
10. Goerigk, L.; Grimme, S., A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011,** *13* (14), 6670.
11. RDKit: open-source. cheminformatics.http://www.rdkit.org.
12. Pracht, P.; Bohle, F.; Grimme, S., Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020,** *22* (14), 7169-7192.
13. Alegre-Requena, J. V.; Sowndarya S. V., S.; Pérez-Soto, R.; Alturaifi, T. M.; Paton, R. S., AQME: Automated quantum mechanical environments for researchers and educators. *WIREs Computational Molecular Science* **2023**.
14. Ardkhean, R.; Fletcher, S. P.; Paton, R. S., Ligand Design for Asymmetric Catalysis: Combining Mechanistic and Chemoinformatics Approaches. In *New Directions in the Modeling of Organometallic Reactions*, Springer International Publishing: 2020; pp 153-189.
15. Bahmanyar, S.; Houk, K. N.; Martin, H. J.; List, B., Quantum Mechanical Predictions of the Stereoselectivities of Proline-Catalyzed Asymmetric Intermolecular Aldol Reactions. *J. Am. Chem. Soc.* **2003,** *125* (9), 2475-2479.
16. Reid, J. P.; Sigman, M. S., Comparing quantitative prediction methods for the discovery of small-molecule chiral catalysts. *Nat. Rev. Chem.* **2018,** *2* (10), 290-305.
17. Martin, Y. C., Exploring QSAR: Hydrophobic, Electronic, and Steric Constants C. Hansch, A. Leo, and D. Hoekman. American Chemical Society, Washington, DC. 1995. Xix + 348 pp. 22 × 28.5 cm.Exploring QSAR: Fundamentals and Applications in Chemistry and Biology. *J. Med. Chem.* **1996,** *39* (5), 1189-1190.

18. Jorner, K.; Brinck, T.; Norrby, P.-O.; Buttar, D., Machine learning meets mechanistic modelling for accurate prediction of experimental activation energies. *Chem. Sci.* **2021,** *12* (3), 1163-1175.
19. Lipkowitz, K. B.; Pradhan, M., Computational Studies of Chiral Catalysts: A Comparative Molecular Field Analysis of an Asymmetric Diels−Alder Reaction with Catalysts Containing Bisoxazoline or Phosphinooxazoline Ligands. *J. Org. Chem.* **2003,** *68* (12), 4648-4656.
20. Zahrt, A. F.; Athavale, S. V.; Denmark, S. E., Quantitative Structure–Selectivity Relationships in Enantioselective Catalysis: Past, Present, and Future. *Chem. Rev.* **2020,** *120* (3), 1620-1689.
21. Poater, A.; Cosenza, B.; Correa, A.; Giudice, S.; Ragone, F.; Scarano, V.; Cavallo, L., SambVca: A Web Application for the Calculation of the Buried Volume of N-Heterocyclic Carbene Ligands. *EurJIC* **2009,** *2009* (13), 1759-1766.
22. Harper, K. C.; Bess, E. N.; Sigman, M. S., Multidimensional steric parameters in the analysis of asymmetric catalytic reactions. *Nat Chem* **2012,** *4* (5), 366-74.
23. Crawford, J.; Sigman, M., Conformational Dynamics in Asymmetric Catalysis: Is Catalyst Flexibility a Design Element? *Synthesis* **2019,** *51* (05), 1021-1036.
24. Andrade, C. H.; Pasqualoto, K. F. M.; Ferreira, E. I.; Hopfinger, A. J., 4D-QSAR: Perspectives in Drug Design. *Molecules* **2010,** *15* (5), 3281-3294.
25. Brethomé, A. V.; Fletcher, S. P.; Paton, R. S., Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters. *ACS Catal.* **2019,** *9* (3), 2313-2323.
26. Verloop, A., THE STERIMOL APPROACH: FURTHER DEVELOPMENT OF THE METHOD AND NEW APPLICATIONS. *Pesticide Chemistry: Human Welfare and Environment* **1983**, 339-344.
27. Piou, T.; Romanov-Michailidis, F.; Romanova-Michaelides, M.; Jackson, K. E.; Semakul, N.; Taggart, T. D.; Newell, B. S.; Rithner, C. D.; Paton, R. S.; Rovis, T., Correlating Reactivity and Selectivity to Cyclopentadienyl Ligand Properties in Rh(III)-Catalyzed C-H Activation Reactions: An Experimental and Computational Study. *J Am Chem Soc* **2017,** *139* (3), 1296-1310.
28. Cartier, A.; Rivail, J.-L., Electronic descriptors in quantitative structure—activity relationships. *Chemometr Intell Lab Syst* **1987,** *1* (4), 335-347.
29. Gonthier, J. F.; Steinmann, S. N.; Wodrich, M. D.; Corminboeuf, C., Quantification of "fuzzy" chemical concepts: a computational perspective. *Chemical Society Reviews* **2012,** *41* (13), 4671.
30. Reed, A. E.; Weinstock, R. B.; Weinhold, F., Natural population analysis. *The Journal of Chemical Physics* **1985,** *83* (2), 735-746.
31. Modak, A.; Alegre-Requena, J. V.; de Lescure, L.; Rynders, K. J.; Paton, R. S.; Race, N. J., Homologation of Electron-Rich Benzyl Bromide Derivatives via Diazo C–C Bond Insertion. *J. Am. Chem. Soc.* **2022,** *144* (1), 86-92.
32. Gallegos, L. C. DISCO. https://github.com/Liliana-Gallegos/DISCO.
33. Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G., A Universal Organic Solvation Model. *J. Org. Chem.* **1996,** *61* (25), 8720-8721.
34. Shen, Y.; Borowski, J. E.; Hardy, M. A.; Sarpong, R.; Doyle, A. G.; Cernak, T., Automation and computer-assisted planning for chemical synthesis. *Nat. Rev. Methods Primers* **2021,** *1* (1), 23.
35. Artrith, N.; Butler, K. T.; Coudert, F.-X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A., Best practices in machine learning for chemistry. *Nature Chemistry* **2021,** *13* (6), 505-508.

36. Kearnes, S. M.; Maser, M. R.; Wleklinski, M.; Kast, A.; Doyle, A. G.; Dreher, S. D.; Hawkins, J. M.; Jensen, K. F.; Coley, C. W., The Open Reaction Database. *J. Am. Chem. Soc.* **2021,** *143* (45), 18820-18826.
37. Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M.; Gonthier, J. F.; James, A. M.; McAlexander, H. R.; Kumar, A.; Saitow, M.; Wang, X.; Pritchard, B. P.; Verma, P.; Schaefer, H. F.; Patkowski, K.; King, R. A.; Valeev, E. F.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D., Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput* **2017,** *13* (7), 3185-3197.
38. Boto, R. A.; Peccati, F.; Laplaza, R.; Quan, C.; Carbone, A.; Piquemal, J.-P.; Maday, Y.; Contreras-Garcia, J. *NCIPLOT4: A new step towards a fast quantification of noncovalent interactions*.
39. Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W., NCIPLOT: A Program for Plotting Noncovalent Interaction Regions. *J. Chem. Theory Comput* **2011,** *7* (3), 625-632.
40. Santiago, C. B.; Guo, J.-Y.; Sigman, M. S., Predictive and mechanistic multivariate linear regression models for reaction development. *Chem. Sci.* **2018,** *9* (9), 2398-2412.
41. 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* (6385), 186-190.
42. Grambow, C. A.; Pattanaik, L.; Green, W. H., Deep Learning of Activation Energies. *J. Phys. Chem. Lett.* **2020,** *11* (8), 2992-2997.
43. Zhou, Z.; Li, X.; Zare, R. N., Optimizing Chemical Reactions with Deep Reinforcement Learning. *ACS Cent. Sc* **2017,** *3* (12), 1337-1344.
44. Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G., Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **2021,** *590* (7844), 89-96.
45. Torres, J. A. G.; Lau, S. H.; Anchuri, P.; Stevens, J. M.; Tabora, J. E.; Li, J.; Borovika, A.; Adams, R. P.; Doyle, A. G., A Multi-Objective Active Learning Platform and Web App for Reaction Optimization. *J. Am. Chem. Soc.* **2022,** *144* (43), 19999-20007.
46. Nielsen, M. K.; Ahneman, D. T.; Riera, O.; Doyle, A. G., Deoxyfluorination with Sulfonyl Fluorides: Navigating Reaction Space with Machine Learning. *J. Am. Chem. Soc.* **2018,** *140* (15), 5004-5008.
47. Dearden, J. C.; Cronin, M. T. D.; Kaiser, K. L. E., How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). *SAR QSAR Environ Res* **2009,** *20* (3-4), 241-266.
48. Golbraikh, A.; Tropsha, A., Beware of q2! *J. Mol. Graph.* **2002,** *20* (4), 269-276.
49. Tropsha, A.; Gramatica, P.; Gombar, V. K., The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* **2003,** *22* (1), 69-77.
50. Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E., Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. *Science* **2019,** *363* (6424), eaau5631.
51. S. V., S. S.; Law, J. N.; Tripp, C. E.; Duplyakin, D.; Skordilis, E.; Biagioni, D.; Paton, R. S.; St. John, P. C., Multi-objective goal-directed optimization of de novo stable organic radicals for aqueous redox flow batteries. *Nat. Mach. Intell* **2022,** *4* (8), 720-730.

52. Häse, F.; Roch, L. M.; Kreisbeck, C.; Aspuru-Guzik, A., Phoenics: A Bayesian Optimizer for Chemistry. *ACS Cent. Sc* **2018,** *4* (9), 1134-1145.

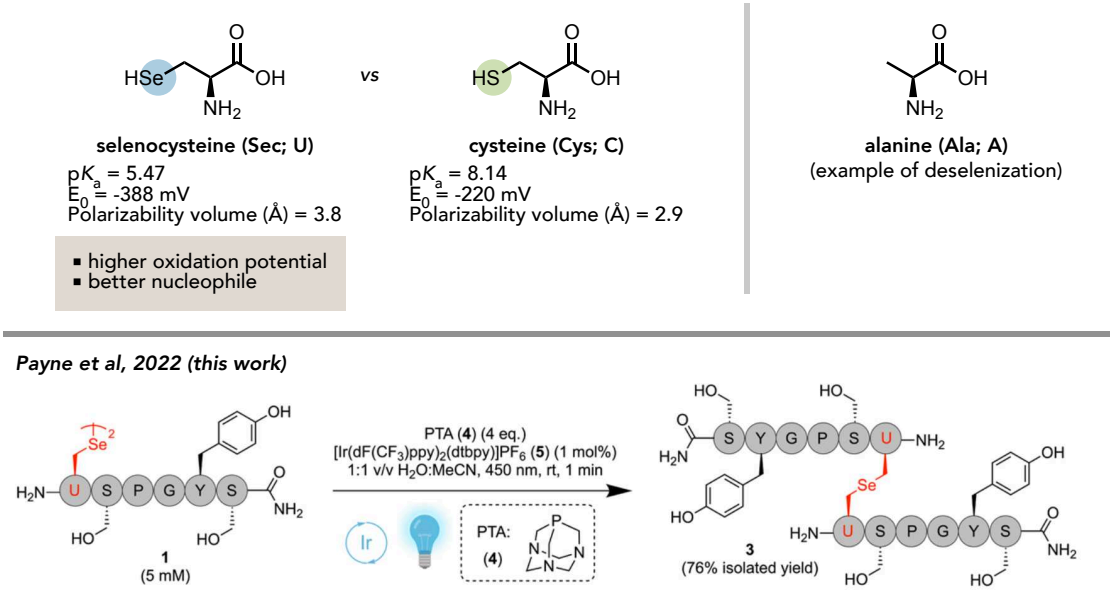
# 2 SITE-SELECTIVE PHOTOCATALYTIC FUNCTIONALIZATION OF PEPTIDES AND PROTEINS AT SELENOCYSTEINE

## *2.1 Chapter Overview*

In this chapter, we tackle the challenge in studying the chemical reactivity found in large biological systems (e.g., peptides and proteins) that was developed by the Payne lab. Specifically, we identify and explain the precise photocatalytic functionalization at selenocysteine using quantum mechanical methods. This collaborative research was published in *Nature Communications.*[1]

## *2.2 Introduction*

The importance of modified peptides and proteins for applications in drug discovery, and for illuminating biological processes at the molecular level, is fueling a demand for efficient methods in late-stage functionalization that facilitate the precise modification of these biomolecules. In 2022, the Payne group developed a photocatalytic method for the rapid and efficient dimerization and site-specific functionalization of peptide and protein diselenides (Scheme 2.1).[1]




**Scheme 2.1. Top:** Comparison of Cys and Sec structures, pKa values, and redox potentials. The expected resulting Alanine amino acid structure after attempting a photocatalytic deselenization

of Sec. **Bottom:** The optimized PDC dimerization reaction on $[H_2N\text{-USPGYS-}NH_2]_2$ model peptide **1**.

Selenocysteine (Sec; U) is commonly referred to as the 21st proteinogenic amino acid involved in the translation process for generating proteins.[2] The unique reactivity of the selenol side chain of Sec (e.g., low p$K_a$ and high oxidation potential and nucleophilicity relative to Cys shown in Scheme 2.1, top)[3] has primarily attracted previous methodologies as a designer modification cite (e.g., fluorescent tags, affinity handle) to execute bioconjugation strategies (e.g., forming a stable covalent link between two molecules)[4] *via* alkylation or arylation reactions which often possess modest chemoselectivity over other nucleophilic side chains (e.g., Cys, Lys), poor kinetics/conversions, and/or use non-biocompatible reaction conditions. Nonetheless, Payne *et al* developed a photocatalytic method for the rapid and high yielding dimerization and site-specific functionalization of peptides and proteins that capitalizes on unique reactivity at Sec *via* a photocatalytic diselenide contraction (PDC) reaction. The PDC reaction involves irradiation at 450 nm (e.g., with blue LED) in the presence of an iridium photocatalyst and a phosphine which results in rapid and clean conversion of diselenides to reductively stable selenoethers. In this collaborative effort, we performed a computational mechanistic investigation to propose a mechanism for this photocatalytic transformation, which is supported by photoluminescence spectroscopy. All experimental work was performed by our experimental collaborators (Payne *et. al*) and all DFT computational work was performed by the present author and Dr. Juan V. Alegre-Requena.

### *2.3 Results and Discussions*

In a serendipitous discovery, the experimental group unexpectedly observed the formation of dimeric selenoether-bridged peptide as a minor product when attempting a photocatalytic deselenization of Sec on a model peptide (**1**) to produce the alanine (Ala) model peptide (**2**) (Scheme 2.1, top). After optimizing for the preferential formation of selenoether dimer over the

deselenization product, the experimental group found that treatment with 4 equivalents of the phosphine 1,3,5-triaza-7-phos-phaadamantane (**PTA**), and 1 mol% of the iridium photocatalyst $[Ir(dF(CF_3)ppy)_2(dtbpy)]PF_6$ (**5**), with LED irradiation at 450nm for 1 min, led to the exclusive formation of the dimeric selenoether **3**, which could be isolated in 76% yield following reverse-phase HPLC purification (Scheme 2.1, bottom). Additional experiments dictated the importance of the following components as the absence of phosphine, photocatalyst, or without 450 nm irradiation, no formation of **3** was observed.

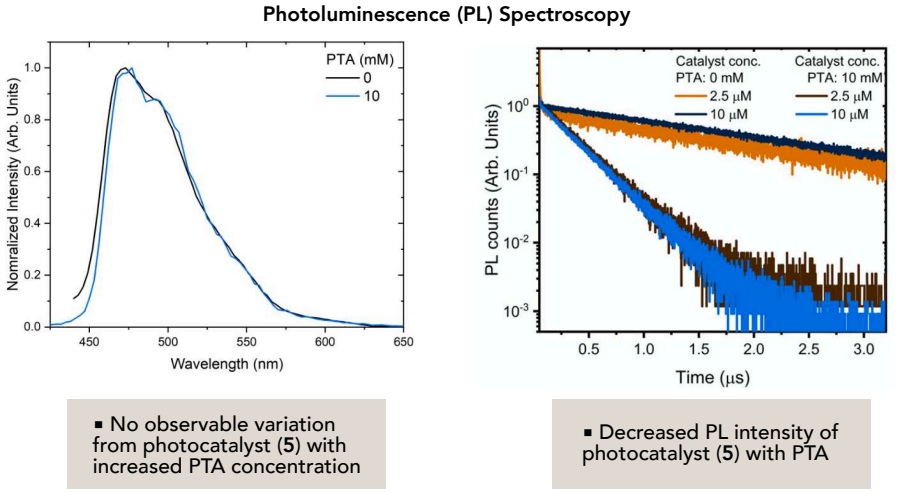



**Figure 2.1.** Photoluminescence spectra obtained from the experimental group. **Left:** of an Argon-sparged 1:1 v/v $H_2O$:MeCN (as per typical PDC reaction conditions) solution of $[Ir(dF(CF_3)ppy)_2(dtbpy)]PF_6$ (**5**) with 0 mM (black line) and 10 mM (blue line) **PTA** added. The excitation wavelength was 415 nm. No significant changes in the PL shape is observed with the addition of PTA to **5**. Arb. Units on the y-axis = arbitrary units. **Right:** Normalized photoluminescence counts of photocatalyst $[Ir(dF(CF_3)ppy)_2(dtbpy)]$ (**5**) (c = 2.5 μM and 10 μM) in argon-sparged 1:1 v/v $H_2O$:MeCN over a period of 3 μs monitored at 470 nm (excitation wavelength 415 nm) with or without **PTA** (0 mM and 10 mM).

Based on the experimental mechanistic studies from photoluminescence (PL) spectroscopy (Figure 2.1), the photocatalyst **5** did not partake in direct covalent chemistry with the phosphine (e.g., no observable variation over the entire time window (0.01–0.30μs) at different concentrations), but the PL lifetime of the photocatalyst diminished (e.g., decreased PL intensity by a factor of two) in the presence of **PTA**.

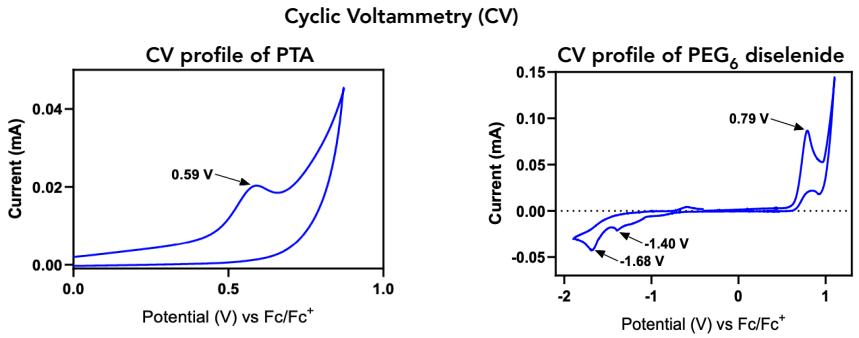



**Figure 2.2.** Cyclic voltammetry results obtained from the experimental group. **Left:** Cyclic voltametric profile of **PTA** in 1:1 v/v MeCN:$H_2O$ with 0.1 M NaCl. Scan rate: 0.1 $V.s^{-1}$. **Right:** Cyclic voltametric profile in 1:1 v/v MeCN:$H_2O$ with 0.1 M NaCl of $PEG_6$ diselenide.

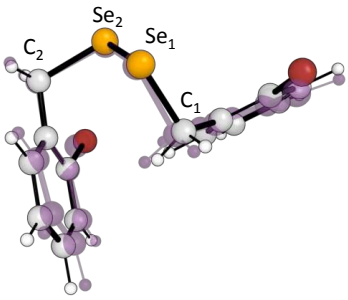
Additionally, cyclic voltammetry (CV) experiments in Figure 2.2 confirmed that **PTA** undergoes more facile single electron oxidation than the diselenide species in the reaction mixture, further supporting phosphine oxidation as the initial step of the reaction.

In order to elucidate and corroborate the feasibility of the new PDC reactivity from the working hypothesis obtained from photophysical experiments, we performed quantum chemical calculations. We used density functional theory (DFT) calculations at the B3PW91-GD3(BJ)/6-311+G(2df,p)//B3PW91-GD3(BJ)/6-31+G(d,p) level of theory[5]. Preliminary mechanistic exploration was executed using $(BnSe)_2$ as the model substrate. This simplified model was then used to select viable pathways to apply a more realistic model for the final energy profile. Using a selenocysteine derivative (e.g., $[N\text{-Boc-L-Sec-OtBu}]_2$ amino acid diselenide system) for the final energy profile, the structure emulates the peptide-like environments employed experimentally, showing activation barriers accessible at room temperature.

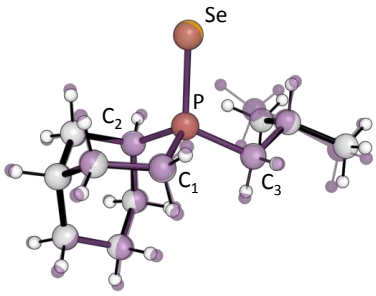
*2.3.1 Benchmarking to X-ray crystallography structures*

We began by evaluating the ability of our chosen level of theory in optimizing structures containing different types of Se atoms by comparison against a structure that has been characterized by X-ray crystallography (Figure 2.3). Considering that we used gas phase in this optimization calculation, which is of different polarity to the bulk crystal, and allowing for the

possibility of crystal packing effects in the X-ray structures, the level of agreement between theory and experiment is high: Se-Se and Se-C bond distances differ by less than 1.0% (Figure 2.3, left)[6], as well as Se=P and P-C bond distances by 1.5% or less (Figure 2.3, right)[7]. All angles and dihedrals are also well reproduced.




| Bond | Bond length (Å) | | Variation |
|---|---|---|---|
| | Crystal | DFT | (%) |
| Se-Se | 2.303 | 2.295 | -0.3 |
| $Se_1$-$C_1$ | 1.988 | 2.001 | 0.7 |
| $Se_2$-$C_2$ | 1.986 | 1.984 | -0.1 |

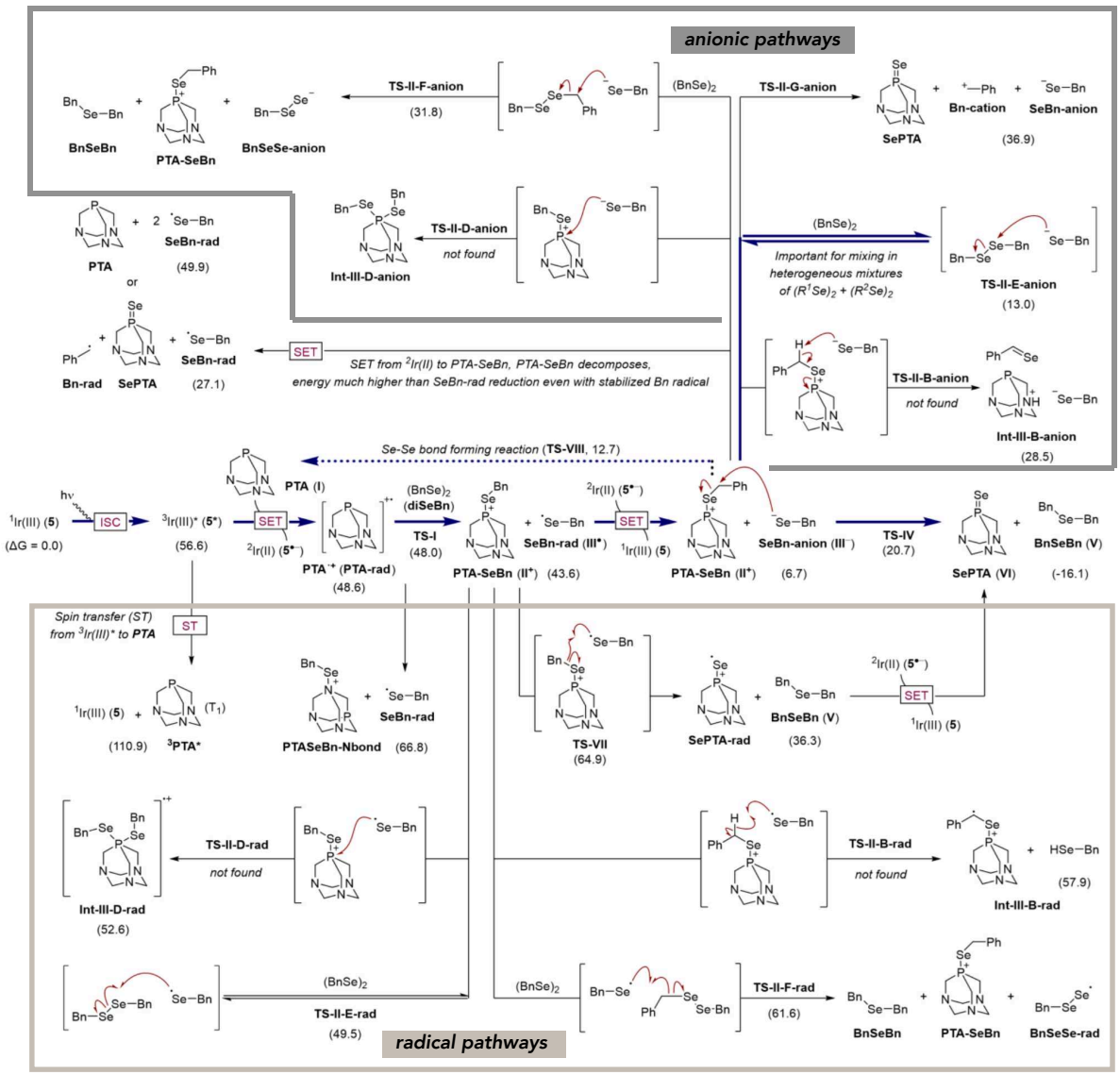



| Bond | Bond length (Å) | | Variation |
|---|---|---|---|
| | Crystal | DFT | (%) |
| Se=P | 2.136 | 2.109 | -1.3 |
| P-$C_1$ | 1.834 | 1.833 | -0.1 |
| P-$C_2$ | 1.822 | 1.850 | 1.5 |
| P-$C_3$ | 1.825 | 1.835 | 0.5 |

**Figure 2.3.** Overlay of the X-ray structure of 1,2-bis(2-bromobenzyl)diselane (Crystallography Open Database, COD, identifier: 2227239) (left) and 2-isobutyl-2-phosphabicyclo[3.3.1]nonane 2-selenide (COD identifier: 2221079) (right). Crystal structures are represented with standard colors and B3PW91-GD3(BJ)/6-31+G(d,p) (gas phase) optimized geometries are shown in purple. Tabulation of multiple bond lengths determined experimentally and computationally is also included.

### *2.3.2 Mechanistic Exploration with Model Substrate: $(BnSe)_2$*

In our mechanistic investigation, we began with a simplified system using $(BnSe)_2$ as the model substrate to explore a variety of possible pathways efficiently. We explored various anionic and radical pathways shown in Scheme 2.2 with estimated barriers for electron-transfer using protocols described previously by Houk and Buchwald[8], Vaissier[9] and a hybrid approach used in the final energy profile shown in section 2.3.3.




**Scheme 2.2.** Preliminary mechanistic exploration using $(BnSe)_2$ as the model substrate. This model was used to select viable pathways for further studies (marked with thick blue arrows). Prefix "SeBn-Model" was included in the names of intermediates and TSs (i.e. **SeBn-Model-TS-IV-anion** for the **TS-IV** step) in the raw thermochemistry data contained in the Appendix section. Free energy values reported in kcal/mol.

Following a photoexcitation of **5** to the triplet state, **PTA** undergoes a single-electron transfer (SET) to form the radical cation, **PTA•+**, species. Through a barrierless and exergonic process, the Se-Se bond is cleaved on diselenide forming a selanyl phosphonium cation (e.g., intermediate, **PTASeBn-Nbound**, bound to the nitrogen atom results higher in energy) and selanyl radical (**II+** and **III•**).

In the radical pathway, we considered forming the selenoether (**V**) product directly *via* (i) homolytic ($S_H2$) deselenization (**TS-VII**) and two radical propagation pathways using a second equivalent of diselenide at (ii) **TS-II-E-rad** and (iii) **TS-II-F-rad**. However, these three pathways proved to be less favorable than the highly exergonic and subsequent single-electron transfer (SET) from the catalyst to reduce the selanyl radical (**III•**) to selenolate anion (**III−**). Additional and unproductive pathways were also explored (e.g., **Int-III-D-rad** and **Int-III-B-rad**); however, they resulted higher in energy and were discarded.

Following the second SET to form the selenolate anion, we next explored the various anionic pathways shown in the gray box in Scheme 2.2. Unsuccessful pathways (e.g., **TS-II-F-anion** and **TS-II-G-anion**) that led directly to the selenoether (**V**) and various byproducts were less stable (e.g., $\Delta G$ = 31.8 and 36.9, respectively) than proceeding *via* the highly exergonic, $S_N2$ nucleophilic substitution process by attacking at the α-carbon center on the selanyl phosphonium species (e.g., $\Delta G$ = -16.1). However, if the selenoate anion attacks instead at the selenium center, this competitive pathway simply reforms the diselenide starting material (e.g., **TS-VIII** with a barrier of 12.7 kcal/mol).

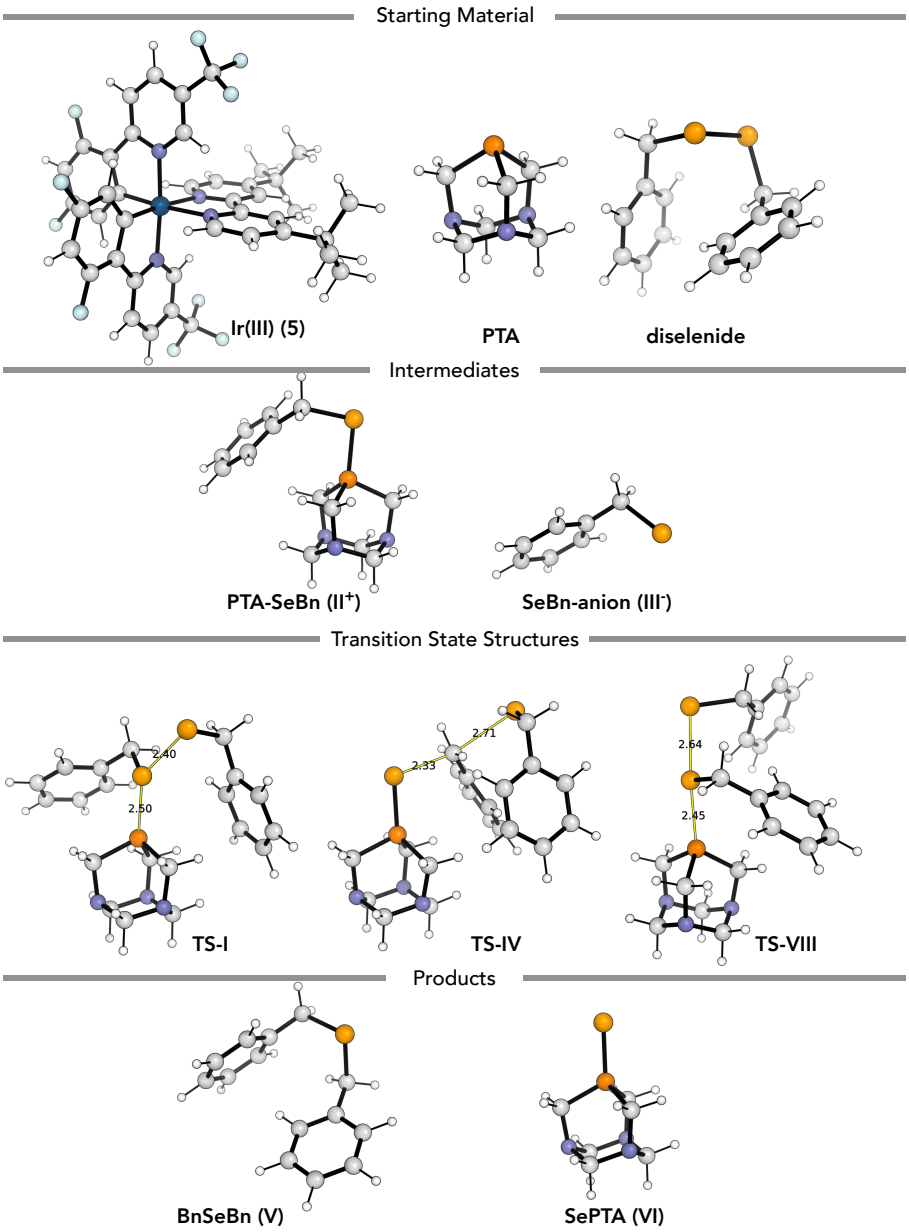
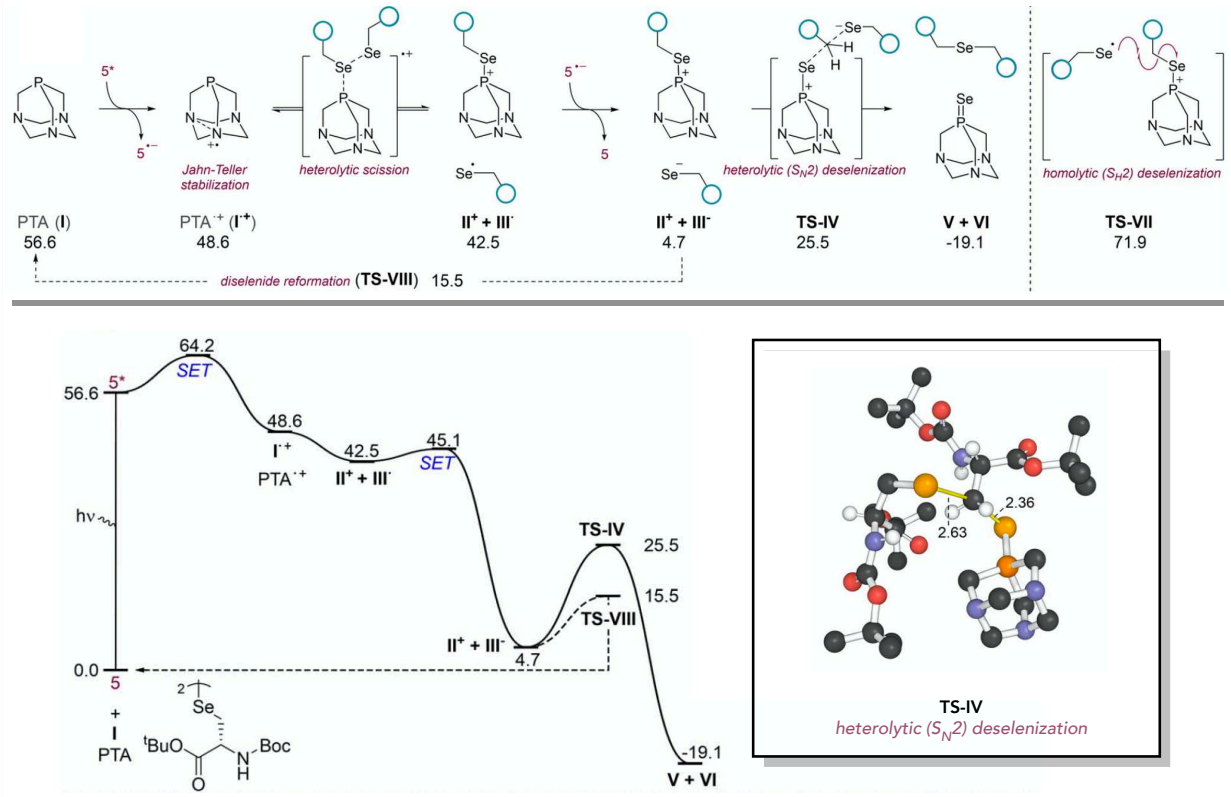



**Figure 2.4.** Representations of the most stable conformers of selected reaction steps for the model system using $(BnSe)_2$ substrate. Yellow thin lines represent bonds involved in the TSs. Distances shown in Å.

From this in-depth analysis, we selected the viable pathway shown in thick blue arrows in Scheme 2.2 as the main pathway to pursue additional mechanistic studies with the geometry optimized structures for initial starting points in Figure 2.4.

*2.3.3 Main and Competitive pathways for the Final Selenocysteine Derivative System: [N-Boc-L-Sec-OtBu]$_2$*

This following section was calculated by Dr. Juan V. Alegre-Requena which includes a summary of the main and competitive reaction pathway with [N-Boc-L-Sec-OtBu]$_2$ for the photocatalytic diselenide contraction reaction (Scheme 2.3) derived from the preliminary mechanistic studies in section 2.3.2.




**Scheme 2.3. Top:** Proposed mechanism for the PDC reaction that converts diselenides to selenoethers. **Bottom:** B3PW91- GD3(BJ)/6-311 + G(2df,p)//B3PW91-GD3(BJ)/6-31 + G(d,p) computed Gibbs free energy profile in kcal/mol (SMD MeCN solvent, 298.15 K) including main and competitive pathways for [N-Boc-L-Sec-OtBu]$_2$.

The long-lived T1 triplet state of catalyst **5** formed following photoexcitation is 56.6 kcal/mol above its ground state. Single-electron oxidation of **PTA** by **5*** is computed to occur favorably ($\Delta G = -8.0$ kcal/mol) with a small barrier, in line with the experimental PL spectroscopy data (Figure 2.1, right). Subsequently, the reaction of radical cation **PTA•+** and diselenide is barrierless

and exergonic by 6.1 kcal/mol, forming a selanyl phosphonium cation and selanyl radical (**II+** and **III•**). We considered a $S_H2$ homolytic substitution pathway involving these species to form the Se-C bond (via the transition state **TS-VII**), however, the barrier is unfavorable ($\Delta G^{\ddagger}$ = 29.4 kcal/mol). In contrast, back electron transfer from the catalyst to reduce the selanyl radical **III•** to selenolate anion **III−** is computed to be more facile ($\Delta G^{\ddagger}$ = 2.6 kcal/mol) and highly exergonic. From the selanyl phosphonium/selenolate ion-pair, $S_N2$ nucleophilic substitution can occur at the α-carbon center to form the contracted product (through **TS-IV**) or at selenium center to reform the diselenide starting material (**TS-VIII**). Additionally, the theoretical steric analysis (performed by Dr. Juan V. Alegre-Requena) at the site of diselenide contraction concluded that relatively small substituents on the other side of the diselenide ($\Delta G^{\ddagger}$ from 17.2 to 17.8 kcal/mol when a Me is replaced with a *t*Bu group) doesn't inhibit dimerization; on the other hand, two sterically demanding substituents cause a substantial increase in barrier height ($\Delta G^{\ddagger}$ of 25.4 kcal/mol). These computational studies raised the exciting possibility that the PDC reaction could be applied to the rapid and site-selective functionalization of large selenoproteins using smaller diselenides.

### *2.4 Conclusion*

In this chapter, we explored a synthetic transformation called the photocatalytic diselenide contraction (PDC) discovered by the Payne lab that enables the rapid and efficient conversion of diselenides to stable selenoethers with the formal extrusion of a selenium atom. Using QM methods, we interrogated various probable reaction pathways for the photocatalytic reactivity from empirical observations (e.g., PL spectroscopy) to substantiate the proposed mechanism. The feasibility and utility of this late-stage functionalization was demonstrated (e.g., within the published article by the experimental group) on various examples with larger polypeptide chains.

## REFERENCES


1. Dowman, L. J.; Kulkarni, S. S.; Alegre-Requena, J. V.; Giltrap, A. M.; Norman, A. R.; Sharma, A.; Gallegos, L. C.; Mackay, A. S.; Welegedara, A. P.; Watson, E. E.; Van Raad, D.; Niederacher, G.; Huhmann, S.; Proschogo, N.; Patel, K.; Larance, M.; Becker, C. F. W.; Mackay, J. P.; Lakhwani, G.; Huber, T.; Paton, R. S.; Payne, R. J., Site-selective photocatalytic functionalization of peptides and proteins at selenocysteine. *Nat. Commun.* **2022,** *13* (1).
2. Driscoll, D. M.; Copeland, P. R., MECHANISM AND REGULATION OF SELENOPROTEIN SYNTHESIS. *Annu. Rev. Nutr.* **2003,** *23* (1), 17-40.
3. (a) Steinmann, D.; Nauser, T.; Koppenol, W. H., Selenium and Sulfur in Exchange Reactions: A Comparative Study. *J. Org. Chem.* **2010,** *75* (19), 6696-6699; (b) Mousa, R.; Notis Dardashti, R.; Metanis, N., Selenium and Selenocysteine in Protein Chemistry. *Angew. Chem., Int. Ed.* **2017,** *56* (50), 15818-15827.
4. (a) Cohen, D. T.; Zhang, C.; Pentelute, B. L.; Buchwald, S. L., An Umpolung Approach for the Chemoselective Arylation of Selenocysteine in Unprotected Peptides. *J. Am. Chem. Soc.* **2015,** *137* (31), 9784-9787; (b) Cohen, D. T.; Zhang, C.; Fadzen, C. M.; Mijalis, A. J.; Hie, L.; Johnson, K. D.; Shriver, Z.; Plante, O.; Miller, S. J.; Buchwald, S. L.; Pentelute, B. L., A chemoselective strategy for late-stage functionalization of complex small molecules with polypeptides and proteins. *Nature Chemistry* **2019,** *11* (1), 78-85; (c) Levengood, M. R.; Van Der Donk, W. A., Dehydroalanine-containing peptides: preparation from phenylselenocysteine and utility in convergent ligation strategies. *Nat. Protoc.* **2006,** *1* (6), 3001-3010.
5. (a) 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; 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 Rev. C.01*, Wallingford, CT, 2016; (b) Luchini, G.; Alegre-Requena, J.; Funes-Ardoiz, I.; Paton, R., GoodVibes: automated thermochemistry for heterogeneous computational chemistry data [version 1; peer review: 2 approved with reservations]. *F1000Research* **2020,** *9* (291); (c) Heverly-Coulson, G. S.; Boyd, R. J., Systematic Study of the Performance of Density Functional Theory Methods for Prediction of Energies and Geometries of Organoselenium Compounds. *J. Phys. Chem. A* **2011,** *115* (18), 4827-4831; (d) Pearson, J. K.; Ban, F.; Boyd, R. J., An Evaluation of Various Computational Methods for the Treatment of Organoselenium Compounds. *J. Phys. Chem. A* **2005,** *109* (45), 10373-10379.
6. (a) Downs, R. T.; Hall-Wallace, M., The American Mineralogist crystal structure database. *American Mineralogist* **2003,** *88*, 247-250; (b) Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A., Crystallography Open Database – an open-access collection of crystal structures. *Journal of Applied Crystallography* **2009,** *42* (4), 726-729; (c) Hua, G.; Fuller, A. L.; Slawin, A. M. Z.;

Woollins, J. D., 1,2-Bis(2-bromobenzyl)diselane. *Acta Crystallographica Section E Structure Reports Online* **2010,** *66* (10), o2579-o2579.
7. Bungu, P. N.; Otto, S., 2-Isobutyl-2-phosphabicyclo[3.3.1]nonane 2-selenide. *Acta Crystallographica Section E Structure Reports Online* **2009,** *65* (3), o560-o561.
8. Jones, G. O.; Liu, P.; Houk, K. N.; Buchwald, S. L., Computational Explorations of Mechanisms and Ligand-Directed Selectivities of Copper-Catalyzed Ullmann-Type Reactions. *J. Am. Chem. Soc.* **2010,** *132* (17), 6205-6213.
9. Vaissier, V.; Barnes, P.; Kirkpatrick, J.; Nelson, J., Influence of polar medium on the reorganization energy of charge transfer between dyes in a dye sensitized film. *Phys. Chem. Chem. Phys.* **2013,** *15* (13), 4804.

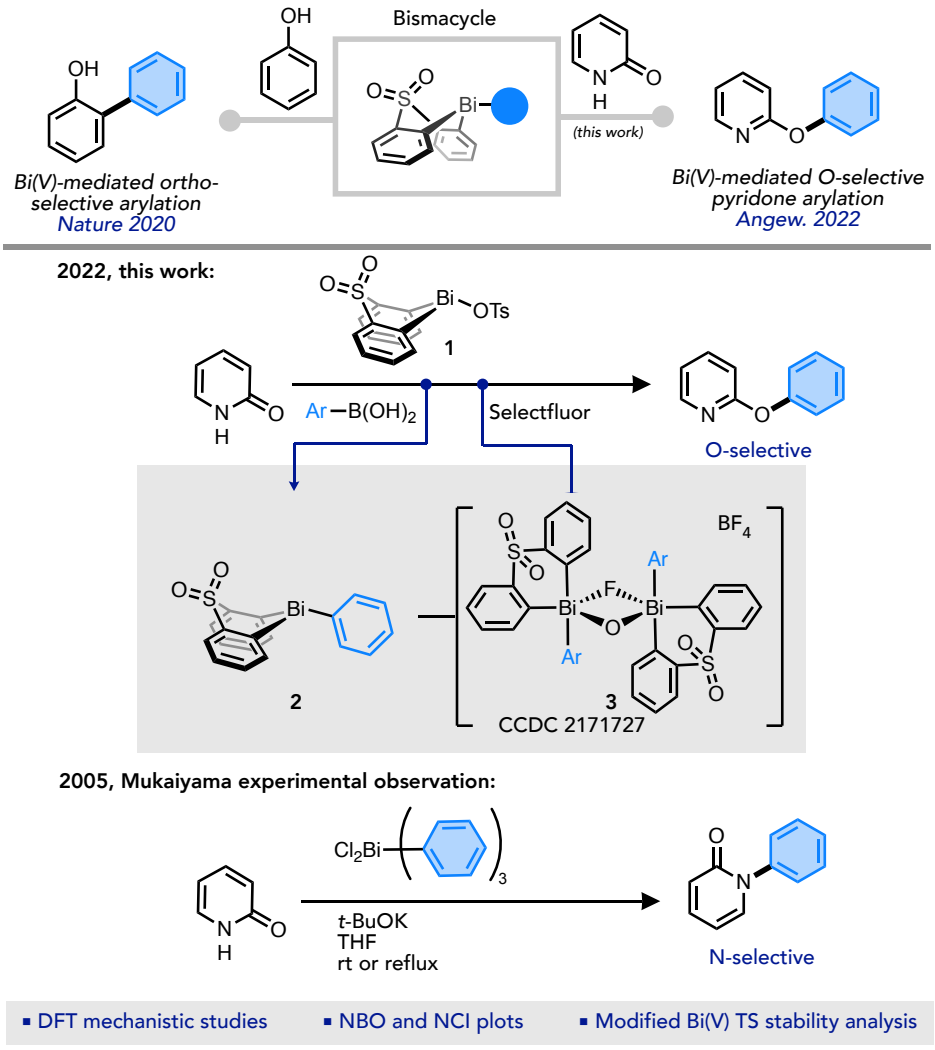
# 3 BISMUTH(V)-MEDIATED SELECTIVE ARYLATION OF 2- AND 4-PYRIDONES

## *3.1 Chapter Overview*

In this chapter, we begin the journey into new bismuth chemistry developed by the Ball group. Their bismuth arylation reaction published in *Nature*[1] transformed the following collaborative work discussed in Chapters 3-4 ranging from the computational protocols implemented in selectivity problems to the versatile chemical reactivity originated from bismuth(V) reagents. Using QM methods, we investigate this new reactivity with organobismuth reagents to propose a reaction mechanism for the selective arylation of 2- and 4-pyridiones. This collaborative research was published in *Angewandte Chemie.*[2]

## *3.2 Introduction*

In 2020, the Ball group developed an approach to synthesize reactive Bi(V) arylating agents that tolerate a variety of synthetically useful functional groups under mild and air-stable conditions from low-cost and commercially available starting materials for the arylation of phenols.[1] Following this discovery, the experimental group applied these reactive organobismuth reagents to 2- and 4-pyridones and developed a convenient strategy for selective O-arylation with arylboronic acids.[2] In this collaborative effort, we performed a computational investigation of (i) the reaction pathways for the Bi(V)-mediated arylation of 2- and 4-pyridones, (ii) the C-O and C-N arylations using the bridged-sulfone bismacycle and triphenylbismuth-derived reagents to confirm the current and Mukaiyama's reversed experimental observations,[3, 4] and (iii) the origin of the selectivity of non-symmetrical substituted 2-pyridones by the different organobismuth reagents using DFT and natural population analysis calculations (Scheme 3.1).




**Scheme 3.1. Top:** Bi(V)-mediated ortho selective and O-selective (this work) pyridone arylation reactions developed by the Ball group using the air, stable Bi(III) reagent. **Bottom:** Modular, O-selective Bi(V)-mediated pyridone arylation and Mukaiyama's N-selective pyridone arylation.

The current synthetic methods for these pyridyl ether motifs are accessed through conventional connections between pyridines and phenols via nucleophilic aromatic substitutions ($S_NAr$) and C-C cross-couplings. Instead, in a complementary and creative strategy to the existing methods, the Ball group reversed the disconnection strategy resulting in a moderately nucleophilic pyridone and a reactive, electrophilic Bi(V) arylating reagent for selectively synthesizing pyridyl ethers via a umpolung strategy. Furthermore, the Bi(V)-mediated O-selective arylation reaction exhibits efficient Bi(III) precursor recovery, and it showcases an efficient synthetic strategy for four

biologically relevant compounds that were previously reported. All experimental work was performed by our experimental collaborators (Ball *et. al*) and all DFT computational work was performed by the present author.

### *3.3 Results and Discussions*

The regioselectivity observed for pyridone O-arylation with the present bismacyclic system stands in stark contrast to the N-selectivity previously reported using simple triarlybismuth(V) reagents. In order to elucidate the origin of these differences, DFT calculations (ωB97XD/def2-QZVPP//ωB97XD/6-31+G(d,p) and def2-SVP (Bi)) with the SMD-solvation model (solvent = acetonitrile, MeCN, for the bismacycle system and solvent = tetrahydrofuran, THF, for the $Ph_3BiCl_2$ system) were applied to all calculations to investigate the key features of each arylation pathway.

#### *3.3.1 Benchmarking to bismuth (III) and (V) species*

We began by evaluating the selected level of theory for describing the structures of bismuth(III) and bismuth(V) species by comparison to the X-ray crystallographic structures of the aryl bismacycle(III) and the bismuth(V) dimer (Figure 3.1).[1] The computed structures for both methods shown in Table 3.1 resulted in a high level of agreement between theory and experiment as all angles and dihedrals are well reproduced, and: (i) with def2-SVP for the Bi atom, the Bi-C bond distances differ by less than 1.5% and Bi-O distances by less than 2.5%; (ii) with def2-TZVP for the Bi atom, the Bi-C and Bi-O bond distances differ by less than 0.5%. Given that using the def2-TZVP basis set for the Bi atom resulted in higher accuracy (relative to the X-ray crystallographic structures) for both the Bi(III) and Bi(V) species, subsequent optimizations were performed using this method to evaluate the selectivity determining steps for the arylation of 2-pyridiones by both

the bismacyclic system and $Ph_3BiCl_2$. These additional Gibbs free energy values with def2-TZVP are reported in parentheses under the relevant reaction steps.

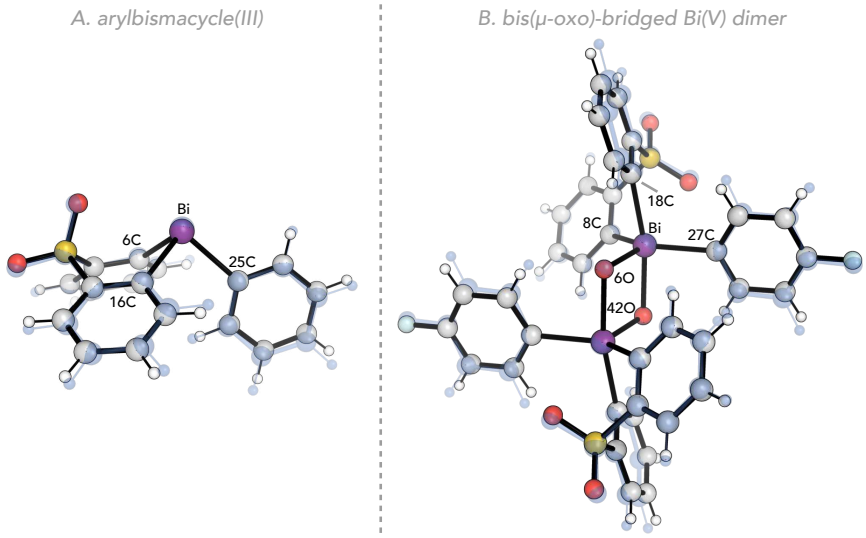



**Figure 3.1.** Overlay of the X-ray structure of aryl bismacycles (Crystallography Open Database, COD, identifier: 1904063) (left) and bis(μ-oxo)-bridged Bi(v) dimer (COD identifier: 1904073) (right). Crystal structures are represented with standard colors and optimized geometries with (SMD-acetonitrile) ωB97X-D/6-31+G(d,p), def2-SVP (Bi atom) are shown in blue.

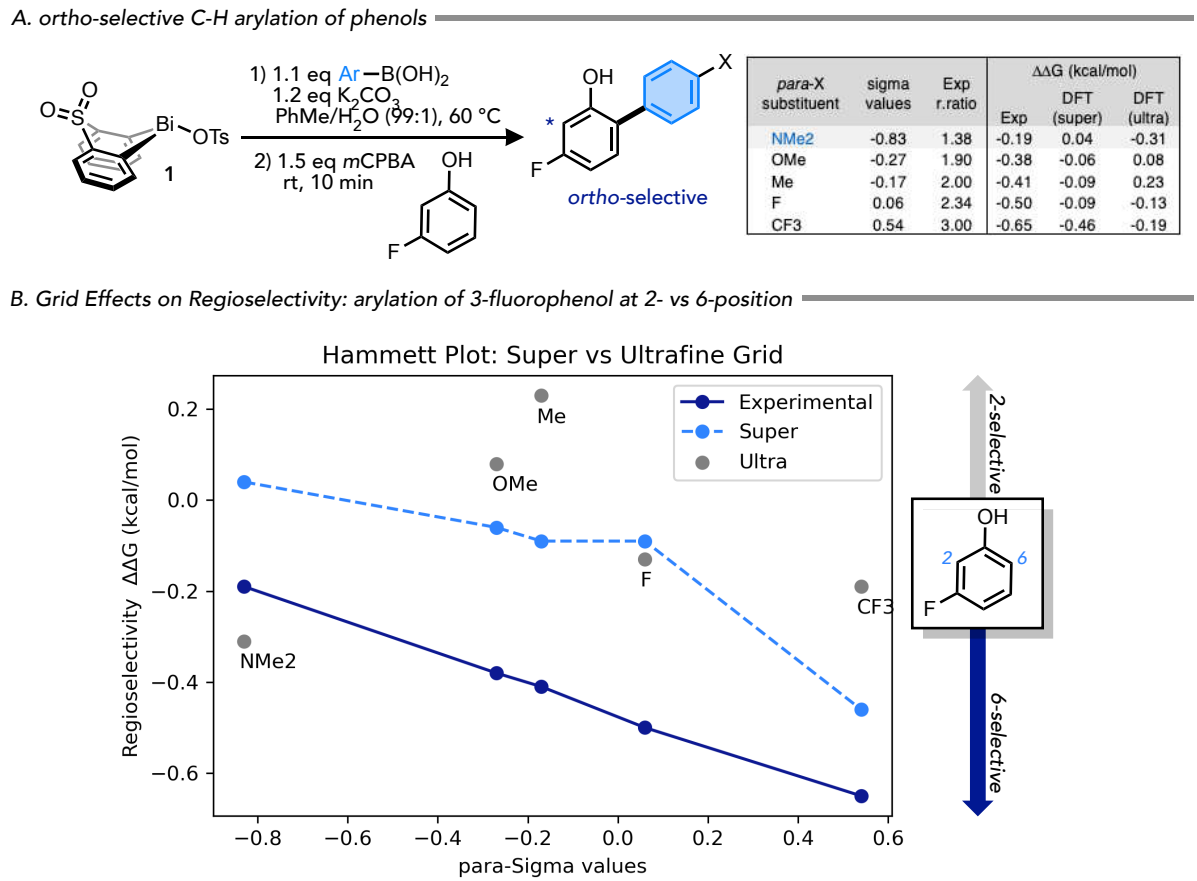
**Table 3.1**. Tabulation of the bond lengths (Å) and the percentage of bond variation from the X-ray crystal structures shown in parenthesis. Bond lengths were determined experimentally from the Bi(III) and Bi(V) dimer X-ray structures and computationally using optimized structures from different levels of theories for bismuth.

| Bonds | Bond length (Å) Bi-crystal | Bond length (Å) and Bond Variation (%) (SMD-MeCN) WB97XD/ 6-31+G(d,p) (C,H,S,O,F) | |
|---|---|---|---|
| | | def2-SVP (Bi) | def2-TZVP (Bi) |
| Bi-25C | 2.249 | 2.270 (0.9) | 2.248 (0.0) |
| Bi-16C | 2.266 | 2.293 (1.2) | 2.271 (0.2) |
| Bi-6C | 2.272 | 2.293 (0.9) | 2.271 (0.0) |
| Bi-27C | 2.201 | 2.218 (0.8) | 2.198 (-0.1) |
| Bi-18C | 2.282 | 2.299 (0.8) | 2.287 (0.2) |
| Bi-8C | 2.240 | 2.248 (0.3) | 2.231 (-0.4) |

| | | | |
|---|---|---|---|
| Bi-42O | 2.202 | 2.249 (2.2) | 2.204 (0.1) |
| Bi-6O | 2.030 | 2.053 (1.1) | 2.029 (-0.1) |

*3.3.2 Benchmarking for DFT integration grid effects*

During the course of our benchmarking investigation, we found inconsistent results between the calculated and the experimentally obtained selectivities for the ortho-selective C-H arylation reaction of 3-fluorophenol (Figure 3.2A) when applying an "ultrafine" integration grid.




| *para*-X substituent | sigma values | Exp r.ratio | ΔΔG (kcal/mol) Exp | DFT (super) | DFT (ultra) |
|---|---|---|---|---|---|
| NMe2 | -0.83 | 1.38 | -0.19 | 0.04 | -0.31 |
| OMe | -0.27 | 1.90 | -0.38 | -0.06 | 0.08 |
| Me | -0.17 | 2.00 | -0.41 | -0.09 | 0.23 |
| F | 0.06 | 2.34 | -0.50 | -0.09 | -0.13 |
| CF3 | 0.54 | 3.00 | -0.65 | -0.46 | -0.19 |

**Figure 3.2. Top:** Hammett *para*-sigma values plotted against the relative free-energy for the *ortho*-selective C-H arylation of 3-fluorophenol coupled with the *para*-substituted aryl boronic acids at the 6-position. **Bottom:** the experimental (dark blue) and computed relative free-energy differences using an "ultrafine" (gray) and "superfine" (light blue) integration grid.

In the previously reported Hammett plot studies for the arylation of 3-fluorophenol (Figure 3.2B), this regioselectivity study produced a strong linear relationship to the *para*-substituted

sigma values indicating the apparent preference for arylating at the 6-position (e.g., more electron-rich and less sterically encumbered site).[1] Therefore, in the attempt to benchmark and replicate this solid regioselectivity trend (e.g., similar bismacycle system to this work) by measuring the Gibbs energy differences of the computed reaction barriers at the selectivity-determining step, we obtained a non-linear relationship with opposite selectivities for 4-methyl and 4-methoxy boronic acids (e.g., Me and OMe on Figure 3.2B) when using the default pruned (99,590) ultrafine grid. We investigated this further by inspecting the TS conformers for 4-$NMe_2$ boronic acid since it resided outside the linear trend to the rest of the data points.

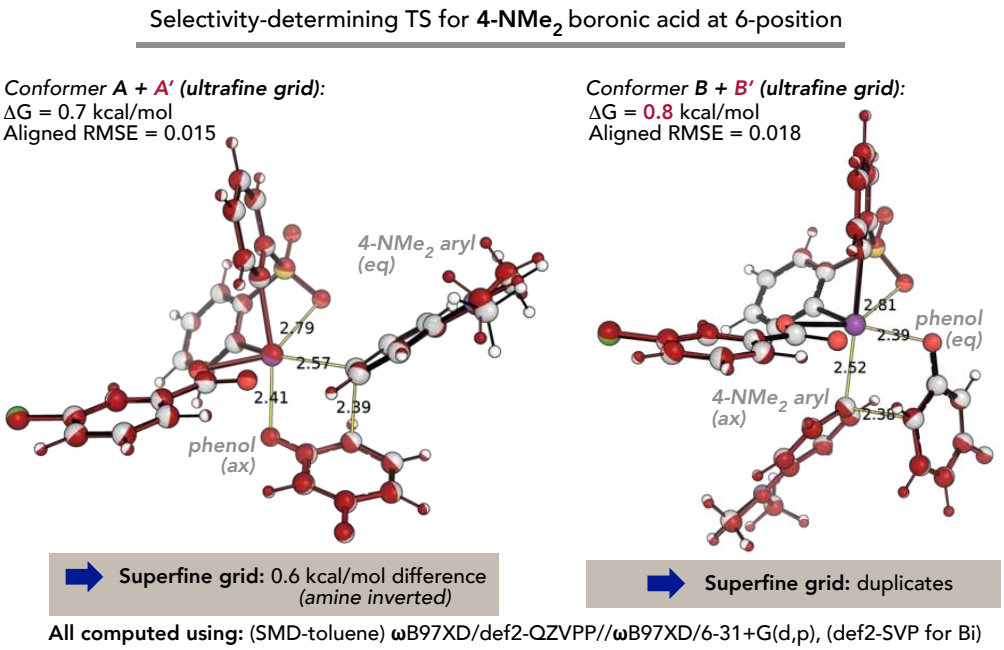



**Figure 3.3.** Most stable TS conformers for the C-H arylation for 3-fluorophenol with 4-$NMe_2$ boronic acid at the 6-position. Conformers in red are higher in energy. **Left:** conformers A and A' (in red) aligned with the phenol in the axial position. **Right:** conformers B and B' (in red) aligned with the 4-$NMe_2$ boronic acid in the axial position.

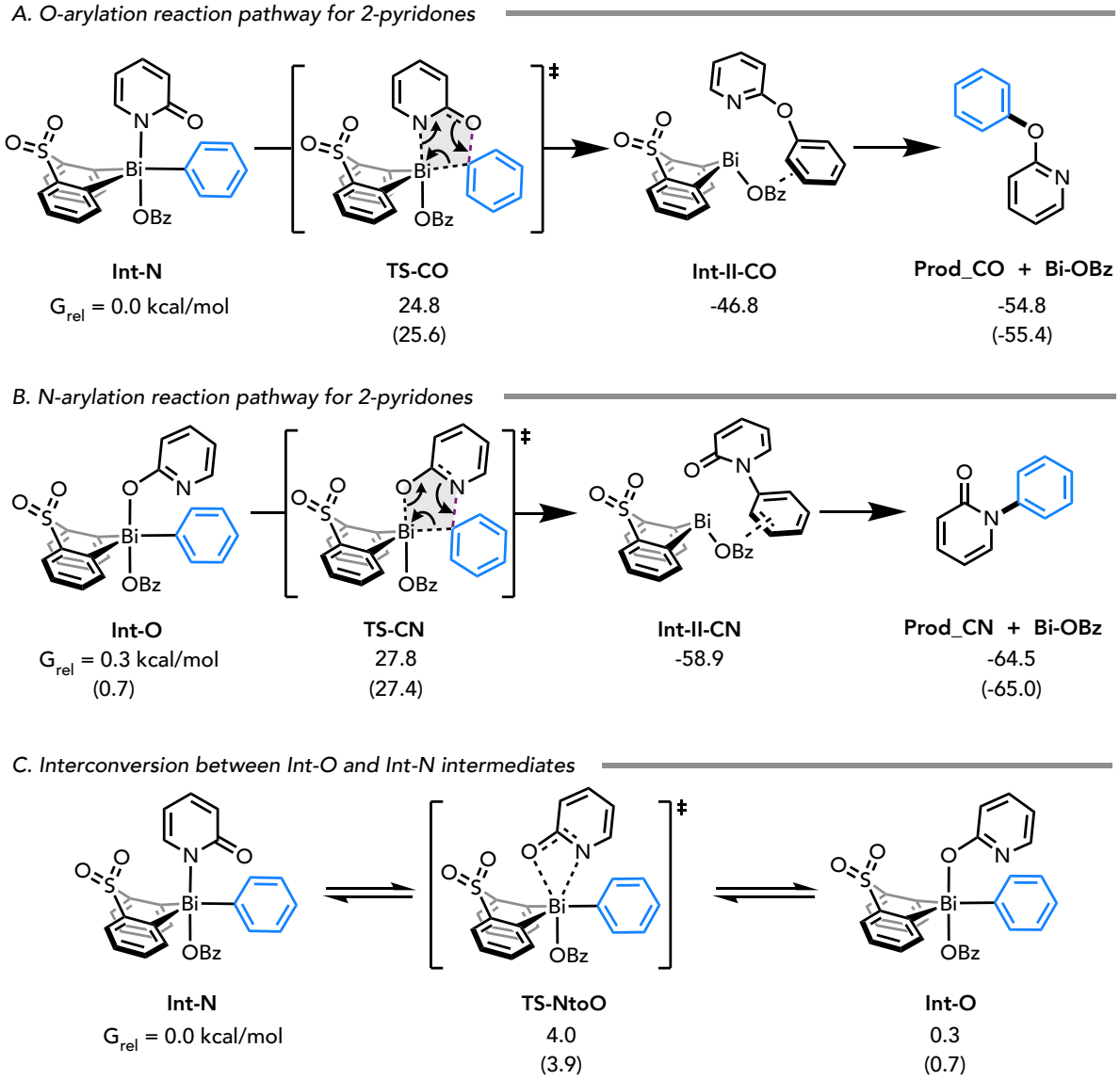
The two most stable TS conformations contributing to the selectivity at the 6-position (Figure 3.3) for the boronic aryl substrate scope consisted of (i) conformers A (or A' shown in red) with the 3-fluorophenol in apical position coupling with the aryl substrate in the equatorial position (e.g., 4-$NMe_2$) and (ii) conformers B (or B' shown in red) with *vice a versa* conformation to A.

However, the Gibbs energy difference for the 4-$NMe_2$ boronic acid resulted into two sets of conformers, finding conformer A' and B' higher in energy by 0.7 and 0.8 kcal/mol, respectively, when using an ultrafine grid. Despite having a similar alignment with a minimal RMSE deviation of 0.015 and 0.018, these two sets of conformers were classified as different conformers due to its high energy difference. After recomputing these conformers with a "superfine" grid, conformers B and B' resulted as duplicates with similar structure alignment and Gibbs energy values. We expected conformers A and A' to have distinct energies as the amine is inverted undergoing the umbrella effect. Wheeler *et al* previously reported this discrepancy for popular integration grid sizes (e.g., pruned (75,302) grid and in some cases a (99,590) grid) showing large uncertainties in computed free energies for selectivity reactions. Similarly, for an Iridium catalyzed C–H functionalization of indoles at either C2 or C7, the relative free energy of the regiocontrolling TS structures, $\Delta\Delta G^{\ddagger}$, resulted in large variations ranging from 0.6 to 2.3 kcal/mol when using an "ultrafine" grid. The authors state that the sensitivity of DFT free energies is related to the sensitivity of vibrational frequencies (e.g., low-frequency vibrational modes) and the magnitude of grid errors can be reduced by its integration size. The regioselectivity drastically improved in our Hammett plot benchmark studies when utilizing the larger "superfine" integration grid (Figure 3.2B in the blue, dashed line) resulting in a similar linear selectivity trend as those obtained experimentally.

Therefore, for this bismuth chemistry, all DFT calculations used a "superfine" pruned (175,974) grid for numerical integration. The default "ultrafine" grid is not recommended for this system due to lack of rotational invariance, resulting in orientational dependence of Gibbs energy differences.[5]

*3.3.3 Bismacycle-based 2-pyridone arylation*

Our mechanistic investigation began by evaluating the reaction pathways for 2-pyridone O- and N-arylation (Scheme 3.2, A and B) starting from the proposed bismuth(V) intermediate that is complexed to 2-pyridone and benzoate.




**Scheme 3.2.** Computed O- (A) and N-arylation (B) reaction pathways with the bismacycle scaffold. Small barrier for interconversion between the O-bound and N-bound coordination modes (C). Boltzmann weighted G of relevant reaction steps calculated using the SMD solvation model (solvent = MeCN) with ωB97XD/def2-QZVPP// ωB97XD/6-31+G(d,p) and def2-SVP (Bi) at 80 °C. Gibbs free energy values in parentheses were calculated using (SMD, MeCN) with ωB97XD/def2-QZVPP// ωB97XD/6-31+G(d,p) and def2-TZVP (Bi) at 80 °C.

For the bismacyclic system, the N- and O-bound Bi(V)-pyridonate isomers both adopt trigonal bipyramidal geometries in which the pyridonate and benzoate ligands are located apically, and the aryl substituents are equatorial (Figure 3.6). The 2-pyridone can coordinate to the bismuth(V) complex through either the nitrogen, **Int-N**, or oxygen atom, **Int-O** (Scheme 3.2 C). While there is a slight preference for the pyridonate intermediate to be O-bound (0.3 kcal/mol), the barrier for conversion between the isomers is small (4.0 kcal/mol) relative to the barriers for arylation. Therefore, a Curtin-Hammett scenario is operative.

Arylation proceeds preferentially via cyclic 5-membered transition structures (TSs) such that the N-bound Bi(V)-pyridonate leads to O-arylation and vice versa (Scheme 3.2, A and B). This step is highly exergonic (e.g., irreversible arylation step) and selectivity-determining. Consistent with experiment, O-arylation is predicted to be kinetically favored ($\Delta\Delta G^{\ddagger}$ = 3.0 kcal/mol).

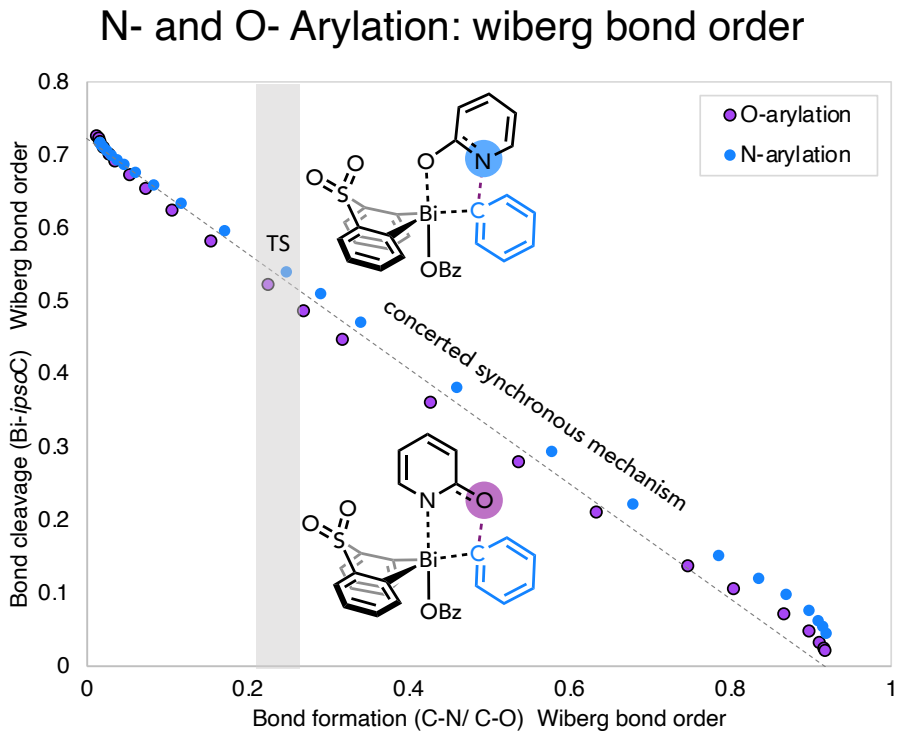



**Figure 3.4.** More O'Ferrall-Jencks plot indicating the nature of the transition states as it proceeds along the reaction coordinates from the ground state 2-pyridone bound intermediate to the products. The calculated Wiberg bond orders reveal a concerted synchronous mechanism as the Bi–*ipso*C bond breaks and the C-N and C-O bond forms to give the N-arylation and O-arylation products, respectively.

In both TSs, cleavage of the Bi-$C_{ipso}$ bond and formation of the new $C_{ipso}$-O or $C_{ipso}$-N bond occur concertedly. As the transition state proceeds into **Int-II**, the Bi-$C_{ipso}$ bond breaking (e.g., y-axis) and the $C_{ipso}$-O or $C_{ipso}$-N bond forming (e.g., x-axis) occurs linearly in the More O'Ferrall-Jencks plot shown in Figure 3.4.

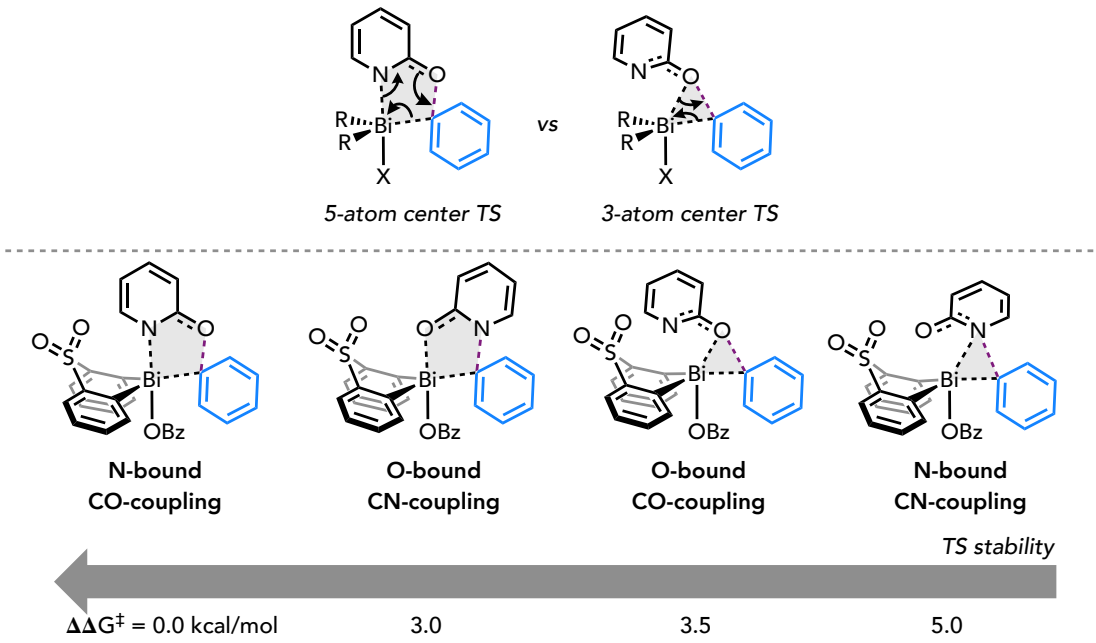



**Figure 3.5.** Relative stabilities of O- and N-arylation TSs with 5- and 3-membered cyclic structures. Boltzmann weighted G values were selected for comparison in the stability analysis.

For the arylation step, we also surveyed 3- and 5-membered transition state structures (Figure 3.5 and Figure 3.6), finding the 5-membered TS most stable for O- and N- arylation pathways (**TS-CO** and **TS-CN**). While this alternative pathway rationalizes the ability of 4-pyridones to undergo arylation, it is less favorable than the 5-membered TS for 2-pyridone (by 3.5 kcal/mol).

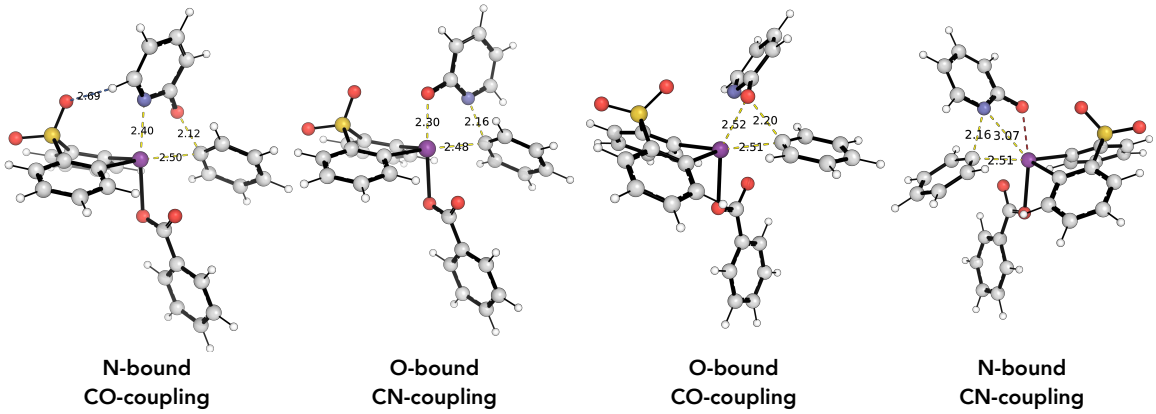
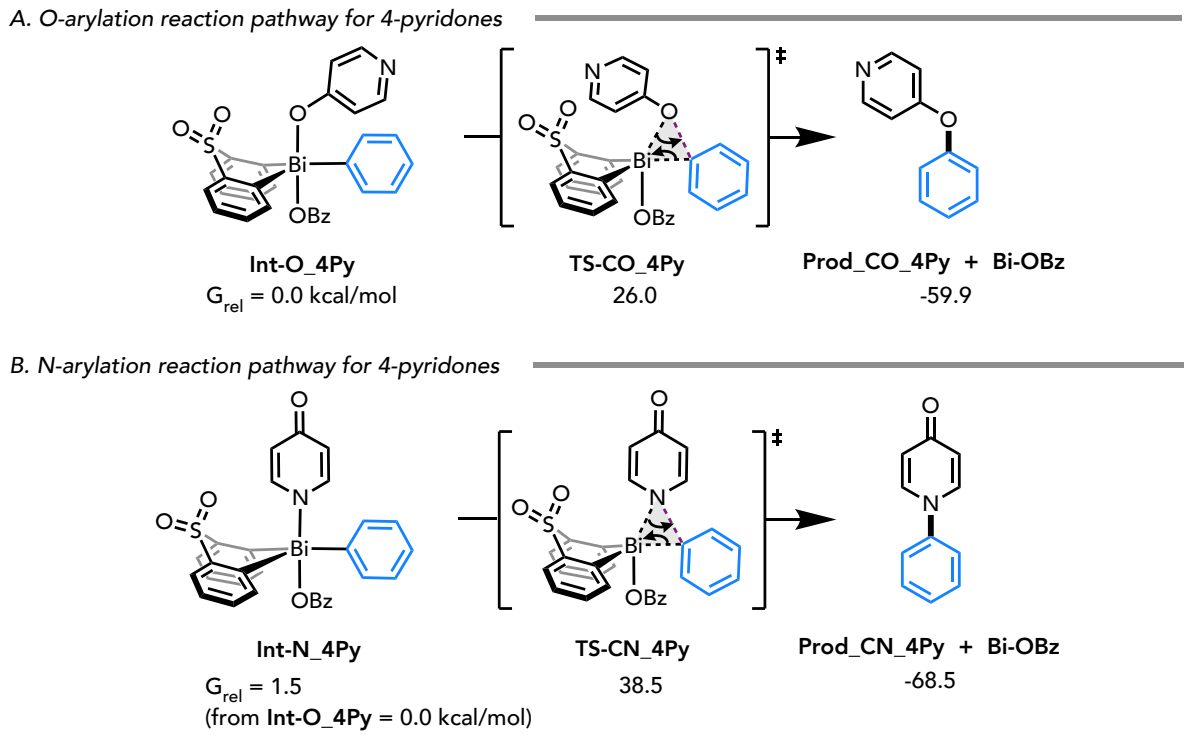



**Figure 3.6.** Most stable TS conformers for the bismacyclic system with 2-pyridone. Dashed yellow lins show breaking/forming bonds (in Å). The blue dashed lines represent hydrogen interaction lengths in Å. The red dash line shows a Bi-O interaction (2.35 Å).

*3.3.4 Bismacycle-based 4-pyridone arylation*

We evaluated the reaction pathways for 4-pyridone O- and N-arylation (Scheme 3.3 and Figure 3.7). Based on the method reported by Mukaiyama and its postulated transformation for 4-pyridones, we calculated the arylation TS *via* a 3-membered TS cyclic structure. We found the N-arylation reaction pathway for 4-pyridiones is a highly unfavored transformation and, consistent with experiment, O-arylation is favored ($\Delta\Delta G^{\ddagger}$ = 12.5 kcal/mol).




**Scheme 3.3.** Computed O- (A) and N-arylation (B) reaction pathways with the bismacycle scaffold for 4-pyridione. Free energies of relevant reaction steps calculated using the SMD solvation model (solvent = MeCN) with ωB97XD/def2-QZVPP// ωB97XD/6-31+G(d,p) and def2-SVP (Bi) at 80 °C.

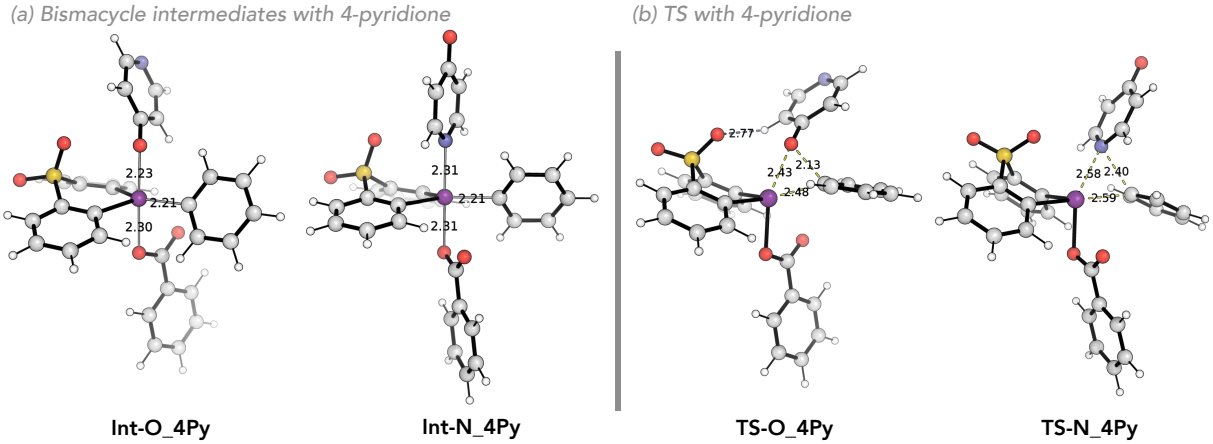
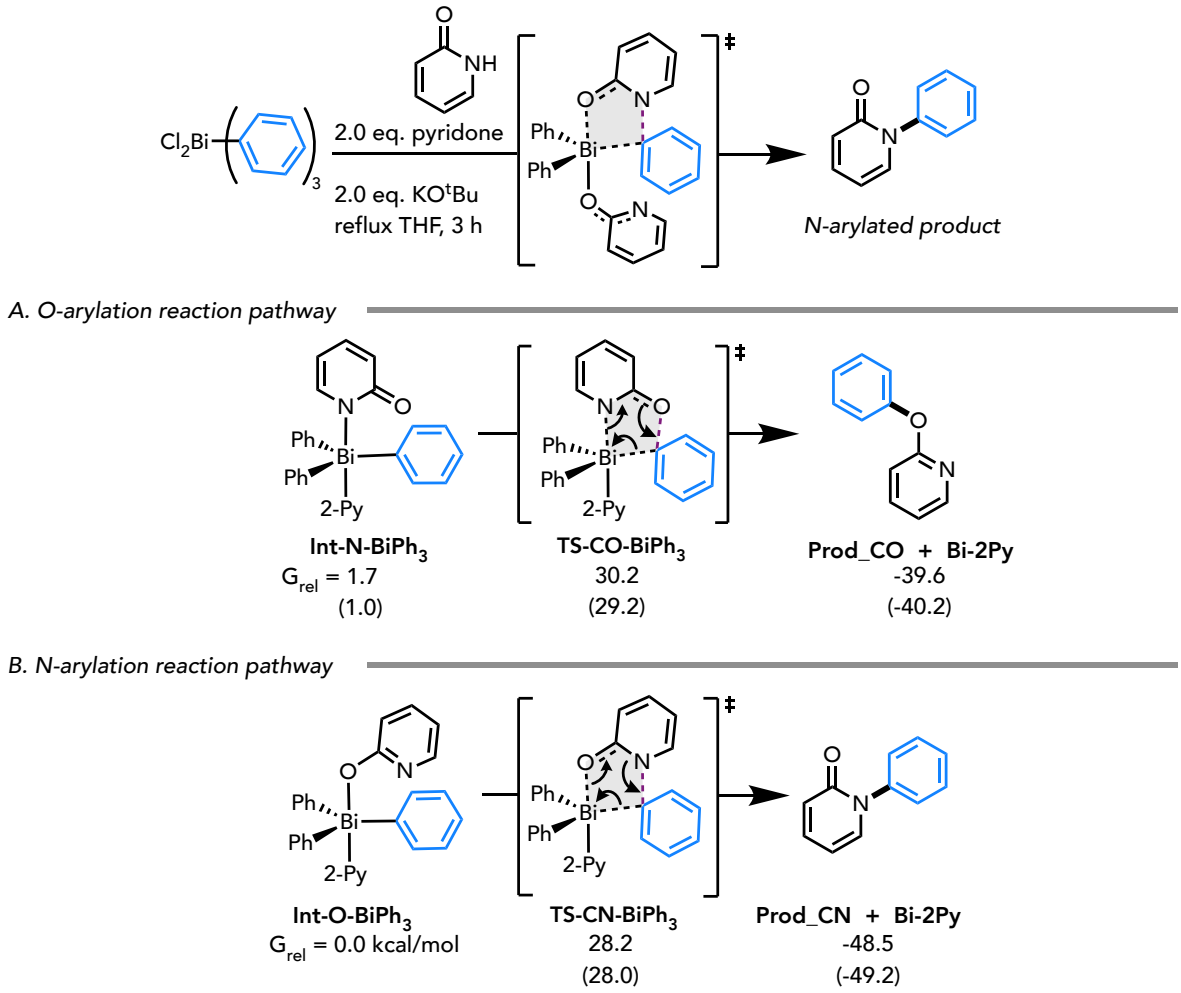



**Figure 3.7.** Representations of the Bi(V)-complex ground-state intermediates and TSs for the bismacyclic system with 4-pyridone. Dashed yellow lines show breaking/forming bonds (in Å). The blue dashed lines represent hydrogen interaction lengths in Å.

*3.3.5 Triphenyl Bi(V)-based 2-pyridone arylation*




**Scheme 3.4.** Computed O- (A) and N-arylation (B) reaction pathways with $Ph_3BiCl_2$. Boltzmann weighted G of relevant reaction steps calculated using the SMD solvation model (solvent = THF) with ωB97XD/def2-QZVPP// ωB97XD/6-31+G(d,p) and def2-SVP (Bi) at 60 °C. Gibbs free energy values in parentheses were calculated using (SMD, THF) with ωB97XD/def2-QZVPP// ωB97XD/6-31+G(d,p) and def2-TZVP (Bi) at 60 °C. 2-Py, 2- pyridyloxy.

The analogous arylation TSs for a simple triphenylbismuth-derived reagent were also calculated (Scheme 3.4, Figure 3.8 and 3.9). While 5- membered TSs are again favored, N-arylation is now kinetically favored over O-arylation (by 2 kcal/mol), consistent with Mukaiyama's observations.

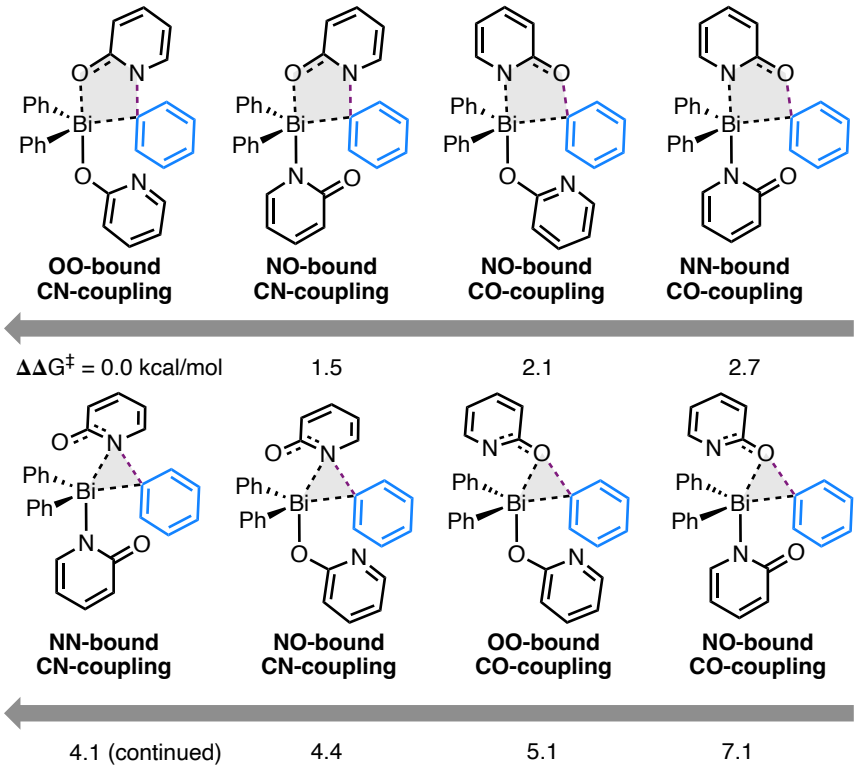



**Figure 3.8.** Relative stabilities of O- and N-arylation TSs with 5- and 3-membered cyclic structures using the triphenylbismuth reagent from Mukaiyama's reaction.

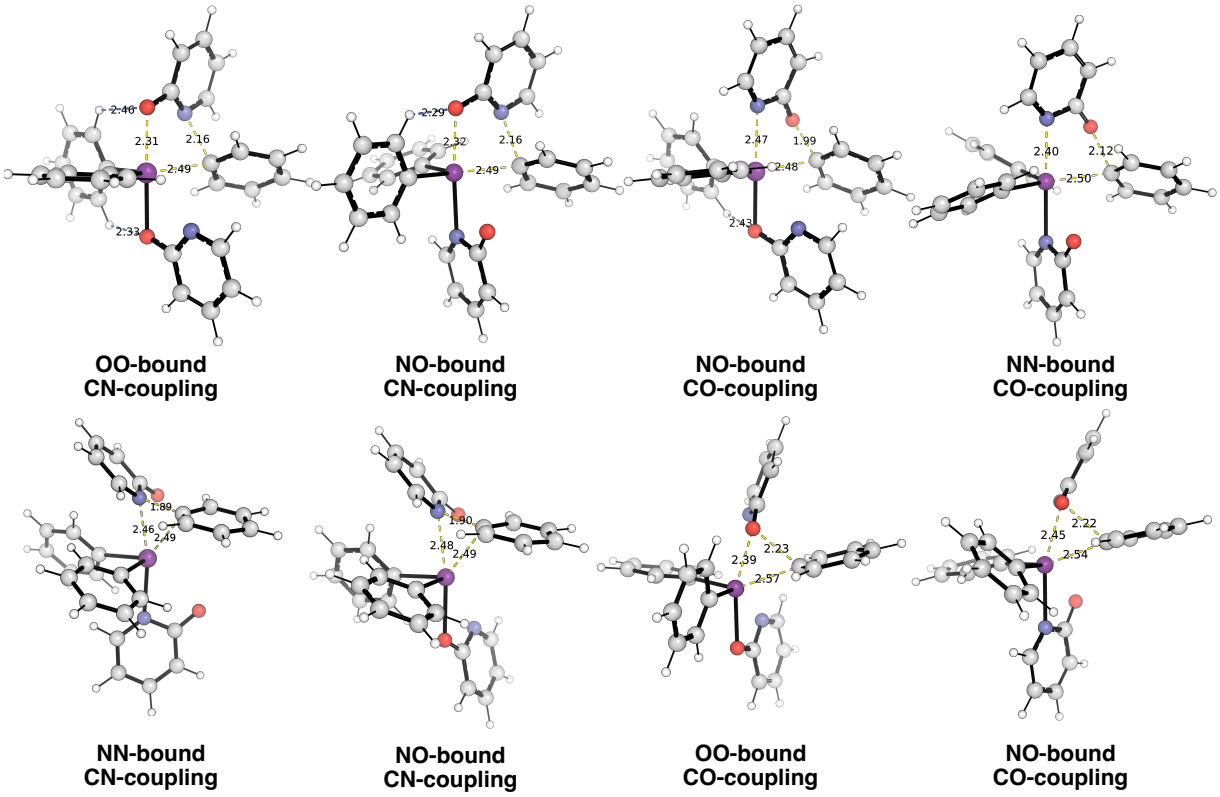



**Figure 3.9.** Representations of the most stable TS conformers for the $Ph_3BiCl_2$ system with 2-pyridone. The yellow dash lines represent bonds forming and breaking with corresponding bond lengths in Å. The blue dashed lines represent hydrogen interaction lengths in Å.

*3.3.6 Favorable INT and TS Interactions Determining Selectivity*

We began by inspecting the stabilities of the intermediate (INT) structures leading up to the transition state structures (TS) at the selectivity-determining step shown in Figure 3.10 and 3.11. The reaction intermediates forming the selective product are favored in both the bismacycle and triphenylbismuth system by 0.3 and 2.1 kcal/mol, respectively.

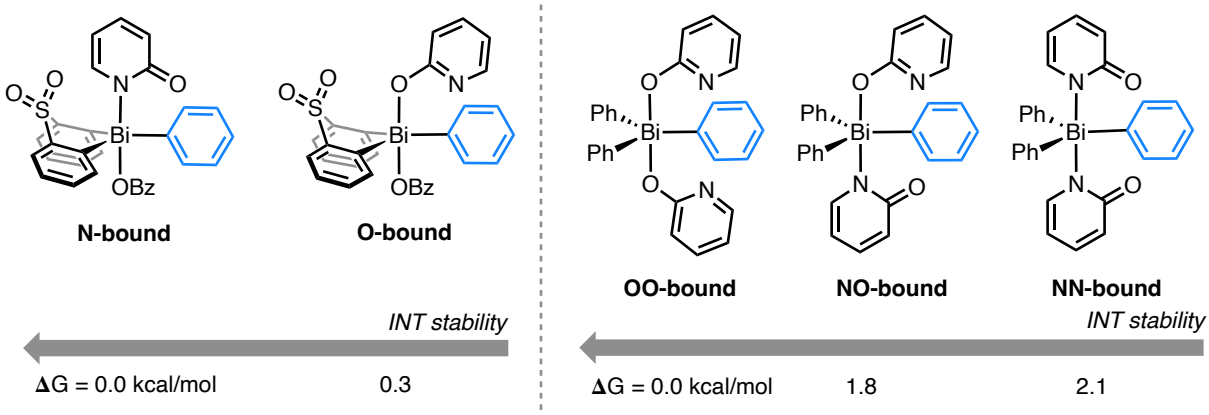



**Figure 3.10.** Ground-state Bi(V)-complex intermediates for both the bismacycle (on left) and the $Ph_3BiCl_2$ (on right). Boltzmann weighted G values were selected for comparisons in the stability analysis.

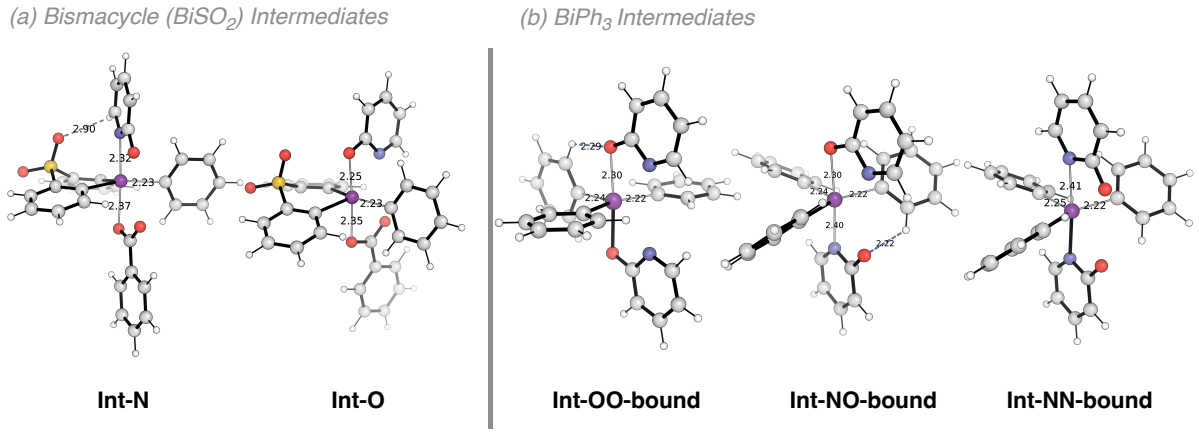



**Figure 3.11.** Representations of the most stable ground-state Bi(V)-complex intermediates for both the bismacycle (on left) and the $Ph_3BiCl_2$ (on right) with corresponding bond lengths in Å. The blue dashed lines represent hydrogen interaction lengths in Å.

Now, to explain the observed selectivity, we compared single-point energy calculations of modified TSs (Figure 3.12) to calculate the influence of the pyridone, spectator ions, and bismacycle geometry in stabilizing the TSs.

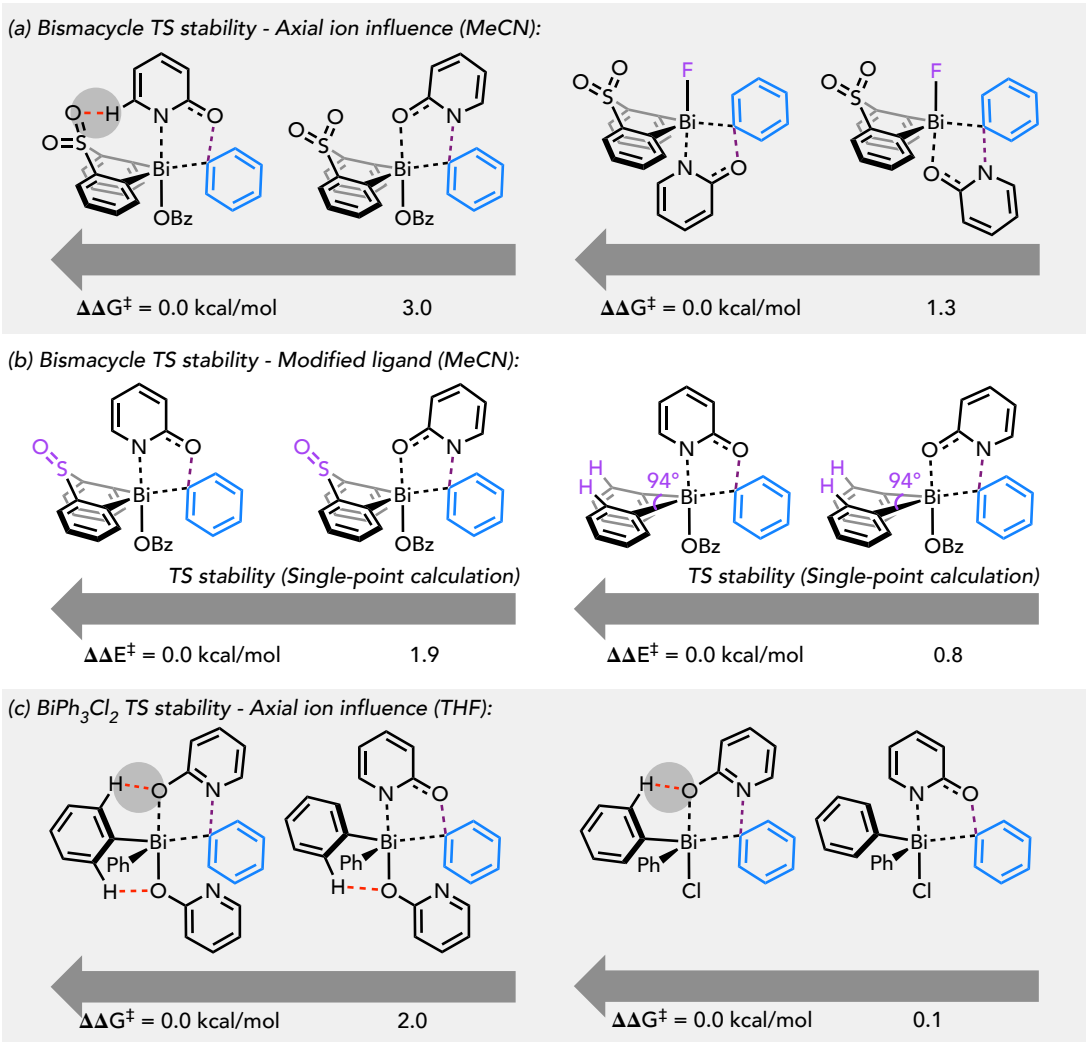



**Figure 3.12.** Comparison of the TS stabilities for both the O- and N-arylated products: (a) with benzoate and fluoride ion; (b) with a modified and removed sulfone group; (c) to the $BiPh_3$ model system with two 2-pyridones and chloride ion. The most stable TS energies were selected for comparison for the optimized structures.

Based on the optimized TS structures, there is a reduction in energy (1.7 kcal/mol) with the fluoride ion (-F) in Figure 3.12A, which can be partially explained by the lack of a hydrogen bond interaction between the 2-pyridone and the sulfone group.

The geometry of the bismacycle is also a major factor (Figure 3.12B); for example, computationally deleting the sulfone group but retaining the geometry of the resulting $BiPh_2$ unit

is still predicted to favor O-arylation. With the sulfone group completely removed, the selectivity is reduced by a total 2.2 kcal/mol.

The spectator ion was also studied in the $BiPh_3$ system between two bound 2-pyridones and the chloride ion, -Cl (shown in Figure 3.12C). In this bismuth system, the second equivalent of 2-pyridone stabilizes and favors the TS of the observed N-selective product by 2 kcal/mol via two strong hydrogen bonding interactions to each 2-pyridone. These interactions are visualized in the ground-state intermediates (Figure 3.11, right) and transition state structures (Figure 3.9).

Therefore, we can conclude that the role of both the sulfone group and the axial spectator ion contribute to the stability of the TS, determining the final observed O-selective product.

*3.3.7 Non-covalent interactions (NCI) influence in each Bismuth system*

NCI isodensity plots were generated to visualize favorable interactions at the TSs for the O-arylation in the bismacycle system and N-arylation in the $BiPh_3Cl_2$ system (shown in Figure 3.13 as blue dash lines). The TS for O-arylation reveals a hydrogen bond interaction between the 2-pyridone and sulfone while the TS for N-arylation reveals two hydrogen bond interactions to 2-pyridone with the phenyl group. Additionally, a relatively stronger O-Bi attractive interaction resides in the O-arylation TS.

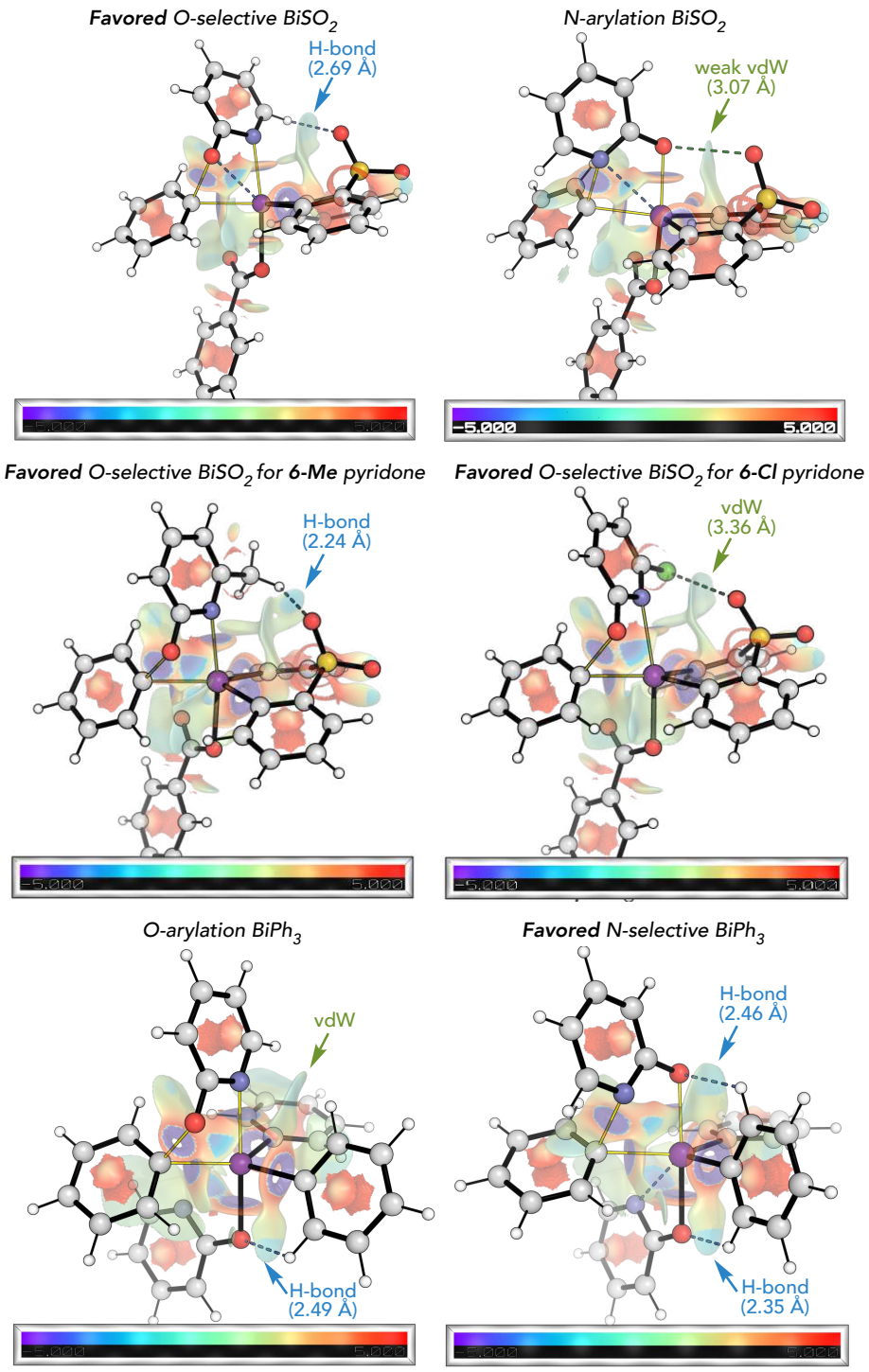



**Figure 3.13.** Non-covalent interactions (NCI) representations for the O-arylation in the bismacycle system and N- arylation in the $BiPh_3Cl_2$ system. The yellow lines represent bonds involved in the TSs and the blue dash lines represent attractive interactions.

Relative to this acyclic system, the bismacyclic scaffold preferentially stabilizes the O-arylation TS by several kcal/mol. Favorable non-covalent interactions between the sulfone bridge and C6-H of the pyridine are present for O-arylation, whereas for N-arylation the carbonyl oxygen is oriented unfavorably towards the $SO_2$ group (Figure 3.13). Similar strategies using a variation of the bismacycle scaffold by Cornella has shown to great effect influence the stability of the octahedral geometry at the TS in $C_{Ar}$-OTf couplings.[6]

### *3.4 Conclusion*

In this chapter, we computationally explored the bismuth reactivity developed by the Ball group for a convenient, telescoped procedure for the Bi(V)-mediated O-arylation of 2- and 4-pyridones with arylboronic acids. Complete O-selectivity was observed across a broad range of substrates, and we can explain the origin of this regioselectivity attributed primarily to the geometry that the cyclic scaffold imposes at the bismuth center visualized by the molecular representations of the transition state structures. We also explored the reactivity using QM methods of the reversed selectivity from previous Bi(V)-mediated arylations by Mukaiyama for comparison. Our initial benchmarking studies (e.g., specifically the grid integration benchmark studies) on the newly discovered bismuth(V) reactivity set the stage for the computational protocols required in bismuth systems for determining reaction selectivity. Wheeler's *et al* work was essential for identifying the computationally observed error and thereby applying the required protocol (e.g., increased integration grid). The bismuth chemistry developed by the Ball group expands to the consecutive work discussed in Chapter 4.

## REFERENCES


1. Jurrat, M.; Maggi, L.; Lewis, W.; Ball, L. T., Modular bismacycles for the selective C–H arylation of phenols and naphthols. *Nature Chemistry* **2020,** *12* (3), 260-269.
2. Ruffell, K.; Gallegos, L. C.; Ling, K. B.; Paton, R. S.; Ball, L. T., Umpolung Synthesis of Pyridyl Ethers by BiV-Mediated O-Arylation of Pyridones. *Angew. Chem., Int. Ed.* **2022,** *61* (51).
3. Kazuhiro, I.; Teruaki, M., Synthesis of N-Aryl Pyridin-2-ones via Ligand Coupling Reactions Using Pentavalent Organobismuth Reagents. *Chem. Lett.* **2005,** *34* (11), 1496-1497.
4. Kazuhiro, I.; Yuzo, N.; Teruaki, M., N-Arylation of Pyridin-2(1H)-ones with Pentavalent Organobismuth Reagents under Copper-Free Conditions. *Bull. Chem. Soc. Jpn.* **2006,** *79* (5), 761-767.
5. Bootsma, A. N.; Wheeler, S., Popular Integration Grids Can Result in Large Errors in DFT-Computed Free Energies. *ChemRxiv* **2019**.
6. Planas, O.; Peciukenas, V.; Cornella, J., Bismuth-Catalyzed Oxidative Coupling of Arylboronic Acids with Triflate and Nonaflate Salts. *J. Am. Chem. Soc.* **2020,** *142* (26), 11382-11387.

# 4 BISMA-STILLE CROSS-COUPLING REACTION

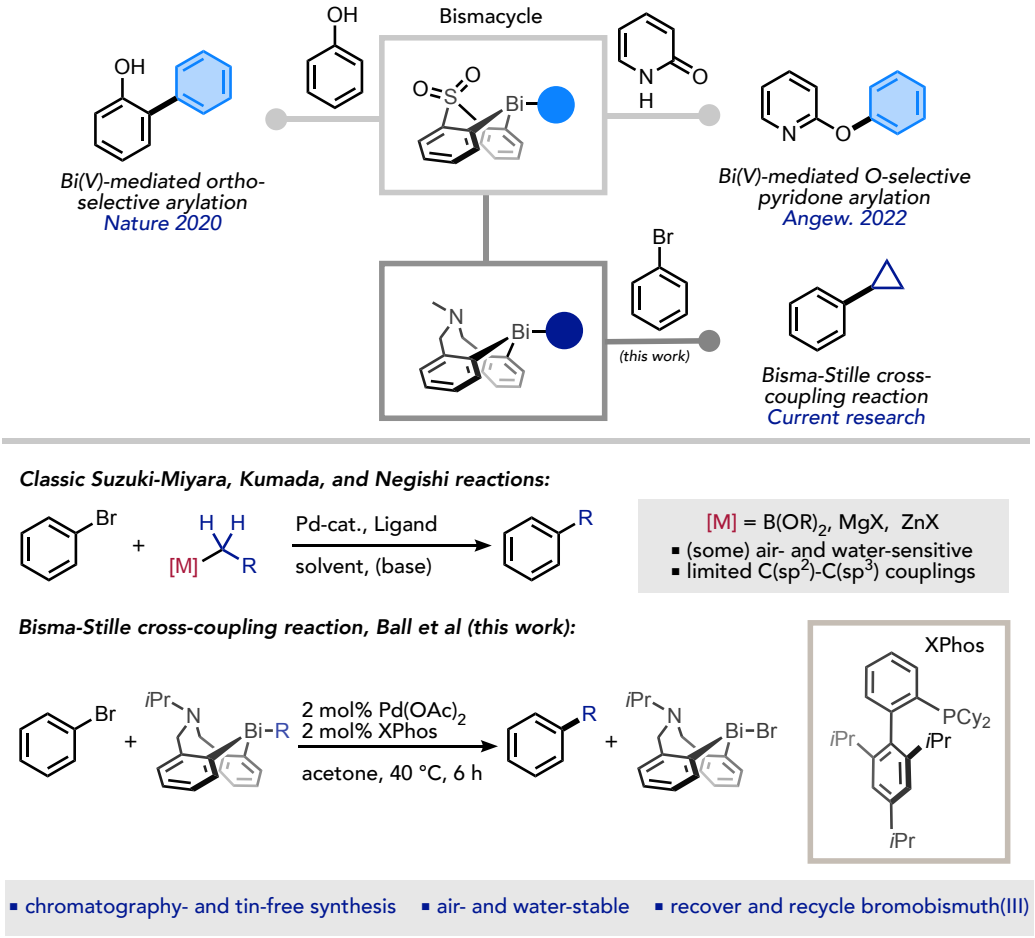
## *4.1 Chapter Overview*

In this chapter, we discuss ongoing experimental and computational work with the Ball group on a palladium-catalyzed cross-coupling reaction to achieve challenging C-C couplings in mild reaction conditions only possible with the amino-bridged bismacycle reagent. The discovery of this reactivity provided the opportunity to investigate the reaction mechanism using QM methods. Herein, we apply similar computational protocols from Chapter 3 to investigate various reaction pathways taken at the elementary steps of the Bisma-Stille catalytic cycle. Finally, we investigate the origin of the reactivity for various coupling substrates and modified organobismuth(V)-reagents *via* a statistical modeling approach using automated workflows discussed in Chapter 7. This collaborative work will be submitted in a peer-reviewed journal in the future.

## *4.2 Introduction*

In 2020, the Ball group developed an approach to synthesize a robust, universal bismacycle(III) precursor (e.g., the sulfone-bridged bismacycle reagent) that could be used as a vehicle to facilitate the transmetallation process of aryl substrates.[1, 2] The reactivity and selectivity of the organobismuth reagent was then extended onto 2- and 4-pyridones for a convenient strategy for selective O-arylation with arylboronic acids.[3] In 2023, Ball *et al* developed a Pd-catalyzed cross-coupling reaction using an evolved, universal precursor (e.g., an amino-bridged bismacycle(III) reagent) to achieve C($sp^2$)-C($sp^3$) couplings (Scheme 4.1). Palladium-catalyzed cross-coupling reactions like the classic Suzuki-Miyara[4], Kumada[5, 6], Negishi[7], and Stille[8-10] reactions provide the synthetic access for C-C bond formation using arylboronic acids, Grignard, organozinc, and organotin reagents with high selectivity, functional group tolerance, and mildness of the reaction conditions. Similar to these classic reactions, the organobismacycle reagent provides a powerful and general method for the formation of carbon-carbon bonds utilizing mild reaction conditions,

commercially available aryl halides, functional group tolerance, insensitive to moisture and oxygen, and avoids any concerns of toxicity from organostannanes.




**Scheme 4.1. Top:** Selected summary of reactivity developed by the Ball group to couple phenols, pyridones, and aryl halides with aryl or alkyl substrates using the developed bismacycle reagents. **Bottom:** the comparison of the classic Stille reaction to the developed Pd-catalyzed cross-coupling of cyclopropyl Bi(III) reagents by the Ball group (this work).

In this collaborative effort, we performed a computational investigation of (i) the full palladium-catalyzed cross-coupling process with cyclopropyl Bi(III) reagents, (ii) the favorable intermolecular interactions between the palladium- and bismacycle-species to explain the origin of reactivity, and (iii) the molecular and atomistic properties of the starting materials: organobismuth(III) reagents and aryl halides, to develop a predictive model that captures the reactivity at the *transmetallation* step using DFT and natural population analysis calculations. All

experimental work was performed by our experimental collaborators (Ball *et. al*) and all DFT computational work was performed by the present author.

### *4.3 Results and Discussions*

The coupling of aryl halides with cyclopropyl substrates with the amino-bridged bismacyclic system was investigated in detail along with the key features to fundamentally explain the observed reactivity. In order to elucidate this new bismuth reactivity in presence of a palladium-catalyst, DFT calculations (ωB97XD/def2-QZVPP//ωB97XD/6-31+G(d,p) and def2-TZVP (Bi, Pd)) with the SMD-solvation model (solvent = acetone) were applied to all final calculations to investigate the overall catalytic cycle. An "ultrafine" pruned (99,590) grid for numerical integration was deemed sufficient for calculating the thermochemistry.

#### *4.3.1 Benchmarking to bismuth (III) and (V) species*

We began by evaluating the selected level of theory for describing the structures of bismuth (III) by comparing to the characterized X-ray crystallography of the trigonal bipyramidal geometry bismacycle (III) with cyclopropane and bromide (Figure 4.1) using computational methods previously reported with the bismacycle system.[3]

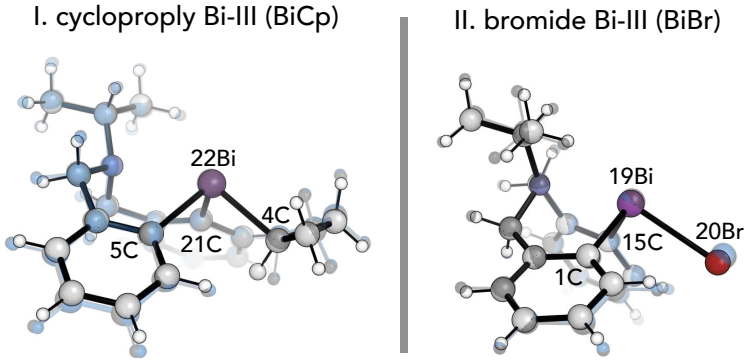



**Figure 4.1.** Overlay of the X-ray structure of the bismacycles (Crystallographic Information file identifier: BICSLH (left) and BICSLC (right)). Crystal structures are represented with standard colors and optimized geometries with (SMD-acetone) ωB97X-D/6-31+G(d,p), with def2-TZVP (shown in blue) and def2-TZVP (gray) for the Bi atom. Structures using LANL2DZ for the Bi atom are omitted.

**Table 4.1.** Tabulation of the bond lengths (Å) and the percentage of bond variation from the X-ray crystal structures shown in parenthesis. Bond lengths were determined experimentally from the Bi(III) X-ray structures and computationally using optimized structures from different levels of theories for bismuth.

| | | | Bond length (Å) and Bond Variation (%) (SMD-acetone) ωB97XD/ 6-31+G(d,p) (C,H,N,Br) | | |
|---|---|---|---|---|---|
| Bi(III) crystal structure | Bonds | Bond length (Å) | def2-SVP (Bi) | def2T-ZVP (Bi) | LANL2DZ (Bi) |
| BiCp | 4C-22Bi | 2.26599 | 2.282 (-0.7) | 2.259 (0.3) | 2.241 (1.1) |
| | 5C-22Bi | 2.26509 | 2.283 (-0.8) | 2.262 (0.1) | 2.243 (1.0) |
| | 21C-22Bi | 2.25436 | 2.281 (-1.2) | 2.260 (-0.2) | 2.240 (0.6) |
| BiBr | 1C-19Bi | 2.24389 | 2.266 (-1.0) | 2.245 (0.0) | 2.227 (0.8) |
| | 15C-19Bi | 2.24918 | 2.270 (-0.9) | 2.249 (0.0) | 2.232 (0.8) |
| | 19Bi-20Br | 2.78133 | 2.830 (-1.7) | 2.809 (-1.0) | 2.817 (-1.3) |

**Table 4.2.** Tabulation of the activation barriers (TS) for the oxidative addition (OA) and reductive elimination (RE) steps carried out at 40 °C in acetone.

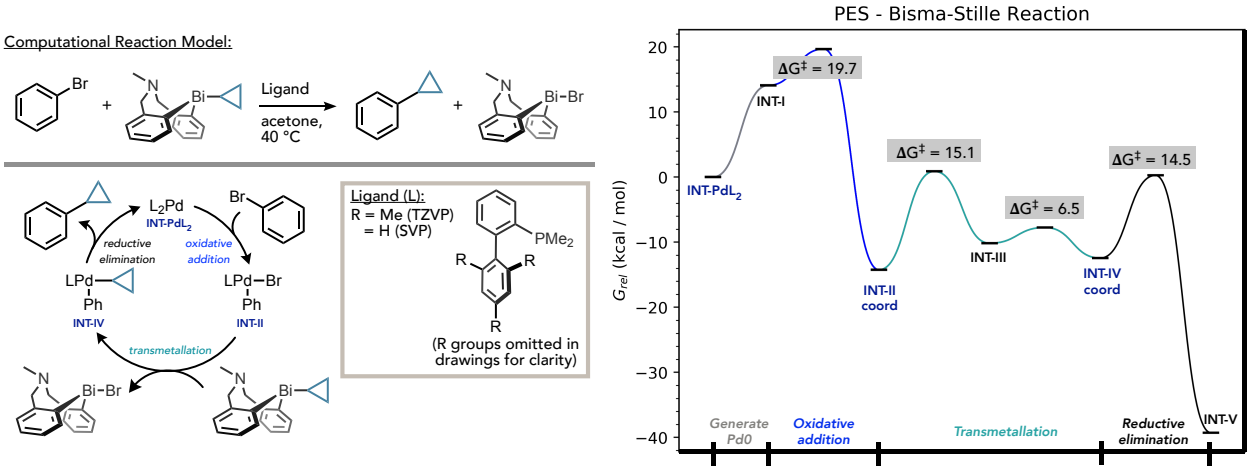
| | | (SMD-acetone) ωB97XD/de2QZVPP// **Functional**/6-31+G(d,p) (C,H,N,Br), def2TZVP (Pd,Bi) | | | |
|---|---|---|---|---|---|
| | | ωB97XD | | B3LYP-D3BJ | |
| Reaction step | GS_Zero (kcal/mol) | TS_OA | TS_RE + BiBr | TS_OA | TS_RE + BiBr |
| 1-OA | INT-PdL2 + BrPh | 19.7 | - | 17.5 | - |
| 2-OA | INT5-coord + BrPh | 7.2 | - | 4.7 | - |
| 3-RE | INT2-coord + BiCp | - | 14.5 | - | 16.1 |

The computed structures with the def2-TZVP method for the Bi atom shown in Table 4.1 resulted in high level of agreement between theory and experiment as all angles and dihedrals are well reproduced, and the Bi-C and Bi-Br bond distances differ by less than 0.5% and 1.5%. Considering that the def2-TZVP basis set for the Bi atom resulted in higher accuracy to the X-ray structure for both Bi(III) species, subsequent benchmark calculations were performed using this

method to evaluate the barrier heights with different functionals at the *oxidative addition* and *reductive elimination* steps (Table 4.2). We obtain an agreement with experimental findings with the two hybrid functionals and the two measured elementary steps indicating that the *oxidative addition* step is the rate-limiting step at 19.7 kcal/mol with ωB97XD which is a reasonable value for a reaction carried out at 40 °C. The ωB97XD functional was used for all subsequent geometry optimization calculations.

### *4.3.2 Newly Proposed Bisma-Stille Reaction Mechanism*

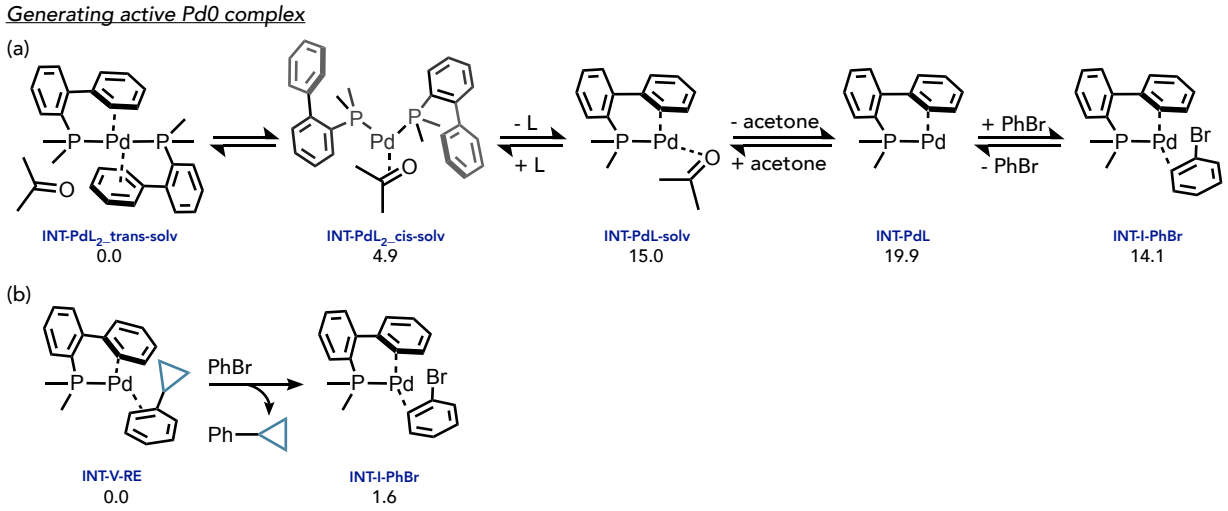
Our mechanistic investigation evaluated the reaction elementary steps of the catalytic cycle – *oxidative addition* (OA), *transmetallation* (TM), and *reductive elimination* (RE), to propose a full potential energy surface for the Bisma-Stille reaction shown in Scheme 4.2. In this mechanistic investigation, a simplified theoretical model with a methylamine-bridged bismacycle and the reduced version of the biphenyl phosphane XPhos ligand was used for the calculations as it sufficiently reproduced the experimental observations.




**Scheme 4.2.** Gibbs Free energy profile for the Bisma-Stille reaction using the shown simplified bismacycle and ligand model system with the most stable ground state and transition structures at the level of theory: (SMD-acetone) ωB97XD/def2-QZVPP//ωB97XD/6-31+G(d,p) (all other atoms), def2-TZVP (Bi, Pd) at 40 °C.

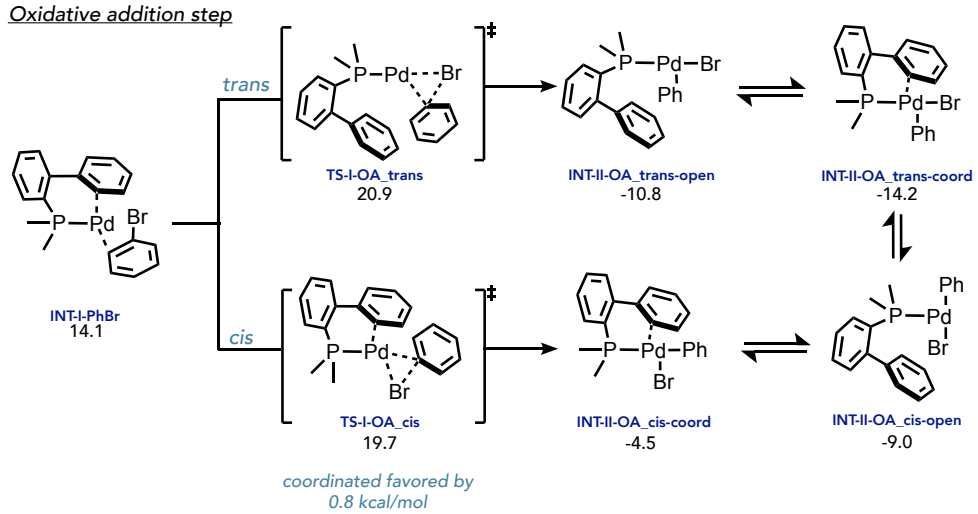
*4.3.3 Generating the Active Palladium Species Step*

A variety of possible ground-state palladium(0) complexes were investigated that undergo *oxidative addition* to calculate relative energies from the most stable species, **INT-PdL2_trans-solv** (Scheme 4.3).




**Scheme 4.3.** Ground-state palladium-complex species for the active Pd0-complex that undergoes oxidative addition, from (a) the intermediate complexed with two bound ligands and solvent or (b) the intermediate generated at the end of the reductive elimination step. Relative energies are given in kcal/mol from INT-PdL2_trans-solv.

*4.3.4 Oxidative Addition Step*

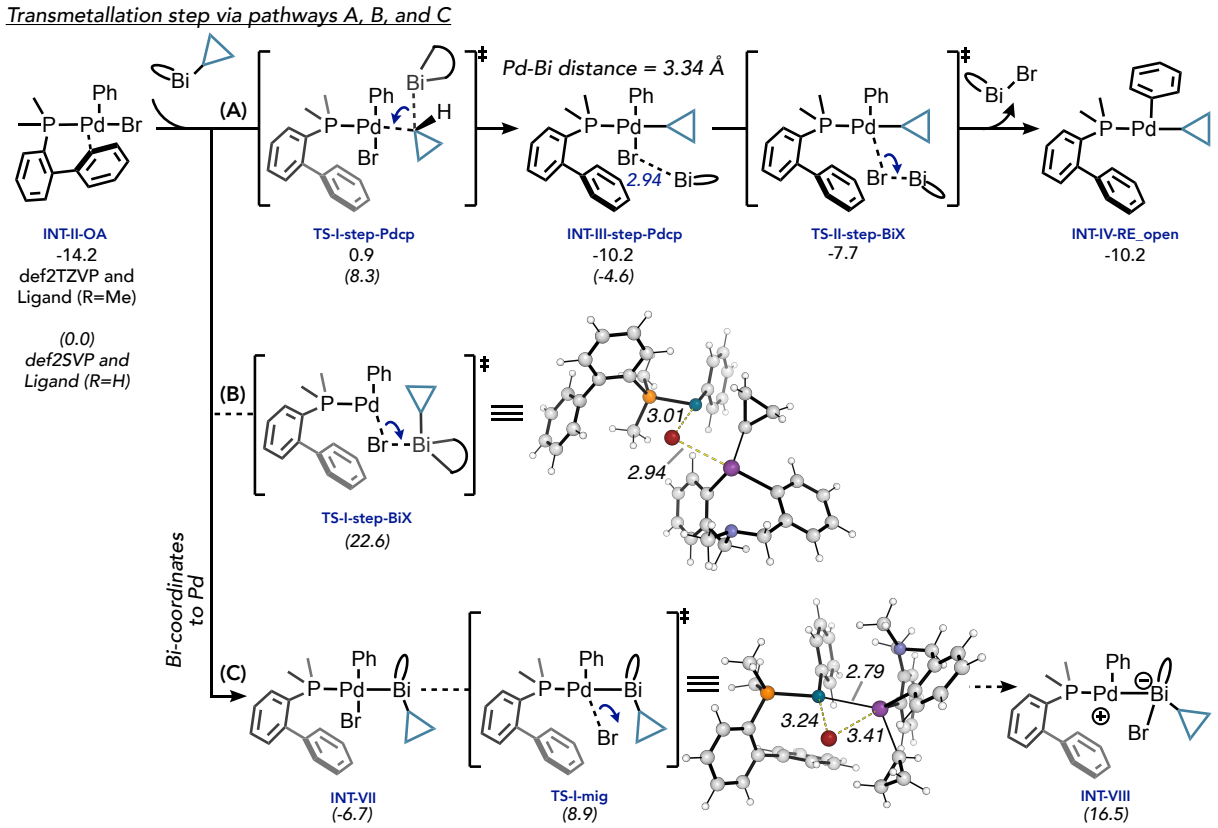



**Scheme 4.4.** Relevant reaction pathways at the oxidative addition step showing the formation of the most stable isomers. Relative energies are given in kcal/mol from INT-PdL$_2$_trans-solv.

The *cis* and *trans oxidative addition* pathways in Scheme 4.4 from the monoligated and coordinated Pd-species, **INT-I-PhBr**, favors the transition state (TS) structure, **TS-I-OA_cis**, with the bromide cis to the ligand (R=Me) by 1.2 kcal/mol in comparison to the most favorable trans-TS conformation, **TS-I-OA_trans**. From the measured elementary steps, the *oxidative addition* step is the rate-limiting step with a barrier height of 19.7 kcal/mol. The exergonic and irreversible *oxidative addition* step creates a variety of stable species that can undergo isomerization to the most stable species, **INT-II-OA_trans_coord**, to calculate the subsequent relative reaction barriers.

*4.3.5 Transmetallation Step*

For the transmetallation step (Scheme 4.5), we performed a preliminary mechanistic investigation using a smaller basis set for bismuth, def2-SVP, and ligand with R=H to model three possible pathways and select the viable pathway for further studies.

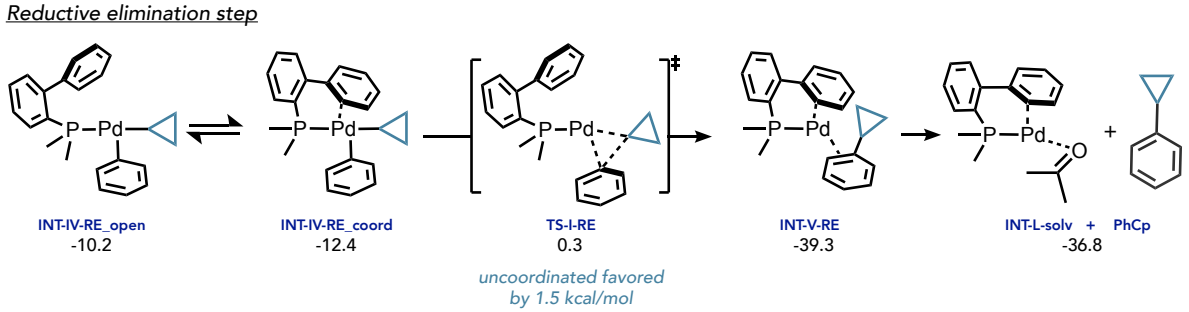



**Scheme 4.5.** Relevant reaction steps at the transmetallation step and reaction pathways (B and C) considered at this step. Relative energies are given in kcal/mol for values computed from INT-PdL2_trans-solv using ligand (R=Me) with ωB97XD/def2-QZVPP//ωB97XD/6-31+G(d,p) (all other atoms), def2-TZVP (Bi, Pd) at 40 °C in acetone. Values shown in parenthesis were computed from INT-II-OA using ligand (R=H) with ωB97XD/6-31+G(d,p) (all other atoms), def2-VP (Bi, Pd) at 40 °C in acetone.

Based on the computed relative energies from **INT-II-OA_trans_coord** shown in parenthesis, the process for first migrating the bromide from palladium to bismuth at **TS-I-step-BiX** ($\Delta G^{\ddagger}$ = 22.6 kcal/mol) or from the bismuth-bound to palladium complex at **TS-I-mig** ($\Delta G^{\ddagger}$ = 15.6 kcal/mol from **INT-VII**), are less favorable compared to first migrating the cyclopropyl at **TS-I-step-Pdcp** ($\Delta G^{\ddagger}_{def2SVP}$ = 8.3 kcal/mol). This is consistent with a stepwise cyclic transmetallation mechanism as dissociating halides are less favorable and endergonic.[11, 12] Additionally, consistent with experimental observations, the stepwise cyclic transition state was proposed due to cyclopropyl's stereo-retentive configuration whereas with an open mechanism the inverted stereochemistry is expected.[13]

Therefore, the proposed transmetallation pathway (A) proceeds via a stepwise process consisting of two facile TSs: (i) the migration of cyclopropyl from bismuth to the Pd-complex (with a barrier height, $\Delta G^{\ddagger}_{def2TZVP}$ = 15.1 kcal/mol from **INT-II-OA_trans_coord**) forming the stable intermediate, **INT-II-step-Pdcp**, followed by (ii) the migration of bromide from palladium to bismuth (with a barrier height, $\Delta G^{\ddagger}_{def2TZVP}$ = 6.5 kcal/mol) to form **INT-IV** and the experimentally observed bismuth(III)-bromide product.

*4.3.6 Reductive Elimination Step*




**Scheme 4.6.** Relevant reaction pathways at the reductive elimination step showing the most stable ground- and transition-state isomers only. Relative energies are given in kcal/mol from INT-PdL2_trans-solv.

The final coupling of cyclopropyl and phenyl occurs feasibly through the *reductive elimination* step (Scheme 4.6) with a monoligated and uncoordinated ligand, **TS-RE**, at 14.5 kcal/mol.

*4.3.7 Transition State Structure Stabilities*

We investigated the stabilities at different isomers and conformers of the transition states at the *oxidative addition*, *transmetallation*, and *reductive elimination* steps (shown in Figure 4.2).

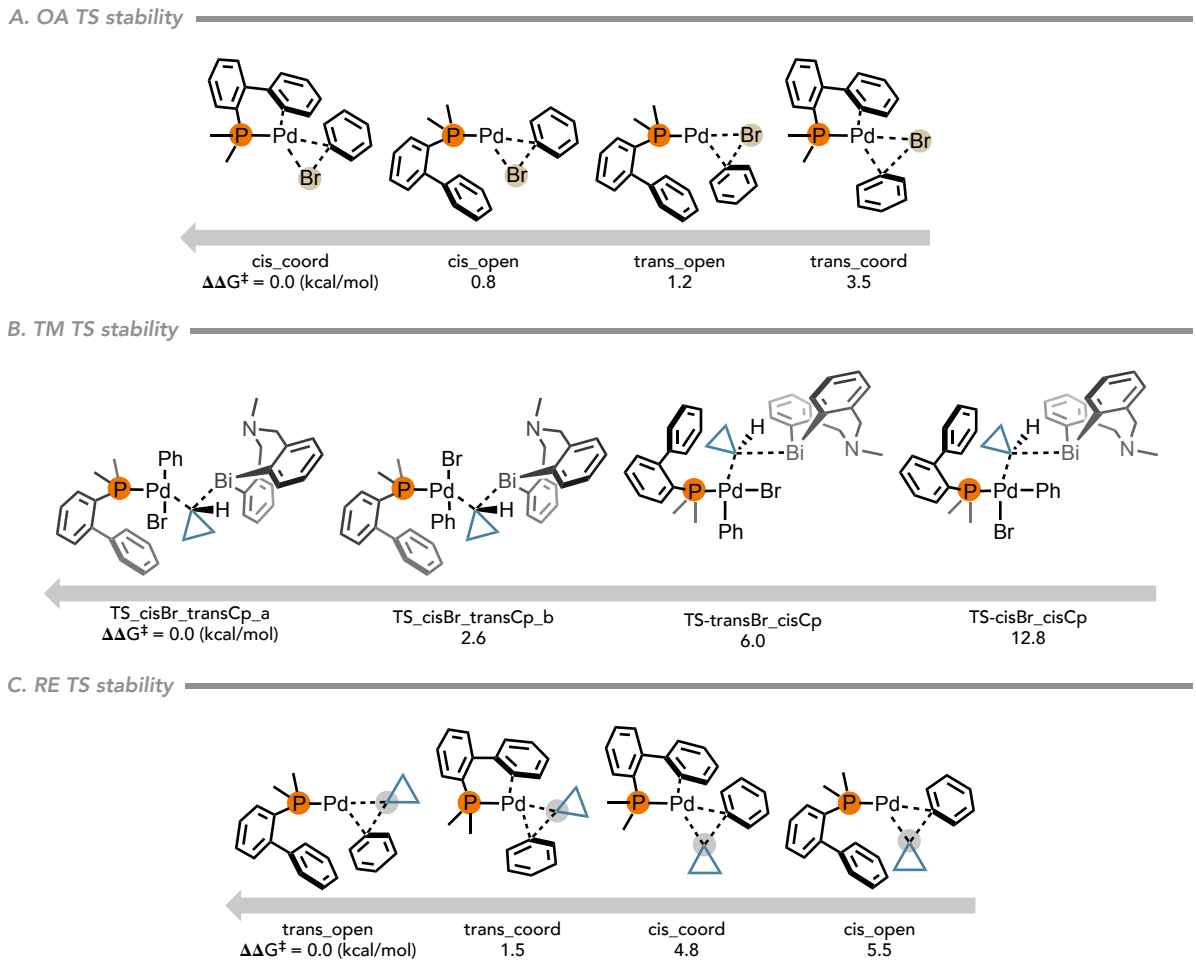



**Figure 4.2.** Relative stabilities of the TS isomers and conformers evaluated in this study at the (A) oxidative addition, (B) transmetallation, and (C) reductive elimination steps. Relative $\Delta\Delta G^{\ddagger}$ energies are given in kcal/mol.

Based on the optimized TS structures at the *oxidative addition* and *transmetallation* steps, the bromide cis to the ligand (R=Me) is favored by 1.2 and 6.0 kcal/mol, respectively. Additionally, the most stable TM configuration favors a trans-cyclopropyl (Cp) to the ligand and bismuth syn to the phenyl substrate by 2.6 kcal/mol in **TS_cisBr_transCp_a** due to phenyl's π-interaction to bismuth's acceptor σ*-orbital. These interactions are visualized in Figure 4.3A and B.

### *4.3.8 Comparing Non-Covalent Interactions (NCI) and Molecular Orbitals for different TM TS structures*

NCI isodensity plots and molecular orbitals were generated to visualize favorable interactions at the *transmetallation* TSs using the bridged N-*i*Pr bismacycle-complex with varying *para*-aryl

substituents: *p*-OMe, *p*-H, and *p*-CN (shown in Figure 4.3). Atomic charges and molecular orbitals were calculated using natural population analysis (NPA) with NBO 7.0[14] interfaced to Gaussian 16. NCIPLOT[15] created the non-covalent interaction plots. Molecular representations were created with PyMol[16] for which the display settings are openly accessible.[17]

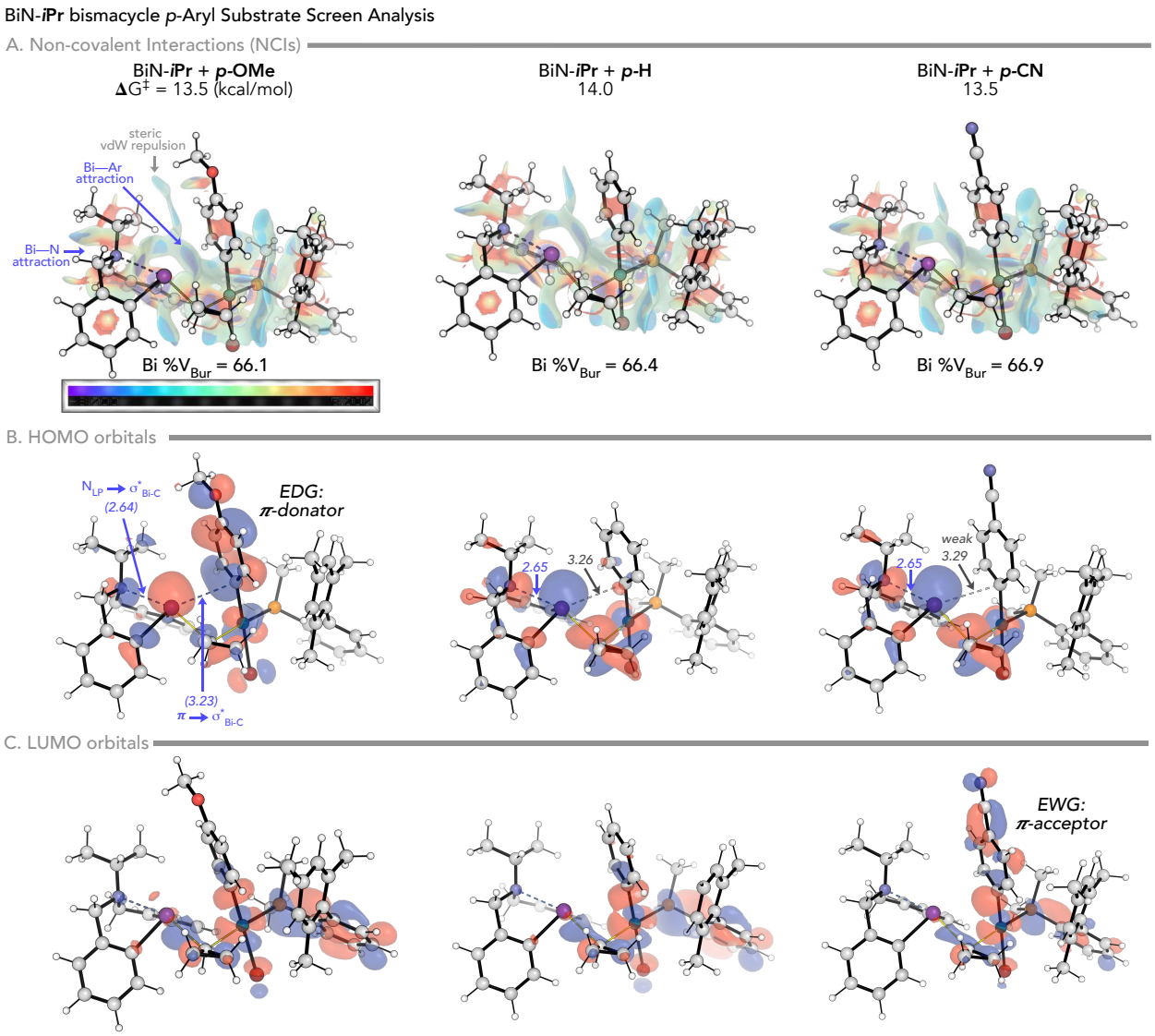



**Figure 4.3.** Non-covalent interaction (NCI) and molecular orbital representations for the bridged *iso*-propyl bismacycle complex with *p*-OMe (left), *p*-H (middle), and *p*-CN (right) aryl substrates. The yellow lines represent bonds involved in the TSs, the blue and gray dash lines represent attractive and weak interactions, respectively. Atom distances are shown in parenthesis in Angstroms.

The NCI plots reveal attraction interactions with bismuth from both nitrogen and the π-interaction from the aryl substrate. The highest-occupied molecular orbitals (HOMO) reveal the origin for the TS's stability stemming from the nitrogen's lone-pair electrons and π-donation into the Bi-C acceptor σ*-orbitals from electron donating groups (Figure 4.3B). Organobismuth(III) compounds have demonstrated capabilities to function as donors to a transition metal fragment while simultaneously acting as weak acceptors to another electronegative group via long- or short-range intramolecular "hypervalent" interactions.[18] These interaction effects can be visualized at each of the *transmetallation* TS structure through the short Bi—N distances formed while an increasing interaction (e.g., shorter Bi—*i*C distances with *p*-OMe) is formulated with proximal EDG aryl substrates. In turn when electron withdrawing groups (e.g., *p*-CN) are utilized, reactivity is unaffected as their π-acceptor orbitals drive the reaction forward (Figure 4.3C).

*4.3.9 Theoretical Studies of the Reactivities between the Bismacycles and Substrate Scope*

We computed the reaction barriers for a variety of *transmetallation* transition states by varying the R-substituent on the bridged bismacycle and the *para*-substituent on the aryl substrate (Figure 4.4).

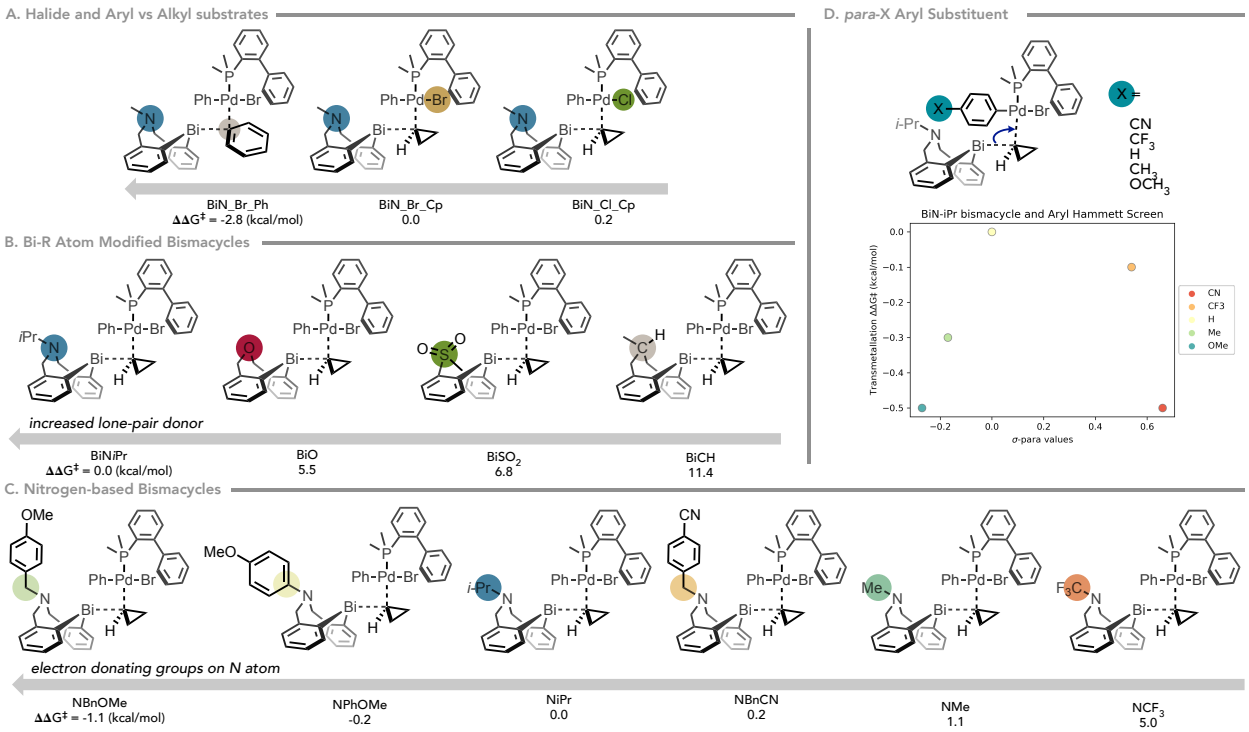



**Figure 4.4.** Relative stabilities for the screened transmetallation TSs comparing the different (A) halides and Bi-bound substrates, aryl vs alkyl, (B) bridged-atom of the bismacycle, (C) electron donating and withdrawing effects on substituents bound to the N-bridged bismacycles, and (D) *para*-aryl substituents. Relative $\Delta\Delta G^{\ddagger}$ energies are given in kcal/mol.

We analyzed different reactivity between phenyl- *vs* cyclopropyl-Bi(III) complexes as well as different halides (e.g., Br vs Cl) in Figure 4.4A. We modified the bismacycles to analyze the lone-pair contribution stemming from the R-atom (e.g., N, O, sulfone group, and alkyl (-CH) group) in Figure 4.4B. The role of the nitrogen in the bridged bismacycle complex was analyzed by modeling electron donating (e.g., BnOMe, PhOMe) and withdrawing groups (e.g., BnCN, $CF_3$) bound to nitrogen shown in Figure 4.4C. The influence of the *para*-substituents on the aryl substrate with the BiN-*i*Pr bismacycle were also surveyed in Figure 4.4D. Based on the optimized TM TS structures, we can observe favorable and increased reactivity trends with the bromine halide, a bridged-nitrogen atom bismacycle with bound electron donating groups (EDGs), and an electronically activated aryl substrate with donating or withdrawing groups. We obtain an increase in stability with an increase of electronic donating or withdrawing effects ($\Delta\Delta G^{\ddagger}$ =-0.5 kcal/mol

with *p*-OMe and *p*-CN groups) as shown in Figure 4.4D. Moreover, the *transmetallation* of a Bi(III)-bound phenyl substrate was also studied, **BiN_Br_PhB**, showing a reduced barrier ($\Delta\Delta G^{\ddagger}$ =-2.8 kcal/mol) compared to the transfer of cyclopropyl.

*4.3.10 Prediction of Reaction Barrier Heights with a Multivariate Linear Regression Model*

A multivariate linear regression model was constructed in terms of atomistic and molecular properties for the starting reactants to capture the reactivity at the computed reaction barrier heights (Figure 4.5) based on the mechanistic understanding of the *transmetallation* step. All supporting data and code are available in a GitHub repository.[19]

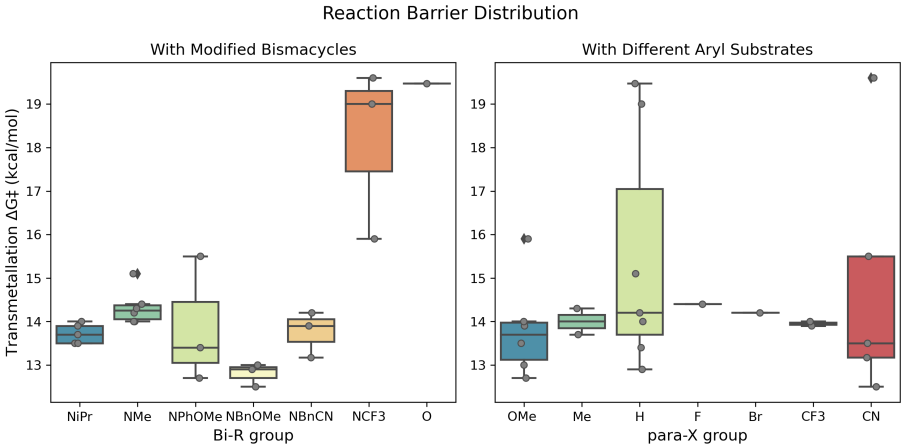



**Figure 4.5.** Box plot distribution of the *transmetallation* activation barrier heights computed with different bismacycle complexes (left) and *p*-aryl substrates (right). Relative energies are given in kcal/mol.

**Table 4.3.** Electronic, steric, and structural descriptors utilized in model development.

| Type | Descriptor |
|---|---|
| Structural | C-Bi-C bite angle (°) – for Bi(III) complex |
| | Bi-R distance (Å) – for bridged-R atom |
| | Bi-iC_Cp distance (Å) |
| | Pd-iC_Cp distance (Å) |
| | Bi-iC_Ar distance (Å) |

| | |
|---|---|
| Electronic:<br>NBO atomic charge | Bi (bismuth)<br>iC-Ar (*ipso*-Aryl)<br>pC-Ar (*para*-Aryl)<br>iC-Cp (Cyclopropyl)<br>Pd (palladium)<br>P (phosphorus)<br>X (Br or Cl)<br>Bridged-Atom (N or O) |
| Electronic:<br>Molecular orbitals | HOMO – for Bi(III)-complex<br>LUMO – for Aryl halide<br>HL Gap – HOMO-LUMO difference |
| Steric | Bismuth's Buried Volume at 3.5Å radius ($\%V_{Bv}$) |

Electronic, steric, and structural descriptors were collected for the ground-state reactants – Bi(III) complex and *p*-aryl substrates, and summarized in Table 4.3. The electronic features were collected from NBO calculations and the steric features were collected for the optimized structures using DBSTEP.[20] The resulting model for $\Delta G^{\ddagger}$ (Figure 4.6 and Figure 4.7) uses Bi and aryl's *para*-C atomic charge and bismuth's percent buried volume and HOMO value giving a mean absolute error (MAE) of 0.5 kcal/mol for a training set of 16 activation barriers and an MAE of 0.6 kcal/mol for a hold-out test set of eight activation barriers.

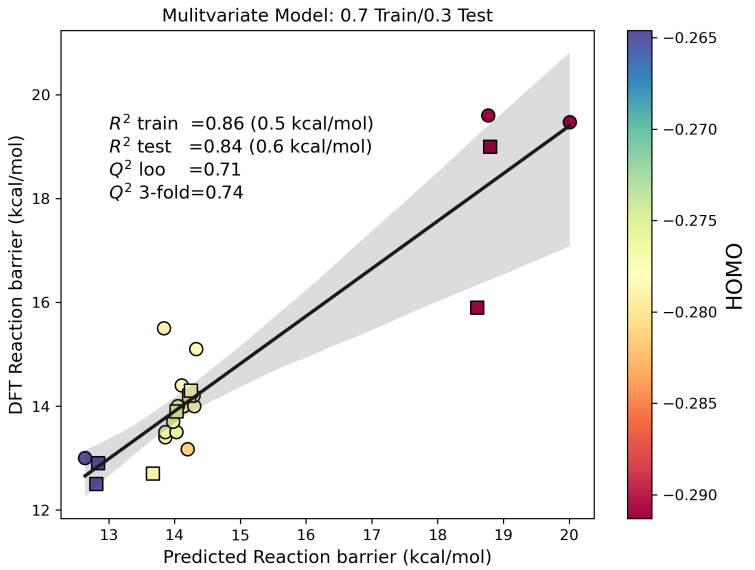



**Figure 4.6.** Multivariate linear regression model: $\Delta G^{\ddagger}$ = 14.652 + 0.988Bi + -1.456HOMO + -0.664Bi_BV + -0.088pC, generated to predict the transmetallation activation barriers in a random 70/30 train and test split.

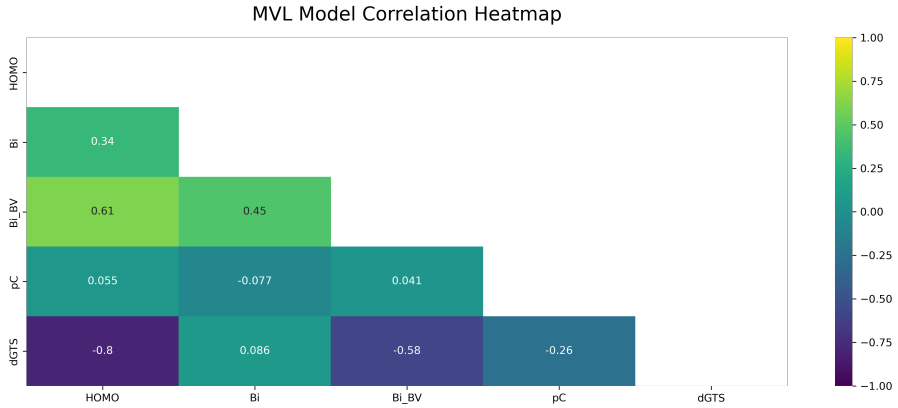



**Figure 4.7.** Pearson correlation heatmap for the descriptors used in the multivariate linear regression model.

Model selection and cross-validation analysis was performed using forward stepwise feature selection and a random 70% Train and 30% Test split using REGGAE, an R-programming language script workflow to generate regression models.[21] The statistical significance of each descriptor to the model was tested by ANOVA with p-values < 0.05 with the exception of *p*C (charge at the *para*C) at 0.69. Leave-one-out (loo) and K-fold (K=3) cross-validation analysis, $Q^2$, was performed using q2 functions for the test data set (cvq2 package with REGGAE). The optimal

regression model was validated by the calculated $R^2$ of the training and test set and the predictive squared correlation coefficient ($Q^2$) for the training set. The optimal model was selected with satisfactory leave-one-out (loo) and 3-fold $Q^2$ values greater than 0.6. The workflow, result analysis, and data of the final MVL model of the DFT reaction barrier vs predicted barriers with assigned split is available in the GitHub repository.

Based on the fundamental understanding at the *transmetallation* step with intramolecular interactions at play, we can describe the reactivity through reaction barrier heights using simple atomistic and molecular terms with a MVL regression model. Using three terms from the Bi(III) complex – bismuth's charge, buried volume, and the HOMO value, and the single *para*-carbon charge from the aryl-halide, we can design a Bi(III)-complex for optimal reactivity at the TM step.

### *4.4 Conclusion*

From the discover of the bismuth chemistry published in *Nature* by the Ball group in 2020, led to new bismuth chemistry discussed in Chapters 3 and 4. In this chapter, we introduce the bismuth reactivity with catalytic amounts of palladium as a new reaction method to form challenging $C(sp^2)$-$C(sp^3)$ bonds *via* a cross-coupling reaction with the amino-bridged bismacycle reagent. Utilizing quantum mechanical, we formulate a proposed reaction mechanism for the Bisma-Stille cross-coupling reaction. Non-covalent interaction (NCI) plots complement the observed product formation *via* non-covalent interactions between the bismuth centered atom and the coupling substrate. Additionally, we perform electronic molecular property calculations for a scope of coupling substrates and build a statistical model to capture these NCIs and explain the origin of reactivity facilitated with automated workflows discussed in Chapter 7.

## REFERENCES


1. Ball, L. T.; Senior, A., Bismuth(V)-Mediated C–H Arylation of Phenols and Naphthols. *Synlett* **2021,** *32* (03), 235-240.
2. Jurrat, M.; Maggi, L.; Lewis, W.; Ball, L. T., Modular bismacycles for the selective C–H arylation of phenols and naphthols. *Nat. Chem.* **2020,** *12* (3), 260-269.
3. Ruffell, K.; Gallegos, L. C.; Ling, K. B.; Paton, R. S.; Ball, L. T., Umpolung Synthesis of Pyridyl Ethers by BiV-Mediated O-Arylation of Pyridones. *Angew. Chem., Int. Ed.* **2022,** *61* (51).
4. Miyaura, N.; Suzuki, A., Stereoselective synthesis of arylated (E)-alkenes by the reaction of alk-1-enylboranes with aryl halides in the presence of palladium catalyst. *J. Chem. Soc., Chem. Commun.* **1979,** (19), 866.
5. Corriu, R. J. P.; Masse, J. P., Activation of Grignard reagents by transition-metal complexes. A new and simple synthesis of trans-stilbenes and polyphenyls. *J. Chem. Soc., Chem. Commun.* **1972,** (3), 144a.
6. Tamao, K.; Kiso, Y.; Sumitani, K.; Kumada, M., Alkyl group isomerization in the cross-coupling reaction of secondary alkyl Grignard reagents with organic halides in the presence of nickel-phosphine complexes as catalysts. *J. Am. Chem. Soc.* **1972,** *94* (26), 9268-9269.
7. Baba, S.; Negishi, E., A novel stereospecific alkenyl-alkenyl cross-coupling by a palladium- or nickel-catalyzed reaction of alkenylalanes with alkenyl halides. *J. Am. Chem. Soc.* **1976,** *98* (21), 6729-6731.
8. Azarian, D.; Dua, S. S.; Eaborn, C.; Walton, D. R. M., Reactions of organic halides with R3 MMR3 compounds (M = Si, Ge, Sn) in the presence of tetrakis(triarylphosphine)palladium. *J. Organomet. Chem.* **1976,** *117* (3), C55-C57.
9. Kosugi, M.; Sasazawa, K.; Shimizu, Y.; Migita, T., REACTIONS OF ALLYLTIN COMPOUNDS III. ALLYLATION OF AROMATIC HALIDES WITH ALLYLTRIBUTYLTIN IN THE PRESENCE OF TETRAKIS(TRIPHENYLPHOSPHINE)PALLADIUM(O). *Chem. Lett.* **1977,** *6* (3), 301-302.
10. Milstein, D.; Stille, J. K., Mechanism of reductive elimination. Reaction of alkylpalladium(II) complexes with tetraorganotin, organolithium, and Grignard reagents. Evidence for palladium(IV) intermediacy. *J. Am. Chem. Soc.* **1979,** *101* (17), 4981-4991.
11. Nova, A.; Ujaque, G.; Maseras, F.; Lledós, A.; Espinet, P., A Critical Analysis of the Cyclic and Open Alternatives of the Transmetalation Step in the Stille Cross-Coupling Reaction. *J. Am. Chem. Soc.* **2006,** *128* (45), 14571-14578.
12. Ariafard, A.; Yates, B. F., Subtle Balance of Ligand Steric Effects in Stille Transmetalation. *J. Am. Chem. Soc.* **2009,** *131* (39), 13981-13991.
13. Casado, A. L.; Espinet, P., Mechanism of the Stille Reaction. 1. The Transmetalation Step. Coupling of R1I and R2 SnBu 3 Catalyzed by trans-[PdR1IL2] (R1= C6Cl2F3; R2. *J. Am. Chem. Soc.* **1998,** *120* (35), 8978-8985.
14. Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Karafiloglou, P.; Landis, C. R.; Weinhold, F. *NBO 7.0*, Theoretical Chemistry Institute, 2018.
15. Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W., NCIPLOT: A Program for Plotting Noncovalent Interaction Regions. *J. Chem. Theory Comput* **2011,** *7* (3), 625-632.
16. Schrödinger, LLC *The PyMOL Molecular Graphics System, version 1.8*, 2015.

17. Paton Lab Workflows. https://github.com/patonlab/paton_group_workflows.
18. Benjamin, S. L.; Reid, G., Neutral organoantimony(III) and organobismuth(III) ligands as acceptors in transition metal complexes – Role of substituents and co-ligands. *Coord. Chem. Rev.* **2015,** *297-298*, 168-180.
19. Gallegos, L. C. Bisma-Stille Analysis. https://github.com/Liliana-Gallegos/bisma-stille_analysis.
20. Luchini, G.; Patterson, T.; Paton, R. S. DBSTEP Release 1.1.0 (1.1.0) Zenodo. patonlab/DBSTEP.
21. Gallegos, L. C. REGGAE. https://github.com/Liliana-Gallegos/REGGAE.

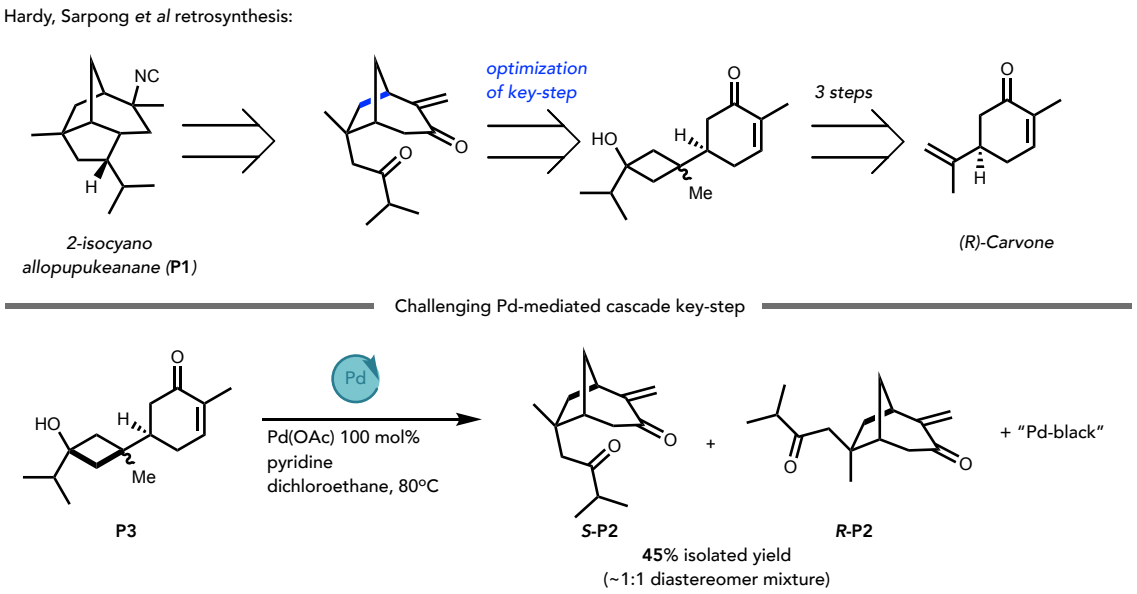
# 5 PALLADIUM(II)-MEDIATED CASCADE REACTION IN THE SYNTHESIS OF 2-ISOCYANOALLOPUPUKEANANE

## *5.1 Chapter Overview*

In this chapter, we merged computational and statistical modeling to optimize for the desired product in a challenging key reaction step in the total synthesis of the natural product of allopupukeanane developed by the Sarpong group. Specifically, we modeled the palladium-mediated cascade reaction using quantum mechanics to identify the key elementary step influencing the reactivity. Then, for a data-driven approach, we screened various ligands and collected mechanistically derived molecular DFT features to incorporate into a Bayesian optimization tool developed by the Doyle lab. Here, we applied automated workflows to collect the features in an automated manner using DISCO (discussed in Chapter 7). This approach proved successful in identifying more suitable and efficient reaction conditions when facing racemic mixture, byproduct formation, and catalyst decomposition challenges. The overall synthesis plan to access multiple natural products *via* the bridged bicycle scaffold highlighted in this chapter is currently an ongoing project by the Sarpong group. The collaborative research discussed in this chapter will be submitted in a peer-reviewed journal in the future.

## *5.2 Introduction*

From the structurally diverse marine sesquiterpenes, the pupukeanane family bears unique bridged scaffolds with potential antimalarial properties.[10] Few synthetic routes to access the novel isocyanopupukeananes scaffolds have been explored.[11] The Sarpong laboratory developed a synthetic route to one of these with naturally occurring isocyanoterpenes, 2-isocyanoallopupukeanane (**P1**, Scheme 5.1), from the naturally found *R*-carvone.




**Scheme 5.1. Top:** Retrosynthesis steps for 2-isocyanoallopupukeanane from *R*-Carvone. **Bottom:** Synthetic route of 2-isocyanoallopupukeanane (**P1**) of the challenging cascade reaction.

The synthetic strategy relied on the development of a novel and key Pd-mediated cascade reaction to rapidly build the 3.2.1-bicyclic framework present in the allopupukeanane core. Despite multiple screening attempts to optimize this key reaction through traditional variable screenings never led to catalytic activity. While the step was valuable to build complexity and furnish the desired compound for completion of the target molecule, significant room for improvement was possible and the step required expensive stoichiometric palladium acetate (Scheme 5.1). To address these challenges through a collaborative effort, the computational analysis along with the experimental reaction conditions defined the chemical space for the recently developed Experimental Design via Bayesian Optimization (EDBO) tool by Doyle *et al*. to navigate this vast reaction space and suggest a new batch of experiments that maximizes the exploration of underexplored space while exploiting conditions that lead to the desired outcome.[12] All experimental work was performed by our experimental collaborators (Hardy, Sarpong *et. al*), all

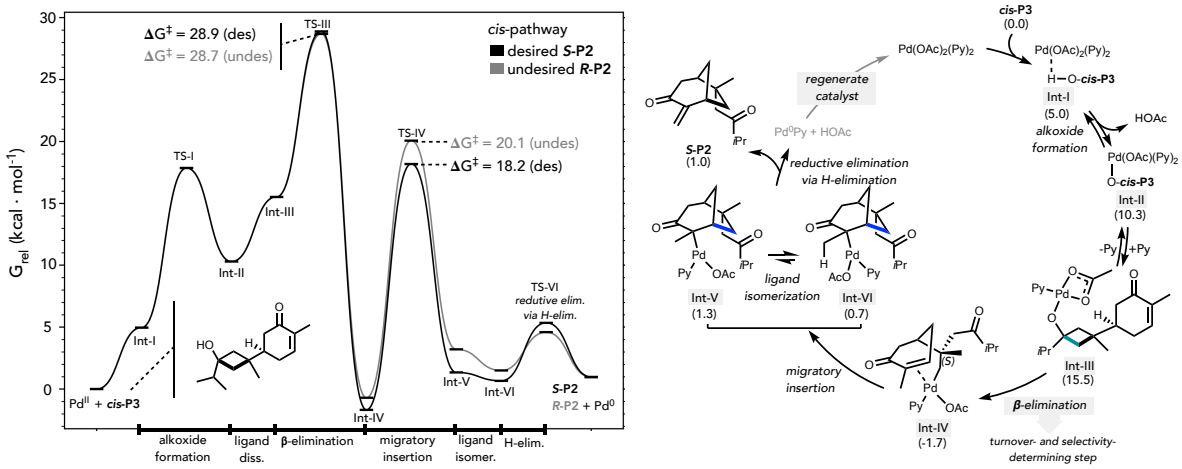
EDBO prediction analysis was employed by our data-chemist experts (Shields, Doyle *et. al*), and all DFT computational work was performed by the present author.

### *5.3 Results and Discussions*

Density functional theory (DFT) calculations were used to rationalize prior experimental observations and to inform feature selection used in Bayesian optimization. The ωB97X-D Gibbs energy profile was computed at 80°C, with def2-QZVPP single-point energetics evaluated at 6-31G(d) (LANL2DZ for Pd) stationary points and an SMD description of 1,2-dichloroethane using Gaussian 16, rev. B.01.[13] We evaluated the transformation of *cis*-**P3** in Scheme 5.2 producing both diastereomers of **P2**. The diastereomeric starting material *trans*-**P3** gives a quantitatively similar energy profile shown in Scheme 5.3. In this study in Section 5.3.2, we also discuss competing reactions at the reductive elimination step to reduce and regenerate the palladium (II) complex. The energy profiles assume a $Pd(OAc)_2$/pyr ratio of 1:2 leading to *in situ* formation of $(pyr)_2Pd(OAc)_2$.[14]

#### *5.3.1 Reaction Mechanism of the cis-Pathway*

The proposed reaction mechanism for *cis*-P3: substrate coordination to this complex forms **Int-I**, from which the loss of acetic acid (*via* **TS-I**) forms alkoxide bound **Int-II**.[15]




**Scheme 5.2.** Computed Gibbs energy profile (1 mol/l, 80°C, shown in kcal·mol$^{-1}$) of the proposed

mechanism for Pd-catalyzed cascade reaction forming the desired and undesired diastereomers, ***S*-P2** and ***R*-P2**, respectively, from cyclobutanol ***cis*-P3**.

Further loss of a pyridine ligand to form **Int-III** generates an open coordinate site at the Pd-center required for β-elimination. The computed TSs reveal an irreversible β-elimination step (ΔG = -17.2 kcal/mol) making **TS-III** diastereoselectivity-determining. However, $\Delta\Delta G^{\ddagger}$ between cleavage of the diastereotopic C-C bonds is just 0.2 kcal/mol, resulting in the low levels of diastereoselectivity obtained experimentally. Migratory insertion then builds the 3.2.1-bicyclic core *via* **TS-IV** with barriers of 18 and 20 kcal/mol for the desired and undesired products, respectively. ***S*-P2** is formed and released after ligand isomerization (**Int-VI**, in which the acetate is *cis* to the β-methyl group) and a concerted reductive β-H-elimination *via* **TS-VI**. We found this TS, forming Pd(0) and acetic acid directly, is more favorable (by 1.3 kcal/mol for the desired ***S*-P2** product) than the four-centered pathway forming an intermediate Pd(II)-hydride.[16] Key TS geometries are shown in Figure 5.1.

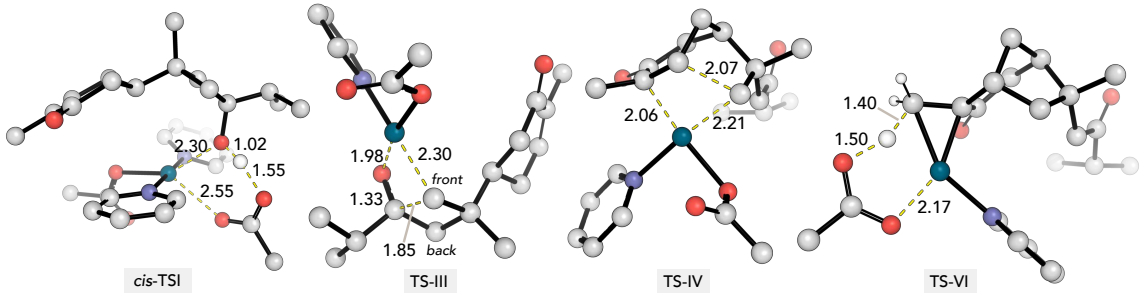
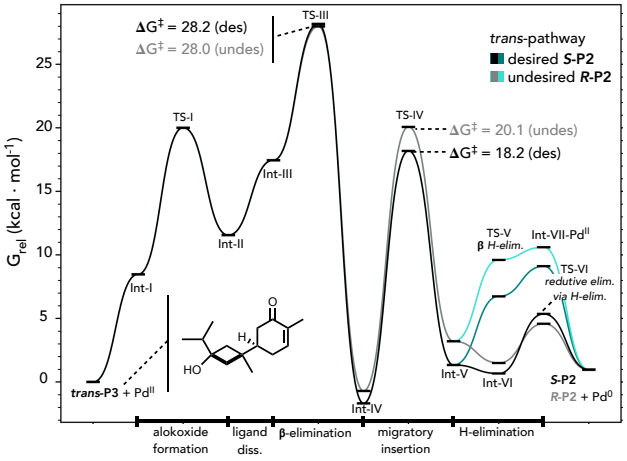



**Figure 5.1.** Key transition state structures for the *cis*-pathway with corresponding bond forming and breaking lengths shown as yellow dashed lines in Angstroms.

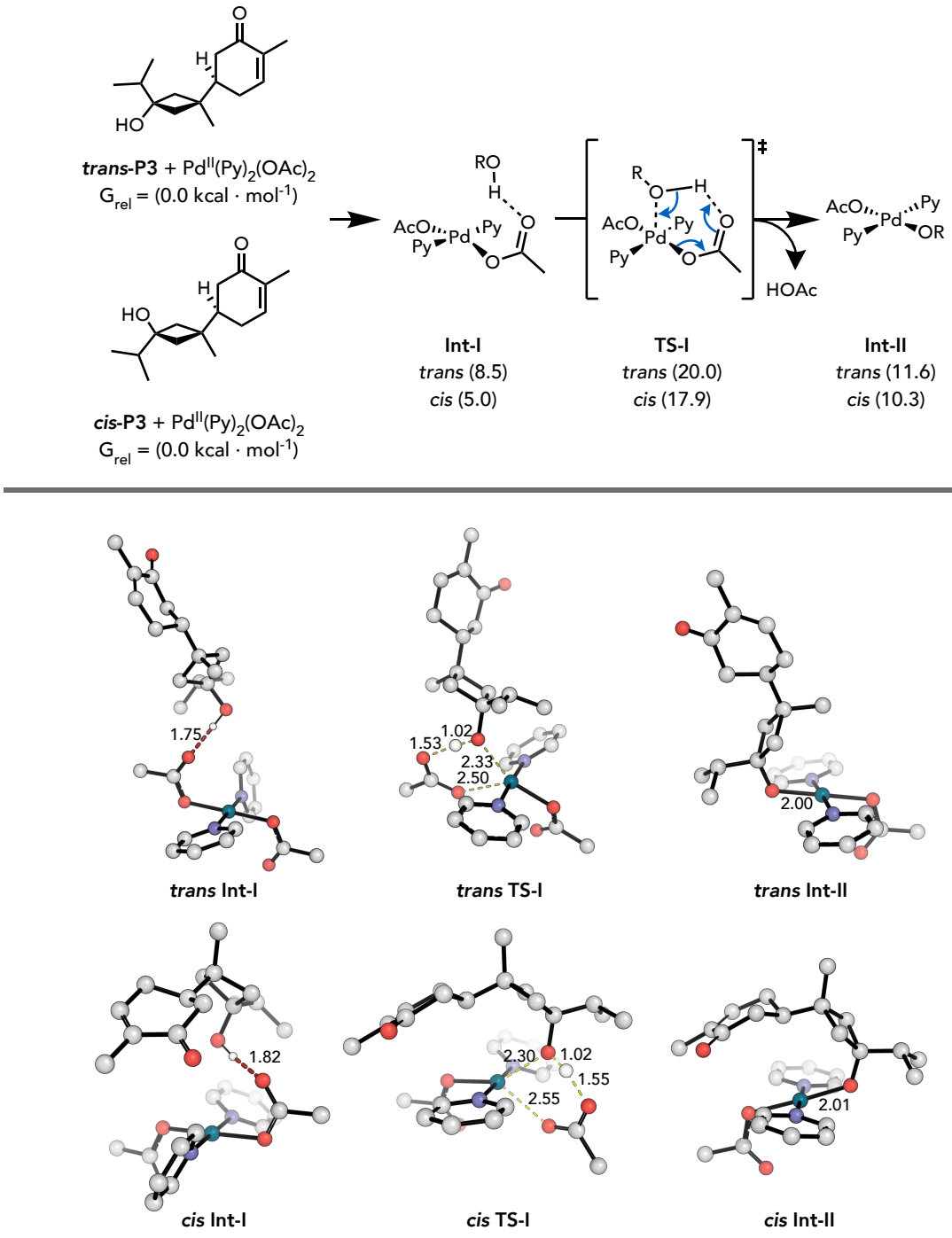
*5.3.2 Reaction Mechanism of the trans-Pathway*

In the analysis of the reaction mechanism from t*rans*-P3: we evaluated the Gibbs energy profile starting from cyclobutanol ***trans*-P3**, including the formation of desired ***S*-P2** and undesired ***R*-P2** bicyclic products **(**Scheme 5.3).



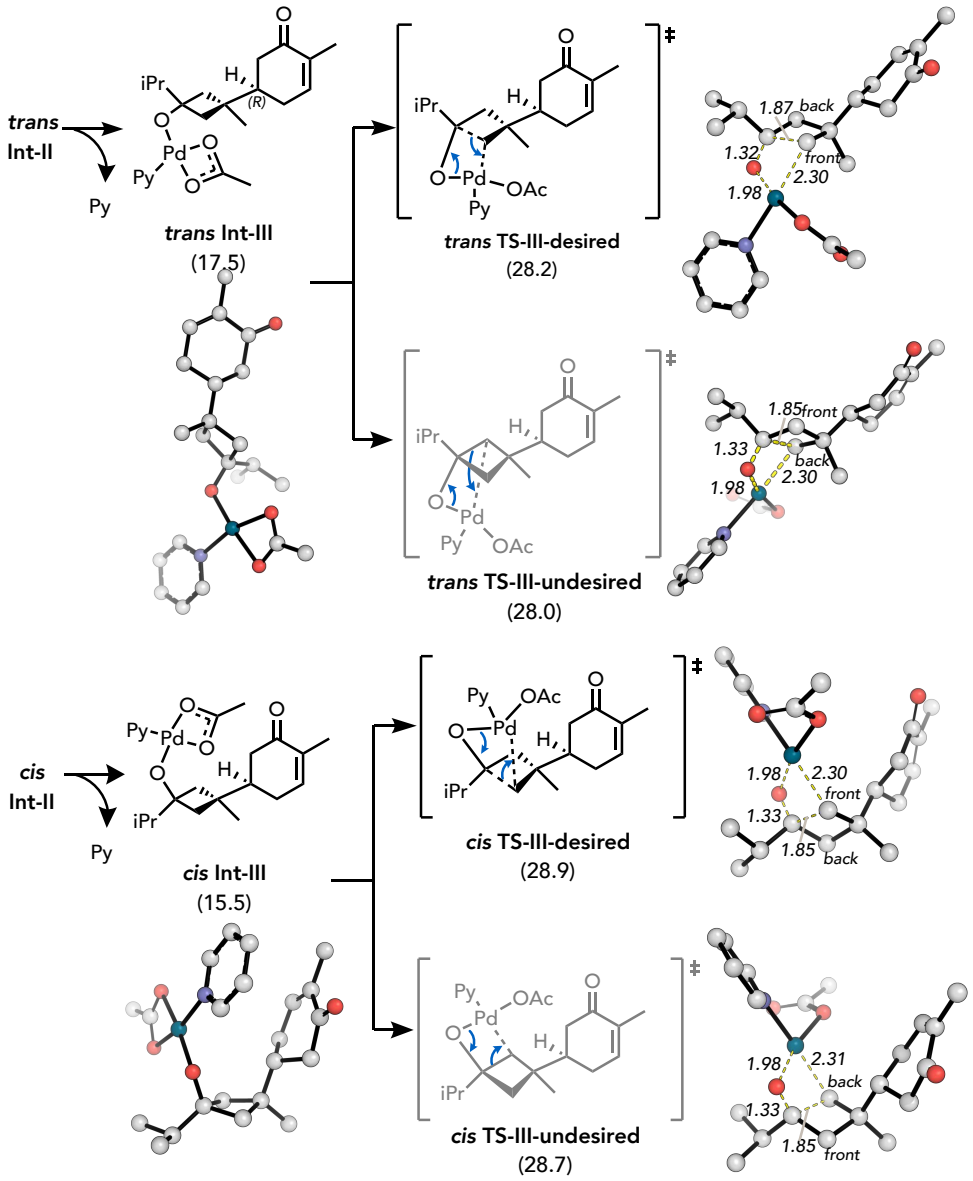

**Scheme 5.3.** Computed Gibbs energy profile (1 mol/l, 80°C, shown in kcal·mol$^{-1}$) of the proposed mechanism for Pd-catalyzed cascade reaction forming the desired and undesired diastereomers, ***S*-P2** and ***R*-P2**, respectively, from cyclobutanol ***trans*-P3**.

Coordination of the cyclobutanol to $Pd(Pyr)_2(OAc)_2$ shown in Scheme 5.4 can occur through **TS-I** with barriers of 20 and 18 kcal/mol for the *cis*- and *trans*-reaction pathways, respectively.



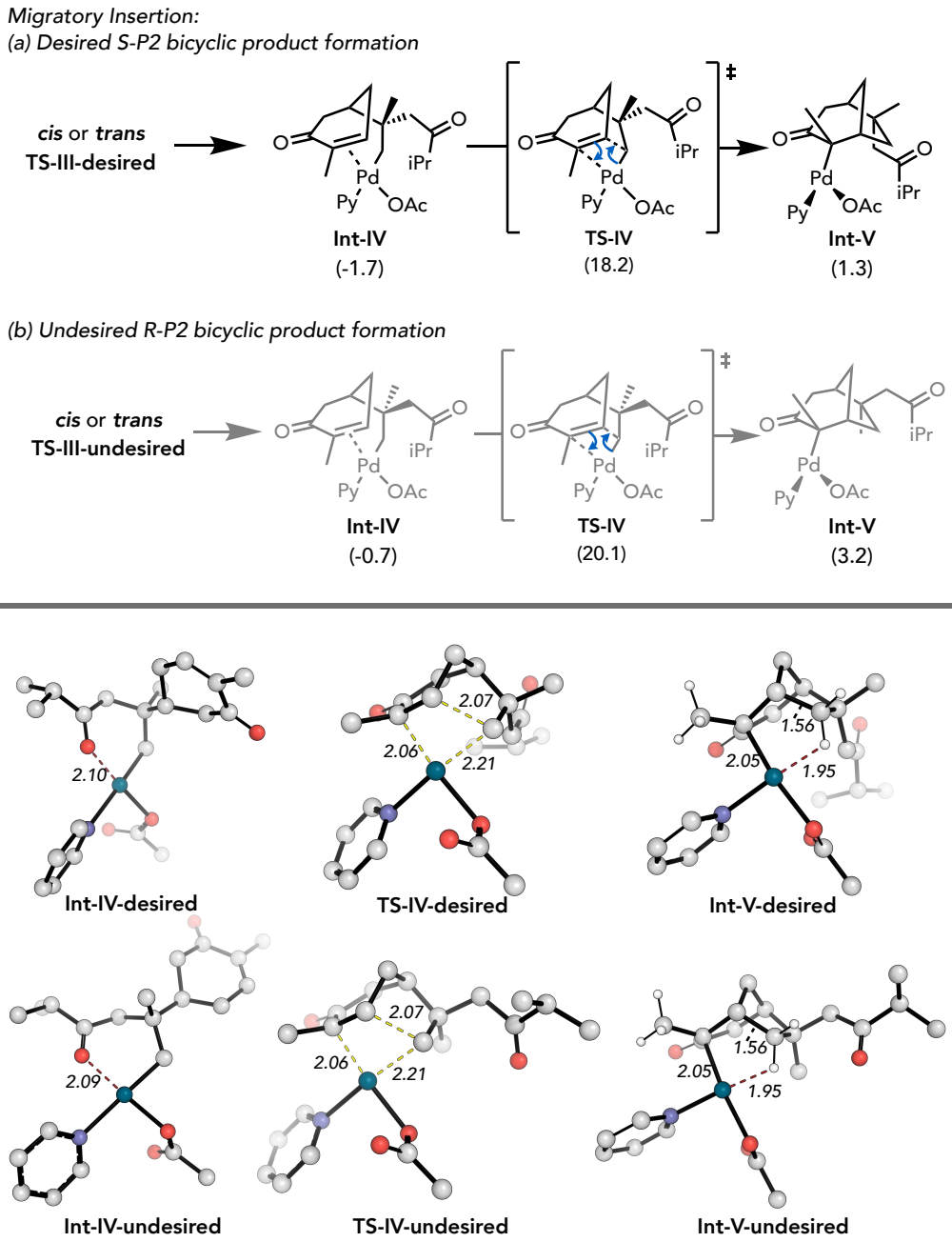

**Scheme 5.4. Top:** Schematic representation of ***cis*-** and ***trans*-P3** binding to Pd-catalyst via an alkoxide formation TS. G values are Boltzmann weighted for each *cis*- and *trans*-reaction pathway. **Bottom:** Most stable conformers for the binding of ***cis*-** and ***trans*-P3** to the Pd-catalyst via **TS-I** showing the distances of bonds forming/breaking and H–bonds as yellow and red dotted lines, respectively. Distances are shown in Å.

To form intermediate **Int-III**, the endergonic loss of a pyridine ligand from **Int-II** ($\Delta G$ = 6 and 5 kcal/mol for *cis*- and *trans*-reaction pathways, respectively) is required. The subsequent β-elimination step (Scheme 5.5) is predicted to occur irreversibly via **TS-III**, making this TS diastereoselectivity-determining. Of the steps computed, this TS is also turnover determining, with a barrier of around 28 kcal/mol in the *trans*-pathway. The calculated levels of selectivity for cleavage of the two β-C–C bonds, resulting in diastereomeric products, is low: $\Delta\Delta G^{\ddagger}$ between the desired and undesired pathways is 0.2 kcal/mol, consistent with the low levels of diastereoselectivity obtained experimentally.




**Scheme 5.5.** Most stable conformers for β-C–C elimination via **TS-III** forming desired (black) and undesired (gray) diastereomeric intermediates with relative Gibbs energies (kcal/mol).

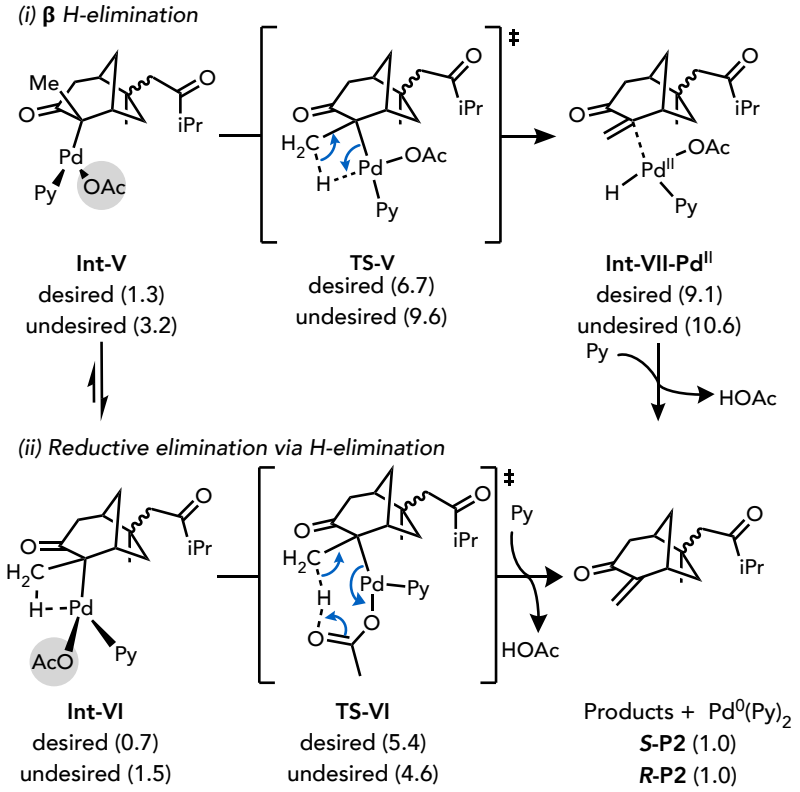
Migratory insertion via **TS-IV** (Scheme 5.6) can proceed with barriers of 18 and 20 kcal/mol for the desired and undesired products, respectively. However, there is no equilibration of **Int-IV** diastereomers, so the selectivity has already been determined by **TS-III**.




**Scheme 5.6.** Most stable conformers for migratory insertion via TS-IV involving desired (black) and undesired (gray) diastereomeric intermediates with relative Gibbs energies (kcal/mol).

The bicyclic product is formed from the Pd-complex **Int-V** through either β-H elimination or an acetate-assisted H-elimination in which $Pd^{II}$ is reduced to $Pd^{0}$ (Scheme 5.7 and Figure 5.2). The latter **TS-VI** is favored after a ligand isomerization gives the **Int-VI** structure (e.g., acetate *cis* to the β-methyl). We found this TS, forming Pd(0) and acetic acid directly, is more favorable (by

1.3 kcal/mol for the desired ***S*-P2** product) than the four-centered pathway forming an intermediate Pd(II)-hydride, **Int-VII**.[16]




**Scheme 5.7.** Competing pathways obtained for β-H-elimination: (i) 4-center TS forming a Pd(II) hydride; (ii) acetate-assisted deprotonation forming Pd(0) and acetic acid in a single step.

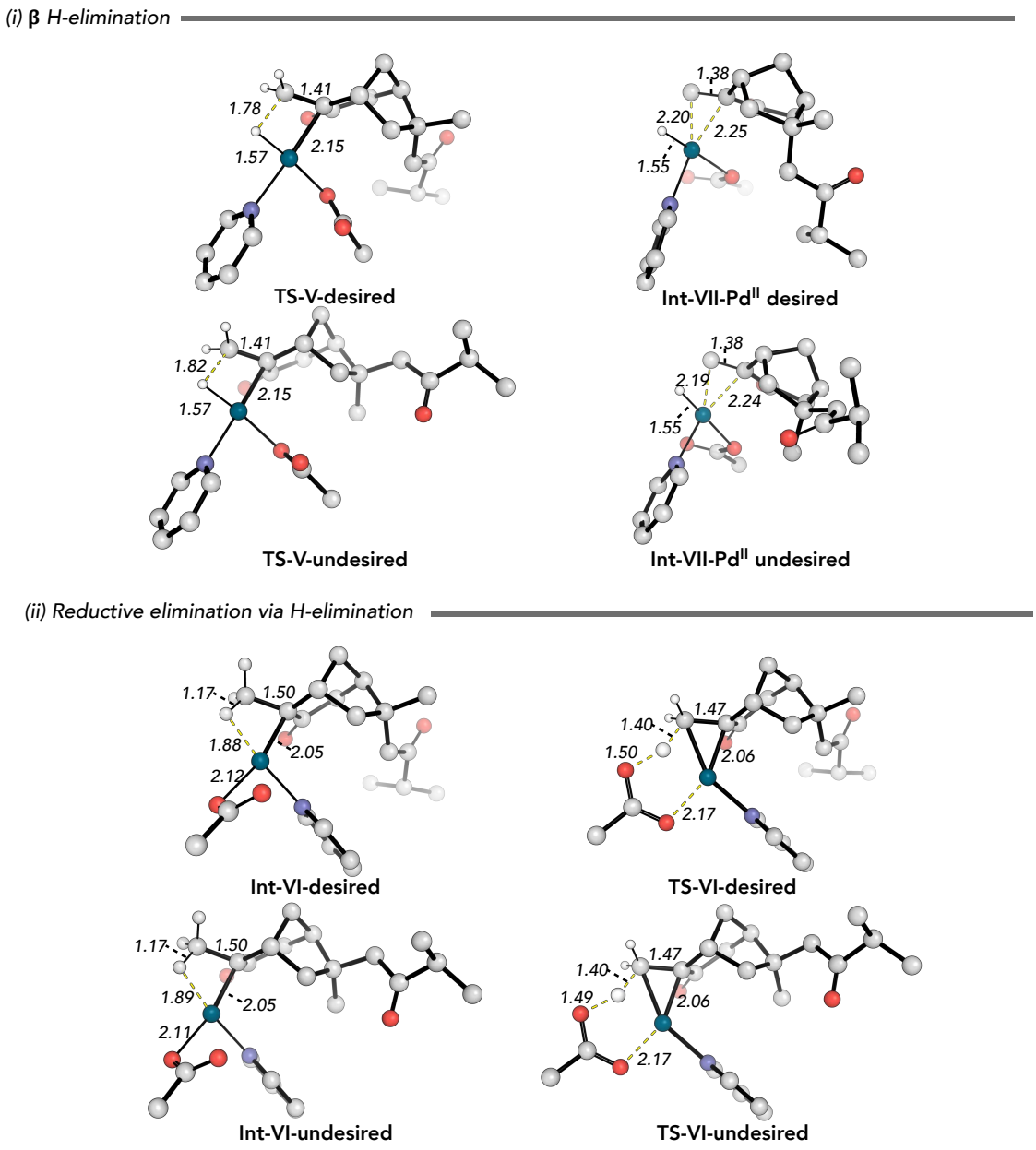



**Figure 5.2.** Most stable conformers for β-H-elimination leading to the formation of the desired ***S*-P2** and undesired ***R*-P2** products *via* the competing pathways.

### *5.3.3 Regenerating the Palladium Catalyst*

In our computed catalytic cycle, the energetic span between catalyst resting state and **TS-III** of 28.9 kcal/mol limits turnover. Accessing a coordinatively unsaturated Pd-center requires ligand dissociation, while a spectator pyridyl ligand remains in the key β-elimination TS. Modifications

to ligand sterics and electronics influence these elementary steps and hence overall catalytic rate. Further, while the regeneration of Pd(II) with molecular oxygen was not computed here, prior mechanistic studies[17] have established ligand identity and stoichiometry play an important role during this stage influencing catalyst decomposition through precipitation of palladium black. Thus, in collecting ligand parameters we focused on shape, steric, and electronic features, along with explicit computation of Pd-ligand dissociation energies (Figure 5.3).

*5.3.4 Parameterization of the Defined Chemical Space*

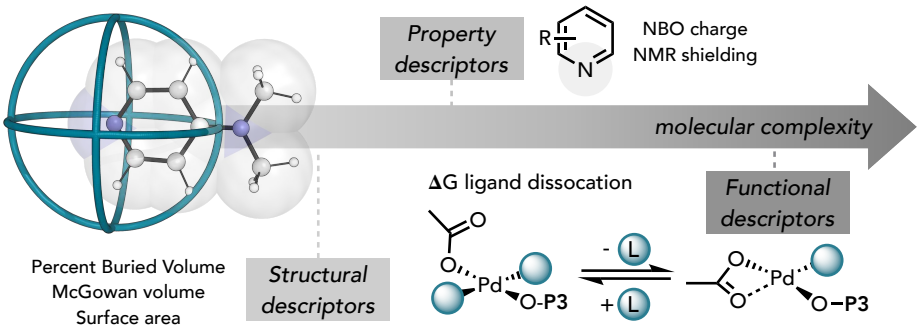



**Figure 5.3.** Mechanistically informed descriptors of varying complexity (structural, electronic and energetic) were calculated for the screened ligands.

Informed by mechanistic insights from the DFT-computed energy profile, we set out to obtain a series of electronic, structural, and energetic parameters for each ligand shown in Figure 5.4 to be used as features used for reaction optimization. We aimed to include distinct ligand classes to see if there was a preference for another class under different conditions. Ultimately the selection of ligands was also guided by what was readily available in the laboratory for ease of screening.

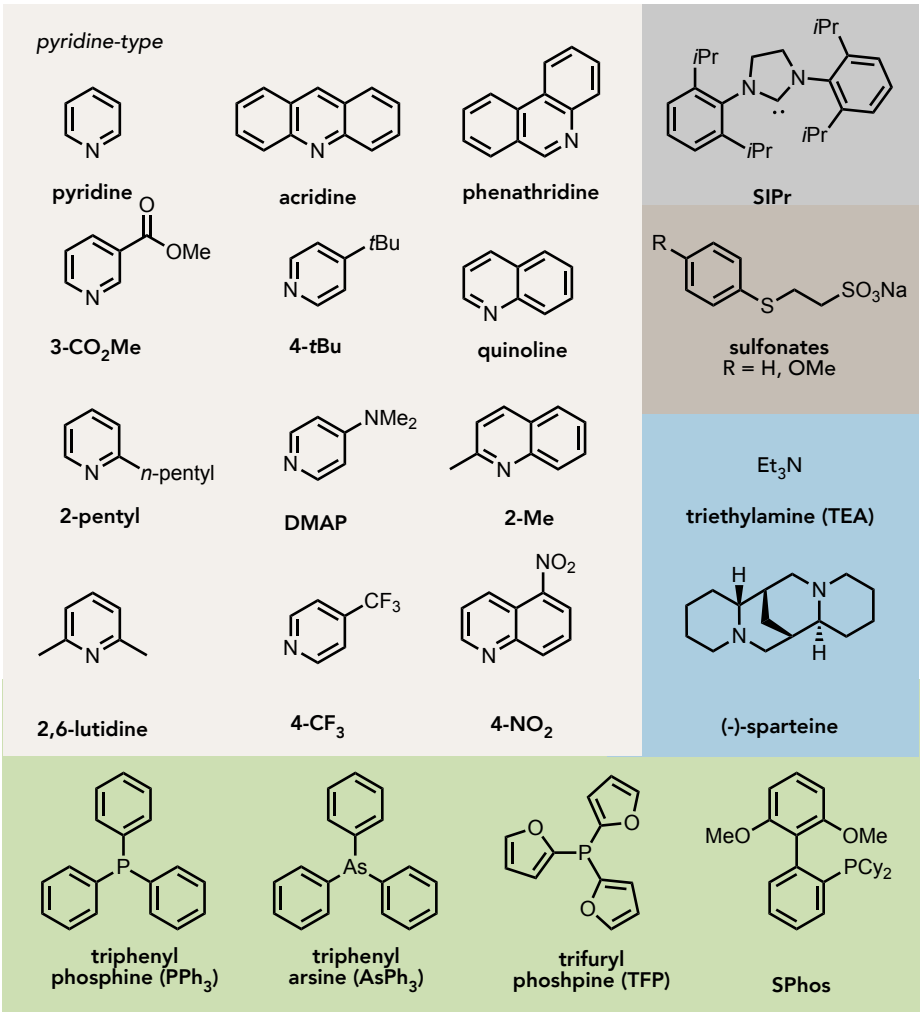



**Figure 5.4.** Ligands parameterized.

For ligand featurization summarized in Table 5.1, we generated a series of QM-derived features, first by optimization at the ωB97X-D/6-31G(d) level of theory, and then using NBO 7.0[18] to obtain natural atomic charges and gauge-independent atomic orbital (GIAO)[19] calculations at the mPW1PW91/6-311+G(d,p)[20] level of theory to obtain isotropic NMR shielding tensors. Using the developed open-source workflow, DISCO[21] (e.g., discussed in Chapter 7), we collected the computed NBO atomic charges, molecular orbital, and NMR chemical tensor values in an automated manner. Molar volumes were obtained from the electronic isosurface in Gaussian16, and buried volumes (centered at the metal-ligating atom) were calculated using the *DBSTEP*

program[22]. Ligand dissociation Gibbs energies (ΔG) were computed for *t*-butanol (as a simplified model substrate) for the dissociation of ligand L to form PdL from PdL2 (e.g., $Pd(OAc)(Py)_2$ → Pd(OAc)(Py) + Py). Using *Goodvibes*, the ligand dissociation energies were extracted from the optimized ground state structures with single-point energy calculations at the different SMD solvents (e.g., toluene, 1,2-dichloroethane, dimethyl sulfoxide, and tetrahydrofuran). In addition to QM-derived electronic descriptors, we included empirical parameters for surface area, volume, polarity, and atomic charges using the Mordred McGowan Volume module and RDKit MolSurf and Estate modules.[23] Experimental solvent parameters were collected for the four different solvents from the Minnesota solvation model database.[24]

**Table 5.1:** Summary of ligand and solvent parameters used within this work.

| Ligand Parameters | Source | Symbol (*units*) |
|---|---|---|
| % Buried Volume | DBSTEP | BV (*%)* |
| Molar Volume | Gaussian 16 | MV (*mL/mol*) |
| Ligand dissociation energy | | ΔG (*kcal/mol*) |
| GIAO NMR shielding tensor | | NMR (*ppm*) |
| Natural atomic charge | NBO 7 | NBO (*a.u.*) |
| McGowan Volume | Mordred McGowan Volume module | V ($mL/\,mol \cdot 100$) |
| Labute's Approximate Surface Area | RDKit MolSurf module | Labute |
| Partial charges and surface area contribution | | PEOE |
| Wildman-Crippen MR and surface area contribution | | SMR |
| Wildman-Crippen LogP and surface area contribution | | SLOGP |
| EState indices and surface area contribution | RDKit Estate module | ESTATE |

| Solvent Parameters | Source | Symbol (*units*) |
|---|---|---|
| Index of refraction | Minnesota Solvent Database | n |
| Index of refraction at optical frequencies at 298 K | | n25 |
| Hydrogen bond acidity | | alpha |
| Hydrogen bond basicity | | beta |
| Surface tension | | gamma |
| Dielectric constant or relative permittivity | | epsilon |
| Aromaticity | | phi |
| Electronegative halogenicity | | psi |

*5.3.5 Summary of the Implementation of the Bayesian Optimization Tool*

Based on our computational findings, we opted to screen electronically diverse ligands, including pyridines and quinolines, sulfonates, phosphines, NHC ligands, as well as basic amines (Figure 5.4) as the key variable to influence the reaction outcome. We collected physically meaningful descriptors to build the reaction space and selected the buried volume, atomic charge, molar volume, NMR shielding, and ΔG of ligand dissociation (Figure 5.3) for these various ligands. Macroscopic and atomistic solvent descriptors for four solvents were also selected to include in the model. Given previous results indicating the success of pyridine, we selected several pyridine-based ligands with a variety of electronic environments to allow the Bayesian optimizer developed by the Doyle lab to capitalize on subtle changes that might have significant effects. While several experiments had already been carried out, the Doyle group trained the EDBO with random initialization using previous experimental data to guide our selection of the optimization space. In turn, the EDBO tool proposed a set of 20 new experiments and carried out by Dr. Melissa Hardy, and the process was iterated. After seven rounds of optimization by EDBO, the best

conditions observed enabled comparable yields to those previously employed using sub-stoichiometric quantities of palladium. Interestingly, in these reactions, the changes between the previous results and those obtained by EDBO were subtle.

### *5.4 Conclusion*

The work in this chapter represents an "in the wild" application of Bayesian algorithms for the optimization of yield in a chemical reaction. The full computational understanding using quantum mechanical methods of the reaction surface performed by the present author guided the selection of atomistic and molecular representations to carry-out a data-driven modeling approach in this reaction optimization challenge. In only 140 experiments, we arrived at conditions that improved the catalytic efficiency of this reaction when previously, the experimental group had not observed any turnover of the Pd-complex. Notably, the proposed changes could not have been determined through a single variable screening as various realistic challenges (e.g., racemic mixtures, byproduct formation, and catalyst decomposition) were faced within a natural products synthetic plan.

## REFERENCES


1. Dowman, L. J.; Kulkarni, S. S.; Alegre-Requena, J. V.; Giltrap, A. M.; Norman, A. R.; Sharma, A.; Gallegos, L. C.; Mackay, A. S.; Welegedara, A. P.; Watson, E. E.; Van Raad, D.; Niederacher, G.; Huhmann, S.; Proschogo, N.; Patel, K.; Larance, M.; Becker, C. F. W.; Mackay, J. P.; Lakhwani, G.; Huber, T.; Paton, R. S.; Payne, R. J., Site-selective photocatalytic functionalization of peptides and proteins at selenocysteine. *Nat. Commun.* **2022,** *13* (1).
2. Driscoll, D. M.; Copeland, P. R., MECHANISM AND REGULATION OF SELENOPROTEIN SYNTHESIS. *Annu. Rev. Nutr.* **2003,** *23* (1), 17-40.
3. (a) Steinmann, D.; Nauser, T.; Koppenol, W. H., Selenium and Sulfur in Exchange Reactions: A Comparative Study. *J. Org. Chem.* **2010,** *75* (19), 6696-6699; (b) Mousa, R.; Notis Dardashti, R.; Metanis, N., Selenium and Selenocysteine in Protein Chemistry. *Angew. Chem., Int. Ed.* **2017,** *56* (50), 15818-15827.
4. (a) Cohen, D. T.; Zhang, C.; Pentelute, B. L.; Buchwald, S. L., An Umpolung Approach for the Chemoselective Arylation of Selenocysteine in Unprotected Peptides. *J. Am. Chem. Soc.* **2015,** *137* (31), 9784-9787; (b) Cohen, D. T.; Zhang, C.; Fadzen, C. M.; Mijalis, A. J.; Hie, L.; Johnson, K. D.; Shriver, Z.; Plante, O.; Miller, S. J.; Buchwald, S. L.; Pentelute, B. L., A chemoselective strategy for late-stage functionalization of complex small molecules with polypeptides and proteins. *Nature Chemistry* **2019,** *11* (1), 78-85; (c) Levengood, M. R.; Van Der Donk, W. A., Dehydroalanine-containing peptides: preparation from phenylselenocysteine and utility in convergent ligation strategies. *Nat. Protoc.* **2006,** *1* (6), 3001-3010.
5. (a) 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; 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 Rev. C.01*, Wallingford, CT, 2016; (b) Luchini, G.; Alegre-Requena, J.; Funes-Ardoiz, I.; Paton, R., GoodVibes: automated thermochemistry for heterogeneous computational chemistry data [version 1; peer review: 2 approved with reservations]. *F1000Research* **2020,** *9* (291); (c) Heverly-Coulson, G. S.; Boyd, R. J., Systematic Study of the Performance of Density Functional Theory Methods for Prediction of Energies and Geometries of Organoselenium Compounds. *J. Phys. Chem. A* **2011,** *115* (18), 4827-4831; (d) Pearson, J. K.; Ban, F.; Boyd, R. J., An Evaluation of Various Computational Methods for the Treatment of Organoselenium Compounds. *J. Phys. Chem. A* **2005,** *109* (45), 10373-10379.
6. (a) Downs, R. T.; Hall-Wallace, M., The American Mineralogist crystal structure database. *American Mineralogist* **2003,** *88*, 247-250; (b) Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A., Crystallography Open Database – an open-access collection of crystal structures. *Journal of Applied Crystallography* **2009,** *42* (4), 726-729; (c) Hua, G.; Fuller, A. L.; Slawin, A. M. Z.;

Woollins, J. D., 1,2-Bis(2-bromobenzyl)diselane. *Acta Crystallographica Section E Structure Reports Online* **2010,** *66* (10), o2579-o2579.
7. Bungu, P. N.; Otto, S., 2-Isobutyl-2-phosphabicyclo[3.3.1]nonane 2-selenide. *Acta Crystallographica Section E Structure Reports Online* **2009,** *65* (3), o560-o561.
8. Jones, G. O.; Liu, P.; Houk, K. N.; Buchwald, S. L., Computational Explorations of Mechanisms and Ligand-Directed Selectivities of Copper-Catalyzed Ullmann-Type Reactions. *J. Am. Chem. Soc.* **2010,** *132* (17), 6205-6213.
9. Vaissier, V.; Barnes, P.; Kirkpatrick, J.; Nelson, J., Influence of polar medium on the reorganization energy of charge transfer between dyes in a dye sensitized film. *Phys. Chem. Chem. Phys.* **2013,** *15* (13), 4804.
10. Fusetani, N.; Wolstenholme, H. J.; Matsunaga, S.; Hirota, H., Two new sesquiterpene isonitriles from the nudibranch, phyllida pustulosa. *Tetrahedron Lett.* **1991,** *32* (49), 7291-7294.
11. (a) Ho, T.-L.; Kung, L.-R., Stereoselective Synthesis of 2-Isocyanoallopupukeanane. *Org. Lett.* **1999,** *1* (7), 1051-1052; (b) Ho, T.-L.; Kung, L.-R.; Chein, R.-J., Total Synthesis of (±)-2-Isocyanoallopupukeanane. *J. Org. Chem.* **2000,** *65* (18), 5774-5779.
12. Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G., Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **2021,** *590* (7844), 89-96.
13. 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; 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 Rev. B.01*, Wallingford, CT, 2016.
14. Emmert, M. H.; Cook, A. K.; Xie, Y. J.; Sanford, M. S., Remarkably High Reactivity of Pd(OAc)2/Pyridine Catalysts: Nondirected C□H Oxygenation of Arenes. *Angew. Chem., Int. Ed.* **2011,** *50* (40), 9409-9412.
15. Steinhoff, B. A.; Guzei, I. A.; Stahl, S. S., Mechanistic Characterization of Aerobic Alcohol Oxidation Catalyzed by Pd(OAc)2/Pyridine Including Identification of the Catalyst Resting State and the Origin of Nonlinear [Catalyst] Dependence. *J. Am. Chem. Soc.* **2004,** *126* (36), 11268-11278.
16. (a) Mekareeya, A.; Walker, P. R.; Couce-Rios, A.; Campbell, C. D.; Steven, A.; Paton, R. S.; Anderson, E. A., Mechanistic Insight into Palladium-Catalyzed Cycloisomerization: A Combined Experimental and Theoretical Study. *J. Am. Chem. Soc.* **2017,** *139* (29), 10104-10114; (b) Zhu, C.; Yang, B.; Mai, B. K.; Palazzotto, S.; Qiu, Y.; Gudmundsson, A.; Ricke, A.; Himo, F.; Bäckvall, J.-E., Highly Selective Palladium-Catalyzed Hydroborylative Carbocyclization of Bisallenes to Seven-Membered Rings. *J. Am. Chem. Soc.* **2018,** *140* (43), 14324-14333; (c) Gellrich, U.; Meißner, A.; Steffani, A.; Kähny, M.; Drexler, H.-J.; Heller, D.; Plattner, D. A.; Breit, B., Mechanistic Investigations of the Rhodium Catalyzed Propargylic CH Activation. *J. Am. Chem. Soc.* **2014,** *136* (3), 1097-1104.

17. Sigman, M. S.; Jensen, D. R., Ligand-Modulated Palladium-Catalyzed Aerobic Alcohol Oxidations. *Acc. Chem. Res.* **2006,** *39* (3), 221-229.
18. Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Karafiloglou, P.; Landis, C. R.; Weinhold, F. *NBO 7.0*, Theoretical Chemistry Institute, 2018.
19. Wolinski, K.; Hinton, J. F.; Pulay, P., Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. *J. Am. Chem. Soc.* **1990,** *112* (23), 8251-8260.
20. Adamo, C.; Barone, V., Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J. Chem. Phys.* **1998,** *108* (2), 664-675.
21. Gallegos, L. C. DISCO. https://github.com/Liliana-Gallegos/DISCO.
22. Luchini, G.; Patterson, T.; Paton, R. S. DBSTEP Release 1.1.0 (1.1.0) Zenodo. patonlab/DBSTEP.
23. Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T., Mordred: a molecular descriptor calculator. *J. Cheminformatics* **2018,** *10* (1).
24. 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* (18), 6378-6396.

# 6 INCORPORATING STASTICAL AND MACHINE LEARNING MODELS FOR REACTION OPTIMIZATION

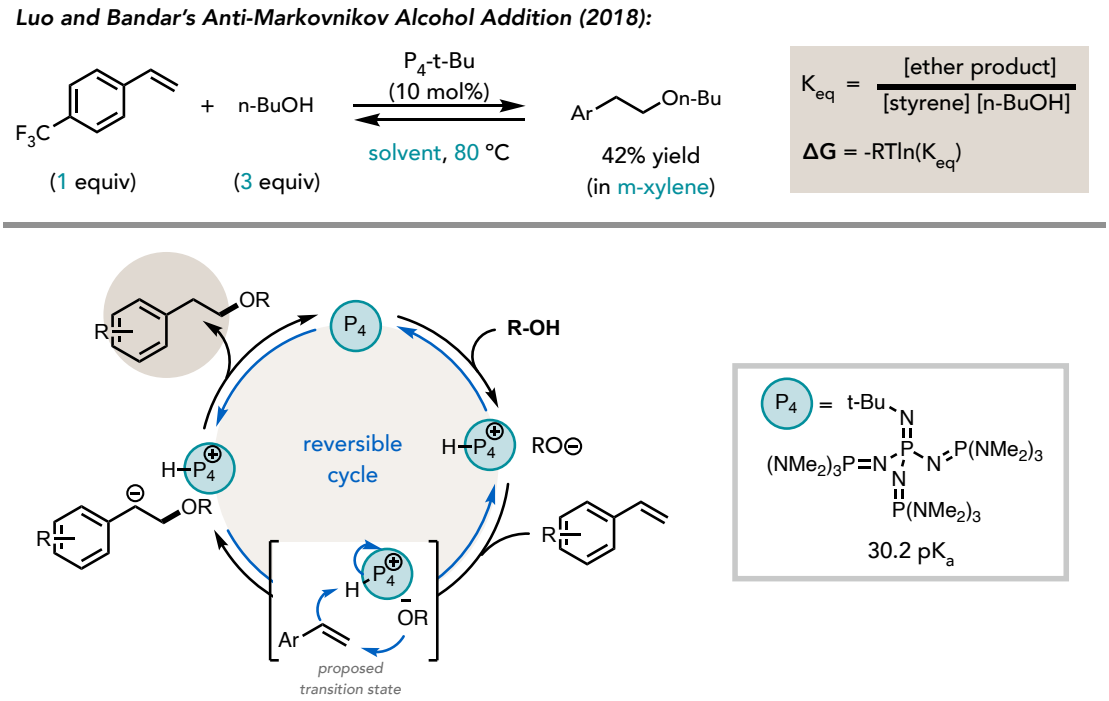
## *6.1 Chapter Overview*

The previous chapters discussed in detail the quantum mechanical approaches for studying reaction mechanisms to predict reaction outcomes. In this chapter, we instead focus on the data-driven approach to formulate statistical relationships sampled over small and large datasets. First, we dive into building multivariate linear regression model with a small dataset to explain the reaction performance in various solvents on the challenging reversible nucleophilic alcohol addition reaction developed by the Bandar group. The collaborative research in section 6.2 was published in the *Journal of the American Chemical Society*.[1] Moving into the second section of this chapter, a machine learning model is trained on a large diverse molecule dataset to predict NMR chemical shifts with high accuracy to DFT-derived NMR values at only a fraction of the cost of DFT methods. Here are two examples where we evaluate predictive model success based on the research goal centered on accuracy or interpretability. The collaborative research in section 6.3 was published in *Chemical Science*.[2]

## *6.2 Statistical Methods Capture Solvent effects in Challenging Nucleophilic Alcohol Addition Reaction*

### *6.2.1 The Base-Catalyzed Anti-Markovnikov Alcohol Addition Reaction*

In 2018, Bandar *et al* proposed the use of the organic superbase $P_4$-*t*-Bu (p$K_{BH+}$ 30.2 in DMSO) as a catalyst for the nucleophilic addition of alcohols to activated styrene derivatives (Scheme 6.1).[3] To extend the reactivity to unactivated styrenes, one of the challenges to address included controlling thermodynamic parameters such that careful reaction conditions can be optimized (e.g., reduce the alcohol concentration and lower the temperature) thereby, maximizing the reaction yield performance.




**Scheme 6.1. Top:** Overview of reported base-catalyzed alcohol addition reactions using organic superbase, P4-*t*-Bu. **Bottom:** the hypothesized reversible catalytic cycle with a proposed P4-*t*-Bu conjugate acid participating in hydrogen bonding that could rapidly protonate alkoxide addition adducts in either a stepwise or concerted process.

In this collaborative project with the Bandar group, we developed a predictive model of this challenging, nucleophilic alcohol addition reaction to (i) predict experimental equilibrium ΔG in a total of 20 different solvents, (ii) explain the solvent effects influencing the reactivity, and (iii) suggest reaction conditions for optimal reaction performance.[1] All experimental work was performed by our experimental collaborators (Bandar *et. al*), all DFT mechanistic calculations were computed by Dr. Juan V. Alegre-Requena, and all statistical analysis was performed by the present author.

### *6.2.2 Solvent Features and Chemical Space*

The python-based package, Scikit-learn, was utilized to build the solvent chemical space using dimension reduction and clustering algorithms. This workflow can be found in the Jupyter

Notebook, Solvent-Analysis-PCA.ipynb along with additional supporting data in the GitHub: Liliana-Gallegos/Solvent_regression_analysis repository.[4]

**Table 6.1.** Solvent descriptors selected in this work were taken from the listed databases.

| Database | Descriptor – abbreviation (*units*) |
|---|---|
| Minnesota solvation | index of refraction – n |
| | hydrogen bond basicity – beta |
| | surface tension – gamma |
| | dielectric constant or relative permittivity – dielectric |
| Hansen solubility | dispersion ($MPa^{1/2}$) |
| | polar ($MPa^{1/2}$) |
| | hydrogen bonding – hydrogen ($MPa^{1/2}$) |
| | Sum of three contributions: dispersion, polar, and hydrogen interactions – sum ($MPa^{1/2}$) |
| DFT-derived | Boltzmann average molar volume – MV_boltz (*mL/mol)* |
| COSMO-RS | area ($Å^2$) |
| | volume ($Å^3$) |
| | dipole (*Debye*) |
| | polarity – Sig2 |
| | polarizability – Sig3 |
| | hydrogen bonding acceptor – Hbond_acc |
| Abraham's solvation | basicity – B |
| | McGowan volume – V ($mL/\,mol \cdot 100$) |

We began investigating for potential solvent properties that could be influencing the reactivity. For the 20 solvents studied experimentally, we constructed a set of 17 molecular descriptors that capture electronic and structural differences. We collected the following solvent descriptors listed in Table 6.1 taken from the Minnesota solvation database, Hansen solubility

database,[5] *COSMO*ments (*COSMO-RS* calculated and Boltzmann averaged values)[6], and selected Abraham's descriptors database.[7]

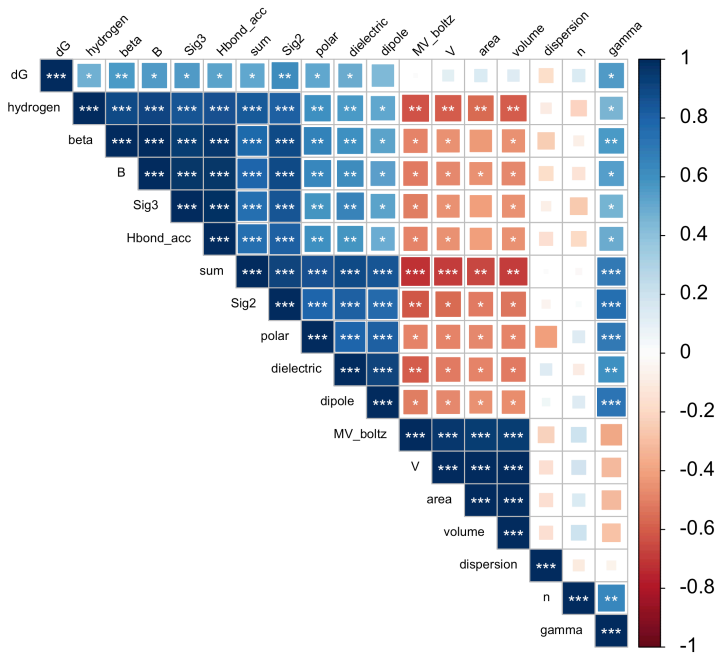



**Figure 6.1.** The Pearson correlation heatmap demonstrating high correlation of the descriptors and response value with dark blue or red color and weak correlations with lighter red or blue shade.

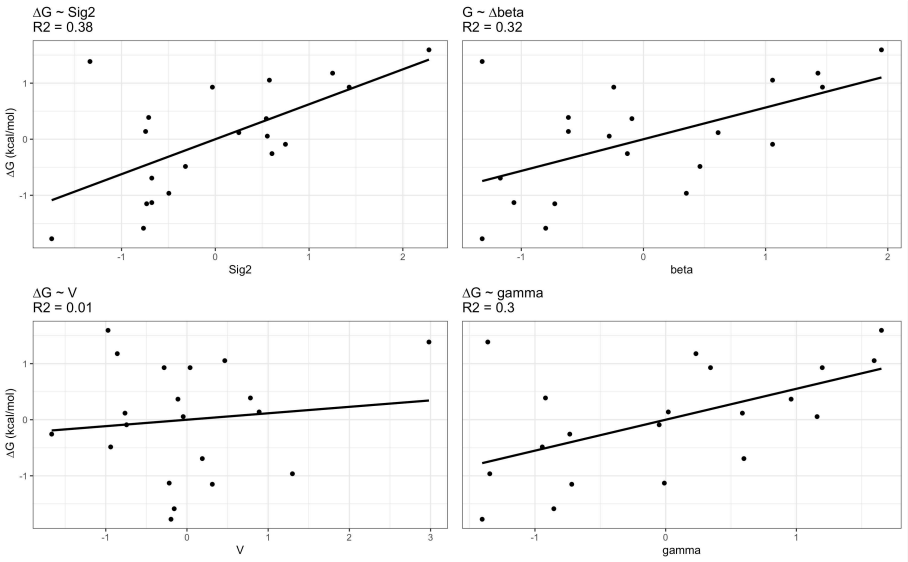



**Figure 6.2.** The one-variable correlations of ΔG with the individual descriptors shown include top three correlated descriptors and the McGowan's characteristic volume (V) descriptor.

Computed Pearson correlations of ΔG with the individual descriptors shown in Figure 6.1 were not high enough in any case to create a robust univariate model. Therefore, attempts to connect the $K_{eq}$ and ΔG values with a single solvent property (e.g., polarity, hydrogen bond basicity, surface tension, or McGowan volume) led to low correlations (Figures 6.2). Therefore, a multivariate analysis was performed to determine whether multiple solvent properties may be related to the ΔG of alcohol addition.[8]

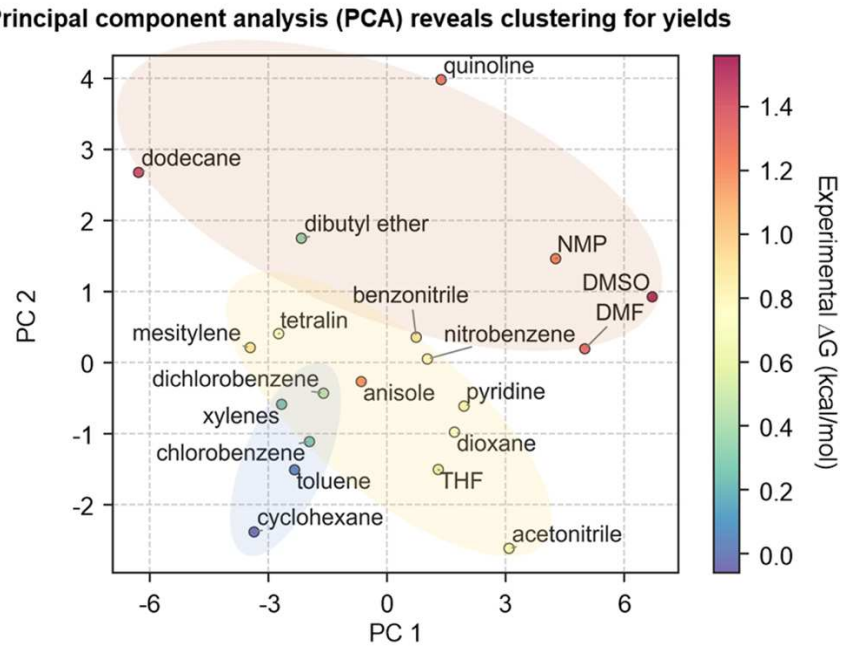



**Figure 6.3.** Principal component analysis and k-means clustering method applied for all 20 solvents colored by its reaction performance in terms of ΔG in kcal/mol.

We initiated this in-depth statistical analysis with a principal component analysis (PCA) to visualize the solvent chemical space (e.g., the variation of ΔG in this feature space with all 20 solvents described by 17 features). Plotting the reaction performance against the first two PCs shows clusters of similarly performing solvents in distinct regions of feature space in Figure 6.3. Based on this analysis, measurement of the equilibrium constant ($K_{eq}$) in various solvents revealed that nonpolar and aromatic solvents generally provide the highest $K_{eq}$ values, while polar solvents

exhibit low $K_{eq}$ values.16 The first two PCs capture a relatively high proportion (75%) of the overall sample variance, with 60% for PC1 and 15% for PC2. Loadings for the first two principal components are summarized in Table 6.2 representing the contribution from individual descriptors.

**Table 6.2.** The descriptors weighted contribution to the two principal components (PC1 and PC2).

| Descriptors | Loadings/weights | |
|---|---|---|
| | PC1 | PC2 |
| *Electronic descriptors* | | |
| sum | 0.300 | 0.017 |
| Sig2 | 0.295 | 0.142 |
| B | 0.276 | 0.175 |
| hydrogen | 0.275 | 0.040 |
| beta | 0.274 | 0.204 |
| Sig3 | 0.268 | 0.149 |
| Hbond_acc | 0.265 | 0.165 |
| dielectric | 0.257 | 0.027 |
| polar | 0.253 | 0.165 |
| dipole | 0.240 | 0.078 |
| gamma | 0.206 | 0.240 |
| dispersion | -0.014 | -0.338 |
| n | -0.028 | 0.251 |
| *Volume descriptors* | | |
| area | -0.228 | 0.395 |
| volume | -0.234 | 0.395 |
| V | -0.236 | 0.387 |
| MV_boltz | -0.246 | 0.354 |

*6.2.3 Process for Feature and Model Selection*

The R-Studio program[9] was utilized for performing most of the statistical analysis to select the features and develop the model which can be found in the GitHub repository.[4, 10]

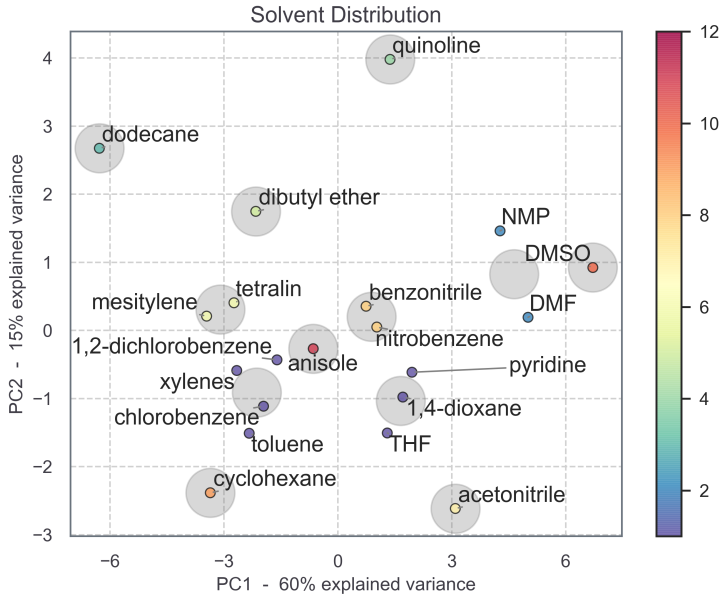



**Figure 6.4.** Twelve k-means clusters for the training set: cyclohexane, chlorobenzene, dibutyl ether, acetonitrile, dioxane, nitrobenzene, mesitylene, anisole, quinoline, DMF, dodecane, and DMSO. Each cluster is shown in a different color.

In order to obtain a uniformed training and test sample split, we applied the clustering method with the k-means algorithm from scikit-learn, where five to fourteen clusters were considered.[11] For each cluster shown in the PCA graph in Figure 6.4, the solvents were selected by its closest proximity to the center of the cluster to generate the training solvent data set (e.g., data compiled in the provided spreadsheet (Keq-solvents.xlsx) under the "PCA clusters" tab). The rest of the solvents not selected composed the test solvent data set.

We processed the dataset for all samples using both (i) forward stepwise and (ii) the dredge function (e.g., function in R-Studio in the MuMIn package to rank all possible models) for feature selection, followed by an Anova analysis (e.g., in the car package).

**Table 6.3.** Best models generated from stepwise feature selection and the dredge function ranked by its corresponding $R^2$ measurement. Highlighted in blue demonstrates the best $R^2$ performance for the least number of descriptors in the model.

| Dataset | Model | Selection process | Feature selection | $R^2$ | RMSE |
|---|---|---|---|---|---|
| All (20 solv) | 1 | Stepwise | ΔG ~ Sig2 + V + sum + dipole | 0.76 | 0.23 |
| | 2 | Stepwise | ΔG ~ Sig2 + V + sum | 0.73 | 0.24 |
| | 3 | Stepwise & dredge | ΔG ~ Sig2 + V | 0.7 | 0.26 |
| | 4 | dredge | ΔG ~ Sig2 + volume | 0.69 | 0.26 |
| | 5 | dredge | ΔG ~ area + Sig2 | 0.68 | 0.26 |
| Without Dodecane (19 solv) | 6 | dredge | ΔG ~ area + Sig2 + V | 0.79 | 0.21 |
| | 7 | dredge | ΔG ~ Sig2 + V | 0.73 | 0.24 |
| | 8 | Stepwise & dredge | ΔG ~ Sig2 + MV_boltz | 0.73 | 0.23 |

**Table 6.4.** Weighted descriptor importance of all possible models generated through the dredge function for the dataset: (i) all 20 solvents and (ii) 19 solvents (excluding dodecane). Descriptors highlighted in blue demonstrate a high variable significance in both datasets.

| Descriptor | Weights | |
|---|---|---|
| | All | Without Dodecane |
| Sig2 | 0.62 | 0.83 |
| sum | 0.57 | 0.3 |
| V | 0.49 | 0.42 |
| volume | 0.37 | 0.32 |
| area | 0.3 | 0.48 |
| dipole | 0.2 | 0.13 |
| hydrogen | 0.16 | 0.19 |
| B | 0.15 | 0.18 |
| MV_boltz | 0.14 | 0.25 |
| Sig3 | 0.13 | 0.12 |
| Hbond_acceptor | 0.12 | 0.15 |

The models generated (summarized in Table 6.3) were considered based on all samples (20 solvents) and without dodecane (19 solvents). The selected model (highlighted in blue in Tables 6.3-6.5) was favored based on the descriptor's importance from the calculated weights of all possible models generated through the dredge function using only two descriptors to capture the solvent effects of this challenging reaction. We also avoided selecting highly correlated variables shown in Figure 6.1 (e.g., with thresholds of $R^2 > 0.8$ between the two variables) to ensure the robustness of the models.

**Table 6.5.** Summary of results for each amount of PCA clusters included in the training sets, showing $R^2$ for the training set and $Q^2$ for the test set, each with its corresponding root mean square error (RMSE). The linear model formula shows coefficients obtained with scaled data.

| PCA clusters | LM formula | Training | | Test | |
|---|---|---|---|---|---|
| | | $R^2$ | RMSE | $Q^2$ | RMSE |
| 5 | ΔG = 1.02 + 0.92 * Sig2 + 0.97 * V | 0.94 | 0.11 | 0.17 | 0.49 |
| 6 | ΔG = 1 + 0.65 * Sig2 + 0.72 * V | 0.93 | 0.11 | 0.58 | 0.35 |
| 7 | ΔG = 0.95 + 0.56 * Sig2 + 0.66 * V | 0.88 | 0.14 | 0.60 | 0.34 |
| 8 | ΔG = 0.96 + 0.59 * Sig2 + 0.66 * V | 0.78 | 0.21 | 0.53 | 0.36 |
| 9 | ΔG = 0.84 + 0.54 * Sig2 + 0.53 * V | 0.92 | 0.14 | 0.43 | 0.35 |
| 10 | ΔG = 0.91 + 0.62 * Sig2 + 0.51 * V | 0.93 | 0.13 | 0.38 | 0.36 |
| 11 | ΔG = 0.94 + 0.56 * Sig2 + 0.47 * V | 0.57 | 0.3 | 0.85 | 0.2 |
| 12 | ΔG = 0.89 + 0.53 * Sig2 + 0.40 * V | 0.67 | 0.29 | 0.76 | 0.22 |
| 13 | ΔG = 0.92 + 0.52 * Sig2 + 0.38 * V | 0.68 | 0.28 | 0.77 | 0.23 |
| 14 | ΔG = 0.9 + 0.48 * Sig2 + 0.34 * V | 0.69 | 0.27 | 0.69 | 0.27 |

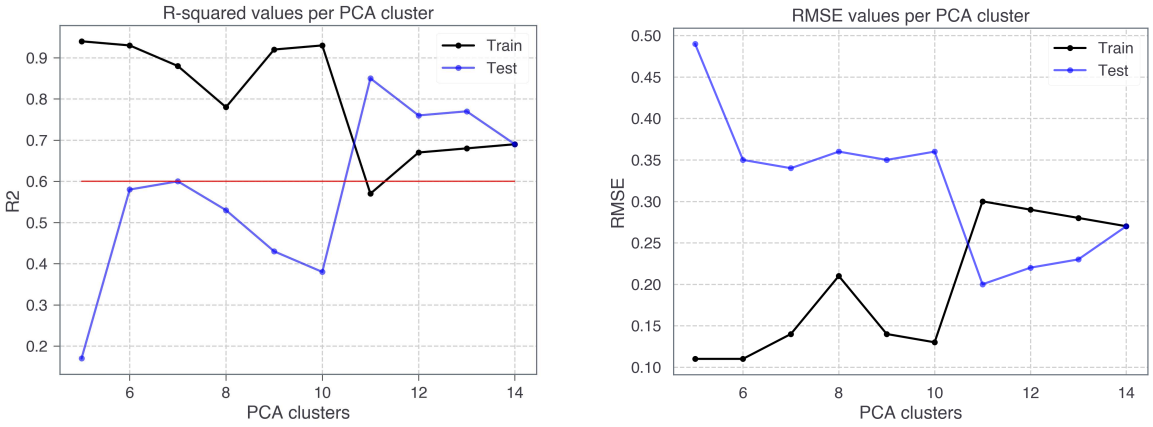



**Figure 6.5.** $R^2$ (for the training set) or $Q^2$ (for the test set) and RMSE values when using different amount of PCA clusters included in the training sets.

For each train and test split, the dataset was scaled based on the mean and standard deviation of the training dataset. Selection of the optimal amount of solvents to use as the training set was based on the $R^2$, $Q^2$ and RMSE values of training and test sets when using different amounts of PCA clusters as training sets (Figure 6.5 and Table 6.5). After 12 clusters, the $R^2$ and $Q^2$ values meet the minimum criteria (e.g, $R^2$ and $Q^2$ values greater than 0.6 shown as a red line in Figure 6.5) giving similar performance values and a satisfactory model. Therefore, we proceeded with the best performing model containing twelve solvents in the training set and eight solvents in the test set (e.g., 0.6 Training: 0.4 Test ratio).

### *6.2.4 Final Model Predicts Equilibrium Gibbs Free-Energies in Various Solvents*

The resulting multivariate linear regression model for ΔG uses two molecular descriptors, giving a mean absolute error (MAE) of 0.22 kcal/mol for a training set of 12 solvents and an MAE of 0.19 kcal/mol for a hold-out test set of eight solvents (Figure 6.6 and Table 6.6). The first of the statistically significant descriptors is Sig2, the second COSMO σ-moment, which describes a molecule's overall electrostatic polarity.[12] The second is the McGowan molar volume (V).[13]

**Table 6.6.** Results for the optimal linear model. Experimental ΔG was calculated with the formula ΔG = -RT ln($K_{eq}$), where $K_{eq}$ was obtained with the equilibrium concentration of reagents and product as $K_{eq}$ = [product]/[*n*-BuOH][styrene].

| Solvent | Experim. ΔG (kcal/mol) | Calcul. ΔG (kcal/mol) |
|---|---|---|
| Training | | |
| Cyclohexane | -0.06 | -0.05 |
| Chlorobenzene | 0.25 | 0.44 |
| Dibutyl ether | 0.33 | 1.02 |
| Acetonitrile | 0.67 | 0.56 |
| 1,4-dioxane | 0.75 | 0.93 |
| nitrobenzene | 0.82 | 1.07 |
| Mesitylene | 0.98 | 0.75 |
| Anisole | 1.24 | 0.82 |
| Quinoline | 1.3 | 1.25 |
| DMF | 1.36 | 1.13 |
| Dodecane | 1.46 | 1.18 |
| DMSO | 1.56 | 1.57 |
| | | MAE = 0.22 |
| | | SD = 0.17 |
| Test | | |
| Toluene | 0.03 | 0.42 |
| Xylene[a] | 0.24 | 0.59 |
| 1,2-dichlorobenzene | 0.46 | 0.57 |
| THF | 0.56 | 0.37 |
| Pyridine | 0.85 | 0.69 |
| Tetralin | 0.86 | 0.77 |
| Benzonitrile | 0.97 | 1.04 |
| NMP | 1.24 | 1.40 |
| | | MAE = 0.19 |
| | | SD = 0.14 |

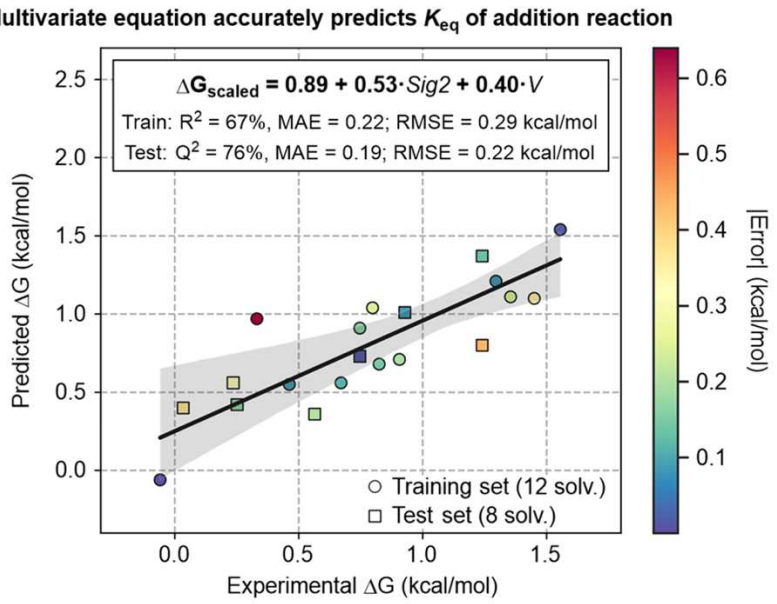



**Figure 6.6.** Final predictive multivariate linear regression (MLR) model in terms of several solvent descriptors resulting in a mean absolute error (MAE) of 0.22 kcal/mol for a training set of 12 solvents and an MAE of 0.19 kcal/mol for a hold-out test set of eight solvents.

The statistical significance of each descriptor to the model was tested by ANOVA. Sig2 and V descriptors were significant descriptors in the model with p-values of $2.20 \cdot 10^{-3}$ and $1.12 \cdot 10^{-2}$, respectively. The optimal regression model was validated by the calculated $R^2$ of the training set and the predictive squared correlation coefficient ($Q^2$) for the test set. These validation measurements were collected for each PCA cluster (train and test split) which resulted into the summarized values represented in Table 6.5. A satisfactory $Q^2$ value was set to greater than 0.6 indicated by the red line on Figure 6.5. The predictive value of the model was also supported by assessing different requirements: $R^2_{test} > 60\%$ (71%), $Q^2_{test} > 50\%$ (76%), 0.85 < slope k < 1.15 (0.91), and $(R^2-R^2_0)/R^2 < 0.1$ (-0.34).[14] Values can be replicated using the provided R-script.[4, 10]

*6.2.5 Predictive Model Interprets Solvent Effects within the Reaction*

We note that both Sig2 and V descriptors are defined computationally and may be generated for any given solvent. Based on the sign of the coefficients in this model, more favorable ΔG values of alcohol addition result from less polar, smaller solvent molecules.

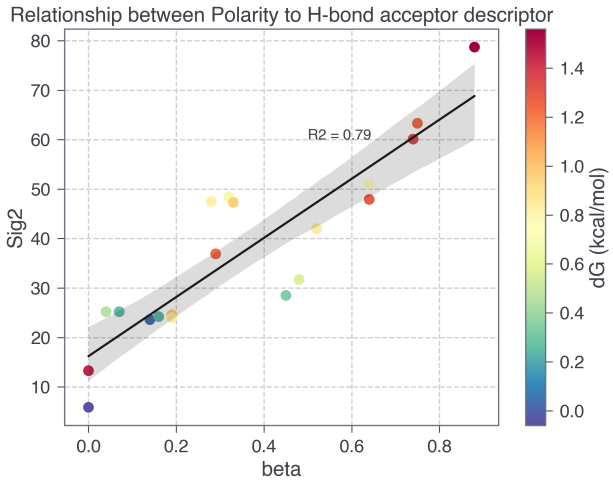



**Figure 6.7.** Correlation of solvent's polarity (Sig2) and H bond acceptor ability (beta) descriptors. The color bar represents the ΔG of equilibrium in the corresponding solvents.

We attribute the success of these solvent descriptors to their ability to describe the solvation of the alcohol pronucleophile. More polar solvents, identified by a larger Sig2 parameter (which is positively correlated with the H-bond acceptor β parameter shown in Figure 6.7), interact more strongly with the alcohol and addition becomes less favorable. While cyclohexane and dodecane both have β values of zero in Figure 6.7 with contrasting reaction outcomes (indicated by its color), the V descriptor differentiates between this reaction outcome in Figure 6.8.

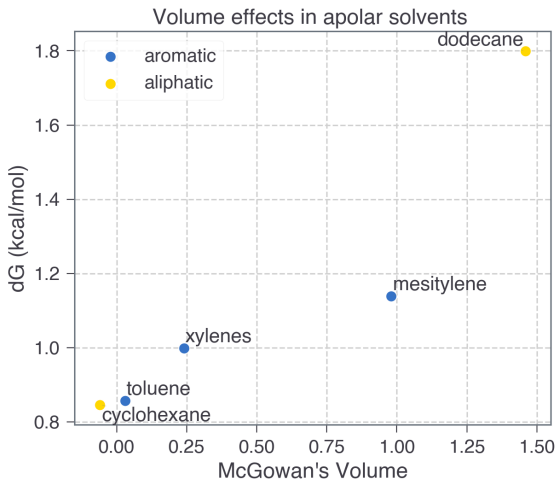



**Figure 6.8.** Relationship between ΔG and McGowan's volume V for apolar solvents considering aromatic and aliphatic solvents with similar electronic properties and different volumes.

For electronically similar solvents (e.g., cyclohexane and dodecane), the partition coefficients of n-BuOH:solvent mixtures further explains the molecular size influence (e.g., or McGowan's volume effects) such that smaller solvent molecules are able to destabilize the H-bonded networks of alcohol clusters (Figure 6.9).

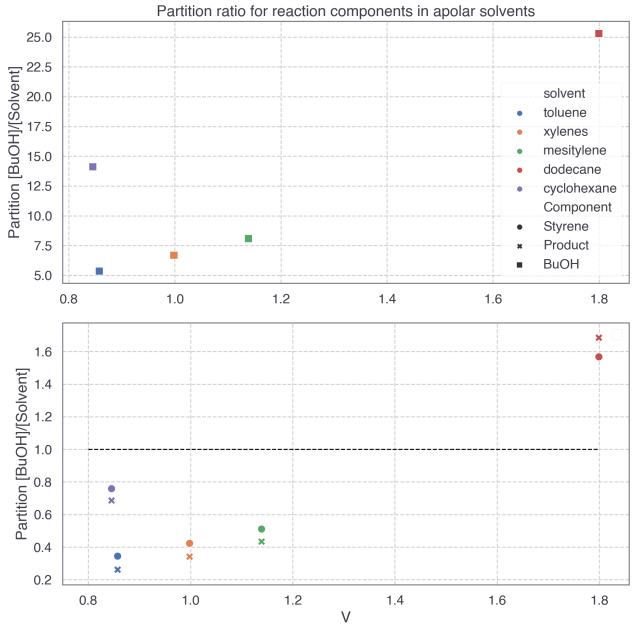
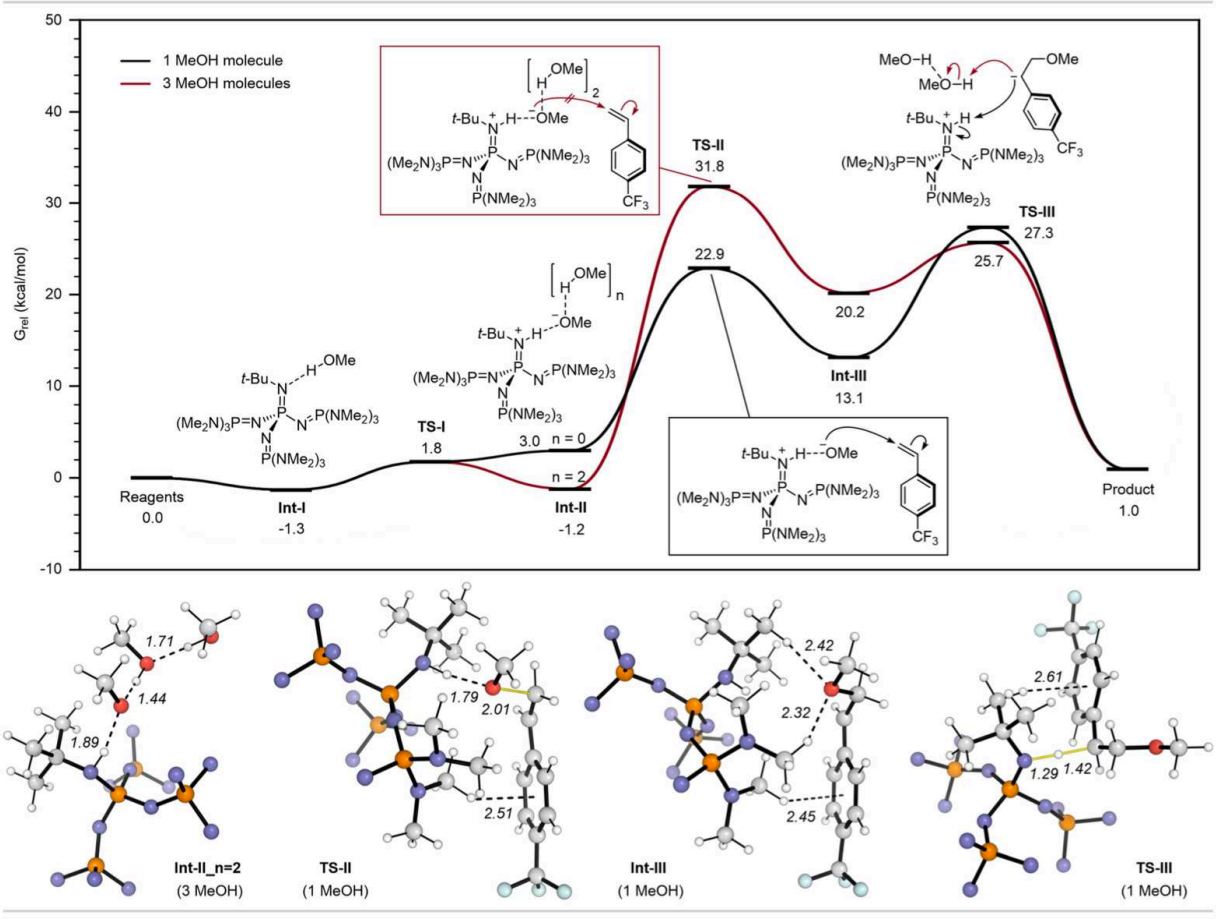



**Figure 6.9.** Representation of partition ratio of reagents and product in a simulated mixture of *n*-BuOH:"solvent" (calculated with *COSMO-RS* at 80 °C with the logP option) with their volume.

The lower the partition [*n*-BuOH]/["solvent"] the more preference of the reaction components to be in the simulated "solvent" phase. The black dotted line represents a partition value (P = 1) below which reaction components are more soluble in the organic phase *vs. n*-BuOH.

A higher partition value (P > 1) represents greater solubility in *n*-butanol while a lower partition value (P < 1) indicates greater solubility in the organic phase in Figure 6.9. Dodecane can be distinguished from other apolar solvents within this study, since all reaction components have low solubility in this solvent (i.e., all reaction components have a partition coefficient ratio > 1). Therefore, considering only weakly interacting apolar solvents (e.g., toluene, xylene, mesitylene, and dodecane), alcohol self-aggregation becomes disadvantageous for reactivity, and a smaller molecular size then dictates favorable ΔG values of alcohol addition with solvents like toluene and cyclohexane.




**Scheme 6.2.** Computed Gibbs energy ($G_{rel}$) profiles for the reaction in PhMe at 80 °C with the

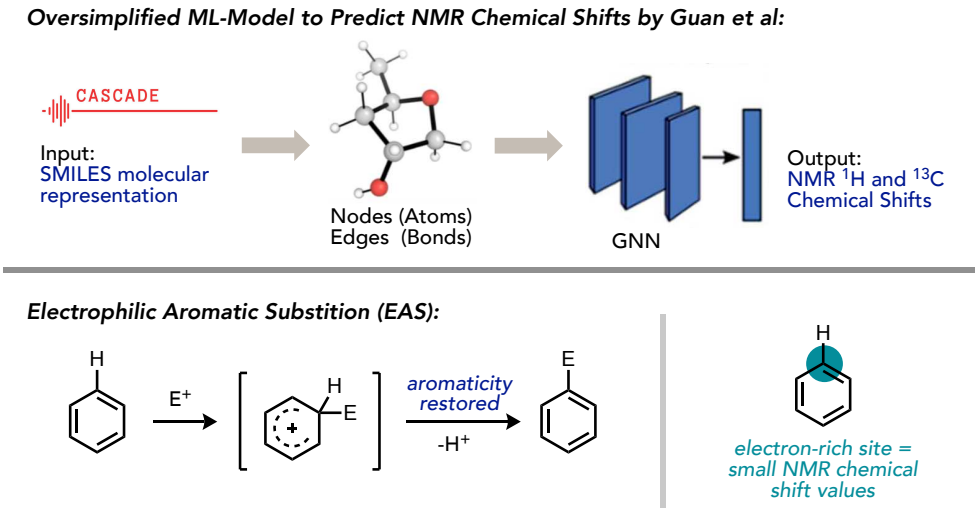
most stable conformers of key stationary points. Nonbonded and TS distances are shown in Å. Distal $CH_3$ groups are omitted for clarity.

These statistical findings provided key solvent features to consider for improving reactivity in the challenging nucleophilic alcohol addition reaction in apolar solvents. For example in toluene, these alcohol aggregates were exposed to inhibit alcohol addition by decreasing methoxide's nucleophilicity and increasing the barrier of the addition step (e.g., $P_4$-*t*-Bu·$(MeOH)_3$ at 31.8 kcal/mol) from DFT computations in Scheme 6.2 (e.g., computed by Dr. Juan V. Alegre-Requena). In agreement to the experimental observations, these computational and statistical findings explained the large negative alcohol rate orders indicative of the sequestered active alkoxide ion pairs from excess alcohol, slowing the reaction rate. Thereby, utilizing (i) less alcohol concentration, (ii) a less polar, smaller solvent molecules enables faster alcohol addition allowing for (iii) lower reaction temperatures for optimal reaction performance.

## *6.3 Graph-Neural Networks Predict NMR Chemical Shifts*

### *6.3.1 Nuclear Magnetic Resonance Chemical Shifts*

Nuclear magnetic resonance (NMR) is one of the primary techniques used to elucidate the chemical structure, bonding, stereochemistry, and conformation of organic compounds. The distinct chemical shifts in an NMR spectrum depend upon each atom's local chemical environment and are influenced by both through-bond and through-space interactions with other atoms and functional groups. The *in silico* prediction of NMR chemical shifts using quantum mechanical (QM) calculations is now commonplace in aiding organic structural assignment since spectra can be computed for several candidate structures and then compared with experimental values to find the best possible match. However, the computational demands of calculating multiple structural- and stereo-isomers, each of which may typically exist as an ensemble of rapidly-interconverting conformations, are expensive.




**Scheme 6.3. Top:** An oversimplified illustration of the GNN architecture developed by Guan *et al*. Molecules are represented according to their atom types and interatomic distances (or bonds). Node features for all $^{1}$H or $^{13}$C atoms pass through a series of dense layers to yield the final chemical shift predictions. **Bottom:** The general mechanism for the electrophilic aromatic substation reaction forming a Meisenheimer–Wheland intermediate complex. The theoretical approach for classifying the most electron-rich site with NMR chemical shift values.

In this work, Dr. Yanfei Guan and Dr. Peter C. St. John developed a rapid machine learning (ML) protocol (Scheme 6.3, top) to predict $^{1}$H and $^{13}$C chemical shifts through an efficient graph neural network (GNN) using 3D structures as input which is available as an automated prediction webserver and graphical interface online at http://nova.chem.colostate.edu/cascade/.[2, 15] To highlight the prediction power of this tool, we used the predicted NMR chemical shifts as descriptors for determination of the sites of electrophilic aromatic substitution (Scheme 6.3, bottom). The regioselectivity predictions were carried out by S. V. Shree Sowndarya and the present author.

### *6.3.2 QM-derived NMR Chemical Shifts*

Computed NMR chemical shifts are derived from computed NMR isotropic shielding constants (σ) for the various nuclei in the molecule using with gauge-independent atomic orbital (GIAO) calculations. Empirical correction schemes for density functional theory (DFT) computed shielding tensors have been instrumental in improving the levels of accuracy: Tantillo and co-

workers[16] derived and compiled linear-scaling parameters for many levels of theory, basis set and solvation models (in the CHESHIRE repository[17]), and have established standardized molecular training and test sets for chemical shift prediction. As a result, contemporary "best practice" DFT protocols boosted by empirical corrections routinely approach accuracies of 2.5 ppm in the prediction of $^{13}C$ shifts, or 0.15 ppm for $^{1}H$ shifts, expressed as root mean square error (RMSD).[16]

Therefore, we applied these recommended methods and calculated isotropic shielding constants with GIAO[18] calculations at the mPW1PW91/6-311+G(d,p)[19] level of theory for 7,455 optimized structures at the M06-2X/def2-TZVP[20] level of theory arriving at 117,997 $^{1}H$ and 9,9105 $^{13}C$ calculated chemical shifts. All molecules contained (i) at least one carbon atom, (ii) a molecular weight less than 500, (iii) C, H, O, N, P, S, F, and Cl atoms, and (iv) a neutral charge. In an automated workflow developed by Dr. Guan, each molecule underwent conformer sampling (e.g., lowest lying conformer from MMFF94s force field[21] minimization selected), DFT optimization (e.g., M06-2X/def2-TZVP level of theory[20] with Gaussian16[22]), structure checking, and a GIAO DFT calculation.

The computed isotropic shielding constants were parsed and collectedusing an automated workflow[23] and converted to chemical shift (δ) values by subtracting the isotropic shielding constants for the nuclei in the molecule of interest ($\sigma_i$) from a shielding constant of a reference sample ($\sigma_{ref}$) such as tetramethyl-silane (TMS). However, an alternative and generally more accurate approach is to obtain chemical shifts $\delta$ by applying empirical linear-scaling relationships[16] obtained from correlation between DFT and experiment. Herein, we applied this linear-scaling relationship to check for possible incorrect assignments of experimental chemical shift based on the presence of statistically significant outliers.

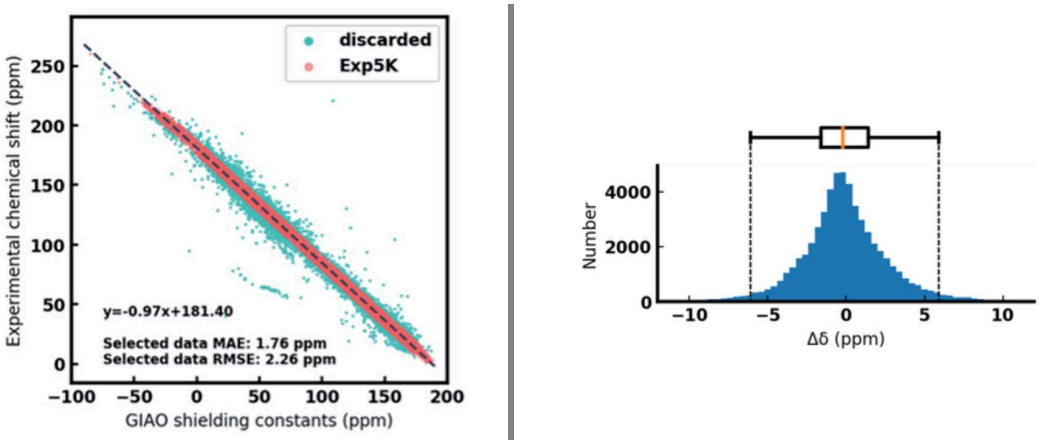



**Figure 6.10. Left:** Correlation between GIAO DFT shielding constants and experimental chemical shifts. **Right:** Histogram shows the error distribution for the correlation. The boxplot shows the lower quartile (Q1) and upper quartile (Q3). Whiskers show the minimum (Q1 – 1.5*IQR) and maximum (Q3 + 1.5*IQR), which corresponds to -6.1 and 5.9 ppm. Data outside the dash line are considered as outliers due to incorrect assignment of experimental chemical shift, which were discarded (colored in green).

To avoid erroneous structures and assignments, the data in Figure 6.10 was cleaned for statistical outliers defined in terms of the interquartile range (IQR) by discarding data with DFT vs. experimental error below (Q1 – 1.5*IQR) or above (Q3 + 1.5*IQR). This corresponded to -6.1 and 5.9 ppm, respectively. As a result, about 8% of data was discard, leaving 5148 structures and 53,334 $^{13}$C chemical shifts. Discarded datapoints are accessible from GitHub (https://github.com/patonlab/CASCADE).[15] With this finalized dataset, we then tested its utility in predicting sites of electrophilic aromatic substitutions.

### *6.3.3 Predicting Sites of Electrophilic Aromatic Substitution*

Electrophilic aromatic substitutions (EAS) reactions introduce electrophile substituents into aromatic rings through a Meisenheimer–Wheland complex followed by deprotonation that restores aromaticity to afford the final substitution product (Scheme 6.3, bottom).[24] In theory, the most electron-rich aromatic carbon would be the most reactive which is related to the lowest (e.g., upfield or higher electron density shielded around the atoms) $^{13}$C and $^{1}$H chemical shift (δ). Previously, the combination of DFT-computed atomic Fukui coefficients, atomic partial charges,

bond orders, and partitioned solvent-accessible surface areas with semi-empirical regioSQM[25] predictions were used to develop a random forest (RF) model with 93% accuracy in predicting the site of substitution using 80/20 train/test splits for 376 molecules.[26] Using three atomic descriptors: (i) the $^{13}C$ chemical shift, (ii) the attached proton $^{1}H$ chemical shift, and (iii) the regioSQM prediction, comparable accuracy is achieved in place of more expensive DFT (e.g., mPW1PW91/6-311+G(d,p)//M062X/def2TZVP) values (Figure 6.11D).

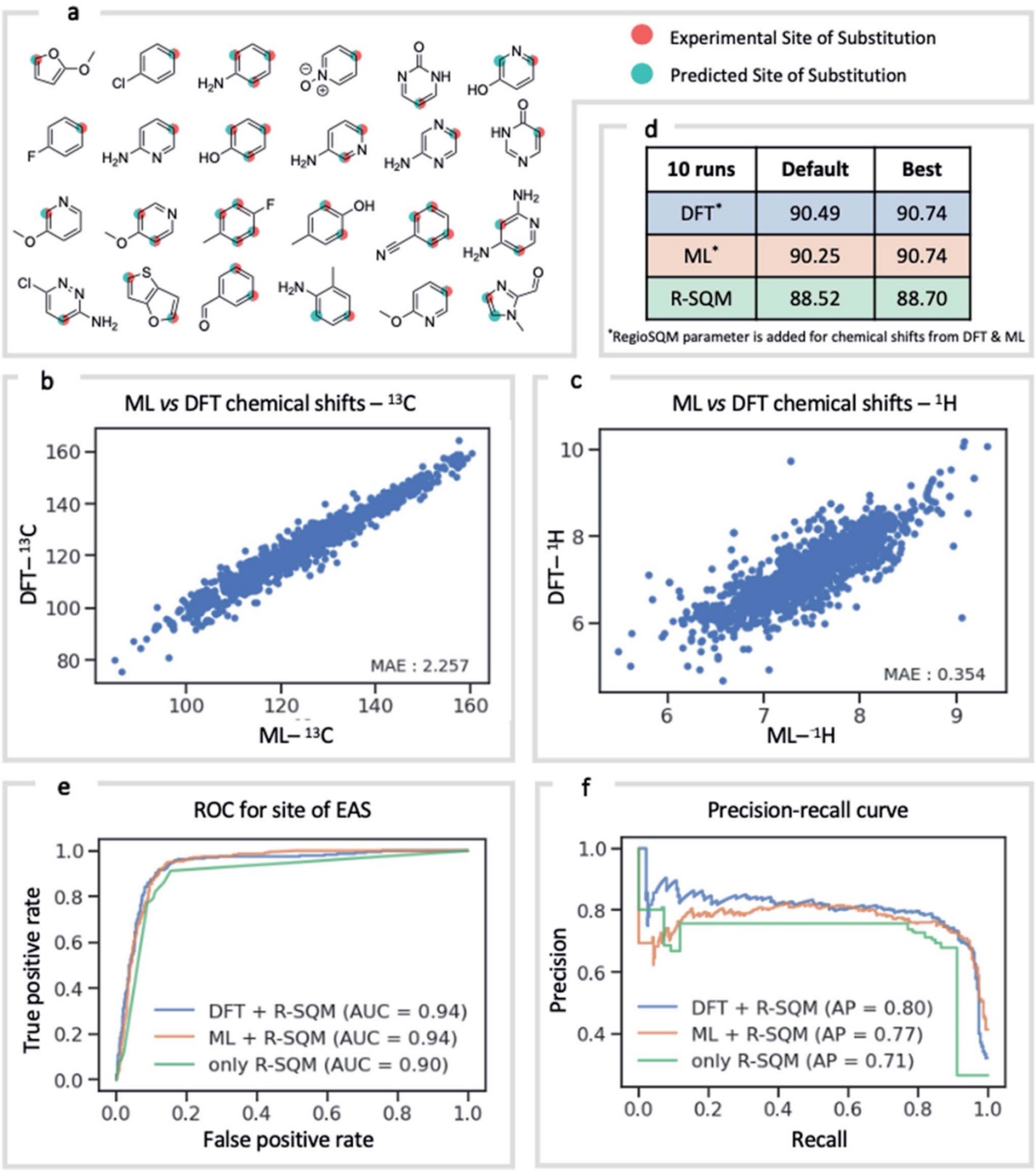



**Figure 6.11.** Regioselectivity prediction of electrophilic aromatic substitutions. (a) Representative molecules present in the EAS dataset. The high- lighted atoms depict the experimental (red) and the predicted (green) site of substitution. (b) DFT computed $^{13}C$ chemical shifts vs. GNN-predictions. (c) DFT computed 1H chemical shifts vs. GNN-predictions. (d) Random forest classifier accuracies in identifying reactive/unreactive ring positions. (e) ROC curves comparing the true positive vs. false positive rate. (f) Precision–recall curves for the different random forest classifiers.

Obtaining the predicted ML chemical shifts and optimization of GNN model hyperparameters was performed by S. V. Shree Sowndarya which resulted in accuracy values of 90.7% with the inclusion of chemical shift descriptors from 88.5% using regioSQM alone (Figure 6.11). ROC and precision–recall plots (Figure6.11E and F) illustrate that the inclusion of chemical shift descriptors increase the performance of an RF classification (i.e., correctly labelling reactive and unreactive positions) from 0.90 to 0.94 and that the average precision is also higher with chemical shift descriptors. NMR chemical shift is influenced by the electron density around a nucleus of interest. It is therefore an attractive choice of physically motivated and interpretable atomic descriptor for use in predictive machine learning models.[27] Although this model depends on the quality of 3D structures generated by MMFF, these efforts resulted in a predictive model of comparable accuracy to DFT against experimental chemical shifts of small organic molecules.

### *6.4 Conclusion*

In this chapter, we focused on the data-driven approach to formulate statistical relationships sampled over small and large datasets to perform a multivariate linear regression (e.g., by the present author) and a graph neural network model (e.g., by Dr. Yanfei Guan). In each of these research projects, a successful predictive model was generated. In the first work discussed in section 6.2, we developed an interpretable statistical model to identify and explain the observed reaction outcome in various solvents. We found that the selected solvent features (e.g., molecular solvent size and polarizability) were critical in identifying the ideal reaction conditions (e.g., decreasing the alcohol equivalents and thereby decreasing the reaction temperature). The influence of these solvent features in the reaction performance from the statistical analysis was confirmed via DFT-studies performed by Dr. Juan V. Alegre-Requena. In the work discussed in section 6.3, we applied the ML-derived $^{13}C$ and $^{1}H$ NMR chemical shifts along with the regioSQM feature to

formulate a GNN model to predict sites for electrophilic aromatic substitutions on various aromatics. The final model achieved comparable accuracy to expensive methods such as DFT. Chemical intuition explains that the most reactive carbon in electrophilic aromatic substitutions is related to electron-rich aromatic carbons, thereby resulting in the lowest chemical shift value. By already having a chemical understanding of this reaction, the research goal was focused on building an accurate model using efficient data-driven methods.

## REFERENCES


1. Luo, C.; Alegre-Requena, J. V.; Sujansky, S. J.; Pajk, S. P.; Gallegos, L. C.; Paton, R. S.; Bandar, J. S., Mechanistic Studies Yield Improved Protocols for Base-Catalyzed Anti-Markovnikov Alcohol Addition Reactions. *J. Am. Chem. Soc.* **2022,** *144* (22), 9586-9596.
2. Guan, Y.; Shree Sowndarya, S. V.; Gallegos, L. C.; St. John, P. C.; Paton, R. S., Real-time prediction of 1H and 13C chemical shifts with DFT accuracy using a 3D graph neural network. *Chem. Sci.* **2021,** *12* (36), 12012-12026.
3. Luo, C.; Bandar, J. S., Superbase-Catalyzed anti-Markovnikov Alcohol Addition Reactions to Aryl Alkenes. *J. Am. Chem. Soc.* **2018,** *140* (10), 3547-3550.
4. Gallegos, L. C. Solvent Regression Analysis. https://github.com/Liliana-Gallegos/Solvent_regression_analysis.
5. Benazzouz, A.; Moity, L.; Pierlot, C.; Sergent, M.; Molinier, V.; Aubry, J.-M., Selection of a Greener Set of Solvents Evenly Spread in the Hansen Space by Space-Filling Design. *Ind. Eng. Chem.* **2013,** *52* (47), 16585-16597.
6. (a) Eckert, F.; Klamt, A., Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE Journal* **2002,** *48* (2), 369-385; (b) Klamt, A., Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995,** *99* (7), 2224-2235; (c) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W., Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998,** *102* (26), 5074-5085; (d) COSMOtherm Version 18.0.0 (Revision 4360), COSMOlogic GmbH & Co KG.
7. Zissimos, A. M.; Abraham, M. H.; Klamt, A.; Eckert, F.; Wood, J., A Comparison between the Two General Sets of Linear Free Energy Descriptors of Abraham and Klamt. *J. Chem. Inf. Comput.* **2002,** *42* (6), 1320-1331.
8. Borhani, T. N.; García-Muñoz, S.; Vanesa Luciani, C.; Galindo, A.; Adjiman, C. S., Hybrid QSPR models for the prediction of the free energy of solvation of organic solute/solvent pairs. *Phys Chem Chem Phys* **2019,** *21* (25), 13706-13720.
9. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/.
10. Gallegos, L. C. REGGAE. https://github.com/Liliana-Gallegos/REGGAE.
11. Henle, J. J.; Zahrt, A. F.; Rose, B. T.; Darrow, W. T.; Wang, Y.; Denmark, S. E., Development of a Computer-Guided Workflow for Catalyst Optimization. Descriptor Validation, Subset Selection, and Training Set Analysis. *J. Am. Chem. Soc.* **2020,** *142* (26), 11578-11592.
12. Klamt, A.; Eckert, F.; Diedenhofen, M., Prediction of soil sorption coefficients with a conductor-like screening model for real solvents. *Environ. Toxicol. Chem.* **2002,** *21* (12), 2562-2566.
13. McGowan, J. C., Molecular volumes and structural chemistry. *Recl. Trav. Chim. Pays-Bas* **1956,** *75* (2), 193-208.
14. (a) Golbraikh, A.; Tropsha, A., Beware of q2! *J. Mol. Graph.* **2002,** *20* (4), 269-276; (b) Tropsha, A.; Gramatica, P.; Gombar, V. K., The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* **2003,** *22* (1), 69-77.
15. Guan, Y.; Paton, R. S. *CASCADE: ChemicAl Shift CAlculation with DEep learning.* *https://github.com/patonlab/CASCADE*.

16. Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J., Computational Prediction of 1H and 13C Chemical Shifts: A Useful Tool for Natural Product, Mechanistic, and Synthetic Organic Chemistry. *Chem. Rev.* **2012,** *112* (3), 1839-1862.
17. M. W. Lodewyk, M. R. S., D. J. Tantillo, P. R. Rablenm and T. B. Cheshire CHEmical SHift REpository with Coupling Constants Added Too. http://cheshirenmr.info.
18. Wolinski, K.; Hinton, J. F.; Pulay, P., Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. *J. Am. Chem. Soc.* **1990,** *112* (23), 8251-8260.
19. Adamo, C.; Barone, V., Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J. Chem. Phys.* **1998,** *108* (2), 664-675.
20. 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 function. *Theoretical Chemistry Accounts* **2008,** *120* (1-3), 215-241.
21. Halgren, T. A., Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* **1996,** *17* (5-6), 490-519.
22. 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; 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 Rev. B.01*, Wallingford, CT, 2016.
23. Gallegos, L. C. DISCO. https://github.com/Liliana-Gallegos/DISCO.
24. Smith, M. B. M., J., *March's Advanced Organic Chemistry:Reactions, Mechanisms, and Structure, 6th ed.* John Wiley & Sons Inc.: Hoboken, NJ, 2007.
25. Kromann, J. C.; Jensen, J. H.; Kruszyk, M.; Jessing, M.; Jørgensen, M., Fast and accurate prediction of the regioselectivity of electrophilic aromatic substitution reactions. *Chem. Sci.* **2018,** *9* (3), 660-665.
26. Tomberg, A.; Johansson, M. J.; Norrby, P.-O., A Predictive Tool for Electrophilic Aromatic Substitutions Using Machine Learning. *J. Org. Chem.* **2019,** *84* (8), 4695-4703.
27. (a) Gordon, C. P.; Raynaud, C.; Andersen, R. A.; Copéret, C.; Eisenstein, O., Carbon-13 NMR Chemical Shift: A Descriptor for Electronic Structure and Reactivity of Organometallic Compounds. *Acc. Chem. Res.* **2019,** *52* (8), 2278-2289; (b) Verma, R. P.; Hansch, C., Use of 13C NMR Chemical Shift as QSAR/QSPR Descriptor. *Chem. Rev.* **2011,** *111* (4), 2865-2899.

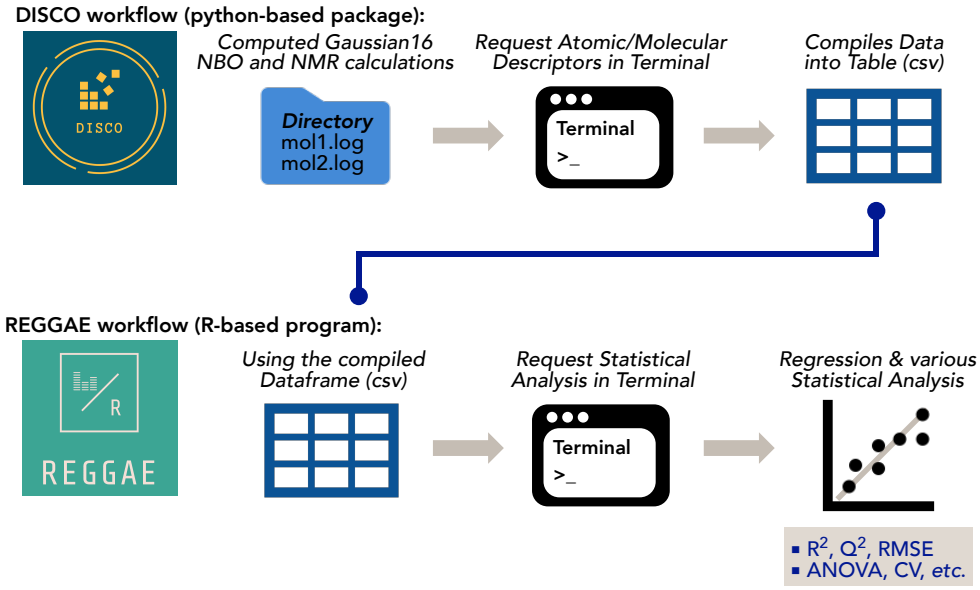
# 7 AUTOMATED WORKFLOWS FOR THE DATA-CHEMIST IN LOW- AND HIGH-THROUGHPUT STUDIES

## *7.1 Chapter Overview*

It can be challenging to reproduce literature studies due to incomplete recording of algorithms and code use, which also makes entering the data-science field non-trivial. In this section, we combine different stages of the research process (e.g., chemical data collection, feature selection and then statistical modeling) in a way that is accessible to non-experts. The following workflows allowed the execution of projects mentioned in Chapters 4 to 6 to be processed and refined in a quick and efficient manner from simple executable commands on the terminal. Both automated workflows, DISCO[1] and REGGAE[2], discussed in this chapter have been released in open-source platforms for use and will be submitted to *Zenodo* in the future.

## *7.2 Introduction*

The intersection of chemistry and statistical modeling along with other high-throughput data-driven approaches (e.g., machine learning methods) have demonstrated the potential in reactivity and selectivity predictions[3], synthesis planning[4], and mechanistic studies[5] in organic chemistry. Several workflows (or tools) have emerged to assist in the automation process in computational chemistry and the data-chemist field since developing the following automated workflows.[6] Herein, we discuss two automated workflows (e.g., DISCO[1]and REGGAE[2]) summarized in Scheme 7.1 in which were implemented in the mechanistic and statistical studies mentioned in Chapters 4-6 to (a) obtain a range of molecular and atomistic properties from DFT computations and (b) develop a statistical relationship and analysis for reaction outcome prediction. These computer-assisted tools were developed to address some of the current challenges associated with (i) accessibility to non-experts in chemistry and statistics, (ii) reduce repetitive tasks in data collection, and (iii) thus, increase the reproducibility and transparency within literature studies.




**Scheme 7.1.** General overview of the two workflows: DISCO (top) and REGGAE (bottom) developed to assist in collecting and performing statistical analysis in low- and high-throughput studies.

Quantum mechanical (QM) features of molecules derived from density functional theory (DFT) calculations achieve a higher accurate molecular representation of its physical and chemical properties (previously discussed in Chapter 1). Therefore, these features serve as ideal candidates within low-throughput studies for building statistical relationships explainable by mechanistic studies. In high-throughput studies, QM features are implemented less in machine learning models due to its relatively large computational cost in comparison to features derived from fast molecular mechanics techniques (e.g., with RDKit, MORFEUS)[6c, 7], semi-empirical methods (e.g., GFNn-xTB)[8], and machine learned models (e.g., CASCADE, ANI-1)[9]. Nonetheless, quantum mechanical featurization of molecules have assisted in successfully building large and small scale, diverse databases for chemical space exploration and predictive modeling.

Building statistical relationships, typically for quantitative structure–activity relationships (QSAR) or quantitative structure–property relationships (QSPR), are focused on drug-lead

identification which generally have been based on multivariate linear regression (MVLR) with limited features.[10] The higher rate for implementation of linear models (e.g., unlike most machine learned models) in small- and large-scale datasets is due to its simplicity to generate models and its interpretability factor. In a linear regression model, the coefficients are directly interpretable as each coefficient describes the effect on the output (i.e., response or y-variable). Utilizing the R-programming language originally developed for statistical computing and graphics, this open-source program provided the basis for developing a streamlined tool for efficient data analysis and reproducible model development for the data-chemist.[11] Other software packages currently in use among chemists include MatLab, Excel, JMP, python, *etc.* As an added factor in our developed statistical workflow (i.e., REGGAE), a researcher with no in-depth knowledge of statistics could develop validated and predictive QSPR models by comparing the resulting analysis to the recommended criteria developed by Tropsha *et al* and Dearden *et al* that is incorporated in this workflow.[12]

Both of these developed open-source programs have detailed workflow examples available on GitHub and the CSU Theory Suite's wiki page.[1-2, 13] All python- and R-based programming code development was performed by the present author.

### *7.3 DISCO*

The python program DISCO - **Di**stributing **C**omputed **O**utputs, was developed to facilitate the extraction of features from Gaussian calculation files and tabulated for statistical modeling.

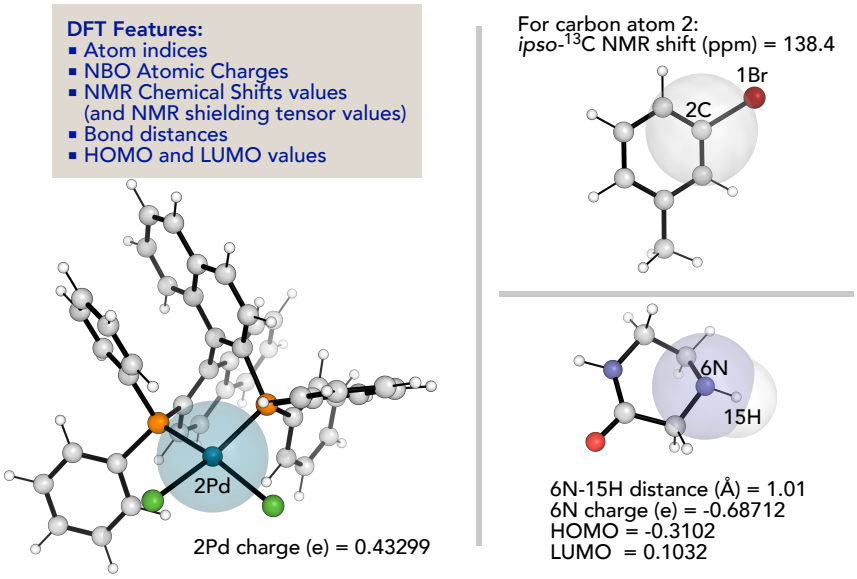



**Figure 7.1.** DFT features collected for the following molecules (i) $PdCl_2$(BINAP) (left), (ii) 1-bromo-3-methylbenzene (top), and (iii) piperazinone (bottom right) from NBO and NMR calculations.

**A. Terminal Window**

```
[~ » python DISCO.py piperazinone_NBO.log --distance '6,15' --mo both --charge 6 --verbose --csv data

o Running: piperazinone_NBO.log
  Atomic Charge: ['6N'] = [-0.68712]
  HOMO and LUMO = [-0.3102] and [0.1032]
  Bond Distance: [['6N', '15H']] = [1.01189]
o Disco DFT features:
                  Name 6 atom index  6 charge    HOMO    LUMO  6N-15H length
0  piperazinone_NBO.log           6N  -0.68712 -0.3102  0.1032        1.01189
(disco_env) ------------------------------------------------------------------------------------------
```

**B. Jupyter Notebook**

```python
import DISCO as cd

#Create DISCO object
mol = cd.disco(file, distance='atom1,atom2', charge=atom, \
               nmr=atom, scale='intercept,slope', \
               mo=True, verbose=True)

#Grab DFT Parameters
charge = mol.Chrg
nmr    = mol.NMR
distance = mol.Dist
atom_index = mol.Atom
homo = mol.HOMO
lumo = mol.LUMO

```

**Figure 7.2.** DFT features collected and stored as “data.csv” for the piperazinone molecule shown

in Figure 7.1 using (A) the Terminal window. (B) Code block cell for grabbing all the listed DFT features using Jupyter Notebook.

The program currently collects features (Figure 7.1) such as NMR chemical shifts (and NMR shielding tensor values), HOMO and LUMO molecular orbital values, atomic charges, and bond distances *via* the Terminal window or a web-based interactive computing platform for running code blocks like Jupyter Notebook as shown in Figure 7.2. The program can handle a large number of computed compounds simultaneously and extract the requested atomistic and molecular properties to compile and store in a CSV file. The collected data can then undergo statistical evaluation. Future feature additions include the calculation of molecular properties over conformer averaged geometric descriptors.

### *7.4 REGGAE*

The regression models were generated using the developed R language program REGGAE - **REG**ression **G**enerator and **A**nalyz**E**r, to facilitate data analysis in a systematic way. The program produces a regression model along with statistical diagnostics and cross-validation tests. The diagnostics include ANOVA analysis, pairwise correlations, outlier, and collinearity testing. The script also produces QSAR criteria for an acceptable model such as $q^2$, leave-one-out cross-validation, and external validation $R^2$ values, amongst others. The program provides a complete model analysis within seconds.

#### *7.4.1 MVLR Model on Solvent Dataset*

When developing a regression model, feature selection can be performed using (i) all of the features (e.g, with "full" option), (ii) Forward-stepwise, or (ii) manual feature selection. In the example shown in Figure 7.3 and Figure 7.4 (also highlighted in section 6.2), we utilized REGGAE to perform a multivariate linear regression along with a cross-validation analysis using the selected

terms (e.g., Sig2 and V features with the -b option) in a predefined split selection to find a relationship to the experimental Gibbs-free energy at equilibrium (e.g., using the -y Exp_dG option). The split was defined as “Train” and “Test” (e.g., with option, -r set to 0) within the column labeled “12kclusters” in the “20solvents.csv” dataset.

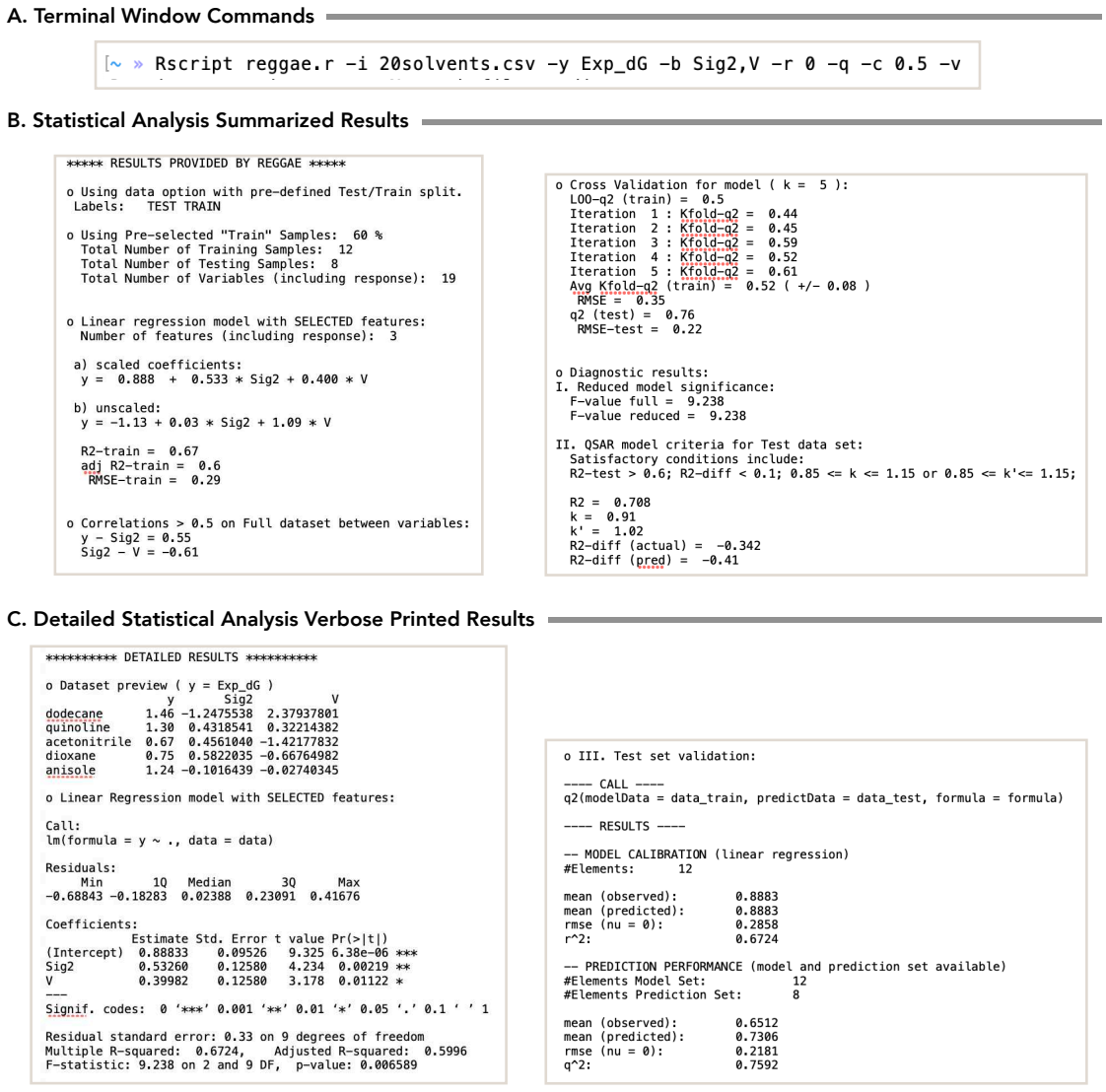



**Figure 7.3.** REGGAE workflow for the solvent analysis discussed in section 6.2. By requesting the (A) commands in a Terminal prompt window, (B) a summarized statistical analysis report and (C) a detailed report is printed out using the verbose option.

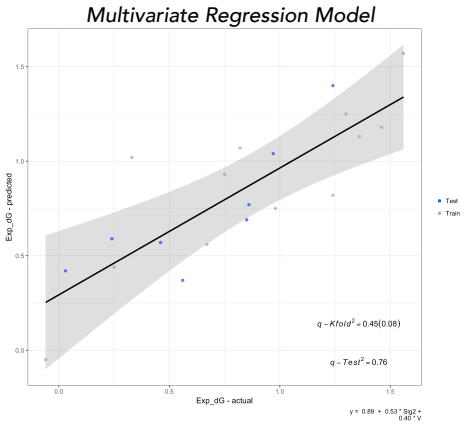
The first portion of the terminal output (e.g., without using the verbose (-v) option) includes a summarized statistical analysis report (Figure 7.3B): the type of split performed, regression formula in both scaled and nonscaled versions, $R^2$ and RMSE values, and Pearson pairwise correlations between features > 0.5 (e.g., using the -c 0.5 option). Additionally, three cross validation analysis: Leave-one-out, K-fold (e.g., set at k=5), and $Q^2$, along with QSAR analysis for the Test dataset were requested using the -q and -d options. Including the verbose option (-v) with the diagnostics (-d) option, the following diagnostic plots: Residual vs Fitted, Normal Q-Q, Cook's distance, and Residuals vs Leverage plots are also saved in the working folder.

For an in-depth statistical analysis report, the verbose (-v) option prints out the following in the Terminal output and in a saved text file (Figure 7.3C and Figure 7.4): preview of the scaled values, ANOVA analysis, a table of predicted y-values, and a matrix of the pairwise correlation values. Additionally, the regression and correlation heatmap plots get saved in working folder as a PNG files as shown in Figure 7.4. This full workflow with additional plots and analyses can also be found on GitHub and the CSU Theory Suite's wiki page.[1-2, 13]

Generated Regression & Pearson Correlation Plots for MVLR Model

```
o Regression plot(s) saved in working file.

o Predicted y-values:
                Y Pred_Y Split
cyclohexane -0.06  -0.05 Train
toluene      0.03   0.42  Test
xylenes      0.24   0.59  Test
chlorobenzene 0.25  0.44 Train
dibutyl ether 0.33  1.02 Train
DCB          0.46   0.57  Test
THF          0.56   0.37  Test
acetonitrile 0.67   0.56 Train
dioxane      0.75   0.93 Train
nitrobenzene 0.82   1.07 Train
pyridine     0.85   0.69  Test
tetralin     0.86   0.77  Test
benzonitrile 0.97   1.04  Test
mesitylene   0.98   0.75 Train
anisole      1.24   0.82 Train
NMP          1.24   1.40  Test
quinoline    1.30   1.25 Train
DMF          1.36   1.13 Train
dodecane     1.46   1.18 Train
DMSO         1.56   1.57 Train
```




```
o Correlation table for all variables:
          y  Sig2     V
y    1.00  0.55  0.14
Sig2 0.55  1.00 -0.61
V    0.14 -0.61  1.00

o Correlation plot saved in working file.
**********END**********
```

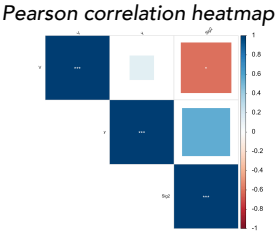



**Figure 7.4.** Additional detailed statistical analysis as a result of using the verbose option within the REGGAE workflow for the solvent analysis discussed in section 6.2. **Top:** The multivariate linear regression plot and the corresponding predicted values labeled as "Train" or "Test". **Bottom:** The Pearson correlation matrix and heatmap plot for the terms used in the model.

### *7.5 Conclusion*

In this chapter, we introduced two open-source automated workflows to streamline research processes. These tools have been applied within the works mentioned in Chapters 4-6 to ensure that computational research and statistical analysis is fully reproducible and transparent. These tools aim to address the current challenges associated with accessibility to non-expert chemists and increasing the reproducibility and transparency of literature studies.

## REFERENCES


1. Gallegos, L. C. DISCO. https://github.com/Liliana-Gallegos/DISCO.
2. Gallegos, L. C. REGGAE. https://github.com/Liliana-Gallegos/REGGAE.
3. (a) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E., Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. *Science* **2019,** *363* (6424), eaau5631; (b) 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* (6385), 186-190; (c) Guan, Y.; Coley, C. W.; Wu, H.; Ranasinghe, D.; Heid, E.; Struble, T. J.; Pattanaik, L.; Green, W. H.; Jensen, K. F., Regio-selectivity prediction with a machine-learned reaction representation and on-the-fly quantum mechanical descriptors. *Chem. Sci.* **2021,** *12* (6), 2198-2208.
4. Coley, C. W.; Rogers, L.; Green, W. H.; Jensen, K. F., Computer-Assisted Retrosynthesis Based on Molecular Similarity. *ACS Cent. Sc* **2017,** *3* (12), 1237-1245.
5. Zhao, S.; Gensch, T.; Murray, B.; Niemeyer, Z. L.; Sigman, M. S.; Biscoe, M. R., Enantiodivergent Pd-catalyzed C–C bond formation enabled through ligand parameterization. *Science* **2018,** *362* (6415), 670-674.
6. (a) Żurański, A. M.; Wang, J. Y.; Shields, B. J.; Doyle, A. G., Auto-QChem: an automated workflow for the generation and storage of DFT calculations for organic molecules. *Reaction Chemistry & Engineering* **2022,** *7* (6), 1276-1284; (b) Alegre-Requena, J. V.; Sowndarya S. V., S.; Pérez-Soto, R.; Alturaifi, T. M.; Paton, R. S., AQME: Automated quantum mechanical environments for researchers and educators. *WIREs Computational Molecular Science* **2023**; (c) Morfeus. https://github.com/kjelljorner/morfeus.
7. RDKit: open-source. cheminformatics.http://www.rdkit.org.
8. (a) Bannwarth, C.; Ehlert, S.; Grimme, S., GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput* **2019,** *15* (3), 1652-1671; (b) Grimme, S.; Bannwarth, C.; Shushkov, P., A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86). *J. Chem. Theory Comput* **2017,** *13* (5), 1989-2009.
9. (a) Guan, Y.; Paton, R. S. *CASCADE: ChemicAl Shift CAlculation with DEep learning. https://github.com/patonlab/CASCADE*; (b) Guan, Y.; Shree Sowndarya, S. V.; Gallegos, L. C.; St. John, P. C.; Paton, R. S., Real-time prediction of 1H and 13C chemical shifts with DFT accuracy using a 3D graph neural network. *Chem. Sci.* **2021,** *12* (36), 12012-12026; (c) Smith, J. S.; Isayev, O.; Roitberg, A. E., ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017,** *8* (4), 3192-3203.
10. (a) Hammett, L. P., Reaction Rates and Indicator Acidities. *Chem. Rev.* **1935,** *16* (1), 67-79; (b) Martin, Y. C., Hansch analysis 50 years on. *WIREs Computational Molecular Science* **2012,** *2* (3), 435-442.
11. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/.
12. (a) Golbraikh, A.; Tropsha, A., Beware of q2! *J. Mol. Graph.* **2002,** *20* (4), 269-276; (b) Tropsha, A.; Gramatica, P.; Gombar, V. K., The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* **2003,** *22* (1), 69-77; (c) Dearden, J. C.; Cronin, M. T. D.; Kaiser, K. L. E., How not to develop

a quantitative structure–activity or structure–property relationship (QSAR/QSPR). *SAR QSAR Environ Res* **2009,** *20* (3-4), 241-266.
13. CSU Theory Suite Wiki Page. https://csu-theory-suite.github.io/theory-suite-wiki/.

# 8 CONCLUDING REMARKS

Computational chemistry has provided the tools to rationalize experimental observations and elucidate reaction mechanisms. At the same time, statistical modeling techniques are essential in developing meaningful and explaining relationships that would have otherwise been overlooked. We introduce various research works where quantum mechanics and statistical models have made synthetic chemistry more predictable. Additionally, open-source automated workflows facilitate the reproducibility and transparency of the research studies mentioned throughout the chapters.

The DFT methods applied to research works discussed in Chapters 2-5 propose reaction mechanisms and explain the experimentally observed reactivity. In Chapter 2, we begin by exploring the challenging chemical reactivity found in large biological systems (e.g., peptides and proteins) discovered by the Payne lab called the photocatalytic diselenide contraction (PDC) to interrogate various probable reaction pathways using QM methods for the photocatalytic reactivity from empirical observations (e.g., PL spectroscopy) to substantiate the proposed mechanism. In Chapters 3 and 4, we embark on QM studies on newly discovered bismuth chemistry developed by the Ball group. More importantly, the initial grid integration benchmarking studies performed on the Ball *et al* 2020 published work in Nature set the stage for the computational protocols required in bismuth systems for determining reaction selectivity. These grid effects pioneered by Wheeler *et al* work was essential for identifying the computationally observed error and thereby applying the required protocol (e.g., increased integration grid). We successfully published this collaborative research on the convenient, telescoped procedure for the Bi(V)-mediated O-arylation of 2- and 4- pyridones with arylboronic acids. Continuing the bismuth work in Chapter 4 with the Ball group, we introduced new computational results on the collaborative work for the developed reaction method to form challenging C(sp2)-C(sp3) bonds via a cross-coupling reaction with the

amino-bridged bismacycle reagent. In this chapter, we begin applying statistical methods of various substrates' electronic and steric molecular properties to build an interpretable, multivariate linear regression model to capture NCIs and explain the origin of reactivity facilitated with automated workflows discussed in Chapter 7.

The work discussed in Chapter 5 represents an "in the wild" application of Bayesian algorithms for yield optimization in the challenging cascade key reaction step of a natural product synthetic plan. We applied a data-driven approach towards building a diverse dataset composed of molecular features derived from complete computational understanding using quantum mechanical methods of the reaction surface. Chapter 6 focused on the data-driven approach to formulate statistical relationships sampled over small and large datasets to perform predictive models. In each of these research projects, a successful predictive model was generated. In the first work discussed in section 6.2, we developed an interpretable statistical model to identify and explain the observed reaction outcome in various solvents. In work discussed in section 6.3, we applied the ML-derived 13C and 1H NMR chemical shifts along with the regioSQM feature to formulate a GNN model to predict sites for electrophilic aromatic substitutions on various aromatics.

Finally, in Chapter 7, we introduced two open-source automated workflows, DISCO and REGGAE, to streamline research processes at various stages in reaction optimization projects. These tools ensure that computational research and statistical analysis are fully reproducible and transparent. Additionally, these tools aim to assist non-expert chemists in these data-chemistry techniques to make synthetic chemistry more predictable.

## APPENDIX

### *A.1 General Computational Methods and Protocols*

**Functionals and Solvents.** Density functional theory (DFT) calculations were carried-out using *Gaussian 16*[1] with the functionals summarized below.[2] TableA.1 summarizes the computational methods utilized in these studies to optimize the geometries of all stationary points followed by single-point energy corrections to refine the calculated electronic energy. We evaluated the ability of our chosen level of theory in optimizing structures by performing benchmark studies based on literature studies using a similar modeling system or by X-ray crystallography. (See chapters for more details). Geometry optimization calculations included the SMD implicit solvation model[3] to account for the solvent effects using similar solvents performed experimentally. In section 6.2, additional calculations using COSMO-RS[4] were performed to obtain the solvent COSMOments descriptors. For section 6.3, the DFT calculations were conducted in gas-phase.

**Table A.1.** Summarized computational methods for DFT studies in Chapters 2-6.

| Chapter/Section | G16 version | Geometry Opt. | Single-point corrections | Grid Size | (SMD=solvent) |
|---|---|---|---|---|---|
| **2** | C.01 | B3PW91/ 6-31+G(d,p), D3BJ, def2-SVP(Ir) | B3PW91/6-311+G(2df,p), D3BJ, def2-TZVPP(Ir) | ultrafine | acetonitrile (MeCN) |
| **3** | C.02 | **ω**B97X-D/6-31+G(d,p), def2-TZVP(Bi) or def2-SVP(Bi for preliminary studies) | **ω**B97X-D/ def2-QZVPP | superfine | MeCN (bismacycle system) and tetrahydrofuran ($Ph_3BiCl_2$ system) |
| **4** | C.03 | **ω**B97X-D/6-31+G(d,p), def2-TZVP(Bi, Pd) or def2-SVP(Bi, Pd for preliminary studies) | **ω**B97X-D/ def2-QZVPP | ultrafine | acetone |
| **5** | B.01 | **ω**B97X-D/ 6-31G(d), LANL2DZ(Pd) | **ω**B97X-D/ def2-QZVPP | ultrafine | dicholoroethane (DCE) |

| | | | | | |
|---|---|---|---|---|---|
| **6.2** | A.03 | **ω**B97X-D/ 6-31+G(d) | DLPNO-CCSD(T)/ cc-pV(DT)Z | ultrafine | toluene |
| **6.3** | B.01 | M06-2X/def2-TZVP | - | ultrafine | gas-phase |

**Conformational Sampling.** In Chapters 3-6 (section 6.2), conformational sampling of ground state structures (e.g., for starting materials, intermediates) and TS structures were mostly conducted manually (e.g., surveyed in terms of rotations about single bonds and dihedral angles), followed by a CREST analysis at the GFN2-xTB level.[5] The lowest energy conformers are then selected. In Chapter 2, the AQME program was used to facilitate the conformational sampling process in an automated manner.[6] In section 6.3, the automated workflow for generating the high-throughput calculations integrated RDKit for conformational searching and sampling to identify the lowest energy conformer using MMFF94s force field.[7]

**Confirming structures.** Vibrational frequency calculations were used to characterize GS and TS structures on the potential energy surface (PES), which possess zero and one imaginary frequency, respectively. Intrinsic Reaction Coordinate (IRC) calculations were performed to connect TS structures with corresponding intermediates on the PES.

**Thermochemical Data Analysis.** Quasi-harmonic corrections were applied to the computed vibrational entropies using the GoodVibes[8] program. Standard state corrections were evaluated at 1M concentration and a reaction temperature of 40 °C in agreement with experimental conditions. Reported Gibbs free energies (kcal/mol) are for the most stable conformer species found in each reaction step. All thermochemical data, including absolute energies and zero-point energies (ZPE) were generated and tabulated separately for creating the potential energy profile. These values are listed in the tables in sections A.2 to A.5.

**Molecular Featurization.** Atomic charges and Wiberg bond indices were calculated using natural population analysis (NPA) with *NBO 7.0*[9] with the similar optimization method. Isotropic NMR shielding tensor values were obtained using gauge-independent atomic orbital (GIAO)[10] calculations at the mPW1PW91/6-311+G(d,p)[11] level of theory. Molar volumes were obtained from the electronic isosurface in *Gaussian16*. Buried volumes (centered at the atom of interest) were calculated using the *DBSTEP*[12] program. Features were compiled using the *DISCO* workflow.[13]

**Figures and Plots Generated.** *NCIPLOT*[14] created the non-covalent interaction plots. Molecular representations were created with *PyMol*[15] for which the display settings are openly accessible.[16]

## *A.2 Summarized Thermochemical Data: Chapter 2 – $(BnSe)_2$ Model Substrate*

### *A.2.1 Raw Absolute Energy Values*

**Table A.2.** Summarized raw absolute energy values for the simple model substrate.

| File | E_SPC | E | ZPE | H_SPC | T.S | T.qh-S | G(T)_SPC | qh-G(T)_SPC | Imag Freq |
|---|---|---|---|---|---|---|---|---|---|
| **Bn-cation** | -270.74231 | -270.68352 | 0.11731 | -270.61835 | 0.03220 | 0.03219 | -270.65055 | -270.65054 | |
| **Bn-rad** | -270.92258 | -270.86253 | 0.11473 | -270.80120 | 0.03269 | 0.03269 | -270.83388 | -270.83389 | |
| **SeBn-anion** | -2672.64713 | -2670.44060 | 0.11898 | -2672.52024 | 0.03792 | 0.03756 | -2672.55816 | -2672.55780 | |
| SeBn-Model-**Int-I-rad_a** | -6086.35991 | -6081.86110 | 0.43462 | -6085.89989 | 0.07921 | 0.07415 | -6085.97910 | -6085.97404 | |
| SeBn-Model-**Int-I-rad_b** | -6086.35401 | -6081.85900 | 0.43438 | -6085.89409 | 0.08125 | 0.07525 | -6085.97534 | -6085.96934 | |
| SeBn-Model-**Int-I-rad_c** | -6086.35105 | -6081.85710 | 0.43539 | -6085.89042 | 0.07845 | 0.07375 | -6085.96887 | -6085.96417 | |
| SeBn-Model-**Int-I-rad_d** | -6086.33608 | -6081.84147 | 0.43334 | -6085.87661 | 0.08426 | 0.07675 | -6085.96087 | -6085.95335 | |
| SeBn-Model-**Int-I-rad_e** | -6086.34616 | -6081.84788 | 0.43338 | -6085.88679 | 0.08271 | 0.07607 | -6085.96951 | -6085.96286 | |
| SeBn-Model-**Int-I-rad_f** | -6086.33677 | -6081.84100 | 0.43380 | -6085.87709 | 0.08352 | 0.07618 | -6085.96061 | -6085.95327 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SeBn-Model-**Int-I-rad_g** | -6086.35479 | -6081.86030 | 0.43568 | -6085.89388 | 0.07917 | 0.07396 | -6085.97306 | -6085.96784 |
| SeBn-Model-**Int-II-anion_a** | -6086.52282 | -6082.02982 | 0.43480 | -6086.06254 | 0.07928 | 0.07390 | -6086.14182 | -6086.13643 |
| SeBn-Model-**Int-II-anion_b** | -6086.52464 | -6082.03063 | 0.43430 | -6086.06474 | 0.07941 | 0.07401 | -6086.14416 | -6086.13876 |
| SeBn-Model-**Int-II-anion_c** | -6086.52780 | -6082.03046 | 0.43420 | -6086.06804 | 0.07946 | 0.07409 | -6086.14750 | -6086.14213 |
| SeBn-Model-**Int-II-anion_d** | -6086.51536 | -6082.00080 | 0.43384 | -6086.05549 | 0.08421 | 0.07639 | -6086.13971 | -6086.13189 |
| SeBn-Model-**Int-II-anion_e** | -6086.51270 | -6081.99812 | 0.43388 | -6086.05272 | 0.08616 | 0.07727 | -6086.13888 | -6086.12999 |
| SeBn-Model-**Int-II-anion_f** | -6086.52777 | -6082.02548 | 0.43407 | -6086.06807 | 0.08267 | 0.07571 | -6086.15074 | -6086.14378 |
| SeBn-Model-**Int-II-anion_g** | -6086.52677 | -6082.02836 | 0.43502 | -6086.06651 | 0.07735 | 0.07302 | -6086.14386 | -6086.13954 |
| SeBn-Model-**Int-II-anion_h** | -6086.52236 | -6082.02332 | 0.43421 | -6086.06244 | 0.08109 | 0.07479 | -6086.14353 | -6086.13722 |
| SeBn-Model-**Int-II-anion_i** | -6086.52767 | -6082.03048 | 0.43435 | -6086.06771 | 0.08027 | 0.07437 | -6086.14799 | -6086.14209 |
| SeBn-Model-**Int-II-anion_k** | -6086.51711 | -6082.01701 | 0.43469 | -6086.05671 | 0.08178 | 0.07511 | -6086.13849 | -6086.13182 |
| SeBn-Model-**Int-II-anion_l** | -6086.52581 | -6082.02900 | 0.43441 | -6086.06588 | 0.08056 | 0.07449 | -6086.14644 | -6086.14037 |
| SeBn-Model-**Int-II-anion_m** | -6086.52760 | -6082.03017 | 0.43455 | -6086.06762 | 0.07897 | 0.07373 | -6086.14659 | -6086.14135 |
| SeBn-Model-**Int-II-anion_n** | -6086.51748 | -6082.00822 | 0.43378 | -6086.05769 | 0.08421 | 0.07644 | -6086.14189 | -6086.13413 |
| SeBn-Model-**Int-II-anion_o** | -6086.52526 | -6082.02233 | 0.43369 | -6086.06573 | 0.08149 | 0.07494 | -6086.14722 | -6086.14067 |
| SeBn-Model-**Int-II-anion_p** | -6086.51787 | -6082.00853 | 0.43367 | -6086.05812 | 0.08667 | 0.07759 | -6086.14480 | -6086.13571 |
| SeBn-Model-**Int-II-anion_q** | -6086.51454 | -6082.02308 | 0.43434 | -6086.05480 | 0.07991 | 0.07411 | -6086.13471 | -6086.12891 |
| SeBn-Model-**Int-II-anion_r** | -6086.52538 | -6082.02429 | 0.43411 | -6086.06561 | 0.07989 | 0.07424 | -6086.14550 | -6086.13986 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SeBn-Model-**Int-II-anion_s** | -6086.51995 | -6082.01811 | 0.43470 | -6086.05970 | 0.08139 | 0.07488 | -6086.14108 | -6086.13458 |
| SeBn-Model-**Int-II-anion_t** | -6086.51983 | -6082.01890 | 0.43574 | -6086.05868 | 0.08004 | 0.07425 | -6086.13872 | -6086.13293 |
| SeBn-Model-**Int-II-rad_a** | -6086.35817 | -6081.86007 | 0.43537 | -6085.89725 | 0.08020 | 0.07474 | -6085.97745 | -6085.97199 |
| SeBn-Model-**Int-II-rad_b** | -6086.35985 | -6081.86282 | 0.43504 | -6085.89937 | 0.07866 | 0.07407 | -6085.97803 | -6085.97344 |
| SeBn-Model-**Int-II-rad_c** | -6086.35572 | -6081.86139 | 0.43614 | -6085.89434 | 0.07867 | 0.07397 | -6085.97301 | -6085.96830 |
| SeBn-Model-**Int-II-rad_d** | -6086.35707 | -6081.86154 | 0.43593 | -6085.89580 | 0.07973 | 0.07445 | -6085.97553 | -6085.97025 |
| SeBn-Model-**Int-II-rad_f** | -6086.35421 | -6081.85949 | 0.43579 | -6085.89283 | 0.08184 | 0.07547 | -6085.97467 | -6085.96831 |
| SeBn-Model-**Int-II-rad_g** | -6086.36473 | -6081.87002 | 0.43465 | -6085.90456 | 0.07968 | 0.07446 | -6085.98424 | -6085.97902 |
| SeBn-Model-**Int-II-rad_h** | -6086.36033 | -6081.85608 | 0.43586 | -6085.89921 | 0.07866 | 0.07398 | -6085.97787 | -6085.97318 |
| SeBn-Model-**Int-II-rad_i** | -6086.35423 | -6081.85623 | 0.43494 | -6085.89372 | 0.07981 | 0.07471 | -6085.97354 | -6085.96843 |
| SeBn-Model-**Int-II-rad_j** | -6086.35926 | -6081.85331 | 0.43537 | -6085.89842 | 0.08116 | 0.07520 | -6085.97958 | -6085.97362 |
| SeBn-Model-**Int-II-rad_l** | -6086.35158 | -6081.84423 | 0.43465 | -6085.89131 | 0.08156 | 0.07536 | -6085.97287 | -6085.96667 |
| SeBn-Model-**Int-II-rad_m** | -6086.35574 | -6081.85215 | 0.43447 | -6085.89573 | 0.08019 | 0.07480 | -6085.97593 | -6085.97053 |
| SeBn-Model-**Int-II-rad_n** | -6086.35623 | -6081.85342 | 0.43482 | -6085.89604 | 0.07904 | 0.07420 | -6085.97507 | -6085.97024 |
| SeBn-Model-**Int-II-rad_o** | -6086.35448 | -6081.85482 | 0.43480 | -6085.89419 | 0.08010 | 0.07454 | -6085.97429 | -6085.96873 |
| SeBn-Model-**Int-II-rad_p** | -6086.35247 | -6081.85383 | 0.43546 | -6085.89164 | 0.07987 | 0.07451 | -6085.97151 | -6085.96615 |
| SeBn-Model-**Int-II-rad_r** | -6086.34818 | -6081.84060 | 0.43336 | -6085.88875 | 0.08544 | 0.07736 | -6085.97418 | -6085.96610 |
| SeBn-Model-**Int-III-anion_a** | -6086.56136 | -6082.05516 | 0.43495 | -6086.10107 | 0.07884 | 0.07352 | -6086.17991 | -6086.17458 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SeBn-Model-**Int-III-anion_b** | -6086.55710 | -6082.04812 | 0.43435 | -6086.09701 | 0.08337 | 0.07575 | -6086.18038 | -6086.17276 |
| SeBn-Model-**Int-III-anion_c** | -6086.55927 | -6082.05476 | 0.43525 | -6086.09873 | 0.07918 | 0.07366 | -6086.17791 | -6086.17239 |
| SeBn-Model-**Int-III-anion_d** | -6086.55877 | -6082.05546 | 0.43612 | -6086.09752 | 0.07903 | 0.07348 | -6086.17655 | -6086.17100 |
| SeBn-Model-**Int-III-anion_e** | -6086.55574 | -6082.05011 | 0.43458 | -6086.09543 | 0.08315 | 0.07555 | -6086.17858 | -6086.17098 |
| SeBn-Model-**Int-III-anion_f** | -6086.55740 | -6082.05767 | 0.43419 | -6086.09778 | 0.07929 | 0.07380 | -6086.17707 | -6086.17157 |
| SeBn-Model-**Int-III-anion_g** | -6086.55706 | -6082.05522 | 0.43411 | -6086.09743 | 0.08139 | 0.07481 | -6086.17882 | -6086.17224 |
| SeBn-Model-**Int-III-anion_h** | -6086.56082 | -6082.05515 | 0.43489 | -6086.10049 | 0.08143 | 0.07459 | -6086.18191 | -6086.17508 |
| SeBn-Model-**Int-III-anion_i** | -6086.55631 | -6082.05256 | 0.43481 | -6086.09600 | 0.08285 | 0.07547 | -6086.17885 | -6086.17147 |
| SeBn-Model-**Int-III-B-anion_a** | -6086.48016 | -6081.98566 | 0.42977 | -6086.02417 | 0.08222 | 0.07552 | -6086.10638 | -6086.09968 |
| SeBn-Model-**Int-III-B-anion_b** | -6086.49215 | -6081.99019 | 0.43448 | -6086.03176 | 0.08322 | 0.07587 | -6086.11498 | -6086.10763 |
| SeBn-Model-**Int-III-B-anion_c** | -6086.49651 | -6082.00537 | 0.43446 | -6086.03636 | 0.08048 | 0.07456 | -6086.11684 | -6086.11092 |
| SeBn-Model-**Int-III-B-anion_d** | -6086.49119 | -6081.99536 | 0.43446 | -6086.03081 | 0.08267 | 0.07570 | -6086.11349 | -6086.10651 |
| SeBn-Model-**Int-III-B-anion_e** | -6086.49371 | -6082.00735 | 0.43468 | -6086.03347 | 0.08047 | 0.07440 | -6086.11394 | -6086.10786 |
| SeBn-Model-**Int-III-B-anion_f** | -6086.47996 | -6081.98598 | 0.42827 | -6086.02473 | 0.08685 | 0.07804 | -6086.11158 | -6086.10278 |
| SeBn-Model-**Int-III-B-anion_g** | -6086.49128 | -6081.99631 | 0.43474 | -6086.03087 | 0.08150 | 0.07512 | -6086.11237 | -6086.10599 |
| SeBn-Model-**Int-III-B-anion_h** | -6086.49413 | -6081.99362 | 0.43468 | -6086.03362 | 0.08274 | 0.07562 | -6086.11636 | -6086.10924 |
| SeBn-Model-**Int-III-B-anion_i** | -6086.49414 | -6082.00042 | 0.43486 | -6086.03345 | 0.08225 | 0.07533 | -6086.11570 | -6086.10878 |
| SeBn-Model-**Int-III-B-anion_j** | -6086.49141 | -6081.99744 | 0.43527 | -6086.03041 | 0.08168 | 0.07499 | -6086.11209 | -6086.10540 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SeBn-Model-**Int-III-B-anion_k** | -6086.47288 | -6081.98173 | 0.42981 | -6086.01666 | 0.08435 | 0.07658 | -6086.10101 | -6086.09324 |
| SeBn-Model-**Int-III-B-anion_l** | -6086.49252 | -6081.99408 | 0.43391 | -6086.03250 | 0.08425 | 0.07639 | -6086.11675 | -6086.10889 |
| SeBn-Model-**Int-III-B-anion_m** | -6086.49611 | -6081.99939 | 0.43440 | -6086.03584 | 0.08196 | 0.07518 | -6086.11779 | -6086.11102 |
| SeBn-Model-**Int-III-B-anion_n** | -6086.49159 | -6081.99768 | 0.43460 | -6086.03114 | 0.08276 | 0.07567 | -6086.11390 | -6086.10681 |
| SeBn-Model-**Int-III-B-anion_o** | -6086.49491 | -6082.00135 | 0.43440 | -6086.03485 | 0.07992 | 0.07436 | -6086.11478 | -6086.10921 |
| SeBn-Model-**Int-III-B-anion_r** | -6086.49144 | -6081.99765 | 0.43452 | -6086.03093 | 0.08421 | 0.07624 | -6086.11514 | -6086.10717 |
| SeBn-Model-**Int-III-B-anion_s** | -6086.49488 | -6081.99647 | 0.43451 | -6086.03441 | 0.08303 | 0.07582 | -6086.11745 | -6086.11023 |
| SeBn-Model-**Int-III-B-anion_u** | -6086.49240 | -6081.99657 | 0.43436 | -6086.03212 | 0.08272 | 0.07556 | -6086.11484 | -6086.10768 |
| SeBn-Model-**Int-III-B-anion_v** | -6086.49505 | -6082.00177 | 0.43498 | -6086.03435 | 0.08078 | 0.07467 | -6086.11514 | -6086.10902 |
| SeBn-Model-**Int-III-B-rad_a** | -6086.32295 | -6081.83148 | 0.43617 | -6085.86113 | 0.08067 | 0.07500 | -6085.94180 | -6085.93613 |
| SeBn-Model-**Int-III-B-rad_b** | -6086.33782 | -6081.82951 | 0.43052 | -6085.88172 | 0.07995 | 0.07477 | -6085.96167 | -6085.95649 |
| SeBn-Model-**Int-III-B-rad_d** | -6086.31521 | -6081.81802 | 0.43560 | -6085.85355 | 0.08367 | 0.07639 | -6085.93722 | -6085.92995 |
| SeBn-Model-**Int-III-B-rad_e** | -6086.31714 | -6081.83159 | 0.43631 | -6085.85533 | 0.08067 | 0.07490 | -6085.93600 | -6085.93023 |
| SeBn-Model-**Int-III-B-rad_f** | -6086.29215 | -6081.79925 | 0.42929 | -6085.83611 | 0.08339 | 0.07714 | -6085.91950 | -6085.91325 |
| SeBn-Model-**Int-III-B-rad_g** | -6086.31745 | -6081.82547 | 0.43666 | -6085.85519 | 0.08168 | 0.07558 | -6085.93688 | -6085.93077 |
| SeBn-Model-**Int-III-B-rad_h** | -6086.28917 | -6081.78827 | 0.43062 | -6085.83229 | 0.08292 | 0.07665 | -6085.91521 | -6085.90894 |
| SeBn-Model-**Int-III-B-rad_i** | -6086.31801 | -6081.82346 | 0.43655 | -6085.85568 | 0.08199 | 0.07559 | -6085.93766 | -6085.93127 |
| SeBn-Model-**Int-III-B-rad_j** | -6086.31560 | -6081.82041 | 0.43606 | -6085.85353 | 0.08407 | 0.07658 | -6085.93760 | -6085.93011 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SeBn-Model-**Int-III-B-rad_k** | -6086.31630 | -6081.81751 | 0.43470 | -6085.85542 | 0.08454 | 0.07686 | -6085.93996 | -6085.93228 |
| SeBn-Model-**Int-III-B-rad_m** | -6086.32048 | -6081.82537 | 0.43566 | -6085.85909 | 0.08107 | 0.07531 | -6085.94016 | -6085.93440 |
| SeBn-Model-**Int-III-B-rad_n** | -6086.31916 | -6081.82709 | 0.43663 | -6085.85690 | 0.08329 | 0.07634 | -6085.94019 | -6085.93324 |
| SeBn-Model-**Int-III-D-rad_a** | -6086.35123 | -6081.86207 | 0.43437 | -6085.89155 | 0.07702 | 0.07323 | -6085.96857 | -6085.96478 |
| SeBn-Model-**Int-III-D-rad_d** | -6086.34656 | -6081.85654 | 0.43455 | -6085.88650 | 0.07938 | 0.07440 | -6085.96589 | -6085.96090 |
| SeBn-Model-**Int-III-D-rad_f** | -6086.34614 | -6081.85301 | 0.43508 | -6085.88577 | 0.07835 | 0.07383 | -6085.96412 | -6085.95960 |
| SeBn-Model-**Int-III-D-rad_g** | -6086.35001 | -6081.85636 | 0.43477 | -6085.89007 | 0.07646 | 0.07296 | -6085.96653 | -6085.96303 |
| SeBn-Model-**Int-III-D-rad_h** | -6086.34073 | -6081.84480 | 0.43541 | -6085.88007 | 0.07913 | 0.07404 | -6085.95919 | -6085.95410 |
| SeBn-Model-**Int-III-rad_a** | -6086.38460 | -6081.89099 | 0.43379 | -6085.92513 | 0.08086 | 0.07495 | -6086.00599 | -6086.00008 |
| SeBn-Model-**Int-III-rad_b** | -6086.36201 | -6081.86660 | 0.43468 | -6085.90153 | 0.08168 | 0.07562 | -6085.98321 | -6085.97714 |
| SeBn-Model-**Int-III-rad_c** | -6086.37499 | -6081.87381 | 0.43659 | -6085.91343 | 0.07851 | 0.07370 | -6085.99194 | -6085.98712 |
| SeBn-Model-**Int-III-rad_e** | -6086.38216 | -6081.88410 | 0.43557 | -6085.92124 | 0.08091 | 0.07497 | -6086.00216 | -6085.99621 |
| SeBn-Model-**Int-III-rad_f** | -6086.38426 | -6081.89271 | 0.43487 | -6085.92404 | 0.07944 | 0.07416 | -6086.00348 | -6085.99820 |
| SeBn-Model-**Int-III-rad_g** | -6086.37751 | -6081.87404 | 0.43678 | -6085.91573 | 0.07883 | 0.07384 | -6085.99456 | -6085.98957 |
| SeBn-Model-**Int-III-rad_h** | -6086.38525 | -6081.89048 | 0.43519 | -6085.92481 | 0.07888 | 0.07390 | -6086.00370 | -6085.99871 |
| SeBn-Model-**Int-III-rad_i** | -6086.37922 | -6081.87804 | 0.43645 | -6085.91773 | 0.07877 | 0.07386 | -6085.99650 | -6085.99159 |
| SeBn-Model-**Int-III-rad_j** | -6086.38147 | -6081.88735 | 0.43575 | -6085.92062 | 0.07883 | 0.07379 | -6085.99945 | -6085.99441 |
| SeBn-Model-**Int-III-rad_k** | -6086.38151 | -6081.88902 | 0.43600 | -6085.92044 | 0.07866 | 0.07371 | -6085.99910 | -6085.99415 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| SeBn-Model-**Int-III-rad_l** | -6086.36049 | -6081.85023 | 0.43466 | -6085.90006 | 0.08282 | 0.07614 | -6085.98288 | -6085.97620 | |
| **SeBn-Model-Reagents_complex_a** | -6086.53667 | -6082.03849 | 0.43342 | -6086.07765 | 0.07980 | 0.07407 | -6086.15745 | -6086.15171 | |
| **SeBn-Model-Reagents_complex_b** | -6086.53136 | -6082.03502 | 0.43338 | -6086.07222 | 0.08210 | 0.07510 | -6086.15433 | -6086.14732 | |
| **SeBn-Model-Reagents_complex_c** | -6086.52396 | -6082.02737 | 0.43395 | -6086.06424 | 0.08349 | 0.07560 | -6086.14773 | -6086.13984 | |
| **SeBn-Model-Reagents_complex_d** | -6086.51842 | -6082.02193 | 0.43419 | -6086.05850 | 0.08374 | 0.07580 | -6086.14224 | -6086.13430 | |
| **SeBn-Model-Reagents_complex_e** | -6086.52927 | -6082.03163 | 0.43319 | -6086.07021 | 0.08271 | 0.07531 | -6086.15292 | -6086.14552 | |
| **SeBn-Model-Reagents_complex_g** | -6086.52418 | -6082.02558 | 0.43351 | -6086.06467 | 0.08503 | 0.07661 | -6086.14970 | -6086.14129 | |
| **SeBn-Model-TS-I_a** | -6086.35802 | -6081.85629 | 0.43448 | -6085.89870 | 0.07922 | 0.07355 | -6085.97792 | -6085.97225 | -46.35 |
| **SeBn-Model-TS-I_b** | -6086.35342 | -6081.85761 | 0.43528 | -6085.89351 | 0.07795 | 0.07286 | -6085.97146 | -6085.96637 | -64.72 |
| **SeBn-Model-TS-I_c** | -6086.35180 | -6081.85426 | 0.43448 | -6085.89262 | 0.07763 | 0.07279 | -6085.97025 | -6085.96541 | -49.91 |
| **SeBn-Model-TS-IV_a** | -6086.50667 | -6082.00479 | 0.43301 | -6086.04828 | 0.07999 | 0.07410 | -6086.12826 | -6086.12238 | -429.52 |
| **SeBn-Model-TS-IV_c** | -6086.50800 | -6082.00693 | 0.43413 | -6086.04889 | 0.07703 | 0.07268 | -6086.12591 | -6086.12157 | -436.48 |
| **SeBn-Model-TS-IV_e** | -6086.50351 | -6082.00088 | 0.43383 | -6086.04442 | 0.07983 | 0.07406 | -6086.12425 | -6086.11848 | -420.17 |
| **SeBn-Model-TS-IV_f** | -6086.50352 | -6081.99853 | 0.43306 | -6086.04507 | 0.08116 | 0.07482 | -6086.12623 | -6086.11989 | -412.3 |
| **SeBn-Model-TS-IV_g** | -6086.50908 | -6082.00161 | 0.43321 | -6086.05057 | 0.07933 | 0.07383 | -6086.12990 | -6086.12440 | -420.21 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **SeBn-Model-TS-VIII_a** | -6086.52408 | -6082.02506 | 0.43461 | -6086.06483 | 0.07780 | 0.07253 | -6086.14263 | -6086.13736 | -38.3 |
| **SeBn-Model-TS-VIII_b** | -6086.51762 | -6082.01815 | 0.43387 | -6086.05865 | 0.08291 | 0.07483 | -6086.14156 | -6086.13348 | -56.51 |
| **BnSe-Model-TS-II-E-anion_a** | -8017.71805 | -8011.14706 | 0.35919 | -8017.33411 | 0.07960 | 0.07355 | -8017.41371 | -8017.40766 | -44.86 |
| **BnSe-Model-TS-II-E-anion_b** | -8017.71419 | -8011.14891 | 0.35850 | -8017.33093 | 0.07911 | 0.07332 | -8017.41004 | -8017.40425 | -44.82 |
| **BnSe-Model-TS-II-E-rad_a** | -8017.54816 | -8010.97672 | 0.35959 | -8017.16376 | 0.08141 | 0.07475 | -8017.24517 | -8017.23851 | -61.56 |
| **BnSe-Model-TS-II-E-rad_b** | -8017.54619 | -8010.96742 | 0.35942 | -8017.16171 | 0.08400 | 0.07613 | -8017.24571 | -8017.23784 | -91.04 |
| **BnSe-Model-TS-II-E-rad_c** | -8017.54965 | -8010.97578 | 0.35970 | -8017.16497 | 0.08313 | 0.07560 | -8017.24810 | -8017.24056 | -75.74 |
| **BnSe-Model-TS-II-E-rad_d** | -8017.54584 | -8010.97695 | 0.35953 | -8017.16145 | 0.08195 | 0.07511 | -8017.24339 | -8017.23655 | -54.2 |
| **BnSe-Model-TS-II-E-rad_e** | -8017.54108 | -8010.96998 | 0.35993 | -8017.15645 | 0.08169 | 0.07483 | -8017.23814 | -8017.23128 | -164.58 |
| **BnSe-Model-TS-II-E-rad_f** | -8017.53848 | -8010.96497 | 0.36075 | -8017.15315 | 0.08263 | 0.07507 | -8017.23578 | -8017.22822 | -130.22 |
| **BnSe-Model-TS-II-E-rad_k** | -8017.54112 | -8010.97187 | 0.36045 | -8017.15615 | 0.08014 | 0.07407 | -8017.23630 | -8017.23022 | -103.72 |
| **BnSe-Model-TS-II-F-anion_a** | -8017.68664 | -8011.11457 | 0.35927 | -8017.30258 | 0.07856 | 0.07336 | -8017.38114 | -8017.37594 | -479.9 |
| **BnSe-Model-TS-II-F-anion_aa** | -8017.67807 | -8011.09521 | 0.35793 | -8017.29470 | 0.08448 | 0.07646 | -8017.37918 | -8017.37117 | -470.18 |
| **BnSe-Model-TS-II-F-anion_ab** | -8017.68201 | -8011.10082 | 0.35811 | -8017.29857 | 0.08225 | 0.07540 | -8017.38082 | -8017.37397 | -461.2 |
| **BnSe-Model-TS-II-F-anion_ac** | -8017.68102 | -8011.10044 | 0.35825 | -8017.29743 | 0.08308 | 0.07569 | -8017.38051 | -8017.37312 | -480.25 |
| **BnSe-Model-TS-II-F-anion_ae** | -8017.68116 | -8011.09753 | 0.35839 | -8017.29752 | 0.08299 | 0.07572 | -8017.38051 | -8017.37324 | -440.9 |
| **BnSe-Model-TS-II-F-anion_af** | -8017.67559 | -8011.09246 | 0.35854 | -8017.29188 | 0.08228 | 0.07536 | -8017.37416 | -8017.36724 | -435.47 |
| **BnSe-Model-TS-II-F-anion_b** | -8017.68597 | -8011.11760 | 0.35915 | -8017.30199 | 0.07816 | 0.07318 | -8017.38016 | -8017.37518 | -471.18 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **BnSe-Model-TS-II-F-anion_c** | -8017.68741 | -8011.11490 | 0.35911 | -8017.30348 | 0.07831 | 0.07330 | -8017.38179 | -8017.37678 | -476.66 |
| **BnSe-Model-TS-II-F-anion_d** | -8017.68343 | -8011.11324 | 0.35934 | -8017.29934 | 0.07929 | 0.07363 | -8017.37863 | -8017.37296 | -473.19 |
| **BnSe-Model-TS-II-F-anion_e** | -8017.68685 | -8011.11016 | 0.35892 | -8017.30312 | 0.07855 | 0.07343 | -8017.38167 | -8017.37655 | -448.45 |
| **BnSe-Model-TS-II-F-anion_g** | -8017.68408 | -8011.10400 | 0.35834 | -8017.30070 | 0.07906 | 0.07388 | -8017.37977 | -8017.37459 | -460.47 |
| **BnSe-Model-TS-II-F-anion_h** | -8017.68310 | -8011.10649 | 0.35942 | -8017.29895 | 0.07838 | 0.07330 | -8017.37732 | -8017.37225 | -459.8 |
| **BnSe-Model-TS-II-F-anion_i** | -8017.68112 | -8011.10452 | 0.35924 | -8017.29699 | 0.07973 | 0.07388 | -8017.37672 | -8017.37087 | -469.83 |
| **BnSe-Model-TS-II-F-anion_k** | -8017.67340 | -8011.10204 | 0.36170 | -8017.28708 | 0.08046 | 0.07399 | -8017.36754 | -8017.36107 | -450.1 |
| **BnSe-Model-TS-II-F-anion_l** | -8017.68176 | -8011.11653 | 0.36023 | -8017.29702 | 0.07827 | 0.07297 | -8017.37529 | -8017.36998 | -470.43 |
| **BnSe-Model-TS-II-F-anion_m** | -8017.67871 | -8011.10866 | 0.35911 | -8017.29472 | 0.07890 | 0.07353 | -8017.37361 | -8017.36825 | -487.08 |
| **BnSe-Model-TS-II-F-anion_n** | -8017.66711 | -8011.08289 | 0.35840 | -8017.28320 | 0.08637 | 0.07731 | -8017.36958 | -8017.36051 | -496.16 |
| **BnSe-Model-TS-II-F-anion_o** | -8017.66860 | -8011.08444 | 0.35858 | -8017.28460 | 0.08542 | 0.07683 | -8017.37003 | -8017.36144 | -486.88 |
| **BnSe-Model-TS-II-F-anion_p** | -8017.68332 | -8011.11774 | 0.36049 | -8017.29829 | 0.07798 | 0.07293 | -8017.37627 | -8017.37121 | -475.06 |
| **BnSe-Model-TS-II-F-anion_q** | -8017.67602 | -8011.10713 | 0.36049 | -8017.29062 | 0.08169 | 0.07471 | -8017.37231 | -8017.36533 | -474.92 |
| **BnSe-Model-TS-II-F-anion_r** | -8017.67508 | -8011.10060 | 0.36028 | -8017.28992 | 0.08124 | 0.07453 | -8017.37116 | -8017.36445 | -448.42 |
| **BnSe-Model-TS-II-F-anion_s** | -8017.67451 | -8011.09050 | 0.35843 | -8017.29054 | 0.08740 | 0.07781 | -8017.37795 | -8017.36835 | -483.32 |
| **BnSe-Model-TS-II-F-anion_t** | -8017.67822 | -8011.10994 | 0.35958 | -8017.29369 | 0.07980 | 0.07402 | -8017.37348 | -8017.36770 | -485.42 |
| **BnSe-Model-TS-II-F-anion_u** | -8017.68009 | -8011.10729 | 0.35920 | -8017.29574 | 0.08094 | 0.07457 | -8017.37668 | -8017.37031 | -480.14 |
| **BnSe-Model-TS-II-F-anion_v** | -8017.67818 | -8011.10993 | 0.35959 | -8017.29363 | 0.07983 | 0.07404 | -8017.37346 | -8017.36767 | -485.5 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **BnSe-Model-TS-II-F-anion_y** | -8017.67697 | -8011.10587 | 0.36002 | -8017.29214 | 0.07952 | 0.07382 | -8017.37165 | -8017.36596 | -444.24 |
| **BnSe-Model-TS-II-F-anion_z** | -8017.67276 | -8011.09554 | 0.35939 | -8017.28842 | 0.08197 | 0.07502 | -8017.37039 | -8017.36343 | -434.4 |
| **BnSe-Model-TS-II-F-rad_a** | -8017.52582 | -8010.95662 | 0.36018 | -8017.14092 | 0.07983 | 0.07421 | -8017.22075 | -8017.21513 | -627.27 |
| **BnSe-Model-TS-II-F-rad_aa** | -8017.52636 | -8010.95865 | 0.35927 | -8017.14225 | 0.07989 | 0.07434 | -8017.22214 | -8017.21659 | -610.64 |
| **BnSe-Model-TS-II-F-rad_ab** | -8017.52648 | -8010.96251 | 0.35962 | -8017.14217 | 0.07886 | 0.07376 | -8017.22102 | -8017.21592 | -599.89 |
| **BnSe-Model-TS-II-F-rad_ac** | -8017.52611 | -8010.95364 | 0.36000 | -8017.14140 | 0.07991 | 0.07424 | -8017.22131 | -8017.21564 | -599.13 |
| **BnSe-Model-TS-II-F-rad_af** | -8017.52426 | -8010.94804 | 0.35833 | -8017.14065 | 0.08433 | 0.07657 | -8017.22498 | -8017.21722 | -607.88 |
| **BnSe-Model-TS-II-F-rad_c** | -8017.53125 | -8010.96173 | 0.35914 | -8017.14728 | 0.07928 | 0.07404 | -8017.22656 | -8017.22132 | -627.64 |
| **BnSe-Model-TS-II-F-rad_d** | -8017.52972 | -8010.96588 | 0.35990 | -8017.14518 | 0.07861 | 0.07351 | -8017.22379 | -8017.21869 | -628.99 |
| **BnSe-Model-TS-II-F-rad_e** | -8017.52665 | -8010.95531 | 0.35961 | -8017.14234 | 0.08033 | 0.07438 | -8017.22267 | -8017.21672 | -599.29 |
| **BnSe-Model-TS-II-F-rad_f** | -8017.52879 | -8010.96534 | 0.36043 | -8017.14390 | 0.07786 | 0.07312 | -8017.22176 | -8017.21702 | -615.07 |
| **BnSe-Model-TS-II-F-rad_g** | -8017.52469 | -8010.95116 | 0.35979 | -8017.14022 | 0.07885 | 0.07379 | -8017.21908 | -8017.21401 | -624.33 |
| **BnSe-Model-TS-II-F-rad_i** | -8017.52034 | -8010.94305 | 0.35903 | -8017.13632 | 0.08145 | 0.07498 | -8017.21778 | -8017.21131 | -650.83 |
| **BnSe-Model-TS-II-F-rad_k** | -8017.52780 | -8010.96351 | 0.35868 | -8017.14420 | 0.08089 | 0.07472 | -8017.22509 | -8017.21891 | -657.38 |
| **BnSe-Model-TS-II-F-rad_m** | -8017.52022 | -8010.95066 | 0.35999 | -8017.13542 | 0.08034 | 0.07450 | -8017.21576 | -8017.20992 | -663.5 |
| **BnSe-Model-TS-II-F-rad_n** | -8017.51555 | -8010.93966 | 0.36045 | -8017.13010 | 0.08282 | 0.07562 | -8017.21291 | -8017.20572 | -638.36 |
| **BnSe-Model-TS-II-F-rad_o** | -8017.51575 | -8010.93367 | 0.35863 | -8017.13169 | 0.08655 | 0.07776 | -8017.21824 | -8017.20945 | -670.79 |
| **BnSe-Model-TS-II-F-rad_s** | -8017.52377 | -8010.95222 | 0.35945 | -8017.13937 | 0.08119 | 0.07490 | -8017.22056 | -8017.21427 | -665.72 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **BnSe-Model-TS-II-F-rad_t** | -8017.51786 | -8010.94390 | 0.35912 | -8017.13358 | 0.08366 | 0.07622 | -8017.21724 | -8017.20980 | -621.09 |
| **BnSe-Model-TS-II-F-rad_u** | -8017.52159 | -8010.95183 | 0.36015 | -8017.13669 | 0.08080 | 0.07461 | -8017.21749 | -8017.21130 | -592.88 |
| **BnSe-Model-TS-II-F-rad_v** | -8017.52288 | -8010.95471 | 0.35956 | -8017.13834 | 0.08220 | 0.07536 | -8017.22054 | -8017.21369 | -624.37 |
| **BnSe-Model-TS-II-F-rad_w** | -8017.51717 | -8010.94501 | 0.36039 | -8017.13192 | 0.08182 | 0.07516 | -8017.21374 | -8017.20707 | -613.34 |
| **BnSe-Model-TS-II-F-rad_y** | -8017.52084 | -8010.95118 | 0.35957 | -8017.13643 | 0.08016 | 0.07449 | -8017.21659 | -8017.21092 | -601.48 |
| **BnSe-Model-TS-II-F-rad_z** | -8017.51379 | -8010.93932 | 0.35990 | -8017.12904 | 0.08229 | 0.07545 | -8017.21132 | -8017.20448 | -605.25 |
| **SeBn-Model-TS-VII_a** | -6086.32634 | -6081.82856 | 0.43370 | -6085.86747 | 0.07916 | 0.07398 | -6085.94663 | -6085.94145 | -612.26 |
| **SeBn-Model-TS-VII_b** | -6086.33055 | -6081.83166 | 0.43383 | -6085.87172 | 0.07756 | 0.07332 | -6085.94928 | -6085.94503 | -612.9 |
| **SeBn-Model-TS-VII_c** | -6086.32141 | -6081.82047 | 0.43390 | -6085.86217 | 0.08103 | 0.07504 | -6085.94321 | -6085.93721 | -601.21 |
| **SeBn-Model-TS-VII_e** | -6086.32308 | -6081.82204 | 0.43440 | -6085.86335 | 0.08144 | 0.07513 | -6085.94479 | -6085.93848 | -579.15 |
| **SeBn-Model-TS-VII_f** | -6086.32972 | -6081.83139 | 0.43451 | -6085.87030 | 0.07794 | 0.07337 | -6085.94824 | -6085.94368 | -604.48 |
| **SeBn-rad** | -2672.47818 | -2670.27274 | 0.11830 | -2672.35187 | 0.03874 | 0.03834 | -2672.39061 | -2672.39021 | |
| **BnSeBn_a** | -2943.49082 | -2941.23709 | 0.23837 | -2943.23771 | 0.05569 | 0.05204 | -2943.29340 | -2943.28975 | |
| **BnSeBn_b** | -2943.49297 | -2941.24097 | 0.23887 | -2943.23953 | 0.05580 | 0.05237 | -2943.29533 | -2943.29190 | |
| **BnSeBn_d** | -2943.49302 | -2941.24045 | 0.23940 | -2943.23932 | 0.05372 | 0.05134 | -2943.29303 | -2943.29066 | |
| **BnSeBn_e** | -2943.49422 | -2941.24094 | 0.23908 | -2943.24070 | 0.05419 | 0.05160 | -2943.29489 | -2943.29230 | |
| **diSeBn_a** | -5345.05996 | -5340.66949 | 0.23985 | -5344.80375 | 0.05905 | 0.05596 | -5344.86280 | -5344.85970 | |
| **diSeBn_b** | -5345.05796 | -5340.66430 | 0.23985 | -5344.80165 | 0.05952 | 0.05634 | -5344.86117 | -5344.85799 | |
| **diSeBn_c** | -5345.05486 | -5340.66326 | 0.24003 | -5344.79833 | 0.06075 | 0.05681 | -5344.85908 | -5344.85514 | |
| **diSeBn_d** | -5345.05181 | -5340.65792 | 0.24031 | -5344.79492 | 0.06100 | 0.05663 | -5344.85592 | -5344.85155 | |
| **diSeBn_e** | -5345.04287 | -5340.64864 | 0.24020 | -5344.78604 | 0.06141 | 0.05681 | -5344.84745 | -5344.84286 | |
| **diSeBn_f** | -5345.05771 | -5340.66525 | 0.24021 | -5344.80112 | 0.05860 | 0.05545 | -5344.85972 | -5344.85657 | |
| **Ir2** | -2943.09938 | -2942.30090 | 0.67959 | -2942.36718 | 0.14460 | 0.12846 | -2942.51178 | -2942.49565 | |
| **Ir3-T1** | -2942.90166 | -2942.10387 | 0.67998 | -2942.16891 | 0.14221 | 0.12774 | -2942.31113 | -2942.29665 | |
| **Ir3** | -2942.99872 | -2942.20164 | 0.68400 | -2942.26270 | 0.13762 | 0.12409 | -2942.40032 | -2942.38679 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **PTA-rad** | -741.27385 | -741.15431 | 0.19188 | -741.07335 | 0.03801 | 0.03802 | -741.11135 | -741.11136 |
| **PTA-SeBn-rad_a** | -3413.96915 | -3411.65223 | 0.31263 | -3413.63937 | 0.06188 | 0.05809 | -3413.70125 | -3413.69745 |
| **PTA** | -741.46256 | -741.34243 | 0.19227 | -741.26207 | 0.03558 | 0.03560 | -741.29765 | -741.29766 |
| **PTASe-Bn-rad_a** | -3413.98558 | -3411.66341 | 0.30972 | -3413.65794 | 0.06369 | 0.05940 | -3413.72163 | -3413.71733 |
| **PTASe-Bn-rad_b** | -3413.98559 | -3411.66341 | 0.30975 | -3413.65793 | 0.06366 | 0.05938 | -3413.72159 | -3413.71731 |
| **PTASeBn-Nbond_a** | -3413.82814 | -3411.52356 | 0.31535 | -3413.49647 | 0.05641 | 0.05453 | -3413.55287 | -3413.55100 |
| **PTASeBn-Nbond_b** | -3413.82587 | -3411.52236 | 0.31507 | -3413.49440 | 0.05738 | 0.05496 | -3413.55178 | -3413.54935 |
| **PTASeBn-Nbond_c** | -3413.82592 | -3411.52198 | 0.31507 | -3413.49419 | 0.05922 | 0.05591 | -3413.55341 | -3413.55009 |
| **PTASeBn-Nbond_d** | -3413.82803 | -3411.52317 | 0.31549 | -3413.49624 | 0.05668 | 0.05462 | -3413.55292 | -3413.55086 |
| **PTASeBn-Nbond_e** | -3413.82549 | -3411.52152 | 0.31489 | -3413.49416 | 0.05760 | 0.05509 | -3413.55176 | -3413.54925 |
| **PTASeBn-Nbond_f** | -3413.82613 | -3411.52199 | 0.31499 | -3413.49451 | 0.05890 | 0.05578 | -3413.55341 | -3413.55030 |
| **PTASeBn_a** | -3413.86483 | -3411.54759 | 0.31461 | -3413.53359 | 0.05756 | 0.05539 | -3413.59115 | -3413.58898 |
| **PTASeBn_b** | -3413.85940 | -3411.54305 | 0.31439 | -3413.52821 | 0.05971 | 0.05641 | -3413.58791 | -3413.58462 |
| **PTASeBn_c** | -3413.85975 | -3411.54341 | 0.31451 | -3413.52853 | 0.05871 | 0.05591 | -3413.58724 | -3413.58444 |
| **SePTA-rad** | -3142.85943 | -3140.59180 | 0.19406 | -3142.65507 | 0.04295 | 0.04288 | -3142.69802 | -3142.69796 |
| **SePTA** | -3143.05419 | -3140.78595 | 0.19448 | -3142.84977 | 0.04062 | 0.04055 | -3142.89040 | -3142.89033 |

### *A.2.2 Reaction Pathways with $(BnSe)_2$ Model Substrate*

**Table A.3.** Summarized reaction pathways for the simple model substrate.

| | **RXN: Initial Ir eq (kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
|---|---|---|---|---|---|---|---|---|---|
| **o** | Ir3+PTA+diSeBn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | Ir3-T1+PTA+diSeBn | 60.9 | 61.35 | -2.52 | 58.85 | 2.88 | 2.29 | 55.97 | 56.56 |
| | | | | | | | | | |
| | **RXN: Anionic Main Pathway** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | Ir3+PTA+diSeBn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o** | Ir3+PTASeBn+SeBn-anion | 6.46 | 14.39 | 0.91 | 7.33 | 0.23 | 0.59 | 7.09 | 6.74 |
| **o** | Ir3+SeBn-Model-Int-II-anion | -3.42 | -9.91 | 1.26 | -1.53 | -7.78 | -9.92 | 6.25 | 8.39 |
| **o** | Ir3+SeBn-Model-TS-IV | 8.42 | 5.59 | 0.69 | 9.53 | -9.66 | -11.17 | 19.2 | 20.7 |
| **o** | Ir3+SeBn-Model-Int-III-anion | -24.08 | -27.45 | 1.72 | -21.85 | -8.45 | -10.48 | -13.4 | -11.37 |
| **o** | Ir3+BnSeBn+SePTA | -16.06 | -9.84 | 0.85 | -15.33 | 0.71 | 0.79 | -16.04 | -16.12 |
| | | | | | | | | | |
| | **RXN: Anionic Competing B pathway (kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | Ir3+PTA+diSeBn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | Ir3+SeBn-Model-Int-III-B-anion | 16.72 | 6.67 | 1.47 | 18.95 | -7.4 | -9.58 | 26.35 | 28.53 |
| | | | | | | | | | |
| | **RXN: Anionic Competitive C pathway (kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | Ir3+PTA+diSeBn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | Ir3+SeBn-Model-TS-VIII | -1.14 | -8.72 | 1.55 | 0.43 | -10.84 | -12.22 | 11.27 | 12.65 |
| **o** | Ir3+SeBn-Model-Reagents_complex | -9.07 | -17.2 | 0.81 | -7.64 | -9.63 | -11.29 | 1.99 | 3.65 |
| | | | | | | | | | |
| | **RXN: Anionic Competitive E and F pathways (kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | Ir3+PTA+diSeBn+diSeBn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o** | Ir3+PTASeBn+SeBn-Model-TS-II-E-anion | -0.59 | -9.38 | 1.12 | 0.78 | -10.98 | -12.24 | 11.75 | 13.02 |
| **o** | Ir3+PTASeBn+SeBn-Model-TS-II-F-anion | 19.07 | 12.28 | 1.02 | 20.4 | -10.73 | -11.44 | 31.13 | 31.84 |
| | | | | | | | | | |
| | **RXN: Anionic Competitive G pathway (kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | Ir3+PTA+diSeBn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | Ir3+Bn-cation+SePTA+SeBn-anion | 49.27 | 63.36 | -0.85 | 48.35 | 9.73 | 11.4 | 38.62 | 36.94 |
| | | | | | | | | | |
| | **RXN: Radical Main Pathway (kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | Ir3+PTA+diSeBn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | Ir2+PTA-rad+diSeBn | 55.25 | 55.76 | -3.01 | 52.86 | 5.9 | 4.27 | 46.97 | 48.59 |
| **o** | Ir2+SeBn-Model-Int-I-rad | 38.67 | 31.81 | -1.21 | 38.33 | -5.64 | -8.5 | 43.97 | 46.83 |
| **o** | Ir2+SeBn-Model-TS-I | 39.84 | 34.82 | -1.29 | 39.06 | -5.66 | -8.9 | 44.72 | 47.96 |
| **o** | Ir2+SeBn-Model-Int-II-rad | 35.65 | 26.28 | -1.18 | 35.41 | -5.34 | -8.31 | 40.75 | 43.71 |
| **o** | Ir2+PTASeBn+SeBn-rad | 49.31 | 57.44 | -2.28 | 47.42 | 5.13 | 3.82 | 42.29 | 43.6 |
| **o** | Ir2+SeBn-Model-TS-VII | 57.21 | 50.35 | -1.62 | 56.2 | -6.31 | -8.7 | 62.51 | 64.9 |
| **o** | Ir2+SeBn-Model-Int-III-rad | 23.13 | 13.06 | -1.49 | 22.6 | -4.45 | -7.7 | 27.05 | 30.3 |
| **o** | Ir2+BnSeBn+SePTA-rad | 42.98 | 49.71 | -2.17 | 41.29 | 6.55 | 5 | 34.74 | 36.28 |
| | | | | | | | | | |

| | RXN: Radical Comp. B and D pathways (kcal/mol) | DE_SPC | DE | DZPE | DH_SPC | T.DS | T.qh-DS | DG(T)_SPC | qh-DG(T)_SPC |
|---|---|---|---|---|---|---|---|---|---|
| o | Ir3+PTA+diSeBn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o | Ir2+SeBn-Model-Int-III-B-rad | 52.5 | 51.63 | -3.78 | 49.71 | -5.21 | -8.14 | 54.92 | 57.85 |
| o | Ir2+SeBn-Model-Int-III-D-rad | 44.24 | 31.74 | -1.32 | 43.72 | -6.79 | -8.84 | 50.5 | 52.55 |
| | | | | | | | | | |
| | **RXN: Radical Comp. E and F pathways (kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| o | Ir3+PTA+diSeBn+diSeBn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o | Ir2+PTASeBn+SeBn-Model-TS-II-E-rad | 42.07 | 36.02 | -1.34 | 41.49 | -4.2 | -7.98 | 45.69 | 49.48 |
| o | Ir2+PTASeBn+SeBn-Model-TS-II-F-rad | 53.71 | 44.58 | -1.66 | 52.72 | -6.38 | -8.83 | 59.11 | 61.55 |
| | | | | | | | | | |
| | **RXN: Radical PTASeBn-Nbond Ints. (kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| o | Ir3+PTA+diSeBn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o | Ir2+PTASeBn-Nbond+SeBn-rad | 72.83 | 72.92 | -1.87 | 71.21 | 5.85 | 4.43 | 65.36 | 66.78 |
| | | | | | | | | | |
| | **RXN: Radical SET to PTASeBn (kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| o | Ir3+PTA+diSeBn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o | Ir3+PTASe-Bn-rad+SeBn-rad | 36.64 | 46.99 | -2.57 | 34.89 | 4.92 | 3.92 | 29.98 | 30.97 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o** | Ir3+SePTA+SeBn-rad+Bn-rad | 42.17 | 56.36 | -2.9 | 39.27 | 10.55 | 12.21 | 28.71 | 27.06 |
| **o** | Ir3+PTA-SeBn-rad+SeBn-rad | 46.96 | 54.01 | -0.75 | 46.54 | 3.38 | 2.69 | 43.16 | 43.85 |
| **o** | Ir3+PTA+SeBn-rad+SeBn-rad | 64.78 | 77.27 | -2.04 | 62.5 | 11.19 | 12.64 | 51.31 | 49.86 |

## *A.3 Summarized Thermochemical Data: Chapter 3*

### *A.3.1 Raw Absolute Energy Values for Bismacycle system: 2-pyridone*

**Table A.4.** Summarized raw absolute energy values for the bismacycle system for 2-pyridione using def2-SVP for bismuth.

| Structure | E_SPC | E | ZPE | H_SPC | T.S | T.qh-S | G(T)_SPC | qh-G(T)_SPC | Imag Freq |
|---|---|---|---|---|---|---|---|---|---|
| **Int-Bz-N-bound-1_a_stable** | -2200.36516 | -2199.66347 | 0.45177 | -2199.86666 | 0.13289 | 0.12134 | -2199.99955 | -2199.98800 | |
| **Int-Bz-N-bound-1_b** | -2200.36477 | -2199.66307 | 0.45206 | -2199.86614 | 0.13136 | 0.12065 | -2199.99750 | -2199.98679 | |
| **Int-Bz-N-bound-1_c** | -2200.36393 | -2199.66198 | 0.45167 | -2199.86556 | 0.13228 | 0.12111 | -2199.99784 | -2199.98667 | |
| **Int-Bz-N-bound-3_b** | -2200.35863 | -2199.65664 | 0.45096 | -2199.86064 | 0.13423 | 0.12195 | -2199.99487 | -2199.98259 | |
| **Int-Bz-N-bound-3_d** | -2200.35876 | -2199.65679 | 0.45130 | -2199.86066 | 0.13269 | 0.12118 | -2199.99336 | -2199.98184 | |
| **Int-Bz-N-bound-3_f** | -2200.35951 | -2199.65747 | 0.45111 | -2199.86149 | 0.13436 | 0.12220 | -2199.99585 | -2199.98369 | |
| **Int-Bz-O-bound-1_a_stable** | -2200.36470 | -2199.66221 | 0.45177 | -2199.86633 | 0.13221 | 0.12077 | -2199.99854 | -2199.98710 | |
| **Int-Bz-O-bound-1_c** | -2200.35905 | -2199.65626 | 0.45080 | -2199.86123 | 0.13661 | 0.12300 | -2199.99784 | -2199.98423 | |
| **Int-Bz-O-bound-1_d** | -2200.36214 | -2199.65904 | 0.45131 | -2199.86397 | 0.13440 | 0.12193 | -2199.99837 | -2199.98590 | |
| **Int-Bz-O-bound-2_d** | -2200.35951 | -2199.65686 | 0.45158 | -2199.86135 | 0.13171 | 0.12051 | -2199.99305 | -2199.98186 | |
| **Int-Bz-O-bound-3_d** | -2200.36324 | -2199.66129 | 0.45120 | -2199.86533 | 0.13178 | 0.12066 | -2199.99711 | -2199.98599 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Int-Bz-O-bound-3_e** | -2200.36398 | -2199.66179 | 0.45154 | -2199.86583 | 0.13155 | 0.12060 | -2199.99738 | -2199.98642 | |
| **Int-II-Bz-CN** | -2200.46073 | -2199.76543 | 0.45296 | -2199.96125 | 0.13340 | 0.12122 | -2200.09465 | -2200.08247 | |
| **Int-II-Bz-CO** | -2200.44113 | -2199.74476 | 0.45277 | -2199.94188 | 0.13385 | 0.12126 | -2200.07574 | -2200.06314 | |
| **Product-Bi-OBz** | -1645.81344 | -1645.32126 | 0.27639 | -1645.50655 | 0.09495 | 0.08943 | -1645.60150 | -1645.59599 | |
| **Product-CN-pyridin-2-one** | -554.63020 | -554.43083 | 0.17545 | -554.44006 | 0.05753 | 0.05527 | -554.49759 | -554.49533 | |
| **Product-CO-phenoxypyridine** | -554.61472 | -554.41476 | 0.17515 | -554.42495 | 0.05760 | 0.05502 | -554.48255 | -554.47997 | |
| **TS-3atom-Bz-CN-N-bound_a-stable** | -2200.31473 | -2199.61640 | 0.44907 | -2199.81920 | 0.13395 | 0.12179 | -2199.95315 | -2199.94099 | -454.93 |
| TS-3atom-Bz-CN-N-bound_i | -2200.30045 | -2199.60566 | 0.44804 | -2199.80567 | 0.13245 | 0.12124 | -2199.93812 | -2199.92691 | -235.78 |
| TS-3atom-Bz-CO-O-bound_a-stable | -2200.31686 | -2199.61858 | 0.44879 | -2199.82150 | 0.13400 | 0.12181 | -2199.95550 | -2199.94331 | -340.59 |
| TS-3atom-Bz-CO-O-bound_b | -2200.30822 | -2199.60962 | 0.44827 | -2199.81282 | 0.14061 | 0.12494 | -2199.95343 | -2199.93776 | -393.38 |
| TS-3atom-Bz-CO-O-bound_c | -2200.30815 | -2199.61015 | 0.44834 | -2199.81300 | 0.13550 | 0.12248 | -2199.94850 | -2199.93548 | -353.62 |
| TS-3atom-Bz-CO-O-bound_d | -2200.30809 | -2199.60893 | 0.44881 | -2199.81263 | 0.13534 | 0.12218 | -2199.94797 | -2199.93481 | -434.85 |
| TS-3atom-Bz-CO-O-bound_i | -2200.31660 | -2199.61763 | 0.44934 | -2199.82085 | 0.13327 | 0.12132 | -2199.95412 | -2199.94217 | -373.26 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| TS-5atom-Bz-CN-O-bound_a-stable | -2200.31763 | -2199.61892 | 0.44931 | -2199.82184 | 0.13517 | 0.12231 | -2199.95701 | -2199.94415 | -442.6 |
| TS-5atom-Bz-CN-O-bound_b | -2200.31483 | -2199.61638 | 0.44899 | -2199.81941 | 0.13310 | 0.12132 | -2199.95251 | -2199.94073 | -439.26 |
| TS-5atom-Bz-CN-O-bound_c | -2200.31394 | -2199.61520 | 0.44919 | -2199.81840 | 0.13234 | 0.12092 | -2199.95074 | -2199.93933 | -420.95 |
| TS-5atom-Bz-CN-O-bound_d | -2200.31382 | -2199.61622 | 0.44882 | -2199.81855 | 0.13209 | 0.12100 | -2199.95064 | -2199.93955 | -448.48 |
| TS-5atom-Bz-CO-N-bound_a-stable | -2200.32266 | -2199.62499 | 0.44936 | -2199.82683 | 0.13245 | 0.12104 | -2199.95928 | -2199.94788 | -413.84 |
| TS-5atom-Bz-CO-N-bound_b | -2200.32270 | -2199.62502 | 0.44948 | -2199.82680 | 0.13228 | 0.12097 | -2199.95908 | -2199.94777 | -413.53 |
| TS-5atom-Bz-CO-N-bound_c | -2200.32168 | -2199.62409 | 0.44904 | -2199.82607 | 0.13363 | 0.12160 | -2199.95970 | -2199.94767 | -434.07 |
| TS-5atom-Bz-CO-N-bound_d | -2200.31382 | -2199.61592 | 0.44944 | -2199.81804 | 0.13235 | 0.12089 | -2199.95039 | -2199.93893 | -399.87 |
| TS-5atom-Bz-CO-N-bound_e | -2200.31795 | -2199.62099 | 0.44931 | -2199.82216 | 0.13339 | 0.12158 | -2199.95555 | -2199.94374 | -427.95 |

**Table A.5.** Summarized raw absolute energy values for the bismacycle system for 2-pyridione using def2-TZVP for bismuth.

| Structure | E_SPC | E | ZPE | H_SPC | T.S | T.qh-S | G(T)_SPC | qh-G(T)_SPC | Imag Freq |
|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Int-Bz-N-bound-1_a_def2TZVP** | -2200.3656 | -2199.6819 | 0.451738 | -2199.8672 | 0.132444 | 0.121139 | -2199.9996 | -2199.9883 | |
| **Int-Bz-O-bound-1_a_def2TZVP** | -2200.3652 | -2199.681 | 0.451921 | -2199.8668 | 0.131769 | 0.120545 | -2199.9985 | -2199.9873 | |
| **Product-Bi-OBz_def2TZVP** | -1645.8139 | -1645.3348 | 0.276296 | -1645.5071 | 0.094999 | 0.089481 | -1645.6021 | -1645.5966 | |
| **Product-CN-pyridin-2-one** | -554.6302 | -554.43083 | 0.175446 | -554.44006 | 0.057533 | 0.055273 | -554.49759 | -554.49533 | |
| **Product-CO-phenoxypyridine** | -554.61472 | -554.41476 | 0.175151 | -554.42495 | 0.057604 | 0.055021 | -554.48255 | -554.47997 | |
| **TS-3atom-Bz-CN-Nbound_a_def2TZVP** | -2200.3008 | -2199.6195 | 0.448137 | -2199.806 | 0.13153 | 0.120827 | -2199.9375 | -2199.9268 | -237.58 |
| TS-3atom-Bz-CO-O-bound_a_def2TZVP | -2200.3166 | -2199.6337 | 0.449095 | -2199.8211 | 0.136571 | 0.122936 | -2199.9577 | -2199.944 | -324.69 |
| TS-5atom-Bz-CN-O-bound_a_def2TZVP | -2200.3186 | -2199.6347 | 0.449161 | -2199.8229 | 0.133938 | 0.121694 | -2199.9569 | -2199.9446 | -465.29 |
| TS-5atom-Bz-CO-N-bound_a_def2TZVP | -2200.3229 | -2199.6397 | 0.449496 | -2199.827 | 0.131404 | 0.120567 | -2199.9584 | -2199.9476 | -437.43 |
| TS-Bz-N-to-O_a_def2TZVP | -2200.3605 | -2199.6764 | 0.451644 | -2199.8632 | 0.129584 | 0.118901 | -2199.9928 | -2199.9821 | -36.72 |

*A.3.2 Raw Absolute Energy Values for Bismacycle system: 4-pyridone with def2-SVP (Bi)*

**Table A.6.** Summarized raw absolute energy values for the bismacycle system for 4-pyridione using def2-SVP for bismuth.

| Structure | E_SPC | E | ZPE | H_SPC | T.S | T.qh-S | G(T)_SPC | qh-G(T)_SPC | Imag Freq |
|---|---|---|---|---|---|---|---|---|---|
| **Int-Bz-N-bound-4-pyridone_a** | -2200.3491 | -2199.6472 | 0.451647 | -2199.8506 | 0.133042 | 0.121509 | -2199.9836 | -2199.9721 | |
| **Int-Bz-O-bound-4-pyridone_a** | -2200.351 | -2199.648 | 0.451604 | -2199.8526 | 0.134298 | 0.121925 | -2199.9869 | -2199.9745 | |
| **Product-Bi-OBz** | -1645.8134 | -1645.3213 | 0.276389 | -1645.5066 | 0.094949 | 0.089434 | -1645.6015 | -1645.596 | |
| **Product-CN-pyridin-4-one** | -554.62354 | -554.42406 | 0.17577 | -554.43319 | 0.055814 | 0.054461 | -554.489 | -554.48765 | |
| **Product-CO-4-phenoxypyridine** | -554.609 | -554.40932 | 0.175353 | -554.41906 | 0.057357 | 0.054906 | -554.47642 | -554.47397 | |
| **TS-3atom-Bz-CN-N-bound_4-pyridone** | -2200.2842 | -2199.5875 | 0.447706 | -2199.7893 | 0.137077 | 0.123912 | -2199.9264 | -2199.9132 | -383.97 |
| **TS-3atom-Bz-CO-O-bound_4-pyridone** | -2200.3068 | -2199.6085 | 0.448967 | -2199.8111 | 0.134828 | 0.122059 | -2199.9459 | -2199.9332 | -377.96 |

*A.3.3 Reaction Pathways for 2- and 4-pyridone with bismacycle*

**Table A.7.** Summarized reaction pathways for the bismacycle system for 2- and 4-pyridiones using def2-SVP and def2-TZVP for bismuth.

| | RXN: 2-Pyridone (def2-SVP) | DE_SPC | DE | DZPE | DH_SPC | T.DS | T.qh-DS | DG(T)_SPC | qh-DG(T)_SPC |
|---|---|---|---|---|---|---|---|---|---|
| **o** | ProdCN+BiOBz | -49.51 | -55.91 | 0.02 | -50.43 | 11.82 | 14.04 | -62.24 | -64.47 |
| **o** | IntII-BZ-CN | -60.24 | -64.28 | 0.73 | -59.62 | -0.16 | -0.69 | -59.46 | -58.92 |
| **o** | TS-BZ-CN | 29.7 | 27.78 | -1.57 | 27.99 | 1.07 | 0.16 | 26.92 | 27.83 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o** | Int-BZ-O | 0.67 | 1.1 | -0.17 | 0.48 | 0.26 | 0.15 | 0.22 | 0.33 |
| **o** | TS-NtoO | 3.2 | 3.86 | -0.23 | 2.39 | -1.71 | -1.58 | 4.1 | 3.97 |
| **o** | Int-BZ-N | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | TS-BZ-CO | 26.66 | 24.08 | -1.57 | 24.96 | 0.3 | 0.15 | 24.66 | 24.81 |
| **o** | IntII-BZ-CO | -47.94 | -51.31 | 0.61 | -47.46 | 0.13 | -0.67 | -47.59 | -46.79 |
| **o** | ProdCO+BiOBz | -39.79 | -45.83 | -0.16 | -40.95 | 11.86 | 13.88 | -52.81 | -54.83 |
| | | | | | | | | | |
| | **RXN: 2-Pyridone (def2-TZVP)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | ProdCN+BiOBz | -49.27 | -52.55 | 0 | -50.19 | 12.61 | 14.82 | -62.79 | -65.01 |
| **o** | TS-BZ-CN | 29.53 | 29.66 | -1.62 | 27.79 | 0.94 | 0.35 | 26.85 | 27.44 |
| **o** | Int-BZ-O | 0.24 | 0.59 | 0.11 | 0.27 | -0.42 | -0.37 | 0.7 | 0.65 |
| **o** | TS-NtoO | 3.19 | 3.45 | -0.06 | 2.51 | -1.79 | -1.4 | 4.3 | 3.91 |
| **o** | Int-BZ-N | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | TS-BZ-CO | 26.82 | 26.48 | -1.41 | 25.22 | -0.65 | -0.36 | 25.87 | 25.58 |
| **o** | ProdCO+BiOBz | -39.56 | -42.47 | -0.18 | -40.7 | 12.65 | 14.66 | -53.35 | -55.37 |
| | | | | | | | | | |
| | **RXN: 4-Pyridone (def2-SVP)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | ProdCN+BiOBz | -53.96 | -61.1 | 0.35 | -54.67 | 10.33 | 13.79 | -65.01 | -68.46 |
| **o** | TS3-BZ-CN | 41.92 | 37.93 | -2.45 | 39.73 | 1.74 | 1.25 | 37.98 | 38.48 |
| **o** | Int-BZ-O | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | Int-BZ-N | 1.18 | 0.48 | 0.03 | 1.26 | -0.79 | -0.26 | 2.05 | 1.52 |
| **o** | TS3-BZ-CO | 27.71 | 24.78 | -1.65 | 26.04 | 0.33 | 0.08 | 25.71 | 25.96 |
| **o** | ProdCO+BiOBz | -44.84 | -51.85 | 0.09 | -45.81 | 11.3 | 14.07 | -57.11 | -59.88 |

### *A.3.4 Raw Absolute Energy Values for $Ph_3BiCl_2$ system*

**Table A.8.** Summarized raw absolute energy values for the triphenyl-organobismuth system for using def2-SVP and def2-TZVP for bismuth.

| Structure: **def2-SVP** | **E_SPC** | **E** | **ZPE** | **H_SPC** | **T.S** | **T.qh-S** | **G(T)_SPC** | **qh-G(T)_SPC** | **Imag Freq** |
|---|---|---|---|---|---|---|---|---|---|
| **Int-Ph3-2Py-NN_a_THF** | -1555.573 | -1555.0636 | 0.439595 | -1555.0956 | 0.113867 | 0.103022 | -1555.2094 | -1555.1986 | |
| **Int-Ph3-2Py-NN_b_THF** | -1555.5816 | -1555.0722 | 0.440317 | -1555.1038 | 0.109878 | 0.100877 | -1555.2137 | -1555.2047 | |
| **Int-Ph3-2Py-NO_a_THF** | -1555.5795 | -1555.0694 | 0.439964 | -1555.102 | 0.111463 | 0.10139 | -1555.2134 | -1555.2033 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Int-Ph3-2Py-NO_b_THF** | -1555.5831 | -1555.0732 | 0.44033 | -1555.1053 | 0.11021 | 0.100758 | -1555.2155 | -1555.2061 | |
| **Int-Ph3-2Py-OO_a_THF** | -1555.5827 | -1555.0715 | 0.43988 | -1555.1052 | 0.113684 | 0.102158 | -1555.2189 | -1555.2073 | |
| **Int-Ph3-2Py-OO_b_THF** | -1555.585 | -1555.0748 | 0.44001 | -1555.1075 | 0.111926 | 0.101421 | -1555.2194 | -1555.2089 | |
| **Product-Bi-Ph3-2Py-Nbound_THF** | -1001.0117 | -1000.7101 | 0.264897 | -1000.7237 | 0.077962 | 0.072858 | -1000.8017 | -1000.7966 | |
| **Product-Bi-Ph3-2Py-Obound_THF** | -1001.0097 | -1000.7072 | 0.264844 | -1000.7218 | 0.079422 | 0.073415 | -1000.8012 | -1000.7952 | |
| **Product-CN-pyridin-2-one_THF** | -554.62781 | -554.42812 | 0.175505 | -554.43913 | 0.052189 | 0.050484 | -554.49132 | -554.48961 | |
| **Product-CO-phenoxypyridine_THF** | -554.61295 | -554.41269 | 0.175183 | -554.42461 | 0.053217 | 0.050729 | -554.47783 | -554.47534 | |
| **TSCN-3atom-Ph3-2Py-CN-NN-bound_THF** | -1555.5335 | -1555.0288 | 0.438843 | -1555.0579 | 0.108266 | 0.099627 | -1555.1662 | -1555.1576 | -179.92 |
| TSCN-3atom-Ph3-2Py-CN-NO-bound_THF | -1555.5323 | -1555.0265 | 0.438764 | -1555.0566 | 0.110107 | 0.100395 | -1555.1667 | -1555.157 | -201.35 |
| TSCN-5atom-Ph3-2Py-CN-NO-bound_THF | -1555.5361 | -1555.0289 | 0.438134 | -1555.0608 | 0.110542 | 0.10087 | -1555.1713 | -1555.1617 | -249.84 |
| TSCN-5atom-Ph3-2Py-CN-OO-bound_THF | -1555.5379 | -1555.0302 | 0.438075 | -1555.0628 | 0.112076 | 0.101325 | -1555.1748 | -1555.1641 | -235.81 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| TSCO-3atom-Ph3-2Py-CO-NO-bound_THF | -1555.5261 | -1555.0187 | 0.437809 | -1555.0508 | 0.112816 | 0.101974 | -1555.1637 | -1555.1528 | -197.65 |
| TSCO-3atom-Ph3-2Py-CO-OO-bound_THF | -1555.5277 | -1555.0204 | 0.437085 | -1555.0531 | 0.11532 | 0.102946 | -1555.1684 | -1555.156 | -187.19 |
| TSCO-5atom-Ph3-2Py-CO-NN-bound_THF | -1555.5341 | -1555.0266 | 0.438024 | -1555.0588 | 0.110324 | 0.10089 | -1555.1692 | -1555.1597 | -263.31 |
| TSCO-5atom-Ph3-2Py-CO-NO-bound_THF | -1555.5351 | -1555.0285 | 0.438096 | -1555.0599 | 0.110814 | 0.100778 | -1555.1707 | -1555.1607 | -471.87 |

| **Structure: def2-TZVP** | **E_SPC** | **E** | **ZPE** | **H_SPC** | **T.S** | **T.qh-S** | **G(T)_SPC** | **qh-G(T)_SPC** | **Imag Freq** |
|---|---|---|---|---|---|---|---|---|---|
| **Int-Ph3-2Py-NN_b_stable_THF_def2TZVP** | -1555.5821 | -1555.0898 | 0.440592 | -1555.1042 | 0.109233 | 0.100522 | -1555.2134 | -1555.2047 | |
| **Int-Ph3-2Py-NO_b_stable_THF_def2TZVP** | -1555.5835 | -1555.091 | 0.440275 | -1555.1058 | 0.110782 | 0.101024 | -1555.2166 | -1555.2068 | |
| **Int-Ph3-2Py-OO_b_stable_THF_def2TZVP** | -1555.5855 | -1555.0931 | 0.440359 | -1555.1077 | 0.110839 | 0.10085 | -1555.2186 | -1555.2086 | |
| **Product-Bi-Ph3-2Py-Nbound_THF_def2TZVP** | -1001.0122 | -1000.7237 | 0.264894 | -1000.7242 | 0.077984 | 0.072864 | -1000.8022 | -1000.797 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Product-Bi-Ph3-2Py-Obound_THF_def2TZVP** | -1001.0103 | -1000.7215 | 0.264864 | -1000.7224 | 0.079312 | 0.073326 | -1000.8017 | -1000.7957 | |
| **Product-CN-pyridin-2-one_THF** | -554.62781 | -554.42812 | 0.175505 | -554.43913 | 0.052189 | 0.050484 | -554.49132 | -554.48961 | |
| **Product-CO-phenoxypyridine_THF** | -554.61295 | -554.41269 | 0.175183 | -554.42461 | 0.053217 | 0.050729 | -554.47783 | -554.47534 | |
| **TSCN-5atom-Ph3-2Py-CN-OO-bound_THF_def2TZVP** | -1555.5366 | -1555.0439 | 0.437652 | -1555.0618 | 0.113884 | 0.102188 | -1555.1756 | -1555.1639 | -454.53 |
| TSCO-5atom-Ph3-2Py-CO-NO-bound_THF_def2TZVP | -1555.5362 | -1555.0438 | 0.438026 | -1555.061 | 0.111522 | 0.101136 | -1555.1725 | -1555.1621 | -471.17 |

### *A.3.5. Reaction Pathways for 2-pyridine for $Ph_3BiCl_2$ system*

**Table A.9.** Summarized reaction pathways for the triphenyl-organobismuth system for using def2-SVP and def2-TZVP for bismuth.

| | **RXN: 2-Pyridone (def2-SVP)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
|---|---|---|---|---|---|---|---|---|---|
| **o** | Prod_CN+Bi2Py | -34.24 | -39.81 | 0.25 | -34.77 | 11.45 | 13.77 | -46.22 | -48.54 |
| **o** | TS-CN | 29.39 | 27.7 | -1.2 | 27.89 | -0.32 | -0.29 | 28.21 | 28.18 |
| **o** | Int-O | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | Int-N | 1.24 | 0.89 | 0.2 | 1.38 | -1.13 | -0.32 | 2.51 | 1.7 |
| **o** | TS-CO | 31.19 | 29.04 | -1.2 | 29.76 | -0.92 | -0.39 | 30.68 | 30.16 |
| **o** | Prod_CO+Bi2Py | -24.91 | -30.13 | 0.05 | -25.66 | 12.09 | 13.92 | -37.76 | -39.58 |
| | | | | | | | | | |
| | **RXN: 2-Pyridone (def2-TZVP)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | Prod_CN+Bi2Py | -33.96 | -36.52 | 0.02 | -34.63 | 12.67 | 14.53 | -47.29 | -49.16 |
| **o** | TS-CN | 30.65 | 30.87 | -1.7 | 28.85 | 1.91 | 0.84 | 26.94 | 28.01 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o** | Int-O | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | Int-N | 1.32 | 1.38 | -0.03 | 1.34 | 0.09 | 0.31 | 1.25 | 1.02 |
| **o** | TS-CO | 30.92 | 30.91 | -1.46 | 29.34 | 0.43 | 0.18 | 28.91 | 29.16 |
| **o** | Prod_CO+Bi2Py | -24.63 | -26.85 | -0.18 | -25.51 | 13.31 | 14.68 | -38.82 | -40.2 |

## *A.4 Summarized Thermochemical Data: Chapter 4*

### *A.4.1 Raw Absolute Energy Values*

**Table A.10.** Summarized raw absolute energy values using ωB97XD/def2-TZVP (Bi,Pd) during optimization with ligand (R=Me).

| Structure | E_SPC | E | ZPE | H_SPC | T.S | T.qh-S | G(T)_SPC | qh-G(T)_SPC | Imag V |
|---|---|---|---|---|---|---|---|---|---|
| **o Br_TZVP_stable** | -2574.435 | -2571.9064 | 0 | -2574.4327 | 0.018536 | 0.018536 | -2574.4512 | -2574.4512 | |
| **o INT-L-solv_TZVP_d_stable_model** | -1323.1431 | -1322.7867 | 0.419784 | -1322.6954 | 0.086197 | 0.080889 | -1322.7816 | -1322.7763 | |
| **o INT-L2_cis-solv_TZVP_a_model** | -2325.1211 | -2324.4918 | 0.755501 | -2324.3178 | 0.127243 | 0.118425 | -2324.445 | -2324.4362 | |
| **o INT-L2_trans-solv_TZVP_b_stable_model** | -2325.1242 | -2324.4941 | 0.754102 | -2324.3208 | 0.13461 | 0.123089 | -2324.4554 | -2324.4439 | |
| **o INT-L_TZVP_model** | -1129.9195 | -1129.6389 | 0.333972 | -1129.5636 | 0.073379 | 0.069417 | -1129.637 | -1129.6331 | |
| **o INT1-PhBr_TZVP_c_stable_model** | -3935.8871 | -3932.6919 | 0.425654 | -3935.4319 | 0.093738 | 0.085943 | -3935.5256 | -3935.5178 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o INT1-TM-step-Pdcp_TZVP-t_stable_model** | -4903.398 | -4899.9427 | 0.764381 | -4902.5841 | 0.134535 | 0.123499 | -4902.7187 | -4902.7076 | |
| **o INT2-OA_cis_coord_TZVP_b_model** | -3935.916 | -3932.7217 | 0.425319 | -3935.4609 | 0.093417 | 0.086476 | -3935.5543 | -3935.5474 | |
| **o INT2-OA_cis_open_TZVP_a_model** | -3935.9271 | -3932.7306 | 0.427205 | -3935.4708 | 0.089511 | 0.08384 | -3935.5603 | -3935.5547 | |
| **o INT2-OA_cis_open_TZVP_b** | -3935.8946 | -3932.7003 | 0.426837 | -3935.4387 | 0.090833 | 0.084349 | -3935.5295 | -3935.523 | |
| **o INT2-OA_trans_coord_TZVP_a_stable_model** | -3935.9336 | -3932.7389 | 0.426259 | -3935.4777 | 0.09116 | 0.085238 | -3935.5689 | -3935.563 | |
| **o INT2-OA_trans_open_TZVP_a_model** | -3935.9294 | -3932.7341 | 0.426998 | -3935.4731 | 0.091081 | 0.0845 | -3935.5642 | -3935.5576 | |
| **o INT3-TM-step_cisBr-transPdcp_TZVP_a_stable_model** | -4903.4057 | -4899.9533 | 0.763691 | -4902.5923 | 0.136602 | 0.124393 | -4902.7289 | -4902.7167 | |
| **o INT3-TM-step_transBr-cisPdcp_TZVP_model** | -4903.3846 | -4899.9288 | 0.765891 | -4902.5696 | 0.13273 | 0.122366 | -4902.7023 | -4902.6919 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o INT4-RE_cis-coord_TZVP_a_model** | -1478.8963 | -1478.4931 | 0.496661 | -1478.3683 | 0.093324 | 0.08764 | -1478.4617 | -1478.456 | |
| **o INT4-RE_cis-open_TZVP_b_model** | -1478.8948 | -1478.4919 | 0.496678 | -1478.3669 | 0.09349 | 0.08742 | -1478.4604 | -1478.4543 | |
| **o INT4-RE_trans-coord_TZVP_d_stable_model** | -1478.8982 | -1478.4946 | 0.496295 | -1478.3706 | 0.093105 | 0.087487 | -1478.4637 | -1478.4581 | |
| **o INT4-RE_trans-open_TZVP_a_stable_model** | -1478.8934 | -1478.4905 | 0.495845 | -1478.366 | 0.094646 | 0.088466 | -1478.4607 | -1478.4545 | |
| **o INT5-RE-coord_TZVP_b_stable_model** | -1478.9433 | -1478.5407 | 0.498945 | -1478.4132 | 0.094328 | 0.087641 | -1478.5075 | -1478.5008 | |
| **o INT7-TM-cis-mig_TZVP_c_stable_model** | -4903.4179 | -4899.9623 | 0.764935 | -4902.6036 | 0.134801 | 0.12345 | -4902.7384 | -4902.727 | |
| **o L_TZVP_a_stable_model** | -1001.9342 | -1001.6699 | 0.332775 | -1001.5814 | 0.067128 | 0.064338 | -1001.6485 | -1001.6457 | |
| **o PROD-Bi_Br_TZVP_stable_model** | -3424.4854 | -3421.4295 | 0.265524 | -3424.2022 | 0.063834 | 0.061854 | -3424.266 | -3424.2641 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o PROD-Ph_cp_TZVP_stable_model** | -348.99037 | -348.8709 | 0.164158 | -348.81788 | 0.040697 | 0.040126 | -348.85858 | -348.858 | |
| **o RXT-Bi_cp_TZVP_stable_model** | -967.45126 | -967.18561 | 0.334964 | -967.09643 | 0.068223 | 0.065546 | -967.16466 | -967.16198 | |
| **o RXT-Br-Ph_stable_model** | -2805.9394 | -2803.3418 | 0.091871 | -2805.8411 | 0.036521 | 0.036528 | -2805.8776 | -2805.8776 | |
| **o TS-OA-L2_cisBr_TZVP_a_stable_model** | -4937.8454 | -4934.3802 | 0.760253 | -4937.036 | 0.132839 | 0.122225 | -4937.1689 | -4937.1582 | -204.91 |
| o TS-OA_cis_coord_TZVP_b_stable_model_opt | -3935.88 | -3932.6851 | 0.425583 | -3935.4258 | 0.088391 | 0.083235 | -3935.5141 | -3935.509 | -173.38 |
| o TS-OA_cis_open_TZVP_a | -3935.8781 | -3932.6836 | 0.425213 | -3935.4239 | 0.089117 | 0.083792 | -3935.513 | -3935.5077 | -172.8 |
| o TS-OA_trans_coord_TZVP_a | -3935.8736 | -3932.6801 | 0.424984 | -3935.4202 | 0.088772 | 0.083145 | -3935.509 | -3935.5034 | -157.54 |
| o TS-OA_trans_open_TZVP_a_model | -3935.8762 | -3932.681 | 0.424708 | -3935.4221 | 0.0917 | 0.084949 | -3935.5138 | -3935.5071 | -171.55 |
| o TS-RE_cis-coord_TZVP_b_model | -1478.8716 | -1478.4691 | 0.496507 | -1478.3446 | 0.090801 | 0.08564 | -1478.4354 | -1478.4303 | -334.92 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| o TS-RE_cis-open_TZVP_d_model | -1478.8688 | -1478.4664 | 0.495823 | -1478.342 | 0.093114 | 0.087134 | -1478.4351 | -1478.4291 | -326.51 |
| o TS-RE_trans-coord_TZVP_a_model | -1478.8751 | -1478.4721 | 0.495831 | -1478.3485 | 0.093466 | 0.08702 | -1478.442 | -1478.4355 | -341.67 |
| o TS-RE_trans-open_TZVP_e_stable_model | -1478.8766 | -1478.4737 | 0.495209 | -1478.3504 | 0.093738 | 0.087455 | -1478.4441 | -1478.4378 | -338.68 |
| o TS1-TM-mig_TZVP_c_stable | -4903.3804 | -4899.9244 | 0.763367 | -4902.5677 | 0.136646 | 0.124037 | -4902.7043 | -4902.6917 | -58.99 |
| o TS1-TM-mig_TZVP_c_stable_model-REDO | -4903.3928 | -4899.9378 | 0.763973 | -4902.5798 | 0.133006 | 0.122267 | -4902.7128 | -4902.7021 | -40.62 |
| o TS1-TM-step_cisBr-cisPdcp_TZVP_model | -4903.3687 | -4899.9144 | 0.763855 | -4902.556 | 0.134128 | 0.122972 | -4902.6901 | -4902.679 | -94.9 |
| o TS1-TM-step_cisBr-transPdcp_TZVP-b_model | -4903.3887 | -4899.9343 | 0.76476 | -4902.5754 | 0.132367 | 0.121579 | -4902.7078 | -4902.697 | -82.47 |
| o TS1-TM-step_cisBr-transPdcp_TZVP_a_stable_model | -4903.391 | -4899.9369 | 0.764552 | -4902.578 | 0.131486 | 0.121142 | -4902.7095 | -4902.6992 | -93.02 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| o TS1-TM-step_transBr-cisPdcp_TZVP_model | -4903.3842 | -4899.9285 | 0.765474 | -4902.5704 | 0.130333 | 0.12074 | -4902.7007 | -4902.6911 | -52.37 |
| o TS2-TM-step_cisBr-BiX_TZVP_e | -4903.3994 | -4899.9473 | 0.762975 | -4902.5874 | 0.134537 | 0.123064 | -4902.7219 | -4902.7104 | -21.96 |
| o TS2-TM-step_cisBr-Pdcp_TZVP-1t_stable_model | -4903.4025 | -4899.9502 | 0.763404 | -4902.5901 | 0.134373 | 0.12273 | -4902.7245 | -4902.7129 | -26.24 |
| o acetone_TZVP_stable_model | -193.19382 | -193.1189 | 0.084038 | -193.10353 | 0.034208 | 0.033916 | -193.13773 | -193.13744 | |

**Table A.11.** Summarized raw absolute energy values using ωB97XD/def2-SVP (Bi,Pd) during optimization with ligand (R=H).

| Structure | E | ZPE | H | T.S | T.qh-S | G(T) | qh-G(T) | Imag V |
|---|---|---|---|---|---|---|---|---|
| **INT1-OA-coord_d_stable** | -3814.7289 | 0.343155 | -3814.3592 | 0.084599 | 0.078084 | -3814.4438 | -3814.4373 | |
| **INT1-TM-step-BiX_SVP_a_stable** | -4781.9546 | 0.680592 | -4781.2251 | 0.133986 | 0.121851 | -4781.359 | -4781.3469 | |
| **INT2-OA-L2_cp_SVP_b_stable** | -4698.5541 | 0.596829 | -4697.9144 | 0.118032 | 0.108586 | -4698.0324 | -4698.023 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **INT2-OA-L2_cp_solv_SVP_stable** | -2320.0563 | 0.683089 | -2319.326 | 0.126099 | 0.11628 | -2319.4521 | -2319.4422 | |
| **INT2-OA-coord_a_stable** | -3814.7711 | 0.34351 | -3814.4011 | 0.082627 | 0.077496 | -3814.4837 | -3814.4786 | |
| **INT2-OA-open_a_stable** | -3814.7689 | 0.343986 | -3814.3985 | 0.082816 | 0.07729 | -3814.4813 | -3814.4758 | |
| **INT2-TM-step-BiX_cp_SVP_b_stable** | -4781.9357 | 0.680478 | -4781.206 | 0.13555 | 0.122578 | -4781.3415 | -4781.3286 | |
| **INT2-TM-step_cisBr-transPdcp_SVP_a_stable** | -4781.9752 | 0.680407 | -4781.2458 | 0.134938 | 0.122067 | -4781.3808 | -4781.3679 | |
| **INT4-RE-coord_d_stable** | -1360.527 | 0.413961 | -1360.0846 | 0.085722 | 0.080531 | -1360.1703 | -1360.1651 | |
| **INT4-RE-open_a_stable** | -1360.5241 | 0.413423 | -1360.0819 | 0.088012 | 0.081581 | -1360.1699 | -1360.1635 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **INT5-RE-coord_b_stable** | -1360.5746 | 0.415941 | -1360.1301 | 0.087952 | 0.081153 | -1360.2181 | -1360.2113 | |
| **INT7-TM-mig_SVP_c_stable** | -4781.9793 | 0.680933 | -4781.2494 | 0.133507 | 0.121886 | -4781.3829 | -4781.3713 | |
| **INT8-TM-mig_SVP_c_stable** | -4781.9429 | 0.681509 | -4781.2124 | 0.133389 | 0.121819 | -4781.3458 | -4781.3342 | |
| **INT9-TM-mig-pos_SVP_stable** | -2210.3414 | 0.680628 | -2209.6143 | 0.125928 | 0.11582 | -2209.7402 | -2209.7301 | |
| **PROD-Bi_Br_stable** | -3421.4165 | 0.265506 | -3421.1317 | 0.065607 | 0.063326 | -3421.1973 | -3421.195 | |
| **PROD-Br_SVP_stable** | -2571.9064 | 0 | -2571.904 | 0.016371 | 0.016371 | -2571.9203 | -2571.9203 | |
| **RXT-Bi_cp_stable** | -967.17243 | 0.334941 | -966.81578 | 0.070336 | 0.067393 | -966.88612 | -966.88318 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **TS-OA-coord_b_stable** | -3814.7182 | 0.342422 | -3814.3498 | 0.082112 | 0.07681 | -3814.4319 | -3814.4266 | -168.5 |
| **TS-OA-open_b_stable** | -3814.7174 | 0.342444 | -3814.3489 | 0.083598 | 0.077444 | -3814.4324 | -3814.4263 | -155.47 |
| **TS-RE-coord_a_stable** | -1360.5051 | 0.412709 | -1360.0643 | 0.085905 | 0.080183 | -1360.1502 | -1360.1445 | -368.76 |
| **TS-RE-open_a_stable** | -1360.5055 | 0.413239 | -1360.0642 | 0.085035 | 0.07974 | -1360.1492 | -1360.1439 | -344.29 |
| **TS-TM-mig_SVP_c_stable** | -4781.9559 | 0.681079 | -4781.2267 | 0.130377 | 0.119629 | -4781.3571 | -4781.3463 | -47.6 |
| **TS-TM-step-BiX_SVP_a_stable** | -4781.9311 | 0.679177 | -4781.2035 | 0.133593 | 0.121105 | -4781.3371 | -4781.3246 | -23.15 |
| **TS-TM-step-Pdcp_SVP-1t_stable** | -4781.9576 | 0.681328 | -4781.2286 | 0.129055 | 0.118584 | -4781.3577 | -4781.3472 | -74.88 |
| **TS2-TM-step-BiX_SVP_e_stable** | -4781.9638 | 0.679291 | -4781.2359 | 0.134437 | 0.121656 | -4781.3703 | -4781.3575 | -42.81 |

**Table A.12.** Summarized raw absolute energy values using B3LYP-D3/def2-TZVP (Bi,Pd) during optimization with ligand (R=Me).

| Structure | E_SPC | E | ZPE | H_SPC | T.S | T.qh-S | G(T)_SPC | qh-G(T)_SPC | Imag V |
|---|---|---|---|---|---|---|---|---|---|
| INT-L2_trans-solv_TZVP_b_stable_B3LYP | -2325.8927 | -2325.2223 | 0.74721 | -2325.0911 | 0.143663 | 0.130854 | -2325.2348 | -2325.222 | |
| INT2-OA-trans_coord_TZVP_a_stable_B3LYP | -3936.3767 | -3933.1648 | 0.42238 | -3935.9219 | 0.096254 | 0.089446 | -3936.0182 | -3936.0113 | |
| INT5-RE-coord_TZVP_b_stable_B3LYP | -1479.4552 | -1479.026 | 0.493808 | -1478.9268 | 0.101196 | 0.092988 | -1479.028 | -1479.0198 | |
| L_TZVP_a_stable_B3LYP | -1002.2681 | -1001.991 | 0.329605 | -1001.9164 | 0.069649 | 0.066637 | -1001.986 | -1001.983 | |
| PROD-Bi_Br_TZVP_stable_B3LYP | -3424.7639 | -3421.6986 | 0.262724 | -3424.4817 | 0.065916 | 0.063639 | -3424.5476 | -3424.5453 | |
| PROD-Ph_cp_TZVP_stable_B3LYP | -349.14885 | -349.02316 | 0.162482 | -348.97714 | 0.040727 | 0.039956 | -349.01787 | -349.0171 | |
| RXT-Bi_cp_TZVP_stable_B3LYP | -967.80547 | -967.52111 | 0.331474 | -967.45204 | 0.070899 | 0.067846 | -967.52294 | -967.51989 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **RXT-Br-Ph_stable_opt_B3LYP** | -2806.0203 | -2803.4268 | 0.090831 | -2805.9223 | 0.035848 | 0.035854 | -2805.9582 | -2805.9582 | |
| **TS-OA_cis_coord_TZVP_b_stable_opt_B3LYP** | -3936.3183 | -3933.1067 | 0.42133 | -3935.8653 | 0.093659 | 0.087535 | -3935.9589 | -3935.9528 | -123.11 |
| **TS-OA_cis_open_TZVP_a_stable_opt_B3LYP** | -3936.317 | -3933.106 | 0.42076 | -3935.8641 | 0.094895 | 0.088458 | -3935.959 | -3935.9526 | -139.75 |
| **TS-OA_trans_open_TZVP_d_opt_B3LYP** | -3936.3147 | -3933.103 | 0.420512 | -3935.8619 | 0.096729 | 0.089256 | -3935.9586 | -3935.9512 | -111.67 |
| **TS-RE_open_trans_TZVP_e_stable_B3LYP** | -1479.3934 | -1478.9637 | 0.490642 | -1478.8686 | 0.098656 | 0.091725 | -1478.9672 | -1478.9603 | -360.36 |
| **acetone_TZVP_stable_B3LYP** | -193.27287 | -193.19525 | 0.08323 | -193.18282 | 0.033839 | 0.033097 | -193.21666 | -193.21592 | |

*A.4.2 Reaction Pathways for Bisma-Stille Cross-Coupling Reaction*

**Table A.13.** Summarized benchmarking results of different levels of theory: ωB97XD *vs* B3LYP.

| **Benchmark Study using ωB97XD/def2-TZVP** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **RXN1** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o** | INT-L2-solv+BrPh | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | TS_OA_cis+L+solv | 34.89 | 227.23 | -2.24 | 32.07 | 10.17 | 12.38 | 21.9 | 19.68 |
| | | | | | | | | | |
| | **RXN2** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | INT5_RE+BrPh | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | TS_OA_cis+PhCp | 7.75 | 204.92 | -0.67 | 6.66 | -1.18 | -0.53 | 7.84 | 7.19 |
| | | | | | | | | | |
| | **RXN3** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | INT2_OA+Bi_cp | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | TS_RE+Bi_Br | 14.31 | 13.38 | -0.31 | 13.51 | -1.22 | -0.99 | 14.73 | 14.5 |

| **Benchmark Study using B3LYP-D3/def2-TZVP** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **RXN1** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | INT-L2-solv+BrPh | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | TS_OA_cis+L+solv | 34.04 | 223.69 | -2.59 | 31.05 | 11.84 | 13.59 | 19.21 | 17.46 |
| | | | | | | | | | |
| | **RXN2** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | INT5_RE+BrPh | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | TS_OA_cis+PhCp | 5.59 | 202.8 | -0.68 | 4.56 | -0.89 | -0.16 | 5.46 | 4.72 |
| | | | | | | | | | |
| | **RXN3** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | INT2_OA+Bi_cp | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | TS_RE+Bi_Br | 15.66 | 14.81 | -0.31 | 14.87 | -1.62 | -1.21 | 16.49 | 16.08 |

**Table A.14.** Summarized reaction pathways for generating Pd0 complex**.**

| Generating active Pd0 Complex | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **RXN: Path A** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **INT-L2-solv_trans + BrPh** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| INT-L2-solv_cis + BrPh | 1.95 | 1.48 | 0.88 | 1.89 | -4.87 | -3.05 | 6.76 | 4.93 |
| INT-L-solv+L + BrPh | 29.43 | 23.54 | -0.97 | 27.57 | 10.25 | 12.58 | 17.32 | 14.99 |
| INT-L+ L + solv + BrPh | 48.13 | 41.69 | -2.08 | 45.23 | 22.28 | 25.33 | 22.95 | 19.9 |
| INT1_PhBr + L + solv | 30.45 | 222.94 | -2.2 | 28.26 | 13.72 | 14.17 | 14.54 | 14.09 |
| | | | | | | | | |
| **RXN: Path B** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **INT5_RE + BrPh** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| INT1_PhBr + PhCp | 3.31 | 200.63 | -0.63 | 2.86 | 2.37 | 1.25 | 0.49 | 1.61 |

**Table A.15.** Summarized reaction pathways for the oxidative addition step.

| | **RXN: *cis*-OA** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
|---|---|---|---|---|---|---|---|---|---|
| **o** | Bi_cp+INT-L2-solv+BrPh | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | Bi_cp+INT1_PhBr+L+solv | 30.45 | 222.94 | -2.2 | 28.26 | 13.72 | 14.17 | 14.54 | 14.09 |
| **o** | Bi_cp+TS_OA_cis+L+solv | 35.16 | 227.43 | -2.29 | 32.32 | 10.6 | 12.79 | 21.72 | 19.53 |
| **o** | Bi_cp+INT2_OA_cis-coord+L+solv | 12.33 | 204.23 | -2.41 | 10.05 | 13.51 | 14.53 | -3.46 | -4.48 |
| **o** | Bi_cp+INT2_OA_cis-open+L+solv | 5.34 | 198.67 | -1.23 | 3.8 | 10.92 | 12.79 | -7.13 | -8.99 |
| | | | | | | | | | |
| | **RXN: *trans*-OA** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o** | Bi_cp+INT-L2-solv+BrPh | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **o** | Bi_cp+INT1_PhBr+L+solv | 30.45 | 222.94 | -2.2 | 28.26 | 13.72 | 14.17 | 14.54 | 14.09 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o** | Bi_cp+TS_OA_trans+L+solv | 37.35 | 229.8 | -2.79 | 34.38 | 12.38 | 13.54 | 22 | 20.84 |
| **o** | Bi_cp+INT2_OA_trans-open+L+solv | 3.94 | 196.44 | -1.36 | 2.38 | 11.96 | 13.22 | -9.58 | -10.84 |
| **o** | Bi_cp+INT2_OA_trans-coord+L+solv | 1.29 | 193.42 | -1.82 | -0.53 | 12.02 | 13.71 | -12.55 | -14.24 |

**Table A.15.** Summarized reaction pathways for transmetallation step *via* pathways A, B, and C.

| **Ligand (R=Me with def2-TZVP); SPC with wB97XD/def2-QZVPP** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **RXN: TM_A (1 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o Bi_cp+INT-L2-solv+BrPh** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o Bi_cp+INT2_OA+L+solv | 1.3 | 193.44 | -1.82 | -0.52 | 12.04 | 13.72 | -12.56 | -14.25 |
| o TS_TM+L+solv | -2.45 | 185.85 | 0.28 | -2.78 | -4.08 | -3.59 | 1.3 | 0.81 |
| o INT3_TM+L+solv | -11.79 | 175.4 | -0.27 | -11.84 | -0.93 | -1.68 | -10.91 | -10.17 |
| o TS2_TM+L+solv | -9.64 | 177.49 | -0.47 | -10.4 | -2.25 | -2.6 | -8.15 | -7.79 |
| o INT4_RE_open+Bi_Br+L+solv | 5.05 | 196.25 | -1.73 | 3.19 | 11.42 | 13.42 | -8.24 | -10.23 |
| | | | | | | | | |
| **Ligand (R=H with def2-SVP); No SPC calculation** | | | | | | | | |
| **RXN: TM_A (2 kcal/mol)** | **DE** | **DZPE** | **DH** | **T.DS** | **T.qh-DS** | **DG(T)** | **qh-DG(T)** | |
| **o Bi_cp+INT2_OA_coord** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **o TS_TM** | -8.51 | 1.86 | -6.99 | -14.26 | -15.76 | 7.27 | 8.77 | |
| **o INT2_TM** | -19.86 | 1.23 | -18.18 | -11.31 | -14.32 | -6.87 | -3.86 | |
| | | | | | | | | |
| **RXN: TM_B (2 kcal/mol)** | **DE** | **DZPE** | **DH** | **T.DS** | **T.qh-DS** | **DG(T)** | **qh-DG(T)** | |
| **o Bi_cp+INT2_OA_coord** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **o TS_TM_BiX** | 7.7 | 0.61 | 8.37 | -11.93 | -14.61 | 20.31 | 22.98 | |
| **o INT2_TM_BiX** | 5.34 | 1.16 | 7.16 | -10.38 | -13.59 | 17.54 | 20.74 | |
| | | | | | | | | |
| **RXN: TM_C (2 kcal/mol)** | **DE** | **DZPE** | **DH** | **T.DS** | **T.qh-DS** | **DG(T)** | **qh-DG(T)** | |
| **o Bi_cp+INT2_OA_coord** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **o INT7_TM** | -22.42 | 1.56 | -20.38 | -12.17 | -14.4 | -8.2 | -5.98 | |
| **o TS_MIG** | -7.78 | 1.65 | -6.19 | -14.17 | -15.85 | 7.98 | 9.66 | |
| **o INT8_TM** | 0.39 | 1.92 | 2.78 | -12.28 | -14.48 | 15.06 | 17.26 | |

**Table A.16.** Summarized reaction pathways for the reductive elimination step.

| RXN: RE (1 kcal/mol) | DE_SPC | DE | DZPE | DH_SPC | T.DS | T.qh-DS | DG(T)_SPC | qh-DG(T)_SPC |
|---|---|---|---|---|---|---|---|---|
| **o Bi_cp+INT-L2-solv+BrPh** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o INT4_RE_open+Bi_Br+L+solv | 5.05 | 196.25 | -1.73 | 3.19 | 11.42 | 13.42 | -8.24 | -10.23 |
| o INT4_RE_coord+Bi_Br+L+solv | 2.04 | 193.67 | -1.45 | 0.32 | 10.4 | 12.77 | -10.08 | -12.44 |
| o TS_RE+Bi_Br+L+solv | 15.6 | 206.8 | -2.13 | 12.98 | 10.8 | 12.72 | 2.18 | 0.26 |
| o INT5_RE+Bi_Br+L+solv | -26.26 | 164.78 | 0.22 | -26.43 | 11.19 | 12.83 | -37.62 | -39.26 |
| o INT-L-solv+PhCp+Bi_Br+L | -23.97 | 166.01 | 0.82 | -24.26 | 10.1 | 12.5 | -34.35 | -36.76 |

### *A.4.3 Raw Absolute Energy Values for Substrate and Bismacycle Screen*

**Table A.17.** Summarized raw absolute energy values for the bismacycle screen.

| Structure | E_SPC | E | ZPE | H_SPC | T.S | T.qh-S | G(T)_SPC | qh-G(T)_SPC | Imag V |
|---|---|---|---|---|---|---|---|---|---|
| **o INT2-OA-Br_Br_coord-trans_TZVP_a_stable** | -6509.6036 | -6503.5752 | 0.416517 | -6509.1534 | 0.09942 | 0.092315 | -6509.2529 | -6509.2458 | |
| **o INT2-OA-Br_CF3_coord-trans_TZVP_a_stable** | -4273.0575 | -4269.7108 | 0.430831 | -4272.5907 | 0.105343 | 0.096953 | -4272.696 | -4272.6876 | |
| **o INT2-OA-Br_CH3_coord-trans_TZVP_a_stable** | -3975.2562 | -3972.0479 | 0.453981 | -3974.7681 | 0.099246 | 0.092091 | -3974.8674 | -3974.8602 | |
| **o INT2-OA-Br_CN_coord-trans_TZVP_a_stable** | -4028.1864 | -4024.9541 | 0.425 05 | -4027.7273 | 0.099518 | 0.092482 | -4027.8268 | -4027.8198 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o INT2-OA-Br_F_coord-trans_TZVP_a_stable** | -4035.1937 | -4031.953 | 0.418267 | -4034.7425 | 0.095988 | 0.089742 | -4034.8385 | -4034.8323 | |
| **o INT2-OA-Br_H_coord-trans_TZVP_a_stable** | -3935.9336 | -3932.7389 | 0.426259 | -3935.4751 | 0.095204 | 0.088719 | -3935.5703 | -3935.5638 | |
| **o INT2-OA-Br_OCH3_coord-trans_TZVP_a_stable** | -4050.4754 | -4047.2327 | 0.459193 | -4049.9813 | 0.1015 | 0.094091 | -4050.0828 | -4050.0754 | |
| **o INT2-OA-Cl_H_coord-trans_TZVP_a_stable** | -1821.889 | -1821.4798 | 0.426768 | -1821.4303 | 0.0945 1 | 0.087807 | -1821.5249 | -1821.5181 | |
| **o RXT-Bi-CH_cp_TZVP_stable_model** | -951.41401 | -951.15562 | 0.346707 | -951.04537 | 0.071113 | 0.067925 | -951.11649 | -951.1133 | |
| **o RXT-Bi-O_cp_TZVP_stable_model** | -948.00052 | -947.73598 | 0.294505 | -947.68595 | 0.067205 | 0.064406 | -947.75315 | -947.75035 | |
| **o RXT-Bi-Ph2_cp_TZVP_stable_model** | -795.34228 | -795.13624 | 0.252655 | -795.07128 | 0.0679 | 0.0625 3 | -795.13918 | -795.13381 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o RXT-Bi-S_cp_TZVP_stable_model** | -1270.9917 | -1270.7157 | 0.290991 | -1270.6799 | 0.068917 | 0.065939 | -1270.7489 | -1270.7459 | |
| **o RXT-BiNCF3_cp_TZVP_stable_model** | -1265.2612 | -1264.8565 | 0.311786 | -1264.9255 | 0.076299 | 0.072146 | -1265.0018 | -1264.9977 | |
| **o RXT-BiNCH2PhCN_cp_TZVP_stable_model** | -1290.7711 | -1290.3871 | 0.416011 | -1290.3264 | 0.088069 | 0.082054 | -1290.4145 | -1290.4085 | |
| **o RXT-BiNCH2PhOMe_cp_TZVP_stable_model** | -1313.0628 | -1312.6683 | 0.450138 | -1312.5832 | 0.090132 | 0.083755 | -1312.6733 | -1312.667 | |
| **o RXT-BiNPhOMe_cp_TZVP_stable_model_a** | -1273.7471 | -1273.366 | 0.421625 | -1273.2976 | 0.085258 | 0.080016 | -1273.3828 | -1273.3776 | |
| **o RXT-BiNiPr_cp_TZVP_stable_model** | -1046.094 | -1045.8014 | 0.392061 | -1045.6774 | 0.076111 | 0.072635 | -1045.7535 | -1045.7501 | |
| **o RXT-BiSO2_cp_TZVP_stable_model** | -1342.8182 | -1342.4603 | 0.242323 | -1342.5559 | 0.06769 | 0.064658 | -1342.6236 | -1342.6205 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **o RXT-BiS_cp_TZVP_stable_model** | -1192.3543 | -1192.1047 | 0.232967 | -1192.103 | 0.064605 | 0.061404 | -1192.1676 | -1192.1644 | |
| **o RXT-Bi_Ph_TZVP_stable_model** | -1081.8071 | -1081.5018 | 0.354615 | -1081.4292 | 0.074481 | 0.070786 | -1081.5037 | -1081.5 | |
| **o RXT-Bi_cp_TZVP_stable_model** | -967.45126 | -967.18561 | 0.334964 | -967.0946 | 0.070317 | 0.067371 | -967.16492 | -967.16197 | |
| **o TS-TM-BiNCF3_Br_CN_TZVP-1t_stable_model** | -5293.4466 | -5289.8159 | 0.739927 | -5292.6495 | 0.149095 | 0.136586 | -5292.7986 | -5292.7861 | -20.18 |
| o TS-TM-BiNCF3_Br_H_TZVP-1t_stable_model | -5201.1951 | -5197.6017 | 0.740922 | -5200.3992 | 0.143645 | 0.132075 | -5200.5428 | -5200.5313 | -42.94 |
| o TS-TM-BiNCF3_Br_OMe_TZVP-1t_stable_model | -5315.7404 | -5312.0986 | 0.773572 | -5314.9086 | 0.153168 | 0.139244 | -5315.0617 | -5315.0478 | -13.34 |
| o TS-TM-BiNCH2PhCN_Br_CN_TZVP_stable_model_a | -5318.967 | -5315.3573 | 0.844161 | -5318.061 | 0.161523 | 0.146264 | -5318.2225 | -5318.2072 | -65.42 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| o TS-TM-BiNCH2PhCN_Br_H_TZVP_stable_model_a | -5226.7134 | -5223.1415 | 0.845721 | -5225.808 | 0.155756 | 0.141705 | -5225.9638 | -5225.9497 | -58.85 |
| o TS-TM-BiNCH2PhCN_Br_OMe_TZVP_stable_model_a | -5341.2561 | -5337.6362 | 0.878999 | -5340.3147 | 0.161886 | 0.146976 | -5340.4766 | -5340.4617 | -58.28 |
| o TS-TM-BiNCH2PhOMe_Br_CN_TZVP_stable_model | -5341.2608 | -5337.6405 | 0.878671 | -5340.3196 | 0.162109 | 0.14719 | -5340.4817 | -5340.4668 | -63.59 |
| o TS-TM-BiNCH2PhOMe_Br_H_TZVP_stable_model | -5249.0073 | -5245.4246 | 0.879965 | -5248.0667 | 0.15866 | 0.143638 | -5248.2253 | -5248.2103 | -68.02 |
| o TS-TM-BiNCH2PhOMe_Br_OMe_TZVP_stable_model | -5363.55 | -5359.9194 | 0.913398 | -5362.5735 | 0.162886 | 0.148135 | -5362.7363 | -5362.7216 | -82.49 |
| o TS-TM-BiNPhOMe-Pdcp_Br_CN_TZVP-1t_stable_model_a | -5301.9394 | -5298.3331 | 0.849372 | -5301.0288 | 0.157258 | 0.143947 | -5301.1861 | -5301.1727 | -24.98 |
| o TS-TM-BiNPhOMe-Pdcp_Br_H_TZVP-1t_stable_model | -5209.6889 | -5206.1194 | 0.851062 | -5208.7785 | 0.156266 | 0.141558 | -5208.9348 | -5208.9201 | -28.14 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| o TS-TM-BiNPhOMe-Pdcp_Br_OMe_TZVP-1t_stable_model | -5324.2321 | -5320.6145 | 0.883556 | -5323.2863 | 0.160863 | 0.146401 | -5323.4472 | -5323.4327 | -28.92 |
| o TS-TM-BiNiPr-Pdcp_Br_CF3_TZVP-1t_stable_model | -5319.1605 | -5315.5274 | 0.826358 | -5318.2745 | 0.154189 | 0.141065 | -5318.4287 | -5318.4156 | -74 |
| o TS-TM-BiNiPr-Pdcp_Br_CN_TZVP-1t_stable_model_b | -5074.2888 | -5070.77 | 0.820113 | -5073.4109 | 0.14958 | 0.137465 | -5073.5605 | -5073.5484 | -67.57 |
| o TS-TM-BiNiPr-Pdcp_Br_H_TZVP-1t_stable_model_b | -4982.0351 | -4978.554 | 0.821056 | -4981.1581 | 0.144801 | 0.133538 | -4981.3029 | -4981.2917 | -73.87 |
| o TS-TM-BiNiPr-Pdcp_Br_Me_TZVP-1t_stable_model | -5021.3582 | -5017.8635 | 0.849157 | -5020.4513 | 0.14933 | 0.137091 | -5020.6006 | -5020.5884 | -74.21 |
| o TS-TM-BiNiPr-Pdcp_Br_OCH3_TZVP-1t_stable_model_b | -5096.5779 | -5093.0489 | 0.854299 | -5095.6651 | 0.151222 | 0.138894 | -5095.8163 | -5095.804 | -81.14 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| o TS-TM-BiS-Pdcp_Br_H_TZVP-1t_b_stable_model | -5128.2805 | -5124.841 | 0.661932 | -5127.5687 | 0.133187 | 0.122696 | -5127.7019 | -5127.6914 | -32 |
| o TS-TM-BiSO2-Pdcp_Br_H_TZVP-1t_b_stable_model | -5278.7502 | -5275.2033 | 0.672348 | -5278.0269 | 0.134689 | 0.124335 | -5278.1616 | -5278.1513 | -14.27 |
| o TS-TM-CH-Pdcp_Br_H_TZVP-1t_stable_model | -4887.3387 | -4883.8917 | 0.77696 | -4886.5087 | 0.138563 | 0.127929 | -4886.6473 | -4886.6367 | -30.92 |
| o TS-TM-N-PdPh_Br_H_TZVP-1t_stable_model | -5017.7507 | -5014.2581 | 0.783452 | -5016.9124 | 0.14351 | 0.131872 | -5017.0559 | -5017.0443 | -39.53 |
| o TS-TM-N-Pdcp_Br_Br_TZVP-1t_stable_model | -7477.062 | -7470.7747 | 0.754375 | -7476.253 | 0.144099 | 0.132088 | -7476.3971 | -7476.3851 | -84.71 |
| o TS-TM-N-Pdcp_Br_CF3_TZVP-1t_stable_model | -5240.5161 | -5236.9098 | 0.768952 | -5239.6902 | 0.150734 | 0.137111 | -5239.8409 | -5239.8273 | -73.68 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| o TS-TM-N-Pdcp_Br_CH3_TZVP-1t_stable_model | -4942.7141 | -4939.2464 | 0.791726 | -4941.8673 | 0.143641 | 0.131997 | -4942.011 | -4941.9993 | -71.19 |
| o TS-TM-N-Pdcp_Br_CN_TZVP-1t_stable_model | -4995.6446 | -4992.1527 | 0.763511 | -4994.8263 | 0.142921 | 0.131593 | -4994.9693 | -4994.9579 | -52.52 |
| o TS-TM-N-Pdcp_Br_F_TZVP-1t_stable_model | -5002.6516 | -4999.1514 | 0.756221 | -5001.8416 | 0.140983 | 0.129763 | -5001.9826 | -5001.9713 | -76.03 |
| o TS-TM-N-Pdcp_Br_H_TZVP-1t_stable_model_a | -4903.391 | -4899.9369 | 0.764552 | -4902.5736 | 0.139462 | 0.128161 | -4902.713 | -4902.7017 | -93.02 |
| o TS-TM-N-Pdcp_Br_OCH3_TZVP-1t_stable_model | -5017.9337 | -5014.4316 | 0.796709 | -5017.081 | 0.145892 | 0.133975 | -5017.2269 | -5017.215 | -88.53 |
| o TS-TM-N-Pdcp_Cl_H_TZVP-1t_stable_model | -2789.3453 | -2788.6751 | 0.764451 | -2788.5279 | 0.13896 | 0.127878 | -2788.6669 | -2788.6558 | -76.89 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| o TS-TM-O-Pdcp_TZVP-1t_stable_model | -4883.9337 | -4880.4805 | 0.72427 | -4883.1582 | 0.135924 | 0.125004 | -4883.2941 | -4883.2832 | -23.67 |
| o TS-TM-Ph2-Pdcp_Br_H_TZVP-1t_stable_model | -4731.2781 | -4727.883 | 0.683364 | -4730.5457 | 0.131423 | 0.120996 | -4730.6771 | -4730.6667 | -18.27 |
| o TS-TM-S-Pdcp_Br_H_TZVP-1t_stable_model | -5206.9261 | -5203.4613 | 0.719788 | -5206.1541 | 0.138021 | 0.127044 | -5206.2921 | -5206.2812 | -48.68 |

*A.4.4 Reaction Barriers for Substrate and Modified Bismacycles Screen*

**Table A.18.** Summarized reaction pathways for the reductive elimination step.

| **Reaction Barriers From Figure 4.4A** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **RXN: BiN_Ph (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br+BiN_Ph** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_N_Ph | -6.29 | -10.87 | 1.62 | -5.07 | -16.43 | -17.34 | 11.35 | 12.27 |
| | | | | | | | | |
| **RXN: BiN_Me (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br+BiN_cp** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_N_Br_H | -3.88 | -7.73 | 2.09 | -2.4 | -16.35 | -17.53 | 13.95 | 15.12 |
| | | | | | | | | |
| **RXN: BiN_Cl (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Cl+BiN_cp** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_N_Cl | -3.15 | -6.06 | 1.71 | -1.89 | -16.23 | -17.13 | 14.34 | 15.24 |
| | | | | | | | | |
| **Reaction Barriers From Figure 4.4B** | | | | | | | | |
| **RXN: BiN_iPr (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br+BiNiPr_cp** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_BiNiPr | -4.73 | -8.56 | 1.72 | -3.49 | -16.64 | -17.45 | 13.14 | 13.96 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **RXN: BiO (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br+BiO** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_O | 0.28 | -3.48 | 2.2 | 1.83 | -16.62 | -17.65 | 18.45 | 19.47 |
| | | | | | | | | |
| **RXN: BiSO2_smallring (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br+BiSO2** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_BiSO2 | 0.99 | -2.55 | 2.36 | 2.56 | -17.7 | -18.22 | 20.26 | 20.78 |
| | | | | | | | | |
| **RXN: BiCH (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br+BiCH** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_CH | 5.6 | 1.8 | 2.51 | 7.37 | -17.42 | -18.02 | 24.79 | 25.39 |
| | | | | | | | | |
| **Reaction Barriers From Figure 4.4C** | | | | | | | | |
| **RXN: BiN_CH2PhOMe (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br+BiNCH2PhOMe** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_BiNCH2PhOMe | -6.81 | -10.91 | 2.24 | -5.24 | -16.74 | -18.09 | 11.5 | 12.85 |
| | | | | | | | | |
| **RXN: BiN_PhOMe (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br+BiNPhOMe_cp** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_BiNPhOMe | -5.12 | -9.06 | 1.99 | -3.65 | -15.18 | -17.05 | 11.54 | 13.41 |
| | | | | | | | | |
| **RXN: BiN_CH2PhCN (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br+BiNCH2PhCN** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_BiNCH2PhCN | -5.53 | -9.67 | 2.17 | -4.06 | -17.27 | -18.24 | 13.21 | 14.18 |
| | | | | | | | | |
| **RXN: BiN_Me (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br+BiN_cp** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_N_Br_H | -3.88 | -7.73 | 2.09 | -2.4 | -16.35 | -17.53 | 13.95 | 15.12 |
| | | | | | | | | |
| **RXN: BiN_CF3_H (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br+BiNCF3_cp** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_BiNCF3 | -0.17 | -3.94 | 1.81 | 0.91 | -17.48 | -18.07 | 18.4 | 18.98 |
| | | | | | | | | |
| **Reaction Barriers From Figure 4.4D** | | | | | | | | |
| **RXN: BiN_iPr_CN (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **o INT2_OA_Br_CN+BiNiPr_cp** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_BiNiPr_Br_CN | -5.28 | -9.06 | 1.88 | -3.88 | -16.35 | -17.35 | 12.47 | 13.48 |
| | | | | | | | | |
| **RXN: BiN_iPr_CF3 (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br_CF3+BiNiPr_cp** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_BiNiPr_Br_CF3 | -5.63 | -9.54 | 2.17 | -4.03 | -17.11 | -17.9 | 13.08 | 13.86 |
| | | | | | | | | |
| **RXN: BiN_iPr_Me (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br_Me+BiNiPr_cp** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_BiNiPr_Br_Me | -5.04 | -8.91 | 1.95 | -3.61 | -16.33 | -17.34 | 12.73 | 13.74 |
| | | | | | | | | |
| **RXN: BiN_iPr_OMe (2 kcal/mol)** | **DE_SPC** | **DE** | **DZPE** | **DH_SPC** | **T.DS** | **T.qh-DS** | **DG(T)_SPC** | **qh-DG(T)_SPC** |
| **o INT2_OA_Br_OMe+BiNiPr_cp** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o TS-TM_BiNiPr_Br_OMe | -5.33 | -9.32 | 1.91 | -3.96 | -16.56 | -17.46 | 12.6 | 13.5 |

Additional reaction barrier heights (along with the collected molecular descriptors) calculated for the Multivariate Linear Regression Model in section 4.3.10 can be found in the GitHub repository: Liliana-Gallegos/bisma-stille_analysis labeled as "Bisma-Stille_descriptors.xlsx".[17] Code for building the MVLR model is also available as Jupyter Notebooks for transparency and reproducibility of plots and figures in the Bisma-Stille Analysis repository.

## *A.5 Summarized Thermochemical Data: Chapter 5*

### *A.5.1 Raw Absolute Energy Values*

**Table A.19.** Summarized raw absolute energy values for the palladium-mediated cascade reaction containing Part 1 of the energy profile.

| Structure | | | E_QZ | E | ZPE | H_QZ | T.S | T.qh-S | G(T)_QZ | qh-G(T)_QZ | Imag V |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ***cis*-pathway** | alkoxide form. | Int-I | -1818.34648 | -1816.32999 | 0.66113 | -1817.62828 | 0.15750 | 0.14207 | -1817.78578 | -1817.77034 | - |
| | | TS-I_a | -1818.32586 | -1816.30826 | 0.65950 | -1817.61018 | 0.15385 | 0.13941 | -1817.76403 | -1817.74959 | -77.30 |
| | | TS-I_b | -1818.32794 | -1816.31511 | 0.66056 | -1817.61224 | 0.14611 | 0.13565 | -1817.75835 | -1817.74789 | -86.52 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | ligand diss. | Int-II | -1589.18390 | -1587.28580 | 0.59611 | -1588.53886 | 0.13388 | 0.12367 | -1588.67274 | -1588.66253 | - |
| | | Int-III_a | -1340.83906 | -1339.05214 | 0.50377 | -1340.29378 | 0.11735 | 0.10917 | -1340.41113 | -1340.40295 | - |
| | | Int-III_b | -1340.83399 | -1339.04972 | 0.50362 | -1340.28887 | 0.11984 | 0.11067 | -1340.40871 | -1340.39954 | - |
| | | Int-III_c | -1340.83918 | -1339.05308 | 0.50321 | -1340.29422 | 0.11977 | 0.11058 | -1340.41399 | -1340.40481 | - |
| | **β**-elim. | TS-III-desired | -1340.81619 | -1339.01902 | 0.50142 | -1340.27302 | 0.12051 | 0.11067 | -1340.39353 | -1340.38369 | -290.05 |
| | | TS-III-undes | -1340.81708 | -1339.02129 | 0.50167 | -1340.27376 | 0.11962 | 0.11026 | -1340.39338 | -1340.38402 | -293.62 |
| ***trans*-pathway** | alkoxide form. | Int-I | -1818.33945 | -1816.31342 | 0.66070 | -1817.62142 | 0.16075 | 0.14333 | -1817.78217 | -1817.76475 | - |
| | | TS-I_a | -1818.32192 | -1816.29958 | 0.65963 | -1817.60608 | 0.15520 | 0.14008 | -1817.76128 | -1817.74616 | -82.08 |
| | | TS-I_b | -1818.32149 | -1816.29991 | 0.65998 | -1817.60546 | 0.15311 | 0.13900 | -1817.75857 | -1817.74446 | -77.57 |
| | ligand diss. | Int-II_a | -1589.17994 | -1587.27530 | 0.59608 | -1588.53463 | 0.13753 | 0.12535 | -1588.67216 | -1588.65998 | - |
| | | Int-II_b | -1589.17972 | -1587.27587 | 0.59609 | -1588.53454 | 0.13709 | 0.12496 | -1588.67162 | -1588.65949 | - |
| | | Int-III_a | -1340.83497 | -1339.04465 | 0.50355 | -1340.28971 | 0.12048 | 0.11082 | -1340.41018 | -1340.40053 | - |
| | | Int-III_b | -1340.83508 | -1339.04447 | 0.50345 | -1340.28991 | 0.12093 | 0.11096 | -1340.41084 | -1340.40087 | - |
| | | Int-III_c | -1340.83525 | -1339.04342 | 0.50344 | -1340.29020 | 0.12009 | 0.11048 | -1340.41029 | -1340.40068 | - |
| | **β**-elimination | TS-III des_a | -1340.81709 | -1339.01868 | 0.50141 | -1340.27398 | 0.11991 | 0.11052 | -1340.39389 | -1340.38450 | -311.05 |
| | | TS-III des_b | -1340.81537 | -1339.01694 | 0.50147 | -1340.27215 | 0.12145 | 0.11114 | -1340.39359 | -1340.38329 | -285.06 |
| | | TS-III-undes_a | -1340.81717 | -1339.01898 | 0.50131 | -1340.27411 | 0.12045 | 0.11075 | -1340.39456 | -1340.38486 | -281.63 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | TS-III-undes_b | -1340.81546 | -1339.01676 | 0.50135 | -1340.27235 | 0.12146 | 0.11116 | -1340.39381 | -1340.38351 | -290.64 |

**Table A.20.** Summarized raw absolute energy values for the palladium-mediated cascade reaction containing Part 2 of the energy profile.

| Structure | | | E_QZ | E | ZPE | H_QZ | T.S | T.qh-S | G(T)_QZ | qh-G(T)_QZ | ImagV |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **desired** | migratory insert. | Int-IV | -1340.86369 | -1339.06492 | 0.50218 | -1340.31860 | 0.12502 | 0.11382 | -1340.44362 | -1340.43242 | - |
| | | TS-IV_a | -1340.83533 | -1339.04616 | 0.50303 | -1340.29122 | 0.11632 | 0.10840 | -1340.40754 | -1340.39962 | -319.53 |
| | | TS-IV_b | -1340.83561 | -1339.04386 | 0.50294 | -1340.29153 | 0.11716 | 0.10873 | -1340.40869 | -1340.40026 | -323.61 |
| | | TS-IV_c | -1340.82872 | -1339.03234 | 0.50243 | -1340.28498 | 0.11874 | 0.10946 | -1340.40372 | -1340.39444 | -356.00 |
| | ligand isom. | Int-V | -1340.86304 | -1339.07331 | 0.50356 | -1340.31796 | 0.11861 | 0.10966 | -1340.43657 | -1340.42762 | - |
| | | Int-VI_a | -1340.85807 | -1339.06291 | 0.50404 | -1340.31260 | 0.11902 | 0.10969 | -1340.43162 | -1340.42230 | - |
| | | Int-VI_b | -1340.86525 | -1339.07137 | 0.50376 | -1340.32042 | 0.11559 | 0.10826 | -1340.43602 | -1340.42868 | - |
| | **β** H-elim. | TS-V_a | -1340.85250 | -1339.05612 | 0.50075 | -1340.31104 | 0.11597 | 0.10788 | -1340.42701 | -1340.41892 | -364.82 |
| | | TS-V_b | -1340.85006 | -1339.05268 | 0.50101 | -1340.30853 | 0.11537 | 0.10759 | -1340.42390 | -1340.41612 | -247.74 |
| | | Int-VII | -1340.85043 | -1339.05925 | 0.50260 | -1340.30690 | 0.11619 | 0.10833 | -1340.42310 | -1340.41523 | - |
| | red. elim. | TS-VI | -1340.85547 | -1339.06388 | 0.50010 | -1340.31542 | 0.11269 | 0.10581 | -1340.42811 | -1340.42122 | -958.54 |
| | | *S*-P2 | -735.40572 | -735.07934 | 0.34769 | -735.03404 | 0.07539 | 0.07268 | -735.10943 | -735.10672 | - |
| **undesired** | migratory insertion | Int-IV | -1340.86269 | -1339.06490 | 0.50241 | -1340.31746 | 0.12378 | 0.11340 | -1340.44125 | -1340.43086 | - |
| | | TS-IV_a | -1340.83083 | -1339.03583 | 0.50314 | -1340.28668 | 0.11705 | 0.10875 | -1340.40373 | -1340.39543 | -312.02 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | TS-IV_b | -1340.83287 | -1339.03748 | 0.50302 | -1340.28873 | 0.11789 | 0.10887 | -1340.40663 | -1340.39760 | -324.06 |
| | | TS-IV_c | -1340.82763 | -1339.03068 | 0.50312 | -1340.28350 | 0.11786 | 0.10888 | -1340.40136 | -1340.39238 | -348.45 |
| | ligand isomerism | Int-V | -1340.85964 | -1339.06368 | 0.50354 | -1340.31451 | 0.11979 | 0.11014 | -1340.43430 | -1340.42465 | - |
| | | Int-VI_a | -1340.86227 | -1339.06490 | 0.50345 | -1340.31743 | 0.11942 | 0.10993 | -1340.43685 | -1340.42735 | - |
| | | Int-VI_b | -1340.85671 | -1339.06173 | 0.50382 | -1340.31129 | 0.12001 | 0.11037 | -1340.43130 | -1340.42166 | - |
| | **β** H-elim. | TS-V_a | -1340.84707 | -1339.04628 | 0.50095 | -1340.30549 | 0.11730 | 0.10840 | -1340.42279 | -1340.41389 | -248.99 |
| | | TS-V_b | -1340.84459 | -1339.04416 | 0.50006 | -1340.30338 | 0.11993 | 0.10998 | -1340.42331 | -1340.41336 | -461.85 |
| | | Int-VII | -1340.84676 | -1339.05245 | 0.50187 | -1340.30372 | 0.11762 | 0.10913 | -1340.42134 | -1340.41285 | - |
| | red. elim. | TS-VI | -1340.85273 | -1339.05791 | 0.49891 | -1340.31288 | 0.12020 | 0.10956 | -1340.43308 | -1340.42244 | -1007.89 |
| | | *R*-P2 | -735.40452 | -735.07482 | 0.34724 | -735.03304 | 0.07751 | 0.07371 | -735.11056 | -735.10676 | - |

**Table A.21.** Summarized raw absolute energy values for the ground state catalysts, ligand, and reactants.

| Structure | | E_QZ | E | ZPE | H_QZ | T.S | T.qh-S | G(T)_QZ | qh-G(T)_QZ | Imag V |
|---|---|---|---|---|---|---|---|---|---|---|
| **Catalyst, ligand, reactants** | *cis*-P3 | -736.60695 | -736.26820 | 0.36891 | -736.21215 | 0.07951 | 0.07671 | -736.29165 | -736.28885 | - |
| | *trans*-P3 | -736.60578 | -736.26671 | 0.36892 | -736.21095 | 0.08050 | 0.07705 | -736.29146 | -736.28800 | - |
| | pyridine | -248.30948 | -248.19559 | 0.09005 | -248.21253 | 0.03666 | 0.03667 | -248.24920 | -248.24920 | - |
| | HOAc | -229.13291 | -229.00043 | 0.06265 | -229.06329 | 0.03612 | 0.03602 | -229.09941 | -229.09931 | - |
| | Pd0L2 | -624.60973 | -623.16024 | 0.18341 | -624.40923 | 0.06599 | 0.06212 | -624.47522 | -624.47135 | - |
| | PdH | -605.43042 | -603.96421 | 0.15245 | -605.26149 | 0.06327 | 0.06005 | -605.32476 | -605.32154 | - |
| | PdII | -1081.72110 | -1080.03240 | 0.29024 | -1081.40060 | 0.09620 | 0.08884 | -1081.49680 | -1081.48940 | - |

**REFRENCES**

1. (a) 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; 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 Rev. C.01*, Wallingford, CT, 2016; (b) 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; 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 Rev. B.01*, Wallingford, CT, 2016; (c) 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; 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 Rev. A.03*, Wallingford, CT, 2016.

2. (a) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C., Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* **1992,** *46* (11), 6671-6687; (b) Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993,** *98* (7), 5648-5652; (c) 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 function. *Theoretical Chemistry Accounts* **2008,** *120* (1-3), 215-241; (d) Purvis, G. D., III; Bartlett, R. J., A full coupled-cluster singles and doubles model: The inclusion of disconnected

triples. *J. Chem. Phys.* **1982,** *76* (4), 1910-1918; (e) Chai, J.-D.; Head-Gordon, M., Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008,** *10* (44), 6615-6620.
3. 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* (18), 6378-6396.
4. (a) COSMOtherm Version 18.0.0 (Revision 4360), COSMOlogic GmbH & Co KG; (b) Eckert, F.; Klamt, A., Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE Journal* **2002,** *48* (2), 369-385; (c) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W., Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998,** *102* (26), 5074-5085.
5. Bannwarth, C.; Ehlert, S.; Grimme, S., GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput* **2019,** *15* (3), 1652-1671.
6. Alegre-Requena, J. V.; Sowndarya S. V., S.; Pérez-Soto, R.; Alturaifi, T. M.; Paton, R. S., AQME: Automated quantum mechanical environments for researchers and educators. *WIREs Computational Molecular Science* **2023**.
7. (a) RDKit: open-source. cheminformatics.http://www.rdkit.org; (b) Halgren, T. A., Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* **1996,** *17* (5-6), 490-519.
8. Luchini, G.; Alegre-Requena, J.; Funes-Ardoiz, I.; Paton, R., GoodVibes: automated thermochemistry for heterogeneous computational chemistry data [version 1; peer review: 2 approved with reservations]. *F1000Research* **2020,** *9* (291).
9. Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Karafiloglou, P.; Landis, C. R.; Weinhold, F. *NBO 7.0*, Theoretical Chemistry Institute, 2018.
10. Wolinski, K.; Hinton, J. F.; Pulay, P., Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. *J. Am. Chem. Soc.* **1990,** *112* (23), 8251-8260.
11. Adamo, C.; Barone, V., Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J. Chem. Phys.* **1998,** *108* (2), 664-675.
12. Luchini, G.; Patterson, T.; Paton, R. S. DBSTEP Release 1.1.0 (1.1.0) Zenodo. patonlab/DBSTEP.
13. Gallegos, L. C. DISCO. https://github.com/Liliana-Gallegos/DISCO.
14. (a) Boto, R. A.; Peccati, F.; Laplaza, R.; Quan, C.; Carbone, A.; Piquemal, J.-P.; Maday, Y.; Contreras-Garcia, J. *NCIPLOT4: A new step towards a fast quantification of noncovalent interactions*; (b) Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W., NCIPLOT: A Program for Plotting Noncovalent Interaction Regions. *J. Chem. Theory Comput* **2011,** *7* (3), 625-632.
15. Schrödinger, LLC *The PyMOL Molecular Graphics System, version 1.8*, 2015.
16. (a) Paton Lab Workflows. https://github.com/patonlab/paton_group_workflows; (b) CSU Theory Suite Wiki Page. https://csu-theory-suite.github.io/theory-suite-wiki/.
17. Gallegos, L. C. Bisma-Stille Analysis. https://github.com/Liliana-Gallegos/bisma-stille_analysis.