DEVELOPMENT AND APPLICATION OF EFFECTIVE QUANTUM CHEMICAL STRATEGIES

By

Prajay Patel

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Chemistry – Doctor of Philosophy

2019

ABSTRACT

DEVELOPMENT AND APPLICATION OF EFFECTIVE QUANTUM CHEMICAL STRATEGIES

By

Prajay Patel

Within the field of computational chemistry, one of the greatest challenges is predicting thermodynamic properties such as enthalpies of formation and interaction energies to understand chemical phenomena throughout the periodic table. To predict these properties at a quantitative level, high-level electronic structure methods, primarily *ab initio* methods, are used. These methods are not utilized as often when increasing molecule size due to the significant computational resources (disk space, memory, CPU time) required. Therefore, effective quantum chemical schemes that take advantage of numerous cost-effective methods are needed and this dissertation showcases their development and application towards main group and transition metal thermochemistry.

In this dissertation, the pK_a of late transition metal hydrides, which are important intermediates in catalytic reactions, were predicted with electronic structure methods including density functional theory (DFT) and *ab initio* methods. Insight into the thermochemistry and binding behavior of these hydrides is key to understanding metal-ligand behavior for inorganic and organometallic complexes.

To utilize *ab initio* methods for high accuracy thermochemistry and circumvent their high computational cost, *ab initio* composite strategies, such as the correlation consistent Composite Approach (ccCA), were developed. In an effort to expand the size limitations of composite methodologies, ccCA was combined with the domain-based local pair natural orbital (DLPNO) methods. Denoted as DLPNO-ccCA, this method was developed for main group thermochemistry and targeted one of the largest molecules examined with composite methodologies. This methodology was expanded to key reaction types in organometallic chemistry, such as olefin insertion in hydroformylation, the largest volume homogeneous chemical reaction in chemical

industry for chemical production, and metal-ligand dissociation. To investigate the vibrational behavior of chemical systems found in the interstellar medium, ccCA was used to generate potential energy surfaces (PESs) characterizing vibrational motion to predict anharmonic frequencies in tandem with vibrational self-consistent field (VSCF) and post-VSCF theory so that there is a reduction in the computational cost associated with generating accurate PESs for anharmonic mode-mode couplings as well as calculating contributions from anharmonic corrections to the potential.

While *ab initio* methods are critical for attaining quality thermochemical predictions, addressing polyatomic molecules of increasing size and complexity, electronic structure methods like DFT are utilized due to the relative computational cost of DFT compared to *ab initio* methods. Applications in this dissertation include the modeling of the frontier orbitals of zinc porphyrin-fullerene supramolecular dyads with DFT to exhibit intramolecular charge transfer and the prediction of the binding energies for several drug-like molecules to polymer-based host compounds that display a binding pocket, which models protein-drug binding interactions, as part of the Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) blind prediction competition.

Copyright by PRAJAY PATEL 2019 This dissertation is dedicated to my family, who supported me on my five-year mission in the final frontier of formal education.

ACKNOWLEDGMENTS

I am truly grateful for everyone who has supported me throughout my academic career. Firstly, I would like to thank Dr. Angela K. Wilson for her guidance over the years. I would like to thank the Wilson group, past and current members, for their support throughout the years for insightful discussions and some fun group bonding activities including but not limited to Becky, John P., Andrew, Kameron, John D., Michael, Zainab, Lucas, Thomas, Yiğitcan, Timothé, Hailey, and Lenin. More so, I would like to thank Dr. Jiaqi Wang for training me when I started on my first project, Dr. Inga Ulusoy for guidance on several projects throughout my graduate career as well as Joseph Chung and Max Bowman, who were high school students that worked in the Wilson group and contributed to one of the projects. I would like to acknowledge my committee at Michigan State University, Ned Jackson, Gary Blanchard, and Ben Levine as well as my former committee at the University of North Texas, Martin Schwartz, Lee Slaughter, and Tom Cundari. I do want to mention Martin Schwartz, who passed away late 2018. His happiness and attitude towards teaching and university life is one that I hope to exude in my career.

I would like to thank all my friends. This includes but is certainly not limited to Alex, Ben, Brooke, Bryan, Carlos, Colton, Danielle, Donny, Eric, Erin B., Erin H., Kat, Karla, Kristin, Matt, Neal, Paul, and Whitney. For all the time we have spent together, whether it was for dinner, undergraduate and/or graduate classes, exercise, or annual gaming tournaments, those moments are something to cherish and I hope there are many more to come.

I want to thank my personal trainers Brian, Becky, Leah, and Lexi as well as my boxing partner Janet for keeping me accountable for my physical fitness and for fostering a positive and motivating gym environment during my time in Michigan. The time spent definitely helped me adjust to life in Michigan after moving from Texas.

Thank you to my large extended family and family friends for all of the support from across the U.S. even if how I explained what I do was not the easiest to understand. I am getting better at that skill everyday, in part of my numerous explanations. This includes every holiday season and occasional vacation/reunion...or for Indian people...weddings. To keep this list succinct: Adam, Arjun, Arti, Chandani, Chris, Dhaval, Jeff, Justin, Katrina, Kelli, Kim, Kyle, Mishaun, Nikki, Nisha, Paayal, Priya, and Robert. I do want to specifically acknowledge my two cousins and grandmother who passed away in 2018. I am grateful to have had the time and memories spent with them, even with language barriers and how often I managed to see them. It is unfortunate when life is taken early; however, celebrating life and not taking it for granted is something I hope to do moving forward.

Finally, I thank my father, Mukesh, mother, Shruti, and sister, Shivani, for keeping me grounded with our weekly conversations and providing perspective on life in general. While I never been the best with words, there are not enough to describe how much love and support my family has given me my whole life, including my decision to pursue graduate school. For that, I am forever grateful.

TABLE OF CONTENTS

LIST OF 7	TABLES xi
LIST OF I	FIGURES
LIST OF S	SCHEMES xxi
KEY TO S	SYMBOLS AND ABBREVIATIONS
CHAPTEI REFEI	R 1 INTRODUCTION 1 RENCES 5
CHAPTEI 2.1 A 2.2 C 2 2 2 2 2 2 2 2 2 2 2 2 2	R 2 THEORETICAL BACKGROUND 8 <i>b initio</i> methods 10 Cost-Saving Wavefunction-based Methods 13 2.1 Local Methods 14 2.2 Resolution of the Identity Approximation 16 2.2.3 Domain-Based Local Pair Natural Orbital Methods 17 2.4 Composite Methods 19 2.2.5 ONIOM 22 2.2.5 ONIOM 25 Density Functional Theory 27 Basis Sets 30 2.4.1 Correlation Consistent Basis Sets 31 2.4.2 Effective Core Potentials 31 2.4.3 Auxiliary Basis Sets 32 2.4.3.1 AutoAux 32 2.4.4 Basis Set Superposition Error 33
2.5 I 2 2 2 2.6 V REFEI	2.4.4 Basis Set Superposition Error 33 mplicit Solvation Models 34 2.5.1 COSMO 35 2.5.2 PCM/C-PCM 36 2.5.3 SMD 37 /ibrational Self Consistent Field Theory 37 RENCES 43
CHAPTEI 3.1 I 3.2 T 3.3 F 3 3 3	R 3 PREDICTION OF PK _A OF LATE TRANSITION METAL HYDRIDES VIA A QM/QM APPROACH 63 ntroduction 63 Cheoretical Methods 68 Results and Discussion 73 3.3.1 Utility of DFT in the Real System 73 3.3.2 Utility of DFT in the Model Layer 77 3.3 Impact of Exact Exchange on the Accuracy of DET 82

	3.4 Impact of Adding Grimme's Empirical Dispersion Correction on the	0.4
		84
	3.5 Impact on the Choice of Basis Set	86
	3.6 Impact of Cavity Models on Implicit Solvation Models	88
	3.7 Impact of the Expansion of the Size of Model System	91
	3.8 Comparison of Different Methodologies	92
3.4	onclusions	95
APF	IDIX	96
REF	ENCES	99
СНАРТ	4 UTILIZATION OF THE DOMAIN-BASED LOCAL PAIR NATURAL	
	ORBITAL METHODS WITHIN THE CORRELATION CONSISTENT	
	COMPOSITE APPROACH	108
4.1	itroduction	108
4.2	omputational Methods	111
4.3	esulte and Discussion	115
т.5	2.1 Energetic Droperties for the Molecule Set	115
	2.2 CDL Timina	113
	$5.2 \text{CPU Immig} \dots \dots \dots \dots \dots \dots \dots \dots \dots $	124
	3.3 Enthalpies and Timing for Linear Alkanes	127
	3.4 Applications of DLPNO-ccCA	130
4.4	onclusions	132
APF	IDIX	134
REF	ENCES	138
211 1 D		
CHAPI	5 COMPUTATIONAL CHEMISTRY CONSIDERATIONS IN	
	CATALYSIS: REGIOSELECTIVITY AND METAL-LIGAND	
	DISSOCIATION	149
5.1	troduction	149
5.2	omputational Methods	153
	2.1 Computational methods for hydroformylation	153
	2.2 Computational methods for ligand dissociation	155
5.3	esults and Discussion	156
	3.1 Regioselectivity in hydroformylation	156
	3.2 Metal-ligand dissociation in organometallics	162
5 /	onclusions	162
J.T DEI		161
KLI		104
СНАРТ	6 VIBRATIONAL POTENTIAL ENERGY SURFACES WITH THE	
	CORRELATION CONSISTENT COMPOSITE APPROACH AND	
	DENSITY EUNCTIONAL THEODY	171
(1		1/1
0.1		1/1
6.2		1/4
	2.1 DFT Calculations	175
	2.2 ccCA Calculations	176
6.3	esults and Discussion	176

	6.3.1 Diatomics
	6.3.3 Hydrocarbons
	6.3.4 Aminophenol
6.4	Conclusions
APP	'ENDIX
REF	$ERENCES \dots \dots$
CHAPT	ER 7 CHARGE STABILIZATION OF HIGH POTENTIAL ZINC
	PORPHYRIN-FULLERENE VIA AXIAL LIGATION OF
	TETRATHIAFULVALENE
7.1	Introduction
7.2	Computational Contributions and Analysis
REF	TERENCES
CUADT	ED 9 CAMDI & HOST CHEST CHALLENCE, DINDING EDEE ENEDCIES
СПАРТ	EK 8 SAMPLO HUST-GUEST CHALLENGE: DINDING FREE ENERGIES VIA A MULTISTED ADDOACH 212
Q 1	VIA A MULTISTEF AFFROACH
0.1	Mathada 216
0.2	Niction Drangentian and Simulation Drate col 216
	8.2.1 System Preparation and Simulation Protocol
02	8.2.2 Qualitum Mechanical Calculations 219 Desults 220
0.3	Results
	8.5.1 CB8
	8.5.2 UA
	8.3.5 TEMOA
0.4	8.5.4 Quantum Mechanical Calculations
0.4	0.4.1 Submission Analysis 224
	8.4.1 Submission Analysis
	8.4.2 Impact of Truncated Basis Sets
	8.4.3 Impact of the Extrapolation Scheme <i>B</i> -parameter $\dots \dots \dots$
0.5	8.4.4 Impact of Representative Geometries
8.5	
APP	'ENDIX
KEF	ekences
CHAPT	ER 9 CONCLUDING REMARKS 251
REF	TERENCES

LIST OF TABLES

Table 2.1:	Summary of ccCA-TM and rp-ccCA steps.	25
Table 3.1:	Summary of the density functionals utilized.	71
Table 3.2:	Theoretical methods for the description of real and model systems within the two-layer ONIOM scheme using C-PCM, COSMO, and SMD for utility of DFT in the real layer.	74
Table 3.3:	Theoretical methods for the description of real and model systems within the two-layer ONIOM scheme using C-PCM, COSMO, and SMD for utility of DFT in the model layer	81
Table 3.4:	MADs in pK _a values of GGA, M-GGA, H-GGA, and HM-GGA Types of Functionals for Comparison of DFT and DFT-D3 Relative to Experiment with SMD.	85
Table 3.5:	MADs in pK_a values relative to experiment for four functionals when changing the basis set used for the model layer.	87
Table 3.6:	MADs of five cavity models in pK _a values relative to experiment using the ONIOM(PBE, M06-L, B3LYP, and M06/aug-cc-pVTZ:B97-D/LANL2DZ) scheme.	91
Table 3.7:	MADs in pK_a values relative to experiment of three expansions of model system of TM hydrides with SMD.	92
Table 3.8:	Predicted pK _a values for Schemes A-E, which are ONIOM (B3LYP-D3/aug-cc-pVTZ:B97-D3/ SDD), B97-D3/SDD, B3LYP-D3/SDD, ONIOM(B3LYP/aug-cc-pVTZ:HF/LANL2DZ), and ONIOM(CCSD(T)/aug-cc-pVTZ :B97-D3/SDD), respectively	94
Table 3.9:	Summary of the basis sets utilized	97
Table 3.10:	MADs in pK _a values of GGA, M-GGA, H-GGA, HM-GGA, and DH-GGA functionals within low-level methods with solvation models relative to experiment, with respect to central TM atoms of the TM Hydrides. All of the results are from calculations with ONIOM(B97-D, M06-L, B3LYP, and M06/ aug-cc-pVTZ:DFT/LANL2DZ) scheme.	97

Table 3.11:	MADs in pK _a values of GGA, M-GGA, H-GGA, HM-GGA, and DH-GGA functionals within low-level methods with solvation models relative to experiment, with respect to ligands of the TM hydrides. All of the results are from calculations with ONIOM(B97-D,M06-L, B3LYP, and M06/ aug-cc-pVTZ:DFT/LANL2DZ) scheme.	98
Table 3.12:	MADs in pK _a values of GGA, M-GGA, H-GGA, HM-GGA, and DH-GGA functionals within high-level methods with solvation models relative to experiment, with respect to central TM atoms of the TM Hydrides. All of the results are from calculations with ONIOM(DFT/aug-cc-pVTZ:B97-D,M06L, and B3LYP/LANL2DZ) scheme.	98
Table 4.1:	Summary of the different variants of ccCA utilized in Chapter 4	113
Table 4.2:	Summary of the approximations, methods, and auxiliary basis sets (ABS) utilized in this work for SCF and post-HF calculations.	114
Table 4.3:	Slope, intercept, and R^2 of the calculated and experimental ΔH_f . The mean signed deviation (MSD), mean absolute deviation (MAD), standard deviation (STDEV), and maximum (MAX) deviation for four variants of ccCA based on the Peterson (P), Schwartz-3 (S3), and Schwartz-4 (S4) extrapolation schemes. The P and S3 extrapolated values are averaged for PS3. Triple and quadruple- ζ level basis sets (TQ) were used for all two-point extrapolations. All deviations are in kcal mol ⁻¹ .	115
Table 4.4:	Mean signed deviation (MSD), mean absolute deviation (MAD), standard deviation (STDEV), and maximum (MAX) deviation for all schemes. All deviations are in kcal mol^{-1}	123
Table 4.5:	Percent CPU time savings for the three schemes of ABS implementation within DLPNO-ccCA and RI-ccCA relative to ccCA. The mean percent difference from ccCA, the most efficient (MAX), and the least efficient (MIN) percent CPU time savings relative to ccCA timings are shown. All timing studies were done with ORCA.	126
Table 4.6:	Deviations in kcal mol ⁻¹ from experimental ΔH_f for linear alkanes ($C_n H_{2n+2}$ $1 \le n \le 8$) using the atomization approach and using isodesmic approaches (shown in parentheses).	128
Table 4.7:	Percent CPU time savings for RI-ccCA and DLPNO-ccCA (FB) relative to ccCA for linear alkanes (C_nH_{2n+2} $1 \le n \le 8$). All timing studies were done with ORCA.	129
Table 4.8:	Interactions energies of select examples from the S66 and L7 molecule sets. All interaction energies are in kcal mol^{-1}	131

Table 4.9:	Component breakdown of the DLPNO-ccCA calculated interaction energies from the S66 and L7 datasets with counterpoise corrections included. All interaction energies are in kcal mol^{-1}	32
Table 4.10:	Molecule list used for full set calculations	35
Table 4.11:	MP2/CBS counterpoise-corrected interaction energies calculated for molecules in the S66 data set used to compare DLPNO-ccCA interaction energies from Reference 126. All interaction energies are in kcal mol^{-1} 1	36
Table 4.12:	MP2/CBS counterpoise-corrected interaction energies calculated for the coronene dimer used to compare DLPNO-ccCA interaction energies from Reference 127. All interaction energies are in kcal mol^{-1}	36
Table 4.13:	DFT-D3/def2-QZVPP non-counterpoise-corrected interaction energies calculated for the coronene dimer used to compare DLPNO-ccCA interaction energies from Reference 127. All interaction energies are in kcal mol^{-1} 1	36
Table 5.1:	A summary of the effect of $\Delta \Delta E^{\ddagger}$ in kcal mol ⁻¹ on the linear-to-branched ratio (l:b) ratio for hydroformylation	52
Table 5.2:	Comparison of several density functionals to linear-to-branched ratios from experiment for ee-[Rh(H)(CO)(L)(olefin)] complexes. $\dots \dots \dots$	58
Table 5.3:	Comparison of the approximate $\Delta \Delta E^{\ddagger}s$ based on the calculated 1:b ratios for ee- [Rh(H)(CO)(L)(olefin)] complexes. Experimental $\Delta \Delta E^{\ddagger}s$ are an approximation of experimental 1:b ratios. All $\Delta \Delta E^{\ddagger}s$ are in kcal mol ⁻¹ 1	59
Table 5.4:	Results using DLPNO methods to predict the linear-to-branched ratio for ee- [Rh(H)(CO)(DIPHOS)(propene)]	61
Table 5.5:	Comparison of the gas-phase ligand dissociation energy of H_2O from the Pt complex calculated with DLPNO-rp-ccCA and RI-DFT-D3/aug-cc-pVnZ. All energies are in kcal mol ⁻¹ and are BSSE-corrected	.63
Table 6.1:	Percent CPU Time relative to CCSD(T,full)/aug-cc-pCV5Z to generate all 17 grid points of the PEC for select diatomics.	80
Table 6.2:	Calculated frequencies in cm^{-1} for B3LYP/aug-cc-pVTZ, ccCA, and CCSD(T, full)/aug-cc-pCV5Z for diatomics in Table 6.1	81
Table 6.3:	VCIPSI-PT2 frequencies using a combination of TPSS and ccCA for single mode and vibrational mode-mode coupling potentials. The use of PECs/PESs is denoted as single:coupled	86

Table 6.4:	Vibrational frequencies predicted with VCIPSI-PT2 for selected vibrations of cis-3-aminophenol and trans-3-aminophenol
Table 6.5:	Calculated frequencies of diatomic and small polyatomic molecules in cm^{-1} obtained with ccCA potentials
Table 6.6:	Calculated frequencies of diatomics in cm^{-1} obtained with TPSS/cc-pVnZ and B3LYP/cc-pVnZ potentials
Table 6.7:	Calculated frequencies of selected diatomics in cm^{-1} with TPSS/aug-cc-pVnZ and B3LYP/aug-cc-pVnZ potentials
Table 6.8:	Calculated vibrational frequencies for H_2O and CO_2 in cm ⁻¹ utilizing VCIPSI-PT2 with TPSS/cc-pVnZ potentials
Table 6.9:	Calculated vibrational frequencies for H_2O and CO_2 in cm ⁻¹ utilizing VCIPSI-PT2 with B3LYP/cc-pVnZ potentials
Table 6.10:	Calculated vibrational frequencies for H_2O and CO_2 in cm ⁻¹ utilizing VCIPSI-PT2 with TPSS/aug-cc-pVnZ potentials
Table 6.11:	Calculated vibrational frequencies for H_2O and CO_2 in cm ⁻¹ utilizing VCIPSI-PT2 with B3LYP/aug-cc-pVnZ potentials
Table 6.12:	Calculated vibrational frequencies for NH ₃ in cm^{-1} utilizing VCIPSI-PT2 with both TPSS and B3LYP potentials with the VTZ and aVTZ basis sets 197
Table 8.1:	Binding free energies for the CB8 host-guest complexes
Table 8.2:	Binding free energies for the OA host-guest complexes
Table 8.3:	Binding free energies for the TEMOA host-guest complexes
Table 8.4:	Binding free energies for the CB8 complexes in kcal mol^{-1} with schemes involving not using the RI approximation, and changing the dielectric constant of the implicit solvent with the truncated correlation consistent basis sets for hydrogen
Table 8.5:	Binding free energies for the CB8 complexes in kcal mol^{-1} with schemes involving not using the RI approximation, changing the dielectric constant of the implicit solvent, and two options for basis set choice when extrapolating to the Kohn-Sham limit
Table 8.6:	Predicted binding energies for OA and TEMOA using MMPBSA and RI- B3PW91 after the removal of mean signed error (MSE)

- Table 8.7: Predicted binding energies when using different values for *B* in Equation 8.1for two-point extrapolations using cc-pVDZ and cc-pVTZ with RI-B3PW91-D3.238
- Table 8.8:Fitting parameter values obtained when using Jensen's extrapolation scheme
for each component in calculating the binding energy (Equation 8.1). The host
and guest are counterpoise-corrected before the extrapolation was performed. . . 243

LIST OF FIGURES

Figure 3.1:	From left to right, the compounds are TM(depe) ₂ , TM(depp) ₂ , TM(PNP) ₂ . (a) The model system (bolded) within the ONIOM-1 QM/QM partitioning scheme for TM hydrides with the TM atom (Ni, Pd, and Pt) and four phosphorous atoms in the layer using the high-level method. (b) ONIOM-2: The QM/QM partitioning scheme for TM hydrides with all the atoms within the chelate rings in the layer using the high-level method. (c) ONIOM-3: The QM/QM partitioning scheme for TM hydrides with all except for the very outside methyl group in the layer using the high-level method	70
Figure 3.2:	MADs in pK _a values for the density functionals within low-level methods relative to experiment. All of the results are from calculations with ONIOM(PBE,M06L,B3LYP,M06/aug-cc-pVTZ:DFT/LANL2DZ) scheme. The results of using the four functionals in the model layer are averaged for the molecule set.	75
Figure 3.3:	MADs in pK _a values for five types of density functionals, GGA, M-GGA, H-GGA, HM-GGA, and DH-GGA functionals, within low-level methods relative to experiment. All of the results are from calculations with ONIOM(PBE,M06L,B3LYP,M06/aug-cc-pVTZ:DFT/LANL2DZ) scheme. The results of using the four functionals in the model layer are averaged for the molecule set.	76
Figure 3.4:	MADs in pK _a values for fourteen GGA, M-GGA, H-GGA, HM-GGA, and DH-GGA functionals within high-level methods relative to experiment. All of the results are from calculations with ONIOM(DFT/aug-cc-pVTZ:B97-D,M06L,B3LYP/LANL2DZ) scheme. The MADs in pK _a values for the three functionals in the real layer are averaged for the molecule set	79
Figure 3.5:	MADs in pK _a values for five types of density functionals, GGA, M-GGA, H-GGA, HM-GGA, and DH-GGA functionals, within high-level methods relative to experiment. All of the results are from calculations with ONIOM(DFT/aug-cc-pVTZ:B97-D,M06L,B3LYP/LANL2DZ) scheme	80
Figure 3.6:	MADs of PBE0 vs. percentage of exact exchange where (a) the average MAD for each metal center and (b) the average MAD for each ligand. All of the results are from ONIOM(PBE0/aug-cc-pVTZ:B97-D/LANL2DZ) scheme with SMD.	83

Figure 3.7:	MADs in pK _a values of DFT and DFT-D3 with SMD relative to experiment, with respect to central TM atoms and ligand size of TM hydrides. The results are from calculations involving the ONIOM(DFT(-D3)/aug-cc-pVTZ:B97-D3/LANL2DZ) scheme		84
Figure 3.8:	MADs of DFT vs. DFT-D3 with SMD for the functionals in the model layer, i.e. ONIOM(DFT(-D3)/aug-cc-pVTZ:B97-D3/LANL2DZ). The MADs are averages of the full molecule set.		86
Figure 3.9:	Mean absolute deviation (MAD) in pK_a values when utilizing different basis sets relative to experiment, with respect to central TM atoms and ligand size of TM hydrides where (a) the cc-pVnZ and aug-cc-pVnZ (n=D,T) are considered for the model layer (HL method) and (b) LANL2DZ and SDD ECPs are considered for the real layer (LL method)		88
Figure 3.10:	Impact of radii models on (a) C-PCM, (b) COSMO, and (c) SMD. The default cavities for C-PCM, COSMO, and SMD are UFF, Klamt, and SMD-Coulomb, respectively. The average MADs are results from calculation with the ONIOM (PBE, M06L, B3LYP, M06/ aug-cc-pVTZ : B97D/ LANL2DZ) scheme and then categorized by metal and ligand.		90
Figure 3.11:	Comparison of the experimental and calculated pK_a values via methodological choices represented by their calculated values and the dotted trend lines. The dashed black line denotes the 1:1 correspondence between experiment and calculated pK_a values. Schemes A-E are ONIOM (B3LYP-D3/ aug-cc-pVTZ : B97-D3/ SDD), B97-D3/ SDD, B3LYP-D3/ SDD, ONIOM (B3LYP/ aug-cc-pVTZ : HF/ LANL2DZ), and ONIOM (CCSD(T)/ aug-cc-pVTZ : B97-D3/ SDD)		93
Figure 4.1:	Differences in electronic energies (mE_h) using Pipek-Mezey (PM) and Foster-Boys (FB) localization schemes using the def2/JK ABS within DLPNO-MP2 for complete basis set extrapolation using a combined Peterson-Schwartz-3 extrapolation scheme (PS3(TQ)). Included subsets are based on the presence of certain elements (hydrocarbons, halogenated, chalcogenated, pnictogenated, and Period 3) and electronic features (aromatic, carbonyl, multiple bonds) as well as the full molecule set. Points within the dashed lines represent differences less than 0.1 mE _h . The box plots depict the distribution of data within each subset where the band in the middle represents the median of the data and data points shown as black circles are more than 3 standard deviations from the median.	1	117

Figure 4.2:	Differences in electronic energies (mE_h) between the Pipek-Mezey (PM) and Foster-Boys (FB) localization methods for all three schemes within DLPNO- CCSD(T) for the same subsets in Figure 4.1. The dashed lines represent differences of less than 0.1 mE _h . The box plots depict the distribution of data within each subset where the band in the middle represents the median of the data and data points shown as black circles are more than 3 standard deviations from the median.	. 118
Figure 4.3:	Differences between the CCSD(T) electronic energies and the DLPNO-CCSD(T) electronic energies in mE_h using the (a) Pipek-Mezey (PM) and (b) Foster-Boys (FB) localization methods for all three schemes for the subsets of the full molecule set shown in Figure 4.1. The dashed lines represent differences less than 0.1 mE_h . The box plots depict the distribution of data within each subset where the band in the middle represents the median of the data and data points shown as black circles are more than 3 standard deviations from the median.	. 119
Figure 4.4:	CPU time of each individual step within (a) ccCA, (b) RI-ccCA, and (c) DLPNO-ccCA for selected species from the molecule set. The Other category represents the timing of the MP2/aug-cc-pVDZ, MP2/aug-cc-pVTZ, MP2/cc-pVTZ, and MP2/cc-pVTZ-DK calculations as these calculations use a small percentage of the total CPU time. All timing calculations were done with the ORCA software package.	. 125
Figure 4.5:	CPU time ratios of DLPNO-ccCA (FB) to (a) ccCA and (b) RI-ccCA. The ratios for Scheme 1 (blue circle), Scheme 2 (black x), and Scheme 3 (green triangle) are shown on a log-log scale. All timing was done with C_1 symmetry enforced and done in ORCA.	. 127
Figure 4.6:	CPU time for ccCA, RI-ccCA, and all 3 schemes for DLPNO-ccCA for the linear alkanes.	. 137
Figure 4.7:	Deviations in ΔH_f for ccCA, RI-ccCA, and all 3 schemes for DLPNO-ccCA for the linear alkanes using the isodesmic approach.	. 137
Figure 5.1:	Hydroformylation reaction converting olefins to linear and branched aldehydes via a Rh catalyst.	. 150
Figure 5.2:	A model of the two reaction pathways for hydroformylation where ΔE_l^{\ddagger} and ΔE_b^{\ddagger} are the reaction barriers for forming the linear and branched product, respectively. The energy difference between the two reaction barriers is denoted as $\Delta \Delta E^{\ddagger}$.	. 151

Figure 5.3:	Computationally determined 3D structures of ee-[Rh(H)(CO)(DIPHOS) (propene)] catalyst complex (top) and the dissociation reaction of H_2O from the cationic (diimine)(aquo)Pt ^{II} complex (bottom)
Figure 6.1:	Mean absolute deviation (MAD) of vibrational frequencies for diatomics using TPSS/VTZ (blue), B3LYP/VTZ (green), TPSS/aVTZ (purple), B3LYP/aVTZ (red), and ccCA-S4 (black)
Figure 6.2:	MAD of vibrational frequencies for H ₂ O, CO ₂ , and NH ₃ using TPSS/V <i>n</i> Z (blue), B3LYP/V <i>n</i> Z (green), TPSS/aV <i>n</i> Z (purple), B3LYP/aV <i>n</i> Z (red), and ccCA-S4 (black). For H ₂ O and CO ₂ , $n = \infty$. For NH ₃ , $n=T$
Figure 6.3:	MAD of vibrational frequencies for C_2H_2 , C_2H_4 , and C_2H_6 using TPSS/VTZ (blue), B3LYP/VTZ (green), TPSS/aVTZ (purple), B3LYP/aVTZ (red), and ccCA-S4 (black).
Figure 6.4:	Infrared spectra for cis-3-aminophenol (top) and trans-3-aminophenol (bottom) obtained with VCIPSI-PT2 frequencies with ccCA potentials and B3LYP/cc-pVTZ harmonic frequencies scaled by 1.0066. All intensities are from the harmonic frequency calculations. A Lorentz broadening of 20 cm ⁻¹ was applied. The experimental frequencies and relative intensities from Ref 8 are shown for comparison
Figure 6.5:	Single mode potential energy curves for vibrational modes 8 (left) and 10 (right) of ethene (C=C and C-H symmetric stretches) generated with ccCA (black) and TPSS/VTZ (red)
Figure 6.6:	Vibrational coupling map for ethene (left) and ethane (right). The vibrational mode couplings shown in black indicate strongly coupled modes that were used for all FASTVCI approaches using ccCA
Figure 6.7:	Vibrational coupling map for cis-3-aminophenol (left) and trans- 3-aminophenol (right). The vibrational mode couplings shown in black indicate strongly coupled modes that were used for all FASTVCI approaches using ccCA
Figure 7.1:	MO6-2X/6-31G* molecular electrostatic potential maps, and the frontier HOMO and LUMO of the optimized structures of (a) $(F_{15}P)Zn-C_{60}$ dyad and (b) C_{60} - $(F_{15}P)Zn$:Py-phTTF triad. The isovalue used for the MO depictions was 0.02 while the density value used was 0.0004
Figure 8.1:	Guest molecules for the cucurbit[8]uril (CB8) host
Figure 8.2:	Guest molecules for the octa-acid (OA) and tetramethyl octa-acid (TEMOA) hosts

Figure 8.3:	Host molecules: cucurbit[8]uril (CB8), octa-acid (OA), and tetramethyl octa- acid (TEMOA).	218
Figure 8.4:	Structures of the CB8 guest molecules inside the binding pocket generated from the clustering analysis.	223
Figure 8.5:	Structures of the OA guest molecules inside the binding pocket generated from the clustering analysis.	225
Figure 8.6:	Structures of the TEMOA guest molecules inside the binding pocket generated from the clustering analysis.	227
Figure 8.7:	Plots for calculated v. experimental results in kcal mol^{-1} for (a) CB8 (b) OA, and (c) TEMOA for MMPBSA (blue), RI-B3PW91-D3 (black), and RI-B3PW91 (green). The dashed lines in each corresponding color refers to the best fit line where the statistical outlier (OA-G2) is removed for (b) and (c). The dashed gray line is the y = x line.	232
Figure 8.8:	Error plots from experimental results in kcal mol ⁻¹ for (a) CB8 (b) OA, and (c) TEMOA for MMPBSA (blue), RI-B3PW91-D3 (black), and RI-B3PW91 (green).	233
Figure 8.9:	Plots for the correlations calculated after the mean signed errors are removed from the results in Tables 8.1-8.3 versus experimental results in kcal mol ^{-1} for (a) OA, and (b) TEMOA for MMPBSA (blue), RI-B3W91-D3 (black). The dashed lines in each corresponding color refers to the best fit line where the statistical outlier (OA-G2) for RI-B3PW91-D3 is removed for (a). The dashed gray line corresponds to the y=x line	244

LIST OF SCHEMES

Scheme 3.1:	The direct thermodynamic scheme		2
-------------	---------------------------------	--	---

KEY TO SYMBOLS AND ABBREVIATIONS

AO atomic orbital

aug-cc-pCVnZ augmented correlation consistent polarized core-valence n-tuple ζ basis set

aug-cc-pVnZ augmented correlation consistent polarized valence n-tuple ζ basis set

aug-cc-pVnZ-PP aug-cc-pVnZ with psuedopotential

CBS complete basis set

CB8 cucurbit[8]uril

- **cc-pCVnZ** correlation consistent polarized core-valence n-tuple ζ basis set
- **cc-pVnZ** correlation consistent polarized valence n-tuple ζ basis set

cc-pVnZ-PP cc-pVnZ with Psuedopotential

ccCA correlation consistent composite approach

CCSD(T) coupled cluster singles-and-doubles with perturbative triples correction

COSMO conductor-like screening model

C-PCM conductor-like polarizable continuum model

DFT density functional theory

DLPNO domain-based local pair natural orbital

 $\Delta \mathbf{H_f}$ enthalpy of formation

HF Hartree-Fock

- **HOMO** highest occupied molecular orbital
- LUMO lowest unoccupied molecular orbital
- MAD mean absolute deviation

MAE mean absolute error

MD molecular dynamics

MM molecular mechanics

MMPBSA molecular mechanics energy combined with Poisson-Boltzmann surface area continuum solvation method

MO molecular orbital

MPPT Møller-Plesset perturbation theory

MP2 Møller-Plesset second-order perturbation method

n-tuple n-multiple basis functions (n = 2, 3, 4, etc.)

OA octa-acid

PAO projected atomic orbitals

PCM polarizable continuum model

PES potential energy surface

PNO pair natural orbital

QM quantum mechanics

 \mathbf{r}^2 correlation coefficient

RI resolution-of-the-identity

RMSE root mean square error

SAMPL statistical assessment of the modeling of proteins and ligands

SCF self-consistent field

SMD solvation model based on density

 τ Kendall's tau

 τ_{crit} critical Kendall's tau

TEMOA tetramethyl octa-acid

TM transition metal

VCI vibrational configuration interaction

VPT2 vibrational second-order perturbation theory

VSCF vibrational self-consistent field

ZPE zero point energy

CHAPTER 1

INTRODUCTION

Computational chemistry emerged as a field of chemistry even before the digital age began. Seminal work of chemists, mathematicians, and physicists such as Erwin Schrödinger and Douglas R. Hartree, as well as Chemistry Nobel Laureates John Pople and Walter Kohn, throughout the 20th century constructed the foundation of computational chemistry even before the digital age began. ^{1–10} With the advent of computers, computational chemistry has become so important because it can provide valuable insight into chemical processes and properties that are difficult to measure experimentally and rationale for mechanistic features within known chemical reactions. There are many sub-fields of computational chemistry functional theory (DFT), semiempirical methods, and molecular dynamics. These approaches are used to investigate the vast numbers of chemical systems ranging from atoms to proteins and semiconductor materials like graphene and TiO₂. Among the more rigorous of these methods are *ab initio* approaches, or methods based on first principles, which focus on modeling the electronic structure of atoms and molecules but often at a high computational cost (memory, disk space, CPU time) relative to DFT and semiempirical methods.

As computing power doubled every two years, following Moore's Law observations, the usage and development of electronic structure methods to account for the rapid increase in computing power continues to grow, though the complexity of the mathematical approximations that still need to be made on modern computing resources inhibits that growth. For *ab initio* methods, numerous approximations and methods have been developed to reduce the computational cost and predict thermodynamic properties within the same range of error as their canonical counterparts.^{15–24}

In this dissertation, a focus is upon the development and integration of a number of strategies –called *ab initio* composite strategies– to enable cost reduction in the prediction of thermodynamic properties for both main group and transition metal species. As described in Chapter 3, the pK_as of

several Group 10 transition metal hydrides were predicted utilizing multilevel approaches due to the molecule size (Section 2.2.5) with DFT and *ab initio* methods, which provides insight into models needed for the prediction of thermodynamic properties of transition metal hydrides and other inorganic complexes that contain sterically bulky ligands.

While density functional methods are more commonly used, particularly for molecules of increasing size and complexity based on their low computational scaling, computations using *ab initio* methods can serve as an effective gauge for computational thermodynamic predictions. For example, *ab initio* composite strategies (see Section 2.2.4), which utilizes a combination of lower cost *ab initio* methods to effectively model a higher cost *ab initio* method at a fraction of the computational cost, can be used to reduce the cost associated with prediction of thermochemical properties. One such composite approach developed in the Wilson group is the correlation consistent Composite Approach, or ccCA, which has successfully been applied for main group and transition metal thermochemistry, predicting thermochemical properties like enthalpies of formation, pK_as , and bond dissociation energies within main group chemical accuracy (1 kcal mol⁻¹) and transition metal chemical accuracy (3 kcal mol⁻¹) on average.^{25–29}

The ccCA variant described in this dissertation in Chapter 4 utilizes the domain-based local pair natural orbital (DLPNO) methods within the ccCA framework for main group thermochemistry, denoted as DLPNO-ccCA, to expand the size limitations of *ab initio* composite methodologies. To evaluate the efficacy of DLPNO-ccCA for main group thermochemistry, the electronic energies and enthalpies of formation generated using ccCA, RI-ccCA, which uses the resolution-of-the-identity (RI) approximation within MP2 to reduce the computational cost, and DLPNO-ccCA are compared. DLPNO-ccCA was utilized for linear alkanes up to octane as well as molecular dimers exhibiting noncovalent interactions such as hydrogen bonding and dispersion including the coronene dimer (72 atoms), which is one of the largest molecule targeted with a composite approach to date. Therefore, showing the effectiveness of composite strategies and their usefulness for larger main group organic species.

DLPNO-ccCA was also used for transition metal species and organometallic complexes, as

shown in Chapter 5. This approach draws on the ccCA variants for the 3*d* and 4*d* transition metalcontaining species, ccCA-TM and rp-ccCA, to predict enthalpies of formation, gas phase ligand dissociation energies, and the regioselectivity of hydroformylation, the largest volume homogeneous chemical reaction in industry for chemical production. This study shows the applicability of *ab initio* composite methods for computational catalysis, which is typically analyzed with density functional methods.

Vibrational spectroscopy is an important approach for characterizing the structural and dynamical properties of molecules. Theoretical methods used for vibrational spectroscopy are often restricted to scaling frequencies within the harmonic approximation or utilizing potential energy surfaces (PES) generated with computationally costly *ab initio* methods that characterize vibrational motion. In Chapter 6, the correlation consistent Composite Approach (ccCA) and density functional theory (DFT) have been used to generate PES for polyatomic molecules (2-15 atoms). Frequencies, dipole moments, and infrared absorbance intensities are predicted in tandem with vibrational self-consistent field (VSCF) and post-VSCF theory to reduce the computational cost associated with generating PESs for anharmonic mode-mode couplings, calculating contributions from anharmonic corrections to the potential, and predicting vibrational frequencies within several cm⁻¹.

Additional research described in this dissertation includes a collaborative effort with Francis D'Souza at the University of North Texas (Chapter 7) to achieve the goal of artificial photosynthesis by modeling the donor-acceptor ability of zinc porphyrin-fullerene dyads and triads with DFT through modeling the frontier orbitals.

In Chapter 8, a combination of molecular dynamics, molecular mechanics, and density functional methods were used for the sixth Statistical Assessments of the Modeling of Proteins and Ligands (SAMPL) blind prediction challenge for host-guest binding. In this challenge, participants are expected to predict binding affinities and other properties for small molecules within a host system. The SAMPL challenge allows for the comparison of methods for binding affinity prediction by using statistical tools and modeling methods that can be essential for

host-guest systems. Empirical dispersion corrections, the RI approximation, and truncated basis sets were utilized to probe how electronic structure approaches that reduce the computational cost contribute to predicting binding affinities, which provides insight into favorable quantum chemical strategies for host-guest binding affinities.

With the wide range of applications, possible applications, and development presented in this dissertation, including biochemical processes, astrochemistry, artificial photosynthesis, and organometallic catalysis, the size and complexity of the molecules examined presents the challenges and successes of modeling organic, inorganic, and organometallic complexes effectively with DFT and *ab initio* composite strategies.

REFERENCES

REFERENCES

- [1] Schrödinger, E. Quantisierung als Eigenwertproblem. Ann. Phys. 1926, 384, 361–376.
- [2] Schrödinger, E. An undulatory theory of the mechanics of atoms and molecules. *Phys. Rev.* **1926**, 28, 1049–1070.
- [3] Hartree, D. R. The Wave Mechanics of an Atom with a non-Coulomb Central Field. Part III. Term Values and Intensities in Series in Optical Spectra. *Math. Proc. Cambridge Philos. Soc.* 1928, 24, 426–437.
- [4] Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Phys.* **1930**, *61*, 126–148.
- [5] Lennard-Jones, J.; Pople, J. A. The Molecular Orbital Theory of Chemical Valency. IV. The Significance of Equivalent Orbitals. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1950**, *202*, 166–180.
- [6] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. Phys. Rev. 1964, 136, B864–B871.
- [7] Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- [8] Honig, B.; Karplus, M. Implications of torsional potential of retinal isomers for visual excitation. *Nature* **1971**, *229*, 558–560.
- [9] Warshel, A.; Karplus, M. Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization. J. Am. Chem. Soc. 1972, 94, 5612–5625.
- [10] Warshel, A.; Karplus, M. Calculation of $\pi\pi^*$ Excited State Conformations and Vibronic Structure of Retinal and Related Molecules. *J. Am. Chem. Soc.* **1974**, *96*, 5677–5689.
- [11] Tsuzuki, S.; Uchimaru, T.; Tanabe, K. Ab Initio Calculations of Intermolecular Interaction Potentials of Corannulene Dimer. *J. Phys. Chem. A* **1998**, *102*, 740–743.
- [12] Kobayashi, R. A CCSD(T) Study of the Relative Stabilities of Cytosine Tautomers. J. Phys. Chem. A 1998, 102, 10813–10817.
- [13] Fox, S. J.; Dziedzic, J.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. Density functional theory calculations on entire proteins for free energies of binding: Application to a model polar binding site. *Proteins Struct. Funct. Bioinforma*. 2014, 82, 3335–3346.
- [14] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. J. Chem. Phys. 2013, 139, 134101.
- [15] Kendall, R. A.; Früchtl, H. A. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theor. Chem. Acc.* **1997**, 97, 158–163.

- [16] Almlöf, J. Elimination of energy denominators in Møller-Plesset perturbation theory by a Laplace transform approach. *Chem. Phys. Lett.* **1991**, *181*, 319–320.
- [17] Häser, M.; Almlöf, J. Laplace transform techniques in Møller-Plesset perturbation theory. J. *Chem. Phys.* **1992**, *96*, 489–494.
- [18] Wilson, A. K.; Almlöf, J. Møller-Plesset correlation energies in a localized orbital basis using a Laplace transform technique. *Theor. Chem. Acc.* **1997**, *95*, 49–62.
- [19] Lambrecht, D. S.; Doser, B.; Ochsenfeld, C. Rigorous integral screening for electron correlation methods. J. Chem. Phys. 2005, 123, 184102.
- [20] Schütz, M.; Hetzer, G.; Werner, H.-J. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. J. Chem. Phys. 1999, 111, 5691–5705.
- [21] Sæbø, S.; Tong, W.; Pulay, P. Efficient elimination of basis set superposition errors by the local correlation method: Accurate ab initio studies of the water dimer. J. Chem. Phys. 1993, 98, 2170–2175.
- [22] Handy, N. C.; Carter, S. Large vibrational variational calculations using 'multimode' and an iterative diagonalization technique. *Mol. Phys.* 2004, 102, 2201–2205.
- [23] Scribano, Y.; Benoit, D. M. Iterative active-space selection for vibrational configuration interaction calculations using a reduced-coupling VSCF basis. *Chem. Phys. Lett.* 2008, 458, 384–387.
- [24] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- [25] DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The correlation consistent composite approach (ccCA): An alternative to the Gaussian-n methods. J. Chem. Phys. 2006, 124, 114104.
- [26] DeYonker, N. J.; Wilson, B. R.; Pierpont, A. W.; Cundari, T. R.; Wilson, A. K. Towards the intrinsic error of the correlation consistent Composite Approach (ccCA). *Mol. Phys.* 2009, 107, 1107–1121.
- [27] Laury, M. L.; DeYonker, N. J.; Jiang, W.; Wilson, A. K. A pseudopotential-based composite method: The relativistic pseudopotential correlation consistent composite approach for molecules containing 4d transition metals (Y-Cd). J. Chem. Phys. 2011, 135, 214103.
- [28] Jiang, W.; DeYonker, N. J.; Determan, J. J.; Wilson, A. K. Toward accurate theoretical thermochemistry of first row transition metal complexes. J. Phys. Chem. A 2012, 116, 870– 885.
- [29] Riojas, A. G.; Wilson, A. K. Solv-ccCA: Implicit solvation and the correlation consistent composite approach for the determination of pK_a. J. Chem. Theory Comput. 2014, 10, 1500– 1510.

CHAPTER 2

THEORETICAL BACKGROUND

The fundamental equation of quantum mechanics is the time-independent Schrödinger equation,¹

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

in which finding an approximate solution is an integral part of computational chemistry. The Hamiltonian operator, \hat{H} , operates on the wavefunction describing the system of interest, Ψ , and returns an energy eigenvalue, E, for the wavefunction, which is an eigenfunction by definition. The Hamiltonian (\hat{H}) is the total energy operator that describes the interactions of the N electrons and the M nuclei with nuclear charge Z via the kinetic energy of the electrons (i, j) and nuclei (A, B), the nuclei-electron interactions at a distance r_{iA} , the electron-electron interactions at a distance r_{ij} , and the nuclei-nuclei interactions at a distance R_{AB} , as shown in Equation 2.2.

$$\hat{H} = \sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{A=1}^{M} \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^{N} \frac{Z_A}{r_{iA}} - \sum_{i=1}^{N} \sum_{j>1}^{N} \frac{1}{r_{ij}} + \sum_{A=1}^{N} \sum_{B>A}^{N} \frac{Z_A Z_B}{R_{AB}}$$
(2.2)

The difficulty with multi-electron systems is the inadequate description of the electron-electron interactions, which implies that the Schrödinger equation is only exactly solvable for one-electron systems. Therefore, approximations must be made to account for the electron-electron interactions present in chemical systems.

An important approximation is the Born-Oppenheimer approximation,² which assumes that the nuclei are stationary relative to the electrons in a system since the nuclei are much heavier than electrons and that electrons move faster than nuclei giving the appearance of stationary nuclei. Through this approximation, the Hamiltonian can be reduced to the electronic Hamiltonian,³ Equation 2.3.

$$\hat{H}_{elec} = -\sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{i=1}^{N} \frac{Z_A}{r_{iA}} - \sum_{i=1}^{N} \sum_{j>1}^{N} \frac{1}{r_{ij}}$$
(2.3)

Equation 2.3 represents the motion of N electrons in a field of M nuclei. The kinetic energy term for nuclei is approximated as zero. The nuclei-nuclei Coulombic energy term is a constant term, not integrated through all space and is thus removed from Equation 2.2. Using the electronic Hamiltonian results in the electronic Schrödinger equation, Equation 2.4.

$$H_{elec}\Psi_{elec} = E_{elec}\Psi_{elec} \tag{2.4}$$

The electronic wavefunction (Ψ_{elec}) is dependent only on the electron spatial coordinates; however, electrons have a spin component that is included in the overall wavefunction.³ Since the wavefunction should not be solely described by neither spatial nor spin components, a suitable principle that combines both descriptions of the electronic wavefunction is required. Starting with the antisymmetry principle, the electronic wavefunction must change signs with respect to electron exchange of both the spatial and spin coordinates, Equation 2.5.³

$$\Psi\left(\vec{x}_1,\ldots,\vec{x}_i,\ldots,\vec{x}_j,\ldots,\vec{x}_N\right) = -\Psi\left(\vec{x}_1,\ldots,\vec{x}_j,\ldots,\vec{x}_i,\ldots,\vec{x}_N\right)$$
(2.5)

Equation 2.5 shows an antisymmetric wavefunction with respect to the coordinate vectors for N electrons. The Hartree product, a many-electron wavefunction that considers non-interacting electrons, is shown in Equation 2.6.³

$$\Psi^{HP}(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N) = \chi_i(\vec{x}_1) \,\chi_j(\vec{x}_2) \cdots \chi_k(\vec{x}_N) \tag{2.6}$$

Equation 2.6 shows the Hartree product where χ is the k spin orbitals for N electrons and their respective spatial and spin coordinates. While the Hartree product does not satisfy the antisymmetry principle outlined in Equation 2.5, a linear combination of these Hartree products is generated to satisfy this principle. The generalized form of this linear combination is known as a Slater

determinant, shown in Equation 2.7a.

$$\Psi(\vec{x}_{1}, \vec{x}_{2}, \dots, \vec{x}_{N}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_{1}(\vec{x}_{1}) & \chi_{2}(\vec{x}_{1}) & \cdots & \chi_{k}(\vec{x}_{1}) \\ \chi_{1}(\vec{x}_{2}) & \chi_{2}(\vec{x}_{2}) & \cdots & \chi_{k}(\vec{x}_{2}) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_{1}(\vec{x}_{N}) & \chi_{2}(\vec{x}_{N}) & \cdots & \chi_{k}(\vec{x}_{N}) \end{vmatrix}$$
(2.7a)
$$= \frac{1}{\sqrt{N!}} \sum_{i}^{N!} (-1)^{p_{n}} \mathbf{P}_{i} \chi_{1}(\vec{x}_{1}) \cdots \chi_{k}(\vec{x}_{N})$$
(2.7b)

Equation 2.7a illustrates a Slater determinant where N is the number of electrons, χ is the k spin orbitals where k is equal to 2N for each position x, the columns are electron orbitals, and the rows represent electrons. In Equation 2.7b, the $\frac{1}{\sqrt{N!}}$ factor is a normalization constant, the $(-1)^{pn}$ represents the parity of the i^{th} term, and **P** is a permutation operator acting on the Hartree product, Equation 2.6. Slater determinants take advantage of the antisymmetry principle by representing a multi-electron wavefunction in the form of a determinant. Useful properties of a Slater determinant include the exchange of any two rows or columns, or interchanging two electrons, resulting in a change of sign of the determinant. Also, any two rows or columns that are identical, indicating two electrons with the same spin orbital, results in the wavefunction to be zero. Determinants can be used to satisfy the antisymmetry principle and the Pauli exclusion principle, respectively.

Other approximations used to solve the Schrödinger equation in molecular quantum chemistry include the combination of a theoretical method and basis set. This work focuses on two major classes of methods that are used to approximate the Schrödinger equation, wavefunction methods, or *ab initio* methods, and Density Functional Theory (DFT).

2.1 *Ab initio* methods

For *ab initio* methods, the fundamental approximation is the Hartree-Fock (HF) approximation, which averages the effects of electron-electron interactions through an average potential $\nu^{HF}(i)$.^{4,5} The Fock operator is an effective one-electron operator, shown in Equation 2.8.

$$f(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \nu^{HF}(i)$$
(2.8)

In Equation 2.8, $\nu^{HF}(i)$ is the average potential of electron *i* in the field of the other electrons in the system. Using the Fock operator reduces the multi-electron Schrödinger equation to numerous one-electron equations. Roothaan-Hall equations are utilized to make the HF equations into a matrix form as shown in Equation 2.9.

$$\mathbf{FC} = \varepsilon \mathbf{SC} \tag{2.9}$$

In Equation 2.9, **F** is the Fock matrix, **S** is the overlap matrix, **C** is the coefficient matrix, and ε is the orbital energy obtained from applying the Fock operator to a wavefunction. The elements of these matrices represent integrals involving basis functions defined by the linear combination of atomic orbitals, or LCAO, approximation, shown in Equation 2.10.

$$\Psi_i = \sum_{\mu=1}^{K} C_{\mu i} \phi_{\mu} \ i = 1, 2, \dots, K$$
(2.10)

Equation 2.10 shows the LCAO approximation through K basis functions, denoted by ϕ , representing the wavefunction of electron *i*. The elements of C are the coefficients $C_{\mu i}$ from Equation 2.10. The elements of **F** are shown in Equation 2.11.

$$F_{\mu\nu} = \int \phi_{\mu}^{*}(1) f(1) \phi_{\nu}(1) d\vec{r_{1}}$$
(2.11)

Equation 2.11 is the matrix representation of the Fock operator, f(1), with a set of basis functions ϕ_{μ} .³ The elements of **S** are shown in Equation 2.12.

$$S_{\mu\nu} = \int \phi_{\mu}^{*}(1) \,\phi_{\nu}(1) \,d\vec{r_{1}}$$
(2.12)

To solve these equations, an initial guess is proposed for the density matrix **P**, defined as C^*C , based on diagonalizing **S**, which in turn is used to generate **F**. Orthogonalizing and diagonalizing **F** results in a new guess for **C** and thus **P**. This procedure is iterated until the change in energy and the change in the density matrix is negligible, and thus is called the self-consistent field (SCF) procedure used to solve for the HF energy.³ The electron correlation energy is defined as the difference between the exact energy and the HF-calculated energy, as shown in Equation 2.13.

$$E_{corr} = E_{exact} - E_{HF} \tag{2.13}$$

While the correlation energy is a small percentage of the total electronic energy, it can have a large magnitude.⁶ Therefore, inclusion of the correlation energy is essential for the accurate prediction of chemical and physical properties. Post-HF methods recover the correlation energy not accounted by the HF method by adding excited determinants. One such method is many-body perturbation theory (MBPT), which utilizes a perturbation expansion with the Hartree-Fock Hamiltonian as the zeroth-order Hamiltonian.³ Using the Rayleigh-Schrödinger expansion of a generalized Hamiltonian, Equation 2.14, Møller and Plesset developed Møller-Plesset perturbation theory (MPPT).

$$\hat{H} = \hat{H}_0 + \lambda V \tag{2.14}$$

In Equation 2.14, is the corrected Hamiltonian, is the sum of one-electron Fock operators, λ is a dimensionless parameter between 0 and 1, and is the perturbation.^{3,7} The *n*th-order electronic energy is termed as the MP*n* methods. The MPPT Hamiltonian uses the Fock operator as the zeroth-order Hamiltonian (MP0), which double-counts electron repulsion when using the HF wavefunction. The MP1 energy eliminates one set of electron-electron interactions through using the operator. Therefore, the HF energy is the sum of the MP0 and MP1 energies.^{3,6,8} The first perturbation that accounts for corrections beyond HF is MP2, which adds a second-order correction and provides a size-extensive correction at a low cost, Equation 2.15.

$$E_1^{(2)} = \frac{V_{1n}V_{n1}}{E_1^{(0)} - E_n^{(0)}}$$
(2.15)

Equation 2.15 uses the operator from Equation 2.14 on all doubly excited determinants. Further expansions via MPPT are not used as often due to the high scaling and computational cost even though the corrections recover more correlation energy.⁶ Another approach to recovering correlation energy is the coupled-cluster method that employs a cluster operator, e^T , to the HF reference wavefunction.⁹ The cluster operator is defined as the sum of the single (\hat{T}_1) , double (\hat{T}_2) , triple $(\hat{T}_3), \ldots$, N-tuple (\hat{T}_N) excitation operator for N electrons, shown in Equation 2.16.

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \ldots + \hat{T}_N$$
 (2.16)

The operator, Equation 2.16, generates all i^{th} excited Slater determinants when operated on the HF reference wavefunction. The coupled cluster wavefunction is the resulting wavefunction after operating on the HF reference wavefunction with the exponential of the cluster operator, Equation 2.17.

$$e^{\hat{T}} = 1 + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right) + \left(\hat{T}_3 + \hat{T}_2\hat{T}_1 + \frac{1}{6}\hat{T}_1^3\right) + \dots$$
(2.17)

For the coupled-cluster methods, Equation 2.11 utilizes the Taylor expansion of the exponential of the cluster operator. This Taylor expansion generates the multiplicative terms $(\hat{T}_2\hat{T}_1)$ and product terms (\hat{T}_1^2) that help account for size inconsistency problems that occur in configuration interaction, or CI, methods.^{8–10} CI methods are defined by specifying configurations of the spin orbitals that construct each Slater determinant in reference to the HF wavefunction.³ By generating all possible excitations for N electrons, full CI is achieved, which is the exact answer to the electronic Schrödinger equation within a given basis set. Using the same excitation operator defined in Equation 2.16, CI wavefunctions can be generated using $(1+\hat{T})$ using intermediate normalization rather than as the excitation operator on the HF wavefunction to produce excited determinants.⁸ One of the more popular coupled cluster methods is CCSD(T), which uses coupled cluster with single excitations, double excitations, and perturbative triple excitations.¹¹ The scaling of a particular method is a key descriptor for determining computational cost. The Hartree-Fock method scales as N^4 , which relies on the relative size of the system, i.e. number of basis functions N.⁸ Post-HF methods, such as MP2 and CCSD(T), have a larger scaling due to the inclusion of correlation energy and more complex operators. MP2 scales at N^5 and CCSD(T) scales iteratively at N^7 .⁸

2.2 Cost-Saving Wavefunction-based Methods

With increasing molecule size and thus an increasing number of basis functions, wavefunction-based methods have decreased practicality. This section outlines approaches, e.g. local correlation methods, composite methods, and multilevel approaches, that reduce the computational cost associated with (CPU time, memory, disk space) higher level *ab initio* methodologies.
2.2.1 Local Methods

The molecular orbitals –canonical MOs– generated from the diagonalization of the Fock matrix are characteristically delocalized, even for smaller molecular systems. The short-range effect of dynamic correlation has a dependence on distance of r^{-6} like dispersion energy.¹² When using canonical MOs to describe electron correlation, its short-range aspect cannot be properly exploited, both in terms of gaining a more qualitatively correct picture of the electron correlation relative to localized MOs and reducing the high computational cost of *ab initio* methods.¹³ Hence, the development of local correlation methods.

The concept of localized MOs was first introduced by Lennard-Jones, Pople, and Hall.^{14–18} Since then, localization techniques^{19–23} and local correlation methods^{24–47} have been developed to utilize localized occupied MOs to localize the dynamic correlation. Localization imposes a mathematical constraint of maximum insensitivity for changes in distant nuclear charges, which allows orbitals to be localized around covalent bonds and atomic lone pairs. Localized MOs are generated through exploiting the invariance of the Hartree-Fock wave function with respect to orthogonal transformations, and are popular in their application to occupied orbitals. ^{19–23,48} For the occupied orbitals, the Foster-Boys (FB)^{19,20} and the Pipek-Mezey (PM)²³ localization techniques are more widely used compared to the Edmiston-Ruedenberg method.^{21–23,49} Both the Foster-Boys and Pipek-Mezey localization techniques scales as N^3 (N is the number of basis functions) since both approaches calculate one-electron dipole integrals and no two-electron integrals, whereas the Edmiston-Reudenberg scheme scales as N^5 , which is caused by the calculation of two-electron integrals.

The Foster-Boys localization approach minimizes the spatial extension of the MOs

$$f_{\rm FB}[\phi] = \sum_{i}^{N} |\langle \phi_i \phi_i | \mathbf{r}_1 - \mathbf{r}_2 | \phi_i \phi_i \rangle|^2$$
(2.18)

or equivalently maximizing the sum of squares between the distances of orbital centroids from the

origin of the coordinate system. ^{19,20}

$$f_{\rm FB}[\phi] = \sum_{i}^{N} |\langle \phi_i | \mathbf{r} | \phi_i \rangle|^2$$
(2.19)

The Pipek-Mezey localization approach uses the operator expectation value definition of the gross Mulliken orbital population to define the localized orbitals

$$f_{\rm PM} = \sum_{i=1}^{N} \sum_{A=1}^{n} (Q_{ii}^A)^2 = \sum_{i=1}^{N} \sum_{A=1}^{n} [\langle \phi_i | \hat{P}_\mu | \phi_i \rangle]^2$$
(2.20)

where the sum of atomic centers A and Q_{ii}^A is the population of orbital $|p\rangle$ on atomic center A.²³ The Hermitian operator \hat{P}_{μ} is defined as

$$\hat{P}_{\mu} = \hat{P}_{\mu}^{\dagger} = \frac{1}{2} (|\tilde{\mu}\rangle\langle\mu| + |\mu\rangle\langle\tilde{\mu}|)$$
(2.21)

where

$$|\widetilde{\mu}\rangle = \sum_{\nu} (\mathbf{S}^{-1})_{\nu\mu} |\nu\rangle \tag{2.22}$$

where **S** is the overlap matrix (Equation 2.12) and $\{|\tilde{\mu}\rangle\}$ are biorthonormal to the atomic basis functions $\{|\mu\rangle\}$. The localized MOs for Pipek-Mezey localization are obtained through maximizing Equation 2.20.²³ The Pipek-Mezey localization method is the use of Mulliken population analysis, which suffers from unphysical behavior, i.e. yielding occupation numbers for individual Mulliken charges that are larger than 1 or less than 0.^{23,48} The unphysical behavior is caused by overlap populations that occur from a non-orthogonal AO basis. An alternative Pipek-Mezey scheme utilizing the Löwdin population analysis has been developed to account for such deficiencies.⁵⁰ However, regardless of which population analysis method is used, the Pipek-Mezey localization properly separates σ - π bonds unlike the Foster-Boys localization approach.

For local correlation methods, the corresponding virtual orbitals are spanned by a set of projected atomic orbitals (PAOs) or pair natural orbitals (PNOs), ^{24,25,51,52} but have been spanned by localized virtual Hartree-Fock orbitals. ^{53,54}

2.2.2 Resolution of the Identity Approximation

The resolution of the identity (RI) approximation enables four-center two-electron repulsion integrals to be expressed as two- and three-centered electron repulsion integrals, reducing the computational cost of calculating the electron repulsion integrals from $O(\zeta^{12})$ to approximately $O(\zeta^9)$.^{55–57} The RI approximation involves the insertion of an approximate resolution of the identity into the Hilbert space of two interacting charge densities ρ and $\tilde{\rho}$, where

$$\rho_{ij} = ij \tag{2.23a}$$

for products of molecular orbitals i, j, and

$$\tilde{\rho}_{ij} = \sum c_{ij,P} P \tag{2.23b}$$

for a linear combination of auxiliary basis functions P with coefficients $c_{ii,P}$.⁵⁶

Through a minimization of the residual density, ρ - $\tilde{\rho}$, the four-center integrals are approximated using Equation 2.24, where the sum is over all functions within a fitted auxiliary basis set (see Section 2.4.3).

$$(ia|jb) = (ia|\hat{1}|jb) \approx \sum_{PQ} (ia|P)(P|Q)^{-1}(Q|jb)$$
 (2.24)

In Equation 2.24, (ai|P) and (P|Q) are the three- and two-center electron repulsion integrals, respectively, a, b denote virtual molecular orbitals, i, j denote occupied molecular orbitals, and P, Q denote auxiliary basis functions.⁵⁸ The RI method has been implemented in post-HF methods as well, reducing the cost of MP2 calculations while recovering a similar amount of correlation energy.⁵⁶ The RI approximation serves a key role in the integral generation for the domain-based local pair natural orbital methods (see section 2.2.3).

In a standard implementation of the RI approximation, the coulomb and exchange integrals are fitted to the auxiliary basis set. However, within the RI framework, variants that create auxiliary basis sets for the coulomb and exchange integrals individually have been developed. Variants included in Chapters 4 and 5 primarily focuses on the contributions from Neese et al., who developed a split-J approximation⁵⁹ and the chain of spheres exchange approximation (COSX)^{60,61} for the

coulomb and exchange portion of the Fock matrix, respectively. The Split RI-J algorithm was a modification of the RI approximation for the Coulomb interaction based on removing redundancies in calculating the Coulomb matrix. The COSX approximation utilizes a semi-numerical integration similar to Friesner's pseudo-spectral method to construct small 'chains' of shells of basis functions with contributions to the exchange matrix above a certain threshold. The combination of the split RI-J and COSX, or RIJCOSX, is mainly used on molecular systems exceeding 50 atoms including open-shell transition metal species due to computational efficiency. The RIJCOSX approximation for the HF wavefunction was considered for 20 closed shell reactions and 9 reactions and yielded a mean absolute error of 0.19 kcal mol⁻¹ with a maximum absolute error of 0.64 kcal mol⁻¹. The RIJCOSX wavefunction has been combined with the RI-MP2 method resulting in errors ranging from 0.01 to 1.2 mE_h for organic systems comprised of 30-57 atoms.⁶⁰ Because of its demonstrated utility, the RIJCOSX approximation is considered alongside the standard RI approximation for SCF energies.

2.2.3 Domain-Based Local Pair Natural Orbital Methods

The domain-based local pair natural orbital (DLPNO) methods, $^{62-66}$ primarily DLPNO-CCSD(T), have been shown to result in a reduced computational cost relative to the cost of CCSD(T) for transition metal-based catalysts and larger organic systems such as complex hydrocarbons and fullerenes while yielding electronic energies within 1 kcal mol⁻¹ from CCSD(T) electronic energies. $^{67-71}$ Recent developments of DLPNO methodology utilize sparse maps that take advantage of the sparsity of data by reducing the number of matrix elements stored and omitting terms in sums that are smaller than a predefined tolerance to achieve linear scaling with respect to *N* basis functions in all major computational steps including the integral transformation and storage, the initial guess, and the pair natural orbital (PNO) construction. 65,66,72 These are improvements to the original version of DLPNO-CCSD(T) that was developed by Riplinger et al, 62,63 which included terms that scaled by N^2 in the screening of electron pairs. DLPNO-MP2 recovers approximately 99.9% of the RI-MP2 correlation energy⁶⁴

and DLPNO-CCSD(T) recovers 92.2% of the canonical triples correction⁶³ and greater than 99.6% of the canonical CCSD(T) correlation energy.^{63,65}

The DLPNO methods use a single determinant reference wave function with the occupied molecular orbitals (MOs) localized. The Fock matrix is then constructed followed by the determination of the projected atomic orbitals (PAOs). The refined electron pair prescreening uses differential overlap integrals (DOI)

$$DOI_{ik} = \sqrt{\int |f_i(\mathbf{r})|^2 |g_k(\mathbf{r})|^2 d\mathbf{r}}$$
(2.25)

where f_i and g_k are basis functions (square integrable one-electron functions), which are calculated via numerical integration techniques to achieve linear scaling.⁶⁴ Domains are selected via normalized DOI between localized MOs and PAOs. This allows both the occupied and unoccupied spaces to be taken into account for domain definition as well as ensuring that all PAOs that have a significant differential overlap with occupied orbital *i* are included in the correlation domain.⁶⁴ This refined prescreening method eliminates fewer electron pairs than the previous implementation based on multipole estimates and pair correlation energies. The integral transformation from atomic orbital (AO) basis three-index integrals ($\mu\nu|K$) to ($i\tilde{\mu}|K$) follows, where μ , ν , λ , σ represent AOs, K, L refer to ABS, i, j, k, l denotes localized MOs, and $\tilde{\mu}$, $\tilde{\nu}$ label the PAOs. The local exchange operators, $K_{\tilde{\mu}\tilde{\nu}}^{ij} = (i\tilde{\mu}|j\tilde{\nu})$, are then constructed to find the semi-canonical local MP2 (SC-LMP2) or local MP2 (LMP2) guess amplitudes

$$T^{ij}_{\tilde{\mu}\tilde{\nu}} = -\frac{K^{ij}_{\tilde{\mu}\tilde{\nu}}}{\varepsilon_{\tilde{\mu}} + \varepsilon_{\tilde{\nu}} - F_{ii} - F_{jj}}$$
(2.26)

where $\varepsilon_{\tilde{\mu}}$ and $\varepsilon_{\tilde{\nu}}$ are energies of quasi-canonical, non-redundant PAOs, and F_{ii} and F_{jj} are diagonal Fock matrix elements in terms of localized orbitals.21,22 These guess amplitudes generate the pair density

$$\mathbf{D}^{ij} = \mathbf{T}^{ij}\tilde{\mathbf{T}}^{ij+} + \mathbf{T}^{ij+}\tilde{\mathbf{T}}^{ij}$$
(2.27)

from which the PNOs are approximated via diagonalization.

Beyond these steps, the DLPNO-MP2 and DLPNO-CCSD(T) methods separate as the DLPNO-CCSD(T) method uses SC-LMP2 as a crude guess for which electron pairs (ij) to include in a subsequent SC-LMP2 calculation and coupled cluster iterations whereas DLPNO-MP2 uses the LMP2 framework to calculate total energies in one iteration. For DLPNO-CCSD(T), an estimate of the pair correlation energy ε_{ij} is computed from the electron pairs that survived the prescreening, which are separated into three classes based on the dimensionless parameters $T_{CutPairs}$ and $T_{CutPairsMP2}$. The first class consists of pairs that are included in CCSD calculations (strong pairs), the second class contains pairs for MP2 that are kept for the triples correction (weak pairs), and the third class includes pairs that are not considered further. Pairs in the first two classes are used in a more accurate SC-LMP2 calculation. Pair correlation energy estimates from the third class are added to the SC-LMP2 energy to approximate the error introduced by local approximations.⁶⁵

The scaling reduction from N^7 to N^5 of DLPNO-CCSD(T) lies in the contribution of orbital triples to the triples energy.⁶³ Orbital triples (ijk) only contribute to the triples energy if the three pairs (ij), (ik) and (jk) survive the pair selection process. The domain for (ijk) is the union of the individual orbital domains (i), (j), and (k). Like the generation of PNOs, the triples natural orbitals (TNOs) are computed from diagonalizing the average of the pair densities for all three pairs (Equation 2.27) generated from a local, non-redundant PAO basis formed from redundant PAOs that span the triples domain. The integrals for TNOs are calculated via the RI approximation through transformations from the redundant PAO basis to the TNO basis for subsets of three-index electron repulsion integrals. The singles and doubles amplitudes are projected into the TNO basis through the triples/pair overlap. Intermediates that enter the triples calculation as well as the actual contribution of a given triple are calculated by the canonical (T) correction.⁶³ (The original publications provide more details about DLPNO methods and their development.^{60,63–65,73,74})

2.2.4 Composite Methods

Ab initio composite methods have been around for a long time and are effective approaches to reduce computational cost. *Ab initio* composite strategies are those that emulate a more computationally demanding method by adding up contributions that describe aspects of modeling quantum chemistry, like relativity, spin-orbit, and contributions from electrons beneath the

valence shell (sub-valence electrons), all of which increase the computational cost of *ab initio* methods. This additive approach can save greater than 90% of the total CPU time. In the case of CO, using a composite strategy only took 2 minutes whereas using the effective *ab initio* method a composite strategy portrays took over 8 hours!

Popularized by John Pople in 1989, *ab initio* composite methods targeted main group thermochemistry with the goal of estimating energies yielded by *ab initio* methods that would require significant computational resources efficiently.⁷⁵ Throughout the 1990s, other composite approaches and improvements of Pople's composite strategies targeting chemical and spectroscopic accuracy, which is defined as 1 kcal mol⁻¹ and 1 kJ mol⁻¹, respectively, appeared and targeted atoms and small polyatomic molecules with more than six non-hydrogen atoms.^{72,76–85} In the early 2000s, simultaneous developments of composite strategies targeting chemical ^{86–90} and spectroscopic ^{91–95} accuracy emerged, primarily for applications in main group thermochemistry. With the present computing power, composite methodologies have targeted molecules as large as buckminsterfullerene (C₆₀).⁹⁶

Though *ab initio* composite methods have primarily targeted main group thermochemistry due to the abundance of well-established reliable experiments, there have been limited studies in the development and application of *ab initio* composite methods towards transition metal and f-block thermochemistry as reliable experimental data is sparse.^{97–107} As well, the chemistry in this region of the periodic table becomes increasingly more complex due to the *d* and *f* electrons, which contribute to low-lying excited states, i.e. excited states that are close in energy to the ground state electronic configuration, as well as the significant effects of relativity and spin-orbit coupling. When formulating composite strategies for transition metal and *f*-block thermochemistry, all of these factors require the usage of relativistic basis sets and pseudopotentials (see Section 2.4.2), multireference methods that accounts for low-lying excited states, and/or a relativistic Hamiltonian that includes spin-orbit coupling and scalar relativity, and if needed, higher orders of relativity. Therefore, the increase in chemical complexity down the periodic table leads to the development of variants with respect to composite methodologies implemented for main group

thermochemistry.^{97–107}

Ab initio composite approaches are utilized to predict thermochemical properties like ΔH_{f} , ionization potentials, and bond dissociation energies (BDE) within chemical accuracy (1 kcal mol⁻¹) and spectroscopic accuracy (1 kJ mol⁻¹) reliably throughout the periodic table at a computational cost significantly less than the effective level of theory. Based on their success for predicting thermochemical properties for main group, transition metal, and to an extent *f*-block molecules, *ab initio* composite strategies could be used as a gateway to understand reaction mechanisms, design catalysts, and characterize heavier elements utilizing wavefunction-based methods.

Ab initio composite approaches are comprised of a reference energy and additive corrections to the reference energy that account for effects like relativity, spin-orbit, the inclusion of interactions between sub-valence and valence electrons, and the energy contributions from numerous simultaneously excited electrons. For the reference, reliable electronic structures (electron configuration and geometry) and scaling vibrational contributions to the energy are needed.¹⁰⁸ Some composite methods utilize higher cost methods to correct for anharmonicity through perturbative methods.^{109,110}

Using the optimized structures, single point calculations are done to obtain the reference energy. This is generally attained through one to three single point calculations involving an *ab initio* method such as MP2, MP4, or CCSD(T) (see Section 2.1) and increasingly larger basis sets (see Section 2.4). While composite methods have utilized one single calculation to obtain a reference energy, most will utilize two or three calculations where the method is the same and the basis sets increase in size. The intention is to find the energy attained with an infinitely large description of the electron space. This is also known as the complete basis set (CBS) limit. Composite schemes that target the CBS limit for the reference energy typically utilize the correlation consistent basis sets to extrapolate to the CBS limit via analytic formulas based on basis set size ^{111,112} and maximum angular momentum.^{113,114}

Additive corrections are then included to supplement the reference energy in a methodical

fashion. These corrections to the energy are arising from scalar relativity, spin orbit (from experimental atomic values), and interactions between valence and sub-valence electrons. When these corrections are added to the reference energy, the total electronic energy is obtained at a much lower computational cost than a calculation for the effective level of theory a composite method achieves.

Many *ab initio* composite methods have been developed targeting either chemical accuracy or spectroscopic accuracy. Composite methodologies that target chemical accuracy include the Gaussian-n (Gn) methods, ^{75–79,89,90,115} Complete Basis Set (CBS-X) method, ^{72,80–84,86,88} and the correlation consistent Composite Approach (ccCA). ^{87,97,116–122} Classes of composite methodologies targeting spectroscopic accuracy for thermochemical properties of diatomics and very small polyatomic molecules (roughly 3-10 atoms) due to the high computational cost of the methods involved include the Weizmann-n (Wn) methods, ^{85,94,95,123–125} High Accuracy Extrapolated *Ab initio* Thermochemistry (HEAT), ^{92,126,127} Feller-Peterson-Dixon (FPD) method, ^{128,129} and the focal point analysis method. ^{130–135}

Composite methodologies that target chemical accuracy are more efficient that those that target spectroscopic accuracy and allow larger molecules to be studied. ^{136–143} There are many variants of composite methodologies including ones modified to describe aqueous phase chemistry or expand to larger molecules.

2.2.4.1 Correlation Consistent Composite Approach

The correlation consistent Composite Approach (ccCA) was created in 2006 by Wilson and co-workers as an alternative to the Gn methods.⁸⁷ While successful for *s*-block and *p*-block thermochemistry in the first four periods, $^{144-146}$ methodological adjustments were made to the ccCA formulation in 2009, which included scaling vibrational contributions and options for the extrapolation scheme for the CBS limit.¹¹⁷

The formulation of ccCA is

$$E_{ccCA} = E_{ref} + \Delta E_{CC} + \Delta E_{CV} + \Delta E_{DK} + \Delta E_{SO} + \Delta E_{ZPE}$$
(2.28)

where E_{ref} is obtained at the MP2/CBS level by combining CBS extrapolations for the SCF energies and MP2 correlation energies with the aug-cc-pVNZ basis sets. The SCF energy is extrapolated with the Feller^{147,148} two-point extrapolation scheme

$$E(n) = E_{\text{HF/CBS}} + Be^{-Cn} \tag{2.29}$$

where *n* indicates the ζ -level of the basis set, E(n) is the energy at the $n^{th} \zeta$ -level, $E_{\text{HF/CBS}}$ represents the Hartree-Fock electronic energy at the CBS limit, *B* is a fitting parameter, and 1.63 is used for *C*.¹¹¹ To extrapolate the MP2 correlation energies, previous ccCA studies considered several different extrapolation schemes, including Peterson's three-point extrapolation scheme¹⁴⁹

$$E(n) = E_{\text{MP2/CBS}} + Be^{-(n-1)} + Ce^{-(n-1)^2}$$
(2.30)

where $E_{\text{MP2/CBS}}$ represents the electronic energy at the CBS limit, and *B* and *C* are fitting parameters. The Peterson (P) three-point extrapolation uses the double-, triple-, and quadruple- ζ correlation consistent basis sets. Other extrapolation schemes used in this work include inverse cubic and quartic equations, commonly referred to as the Schwartz-3 (S3) and Schwartz-4 (S4) two-point extrapolation schemes, respectively.^{112,113,150–153}

$$E(l_{max}) = E_{\text{MP2/CBS}} + \frac{B}{(l_{max})^3}$$
(2.31)

$$E(l_{max}) = E_{\text{MP2/CBS}} + \frac{B}{\left(l_{max} + \frac{1}{2}\right)^4}$$
(2.32)

In Equations 2.31 and 2.32, l_{max} is the highest angular momentum function included in the basis set, which differs for main group and transition metals. Both of the S3 and S4 two-point extrapolation schemes use the l_{max} of the triple- and quadruple- ζ level basis sets, denoted as S3(TQ) and S4(TQ). Since the S3 scheme tends to overestimate the energy at the CBS limit due to a slower convergence rate and the Peterson scheme tends to underestimate the energy at the CBS limit due to more rapid convergence, the average of both schemes, denoted PS3(TQ), is considered. ^{117,154} The core-core (CC) correlation (ΔE_{CC}) accounts for higher levels of correlation beyond the MP2

level by using CCSD(T) at the cc-pVTZ level.

$$\Delta E_{CC} = E \left[\text{CCSD(T)/cc-pVTZ} \right] - E \left[\text{MP2/cc-pVTZ} \right]$$
(2.33)

The core-valence (CV) correction accounts for the n and n-1 orbital shells, where $n \ge 2$. This correction accounts for the interactions between valence and sub-valence electrons whereas the other composite steps only include valence-valence interactions. The FC1 notation indicates the inclusion of the n-1 orbital shell.

$$\Delta E_{CV} = E \left[\text{MP2(FC1)/aug-cc-pCVTZ} \right] - E \left[\text{MP2/aug-cc-pVTZ} \right]$$
(2.34)

The scalar relativistic correction uses the second-order spin-free Douglas Kroll Hess Hamiltonian to account for scalar relativistic effects.^{155–157}

$$\Delta E_{DK} = E \left[\text{MP2/cc-pVTZ-DK} \right] - E \left[\text{MP2/cc-pVTZ} \right]$$
(2.35)

Experimental spin-orbit corrections for atoms are applied from tables provided by Moore.¹⁵⁸

This formalism is used as the base model and is altered to accommodate the need for relativistic corrections and effective core potentials for transition metals. For ccCA-TM, ¹¹⁶ developed for 3*d* transition metals, the modifications from ccCA include the use of scalar relativistic basis sets and the use of CCSD(T) and an augmented double- ζ core-valence basis set. For rp-ccCA, ⁹⁷ developed for 4*d* transition metals, effective core potentials (ECPs) are used in all steps of the ccCA-TM formulation. Variants where the fundamental aspects of ccCA remains the same but the steps are modified have been developed to adapt to chemical problems as well, such as organic acid/base chemistry (Solv-ccCA), ¹²² active-site chemistry (ONIOM-ccCA, ONIOM-rp-ccCA), ^{119,159} and modeling open-shell organic species, such as radicals (MR-ccCA, ccCA-CC(2,3)). ^{118,120} For example, in Solv-ccCA, all methodological steps within ccCA remain the same except for including an implicit solvent model to describe long-range solvent effects.

	ccCA-TM	rp-ccCA
Geometry Optimization	B3LYP/cc-pVTZ-DK	B3LYP/cc-pVTZ-PP
E _{ref}	HF/aug-cc-pVTZ-DK HF/aug-cc-pVQZ-DK	HF/aug-cc-pVTZ-PP HF/aug-cc-pVQZ-PP
Extrapolations	Equation 2.29	Equation 2.29
MP2/CBS	MP2/aug-cc-pVDZ-DK MP2/aug-cc-pVTZ-DK MP2/aug-cc-pVQZ-DK	MP2/aug-cc-pVDZ-PP MP2/aug-cc-pVTZ-PP MP2/aug-cc-pVQZ-PP
Extrapolations	Equations 2.30 -2.32	Equations 2.30 - 2.32
ΔCC	CCSD(T)/cc-pVTZ-DK - MP2/cc-pVTZ-DK	CCSD(T)/cc-pVTZ-PP - MP2/cc-pVTZ-PP
ΔCV	CCSD(T,FC1)/aug-cc-pCVDZ-DK – CCSD(T)/aug-cc-pCVDZ-DK	CCSD(T,FC1)/aug-cc-pCVDZ-PP - CCSD(T)/aug-cc-pCVDZ-PP
ΔDK	Included in previous steps	Included in previous steps
Δ SO ZPE	Experimental atomic values Vibrational ZPE scaled by 0.989	Experimental atomic values Vibrational ZPE scaled by 0.989

Table 2.1: Summary of ccCA-TM and rp-ccCA steps.

Numerous routes have been utilized to reduce the computational cost associated with ccCA to expand the range of molecules in terms of size that can be examined with this approach.^{119,121,159–161} RI-ccCA and ccCA-F12 implemented mathematical approximations to mitigate the cost of calculating four-center-two-electron repulsion integrals and using the aug-cc-pVQZ basis set, which both are major contributions to the overall computational cost of ccCA.^{121,161}

ccCA and its adaptations are suitable for applications targeting chemical accuracy for chemical systems ranging from atoms and diatomics to organometallic complexes and biomolecules. Some of these applications are presented in Chapters 4-6.

2.2.5 **ONIOM**

Multilayer methods provide additional routes to reduce computational cost. For these approaches, a molecular syste is divided into multiple layers and each layer is treated with a

different theoretical approach. This enables the chemistry of greatest interest to be targeted with a high-level method, while the overall molecular system is treated with a more approximate, albeit more efficient, approach. One of the earlier uses of multilayer methods combined quantum mechanical (QM) methods with molecular mechanics (MM) force fields to measure the torsional potential energy surface of the retinal molecule.¹⁶² The use of this hybrid methodology was extended to describe ground and excited-state potential energy surfaces in tandem with a Pariser-Parr-Pople SCF-CI method^{163,164} for π electrons and empirical functions for σ electrons, respectively.^{165,166} This method was later generalized into the QM/MM method, which includes a model system and the real system.¹⁶⁷ The model system describes the chemically significant portion of the system and uses QM methods for higher accuracy, whereas the real system is described by a less accurate but more computationally efficient MM force field. The total energy of the whole system is shown in Equation 2.36.

$$E_{\rm QM/MM} = E_{\rm QM} + E_{\rm MM} + E_{\rm QM-MM}$$
(2.36)

Equation 2.36 is an additive scheme ¹⁶⁸ combining the energy of the two systems, $E_{\rm QM}$ and $E_{\rm MM}$, and the energy of the interaction between the two systems, $E_{\rm QM-MM}$. In contrast to this additive scheme employed for the QM/MM method, our Own N-Integrated molecular Orbital molecular Mechanics, or ONIOM^{169–178} method is an extrapolative scheme that can utilize a QM/QM or a QM/MM scheme. The development of the ONIOM methodology started with the development of an alternative QM/MM scheme known as IMOMM, or Integrated Molecular Orbital + Molecular Mechanics, shown in Equation 2.37.¹⁶⁹

$$E_{\text{IMOMM}} = E_{\text{ONIOM2}(\text{QM:MM})} = E_{\text{QM,model}} + E_{\text{MM,real}} - E_{\text{MM,model}}$$
(2.37)

In Equation 2.37, the total energy of this extrapolative scheme is evaluated as the MM method for the model system is subtracted from the sum of the energies obtained through the QM method for the model system and the MM method for the real system. The main difference between $E_{\text{QM/MM}}$ and E_{IMOMM} is that the subtractive operation for E_{IMOMM} removes the doubly-counted MM contributions to the total energy in Equation 2.36.^{169,179} The IMOMM scheme was extended to QM/QM systems in the Integrated Molecular Orbital + Molecular Orbital formalism, which is denoted as IMOMO or ONIOM2(QM1:QM2).¹⁷² The total energy of the system is calculated in the same manner as the IMOMM method except for the use of a second QM method replacing the MM force field, shown in Equation 2.38.

$$E_{\text{IMOMO}} = E_{\text{ONIOM2}(\text{QM1:QM2})} = E_{\text{QM1,model}} + E_{\text{QM2,real}} - E_{\text{QM2,model}}$$
(2.38)

ONIOM is not limited to two layer systems. A combination of the IMOMM and the IMOMO methods yield a three-layer ONIOM method denoted as ONIOM3(QM1:QM2:MM) as utilized in Equation 2.39.¹⁷⁵

$$E_{\text{ONIOM3}(\text{QM1:QM2:MM})} = E_{\text{QM1,model}} + E_{\text{QM2,intermediate}} - E_{\text{QM2,model}} + E_{\text{MM,real}} - E_{\text{MM,intermediate}}$$
(2.39)

For ONIOM3, three layers, model, intermediate, and real with a different level of theory used to describe each layer. In general, the high level method, QM1, is an *ab initio* method, the intermediate level method, QM2, is a DFT method, and the real level method is a MM force field. Based on the formulations for ONIOM2 and ONIOM3, the ONIOM method can be generalized to an arbitrary *n*-layer *n*-level method, Equation 2.40.

$$E_{\text{ONIOMn}} = \sum_{i=1}^{n} E[\text{level}(i), \text{model}(n+1-i)] - \sum_{i=2}^{n} E[\text{level}(i), \text{model}(n+2-i)]$$
(2.40)

The n=2 (ONIOM2) and n=3 (ONIOM3) forms of n-layer ONIOM are most commonly used, as $n \ge 3$ approaches become impractical. Overall, the ONIOM method is most commonly used for large biological macromolecules, ^{162,165,166} transition metal complexes, ^{180,181} and organometallic catalysts. ^{182–185}

2.3 Density Functional Theory

Density functional theory (DFT) originates from the Hohenberg-Kohn theorems.^{186,187} In 1964, an existence proof showed that the charge density ($\rho[r]$) determines the electronic properties of the ground state including energy. DFT utilizes the electron density as a variable to approximate the solution to the Schrödinger equation. Analogous to the Roothaan-Hall Equations in the Hartree-Fock formalism, the DFT equivalent –the Kohn-Sham equations– were derived by Kohn and Sham in the early 1960s.^{186,187} The DFT energy is shown in Equation 2.41.

$$E[\rho] = T_s[\rho] + V_{ne}[\rho] + J[\rho] + E_{xc}[\rho]$$
(2.41)

Equation 2.41 is dependent on the kinetic energy (T_s) of non-interacting electrons, the energy term for nuclear-electron interactions (V_{ne}) , electron-electron repulsion interactions (J), and the exchange-correlation energy term (E_{xc}) . In principle, the exact form of the exchange-correlation functional makes DFT an exact and *ab initio* method; however, the exact form of the exchangecorrelation functional is not known based on the inhomogeneity of the charge density.¹⁸⁸ Therefore, the implementation of DFT is the development of functionals that approximate the exchangecorrelation functional. Density functionals are sorted into a hierarchy based on the complexity of the functional. As defined by Perdew as the "Jacob's ladder" for DFT, the tiers of functionals from least to most complex are the local spin density approximation (LDA), the generalized gradient approximation (GGA), meta-GGA, hybrid-GGA, hybrid-meta GGA, and double hybrid GGA.¹⁸⁹

The local spin density approximation (LDA or LSDA) is based on the uniform electron gas model and was first introduced by Kohn and Sham.¹⁸⁷ LDA uses the exchange for the uniform electron gas to create a functional solely dependent on the spin density.¹⁹⁰

$$E_{XC}^{LDA} = \int \rho(\mathbf{r}) \varepsilon_{XC}[\rho] \mathbf{r}$$
(2.42)

Equation 2.42 represents the exchange-correlation (XC) energy for LDA functionals, which is dependent on the single particle density $\rho(\mathbf{r})$ and the XC energy per particle, $\varepsilon_{XC}[\rho(\mathbf{r})]$. LDA is known to be more effective at describing solid state lattice parameters than more complex DFT functionals due to the similarity of metallic systems to the homogeneous electron gas.^{188,191,192} GGA functionals incorporate the gradients of the spin densities in the expression for exchangecorrelation energy and are therefore a correction to the LDA, shown in Equation 2.43.

$$E_{XC}^{GGA} = \int e_{XC}^{GGA}(n_{\uparrow}(\mathbf{r}), n_{\downarrow}(\mathbf{r}), |\nabla n_{\uparrow}(\mathbf{r})|, |\nabla n_{\downarrow}(\mathbf{r})|) d^{3}\mathbf{r}$$
(2.43)

Meta-GGA functionals, which decrease the amount of self-interaction error introduced by GGA functionals, use the Laplacian of the spin densities as two additional variables and the kinetic energy densities, τ , as shown in Equation 2.44.^{193,194}

$$E_{XC}^{M-GGA} = \int e_{XC}^{GGA}(n_{\uparrow}(\mathbf{r}), n_{\downarrow}(\mathbf{r}), |\nabla n_{\uparrow}(\mathbf{r})|, |\nabla n_{\downarrow}(\mathbf{r})|, |\nabla^{2}n_{\uparrow}(\mathbf{r})|, |\nabla^{2}n_{\downarrow}(\mathbf{r})|, \tau_{\uparrow}, \tau_{\downarrow})d^{3}\mathbf{r}$$
(2.44)

LDA, GGA, and meta-GGA functionals are referred to as local functionals because the electronic energy density at a single point is dependent on the behavior of the density in proximity to that point. ^{193,195–197}

Hybrid functionals combine the GGA exchange correlation functional with the exact exchange defined in the Hartree-Fock method using the Kohn-Sham orbitals in order to address the shortcomings of the self-exchange of DFT functionals as shown in Equation 2.45.

$$E_{XC}^{hybrid} = a(E_{X,exact} - E_X^{GGA}) + E_{XC}^{GGA}$$
(2.45)

Equation 2.45 is applied to both GGA and meta-GGA functionals and thus named hybrid-GGA and hybrid-meta-GGA functionals, respectively. Double hybrid functionals utilize the PT2 correlation energy into the correlation functional. ^{198,199} Due to the addition of a percentage of exact exchange into the functional, hybrid-GGA, hybrid-meta-GGA, and double-hybrid functionals are referred to as non-local functionals.

Based on the "Jacob's ladder" model for DFT by Perdew,¹⁸⁹ for each rung of the ladder, additional "factors" are appended to the functionals of the rung below, as illustrated from Equations 2.42-2.45. As a result, the increasing complexity of functionals progressing up Jacob's ladder implies an assumption of greater accuracy. However, greater accuracy cannot be presumed, i.e. local functionals may be more effective at describing a system than non-local functionals. Therefore, the rational choice of DFT functionals should be determined by carefully considering the calibration of DFT functionals with experiments or high accuracy *ab initio* methods for a particular application.

DFT is able to yield results for thermodynamic properties comparable to post-HF methods at a reduced computational cost as DFT scales at approximately N^4 or N^5 depending on the complexity of the functional where *N* is the number of basis functions. DFT has the inability to properly account for the weak interactions due to dispersion forces that arises from local exchange-correlation, and systems with long-range interactions–dissociation of radials and other charged odd-electron systems, and self-interaction error. Also, the exchange-correlation functional is local, which is unsuitable for charge transfer reactions. Attempts have been made to overcome the inability to accurately describe long-range interactions through dispersion-corrected density functionals (DFT-D methods),^{200,201} which use a semi-empirical parameterization to correct for the lack of dispersion, and the double-hybrid-GGA functionals.

2.4 Basis Sets

A basis set consists of mathematical functions that are used to describe the electronic wavefunction. Gaussian basis functions, shown in Equation 2.46, are the most common functions used in basis sets.

$$\phi(\zeta, r) = N e^{-\zeta r^2} \tag{2.46}$$

For the gaussian-type orbital (GTO) or a Gaussian primitive (Equation 2.46), N is the normalization constant, ζ is the exponent and r is the electron-nucleus distance. Gaussian-type functions were chosen since the product of two GTOs is another GTO, which greatly simplifies calculating the four-center two-electron repulsion integrals – the most computationally expensive step in the SCF procedure (Section 2.1).

A linear combination of these gaussian primitives (Equation 2.10) minimizes the number of basis functions needed for an accurate representation of the MOs. Gaussian basis sets are designed in hierarchies of increasing size (ζ -level). While increasing the ζ -level of a basis set increases the computational cost, a systematic way to obtain higher quality results is attained. Basis sets commonly utilized in electronic structure calculations are atom-centered and energy-optimized, i.e. the exponents are optimized to minimize the electronic energy, thus allowing a more widely applicable basis set.

Two popular styles of basis sets include the Pople-style basis sets developed by Pople^{202–205}

and the correlation consistent basis sets developed by Dunning and co-workers.^{206–215}

2.4.1 Correlation Consistent Basis Sets

The correlation consistent basis sets are referred to as correlation consistent, polarized, valence, $n-\zeta$, or cc-pV*n*Z where $n = double-\zeta$ (DZ), triple- ζ (TZ), quadruple- ζ (QZ), etc. level of basis set. The correlation consistent basis sets can be augmented through the addition of low-exponent diffuse functions, noted as the aug-cc-pV*n*Z basis sets. The correlation consistent family of basis sets also includes cc-pCV*n*Z basis sets that account for the correlation energy from the interaction of core-core and core-valence electrons as well as the valence-valence correlation energy, ^{209,213} and the cc-pV*n*Z-DK set accounts for scalar relativistic effects and is implemented for main group, 3*d* transition metal, and lanthanide atoms.^{212,214–216} The cc-pV(*n*+*d*)Z basis sets were developed for second-row atoms (Al-Ar) through the inclusion of an additional tight-*d* function and reoptimization of the *d*-function in the basis set to address deficiency in the original correlation consistent basis sets.²¹¹

For *ab initio* methods, one of the main advantages to these basis sets is their unique construction, which enables the extrapolation of some properties like energies, to the complete basis set (CBS) limit, ¹⁴⁷ which eliminates the basis set incompleteness error. At the CBS limit, the electronic energy is not changed by the addition of extra basis functions since the basis set completely spans the space of molecular orbitals, making an infinite or complete basis set.

2.4.2 Effective Core Potentials

When describing chemical systems with elements beyond the first-row transition metals, many basis functions are required to define all of the electrons, which causes a significant increase in the computational time needed relative to earlier main group species. In addition, any basis set that describes these TM systems needs to account for the effects of relativity that can manifest in elements beyond the first-row transition metals. Therefore, the concept of the effective core potential (ECP), or pseudopotential (PP), was developed.²¹⁷ An ECP portrays the core electrons

with a potential that is fitted from relativistic calculations and treats the remainder of the electrons explicitly, which reduces the computational cost relative to their all-electron counterparts and generally has a negligible effect on accuracy.⁸ cc-pVnZ-PP is a form of basis sets that have been developed that pair ECPs with correlation consistent basis sets for the valence space.^{214,215}

2.4.3 Auxiliary Basis Sets

Auxiliary basis sets (ABS) were designed to offset the increase in computational cost arising from the calculation of four-center two-electron repulsion integrals in methods such as MP2 (i.e. RI-MP2) by using the resolution of the identity (RI) approximation. To provide details, ABS can be obtained through fitting procedures involving the coulomb integrals (basis/J or J-fit), or both the coulomb and exchange integrals (basis/JK or JK-fit), as discussed below.^{57,218,219} For a J-fit auxiliary basis set, the coefficients are fitted to a linear combination of three center (ij|a) and twocenter coulomb integrals whereas the K-fit auxiliary basis set is obtained through the difference between the exact exchange and the approximate exchange generated.

The auxiliary basis sets for MP2 were optimized by minimizing the quantity

$$\delta_{RI} = \frac{1}{4} \sum_{iajb} \frac{(\langle ab || ij \rangle - \langle ab || ij \rangle_{RI})^2}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j}$$
(2.47)

with respect to the auxiliary basis set exponents, where $\langle ab||ij\rangle = (ai|bj) - (aj|bi)$. The auxiliary basis sets are constructed so that the number of auxiliary basis functions are not greater than four times the number of basis functions in the standard basis set, as this could negate the advantage gained for computational cost reduction. Also, the quantity in Equation 2.47 must be less than 10^{-6} when divided by E_{MP2} , and $|E_{MP2} - E_{RI-MP2}|$ must be less than 20 μE_h for auxiliary basis sets in reproducing MP2 energies, but at a fraction of the cost. ^{57,218,219}

2.4.3.1 AutoAux

For lower parts of the periodic table, there are many atoms for which optimized auxiliary basis sets are not available. For instance, auxiliary basis sets for cc-pCVnZ basis sets optimized

for *ab initio* correlated methods are not available. To expand the availability of auxiliary basis sets, Stoychev et al.²²⁰ developed a generation scheme called AutoAux within the ORCA software package.²²¹ Their scheme was used to generate ABS for def2-SVP, def2-TZVP, def2-QZVPP, and cc-pwCV*n*Z where n = D, T, Q, and 5, for H-Rn. They calculated both absolute and relative energies via several reaction sets. For the cc-pwCV*n*Z basis sets, the average RI error was within 175 μ E_h relative to absolute HF and MP2 energies calculated with the AutoAux feature. While AutoAux is useful for generating auxiliary basis sets on-the-fly in a calculation, these sets are often twice the size of optimized auxiliary basis sets but can still benefit from the RI approximation. The AutoAux scheme is utilized for the transition metal species in Chapter 5.

2.4.4 Basis Set Superposition Error

Basis set superposition error (BSSE) arises for interaction energies of molecular complexes via an improved description of each fragment in the presence of the basis set of the other fragment. The interaction energy (ΔE_{AB}) between two molecular fragments A and B is

$$\Delta E_{AB} = E_{AB}^{AB}(AB) - E_{A}^{A}(A) - E_{B}^{B}(B)$$
(2.48)

To overcome this error, when describing the energy of fragment A, the presence of a ghost fragment B, i.e. the inclusion of the basis functions of fragment B without the atoms of fragment B present, serves to counterbalance the effect that basis set B has on fragment A in the calculation of complex AB.¹⁵² The counterpoise-corrected interaction energy is

$$\Delta E_{AB}^{CP} = E_{AB}^{AB}(AB) - E_A^{AB}(AB) - E_B^{AB}(AB)$$
(2.49)

where E_{AB}^{AB} is the energy of complex AB calculated with the basis set for AB, $E_{A}^{AB}(AB)$ and $E_{B}^{AB}(AB)$ are the energies of fragments A and B, respectively, calculated with the basis set for AB. When substituting Equation 2.49 into Equation 2.48, the counterpoise correction to the interaction energy is obtained.

$$\Delta E_{corr}^{CP} = \Delta E_{AB}^{CP} - \Delta E_{AB} = \left(E_A^A(A) - E_A^{AB}(AB) \right) + \left(E_B^B(B) - E_B^{AB}(AB) \right)$$
(2.50)

Therefore, for variational wavefunctions, the counterpoise correction is always positive since $E_A^A(A) > E_A^{AB}(AB)$ and $E_B^B(B) > E_B^{AB}(AB)$.⁶

Depending on the nature of the interaction, molecular interaction energies vary considerably in magnitude. Interaction energies range from 100-500 mE_h for covalent bonds to 50-500 μ E_h for dispersion-bound complexes.¹⁵² While BSSE is present in all electronic structure calculations, the effects of BSSE are more prevalent for weakly-bound interactions, i.e. van der Waal interactions.

2.5 Implicit Solvation Models

As numerous chemical reactions are performed in solution, appropriate computational models are needed to characterize solute-solvent interactions and describe other properties such as charge distribution and solvation free energies. Two types of models that are used to incorporate the effects of solvation explicit solvation models where all of the solvent molecules are explicitly represented in the calculation and implicit solvation models that represents the solvent molecules as a continuum. While explicit models can describe short range solute-solvent interactions, these models are computationally expensive as they require 100-1000 solvent molecules for a single QM calculation. Using implicit solvation models yields a lower computational cost relative to using explicit solvent molecules but neglect detailed descriptions of the solute-solvent interactions.

Implicit solvation models are based on the approximation of a liquid medium as a dielectric unstructured fluid through the use of a quantum mechanical description of the solute. Implicit solvation models provide an extension of the Born and Onsager models previously used to describe fundamental properties of solutions.^{222,223} The general formulation of the solute-solvent system^{223,224} in implicit models is that the solute is represented by a permanent point dipole μ and a polarizability α with a radius *a*, and the solvent molecules are modeled as the average of the charge distribution represented as a continuum dielectric medium with a fixed dielectric value (ε). The Poisson equation, shown in Equation 2.51, is used to define the electrostatic potential as a function of charge density.¹⁰

$$\nabla \varepsilon(r) * \nabla \phi(r) = -4\pi \rho(r) \tag{2.51}$$

In Equation 2.51, the solute is described explicitly within a cavity of vacuum while the solvent is described implicitly via charge distribution.

The initial shapes of the vacuum cavity are spheres and ellipsoids, for which the Poisson equation was solved analytically by Born²²² and Onsager,²²³ respectively. The Born solvation model creates a single point charge inside a spherical cavity.²²² The Onsager model calculates the dipole moment of the solute by using the point-dipole approximation, and thus, is only applicable to molecules with dipole moments.²²³ When describing multipolar systems, the reaction field has a poor description if the molecule is not spherical since the Onsager model uses an elliptical or near-spherical cavity. Therefore, arbitrary cavities that use the overlap of atomic spheres defined by their van der Waals radii and utilize a numerical solution to the Poisson equation are essential for the development of accurate QM solvation models. A procedure used for calculations with a solvation model is the self-consistent reaction field, which originates from using the solutions to the Poisson equation as a perturbation to the gas phase Hamiltonian used for *ab initio* (Section 2.1) and density functional methods (Section 2.3), as previously discussed.

Most implicit models are parameterized to describe aqueous solvation free energies at room temperature; SM8,²²⁵ SMD,²²⁶ and COSMO-RS²²⁷ are solvation models that also describe non-aqueous solvents and elevated temperatures. Implicit models are used extensively in pK_a studies due to the pK_a depending on the solvation free energy.^{228,229}

2.5.1 COSMO

The COnductor-like Screening MOdel (COSMO) was developed by Klamt in 1993 and is based on the screening in conductors, which are infinitely strong dielectrics.²³⁰ This approach uses arbitrarily-shaped cavities and a boundary element method to describe apparent surface charges that define the same electrostatic potential as a numerical solution to the Poisson equation. COSMO is used to calculate the energies of a molecule within a dielectric medium. The dielectric screening energies for a given geometry scale with the dielectric permittivity of ε of the screening medium as shown in Equation 2.52; x is 0.5 for the COSMO model.

$$\frac{\varepsilon - 1}{\varepsilon + x} \text{ where } 0 \le x \le 2 \tag{2.52}$$

This is due to the response of a conductor to a solute charge distribution compared to the response from a dielectric medium. 10 The COSMO approach allows the calculation of analytical gradients within an arbitrarily-shaped cavity. Therefore, geometry optimizations in the solvation phase are practical as numerical gradients, which increase the computational cost, are typically used; however, one of the difficulties includes finding the optimum parameters such as a set of van der Waal radii to create the solvent accessible surface.²³⁰

2.5.2 PCM/C-PCM

The Polarizable Continuum Model (PCM) was developed in 1981 and employs an apparent surface charge on the cavity surface.^{231,232} When using PCM and its variants, arbitrary cavity shapes are used, unlike Onsager models; this provides better electronic energy results through a more realistic description of the solute in solution. The basic PCM definition, Equation 2.53, utilizes a continuous surface charge, $\sigma(s)$, with the gradient on the internal (*in*) part of the surface to describe the apparent surface charge distribution.^{231–233}

$$\sigma(s) = \frac{\varepsilon - 1}{4\pi\varepsilon} \frac{\partial}{\partial \vec{n}} (V_m + V_\sigma)_{in}$$
(2.53)

In Equation 2.53, ε is the dielectric constant, \vec{n} indicates the unit vector perpendicular to the cavity surface pointing outward, V_M , is the potential generated by the charge distribution, and V_{σ} is the potential over the whole space generated by the polarization of the dielectric medium.

Unfortunately, using arbitrarily-shaped cavities are rather expensive because of the requirement of numerical solutions for the derivatives and gradients. The difficulties with these models are the sharp edges created by the overlapping spheres on the solvent accessible surface. Therefore, the surface is smoothed by other spheres not centered on atoms to simulate the solvent excluded surface. The C-PCM²³⁴ formulation was adapted in 1998 based on COSMO and used a conductor-like setting within the PCM model. As described with COSMO, the solvent was treated as a

conductor and the polarizability of the system becomes zero, which decreases the complexity of solving the Poisson equation. C-PCM utilizes the same equation used by PCM, Equation 2.53, with the key assumption of using the scaling factor, Equation 2.52, to describe the polarization charges such that the Gauss law is obeyed; in reference to Equation 2.52, x is 0 for the C-PCM model.²³⁴ The C-PCM model is amongst the most widespread implicit solvation models used for studies on organometallic systems,^{180,235–237} and the development of hybrid QM/QM schemes for solvation.^{176,238}

2.5.3 SMD

The universal implicit solvation model SMD^{226} was developed by Truhlar where the full solute electron density is used without defining partial atomic charges. This density-based model separates the solvation free energy into two components: the bulk electrostatic energy calculated through the integral equation formalism PCM model, which replaces the molecular electric field on the surface with the electrostatic potential; and the cavity-dispersion-solvent-structure component. The first component uses the SCRF treatment and the solution for the nonhomogeneous Poisson equation, Equation 2.54.

$$\nabla(\varepsilon\nabla\phi) = -4\pi\rho_f \tag{2.54}$$

The second component is the contribution arising from short-range solute-solvent interactions in the first solvation shell. SMD describes the solvent accessible surface via a superposition of nuclear-centered spheres with intrinsic Coulomb radii. SMD focuses on the standard solvation free energies and was parameterized using 2821 solvation data points including free energies in 90 non-aqueous solvents and water.

2.6 Vibrational Self Consistent Field Theory

Vibrational spectroscopy is a useful approach to characterize intermolecular interactions for reaction pathways and vibrational motion. In the theoretical treatment of vibrational spectra, accurate potential energy surfaces (PES) are necessary to describe nuclear dynamics, reaction

dynamics, and quantum rate constants.^{239–241} However for vibrational calculations, electronic structure methods are often restricted to the harmonic oscillator approximation since the vibrational Hamiltonian can be partitioned into a set of one-dimensional harmonic oscillators using normal mode coordinates within the harmonic oscillator approximation.²⁴² The errors inherent in both the harmonic oscillator approximation and electronic structure methods accumulate to yield deviations over 100 cm⁻¹ for vibrational frequencies in some cases although the individual contributions of the harmonic approximation and electronic structure method to the error are unknown.

Computationally, the harmonic oscillator approximation is conceptually simpler than fully anharmonic calculations but results in a loss of accuracy for vibrational properties. One way to correct for anharmonicity in molecular vibrations is to apply an empirical scaling factor to harmonic vibrations, commonly called a frequency scaling factor.¹⁰⁸ The scaling factor is determined through a least-squares fitting to corresponding experimental frequencies; thus, this approach is an underlying potential for addressing computing observables of the anharmonic PES. However, scaling factors for DFT are approximately 1.00 ± 0.05 whereas those for *ab initio* methods are lower (0.95 \pm 0.05),¹⁰⁸ implying that DFT yields more accurate vibrational frequencies with the harmonic approximation and introduces uncertainty into which aspects of DFT contribute to predicting vibrational frequencies.

Directly calculating the anharmonic PESs for vibrations provides better insight about removing uncertainty arising from both the harmonic approximation and the use fo common global frequency scaling factors. One of the *ab initio* methods developed for anharmonic vibrational spectroscopy is vibrational self consistent field (VSCF) theory, which was developed in the late 1970s.^{243–247} The vibrational Schrödinger equation with mass-weighted normal coordinates Q_i ,

$$\left[-\frac{1}{2}\sum_{j=1}^{N}\frac{\partial^{2}}{\partial Q_{j}^{2}}+V\left(Q_{1},\ldots,Q_{N}\right)\right]\Psi_{n}\left(Q_{1},\ldots,Q_{N}\right)=E_{N}\Psi_{n}\left(Q_{1},\ldots,Q_{N}\right)$$
(2.55)

where V is the potential energy function of the system, n is the state number, and N is the number of vibrational degrees of freedom (normal modes), utilizes the Born-Oppenheimer approximation and neglects rotational coupling effects to vibration. VSCF theory is similar to the Hartree-Fock theory (see Section 2.1) since each vibrational mode is characterized in the mean field of the other vibrational motions. Unlike in Hartree-Fock theory, the total wavefunction of VSCF approximation is a product of single mode wavefunctions akin to a Hartree product

$$\Psi(Q_1, \dots, Q_N) = \prod_{i=1}^N \psi_i^{(n)}(Q_i)$$
(2.56)

where the single mode wavefunctions $\psi_i^{(n)}$ are called the modals and Q_N are mass-weighted normal coordinates since vibrations are distinguishable. Error due to introducing the separability approximation depends on the coordinate system used.²⁴⁸

Using a variational principle for the ansatz in Equation 2.56 leads to the single mode VSCF equation

$$\left[-\frac{1}{2}\frac{\partial^2}{\partial Q_i^2} + \overline{V_i^{(n)}}(Q_i)\right]\psi_i^{(n)} = \varepsilon_i^{(n)}\psi_i^{(n)}(Q_i)$$
(2.57)

where the mean effective potential $\overline{V_i^{(n)}}(Q_i)$ for mode Q_i is given by

$$\overline{V_i^{(n)}}(Q_i) = \left\langle \prod_{j \neq i}^N \psi_j^{(n)}(Q_j) \middle| V(Q_1, \dots, Q_N) \middle| \prod_{j \neq i}^N \psi_j^{(n)}(Q_j) \right\rangle$$
(2.58)

To examine the full potential $V(Q_1, Q_2, ..., Q_N)$, the potential can be expanded via a multimode expansion

$$V(Q_1, Q_2, \dots, Q_N) = \sum_i V_i^{(1)}(Q_i) + \sum_{ij} V_{ij}^{(2)}(Q_i, Q_j) + \sum_{ijk} V_{ijk}^{(3)}(Q_i, Q_j, Q_k) + \dots$$
(2.59)

where $V_i^{(1)}(Q_i)$ are the single-mode diagonal terms

$$V_i^{(1)}(Q_i) = V(0, \dots, Q_i, \dots, 0)$$
(2.60)

the pair-wise interactions $W_{ij}^{(2)}(Q_i, Q_j)$ from the expansion of $V(Q_1, Q_2, \dots, Q_N)$ are

$$W_{ij}^{(2)}(Q_i, Q_j) = V_{ij}^{(2)}(Q_i, Q_j) - V_i^{(1)}(Q_i) - V_j^{(1)}(Q_j)$$
(2.61a)

$$= V(0, \dots, Q_i, \dots, Q_j, \dots, 0) - V_i^{(1)}(Q_i) - V_j^{(1)}(Q_j)$$
(2.61b)

and so forth with higher order expansions. N-order expansions of the potential are not feasible for N larger than six since the integration over the potential is a N-1 dimensional integral. Therefore, the expanded potential is usually truncated in terms of a quartic or sextic force field.²³⁹

Equations 2.57 and 2.58 are solved self-consistently for the single mode wavefunctions, energies, and effective potentials. Several methods can be applied for the solution of Equation 2.57 to get both the ground and excited VSCF states of the system. Due to this approximation, the total energy is given by

$$E_n = \sum_{i=1}^{N} \varepsilon_i^{(n)} + (n-1) \left\langle \prod_{j \neq i}^{N} \psi_j^{(n)}(Q_j) \middle| V(Q_1, \dots, Q_N) \middle| \prod_{j \neq i}^{N} \psi_j^{(n)}(Q_j) \right\rangle$$
(2.62)

The major computational difficulty is due to the evaluation of multidimensional integrals inherent in Equations 2.57, 2.58, and 2.62, especially for large systems, and that depends on the mathematical form of the potential. Hence, the choice of potential plays a key role in the VSCF approximation.

Since VSCF describes the effect of a singular vibrational mode in the mean field of all other vibrational modes as the Hartree-Fock method does for electrons, the effects of correlation between modes need to be described. For post Hartree-Fock methods such as MP2 and CI, there are complementary vibrational equivalents that correlate vibrational motion. For example, for perturbation theory, the full vibrational Hamiltonian is written in the form

$$H = H^{SCF,(n)} + \Delta V(Q_1, \dots, Q_N)$$
(2.63)

where $H^{SCF,(n)}$ is the Hamiltonian used in Equation 2.57 and the equation for ΔV is given by

$$\Delta V(Q_1, \dots, Q_N) = V(Q_1, \dots, Q_N) - \sum_{i=1}^N \overline{V_i^{(n)}}(Q_i)$$
(2.64)

where ΔV represents all correlation effects between vibrational modes. Considering the pair-wise approximation terms in Equations 2.61a, ΔV can be rewritten as

$$\Delta V(Q_1, \dots, Q_N) = \sum_{i=1}^N V_i^{(1)}(Q_i) + \sum_j \sum_{i>j} W_{ij}^{(2)}(Q_i, Q_j) - \sum_{i=1}^N \overline{V_i^{(n)}}(Q_i)$$
(2.65)

where the potential at the minimum is taken at zero, leaving diagonal terms and pair-wise terms as shown in Equations 2.60 and 2.61b. Methods that account for correlation effects of vibrational modes are known as post-VSCF methods.

Post-VSCF methods include VSCF-PT2, which is a second-order perturbation to account for correlation effects between vibrational modes, as well as vibrational coupled cluster (VCC) and vibrational configuration interaction (VCI) methods, and a combination of VCI with perturbatively selected interactions (VCIPSI-PT2).^{239,249–254} The idea is that ΔV , which is the difference between the true Hamiltonian and the VSCF Hamiltonian, must be small as VSCF is a good approximation. VCI yields the best possible results variationally given the basis set limits.^{239,250} Analogous to CI, every possible contribution of a complete set of functions is considered and thus full VCI with an infinite basis set is the exact solution to the vibrational time independent Schrödinger equation (Equation 2.55) given the constraints (BO approximation and neglecting rotational effects on vibration).

For N normal modes, there are N(N-1)/2 coupling potentials. Each coupling potential is computed with electronic structure methods on a grid of $N_{\rm grid} \times N_{\rm grid}$ points ($N_{\rm grid}$ = 16 in Chapter 6). For example, C₆H₆, which has 30 normal modes, would require 111,360 single point calculations for all 435 pair-wise coupling potentials assuming $N_{\rm grid}$ = 16. This requires additional approximations such as the vibrational configuration interaction with perturbation selected interactions (VCIPSI) algorithm developed by Scribano and Benoit²⁵⁴ to iteratively select the VCI active space based on previous implementations of this algorithm for ab initio electronic structure calculations²⁵⁵ and VCI methods.^{251–253} The active space is treated variationally and then increased iteratively using a vibrational Møller-Plesset barycentric (VMPB) partition scheme to improve the representation of the complete VCI wavefunction. The VCIPSI-PT2 method utilizes the final VMPB correction in the VCI active space. Implementation of this algorithm led to a savings of 70-80% while only deviating from VCI by approximately 0.01 cm⁻¹ for all vibrations of CH₄ when using MP2/aug-cc-pVTZ for generating the vibrational PES and a savings of 85% and a deviation of 0.07 cm⁻¹ for the OH stretching frequency of benzoic acid while using 0.49% of the disk space that VCI used for the same vibration.²⁵⁴

Other measures to reduce the computational cost includes screening weakly coupled pair-wise coupling interactions via a threshold established from calculating the coupling strength (Equation

2.66), which can be calculated with only the VSCF potential. 256,257

$$\xi(q_i, q_j) = \frac{1}{N_{\text{grid}}^2} \sum_{n_i=1}^{N_{\text{grid}}} \sum_{n_j=1}^{N_{\text{grid}}} |V_{ij}^{(k)}(n_i, n_j)|$$
(2.66)

By removing non-essential vibrational coupling elements from the potential, a fast-VSCF approach is attained. This can greatly reduce the computational cost to generate fully anharmonic PESs for polyatomic molecules of increasing size and complexity.

The use of VSCF and VCI methods are pertinent in Chapter 6, as these methods are used to analyze anharmonic PESs to predict anharmonic vibrations for diatomic and polyatomic molecules.

REFERENCES

REFERENCES

- [1] Schrödinger, E. Quantisierung als Eigenwertproblem. Ann. Phys. 1926, 384, 361–376.
- [2] Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. Ann. Phys. **1927**, 389, 457–484.
- [3] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Strucutre Theory*, 1st ed.; Dover Publications, Inc.: Mineola, New York, 1996; pp 40–43,45.
- [4] Hartree, D. R. The Wave Mechanics of an Atom with a non-Coulomb Central Field. Part III. Term Values and Intensities in Series in Optical Spectra. *Math. Proc. Cambridge Philos. Soc.* 1928, 24, 426–437.
- [5] Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Phys.* **1930**, *61*, 126–148.
- [6] Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; John Wiley & Sons, Ltd: Chichester, UK, 2000; pp 1–908.
- [7] Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab initio molecular orbital theory*; John Wiley & Sons, Inc.: New York, NY, 1986; Vol. 33.
- [8] Jensen, F. Introduction to Computational Chemistry; John Wiley & Sons, Ltd: USA, 2006.
- [9] Bartlett, R. J. Coupled-cluster approach to molecular structure and spectra: A step toward predictive quantum chemistry. *J. Phys. Chem.* **1989**, *93*, 1697–1708.
- [10] Cramer, C. J. Essentials of Computational Chemistry, Theories and Models, 2nd ed.; John Wiley & Sons, Ltd: Chichester, UK, 2004; Vol. 43; pp 1720–1720.
- [11] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- [12] Schütz, M.; Hetzer, G.; Werner, H.-J. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. J. Chem. Phys. 1999, 111, 5691–5705.
- [13] Heßelmann, A. Local Molecular Orbitals from a Projection onto Localized Centers. J. Chem. Theory Comput. **2016**, *12*, 2720–2741.
- [14] Lennard-Jones, J. The Molecular Orbital Theory of Chemical Valency. I. The Determination of Molecular Orbitals. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1949**, *198*, 1–13.
- [15] Lennard-Jones, J. The Molecular Orbital Theory of Chemical Valency. II. Equivalent Orbitals in Molecules of Known Symmetry. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1949**, *198*, 14–26.
- [16] Hall, G. G.; Lennard-Jones, J. The Molecular Orbital Theory of Chemical Valency. III. Properties of Molecular Orbitals. Proc. R. Soc. A Math. Phys. Eng. Sci. 1950, 202, 155–165.

- [17] Hall, G. G. The Molecular Orbital Theory of Chemical Valency. VI. Properties of Equivalent Orbitals. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1950**, *202*, 336–344.
- [18] Lennard-Jones, J.; Pople, J. A. The Molecular Orbital Theory of Chemical Valency. IV. The Significance of Equivalent Orbitals. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 1950, 202, 166–180.
- [19] Boys, S. F. Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Rev. Mod. Phys.* **1960**, *32*, 296–299.
- [20] Foster, J. M.; Boys, S. F. Canonical Configuration Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300–302.
- [21] Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457–464.
- [22] Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. II. J. Chem. Phys. **1965**, *43*, S97–S116.
- [23] Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. J. Chem. Phys. 1989, 90, 4916–4926.
- [24] Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* 1983, 100, 151–154.
- [25] Sæbø, S.; Pulay, P. Local configuration interaction: An efficient approach for larger molecules. *Chem. Phys. Lett.* **1985**, *113*, 13–18.
- [26] Sæbø, S.; Pulay, P. Fourth-order Møller–Plessett perturbation theory in the local correlation treatment. I. Method. J. Chem. Phys. **1987**, *86*, 914–922.
- [27] Sæbø, S.; Pulay, P. The local correlation treatment. II. Implementation and tests. *J. Chem. Phys.* **1988**, 88, 1884–1890.
- [28] Sæbø, S.; Tong, W.; Pulay, P. Efficient elimination of basis set superposition errors by the local correlation method: Accurate ab initio studies of the water dimer. *J. Chem. Phys.* 1993, 98, 2170–2175.
- [29] Sæbø, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, 44, 213–236.
- [30] Schütz, M.; Werner, H.-J. Local perturbative triples correction (T) with linear cost scaling. *Chem. Phys. Lett.* **2000**, *318*, 370–378.
- [31] Schütz, M. Low-order scaling local electron correlation methods. III. Linear scaling local perturbative triples correction (T). *J. Chem. Phys.* **2000**, *113*, 9986–10001.
- [32] Schütz, M.; Werner, H.-J. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J. Chem. Phys.* **2001**, *114*, 661–681.

- [33] Pulay, P.; Sæbø, S. Orbital-invariant formulation and second-order gradient evaluation in Møller-Plesset perturbation theory. *Theor. Chem. Acc.* **1986**, *69*, 357–368.
- [34] Almlöf, J. Elimination of energy denominators in Møller-Plesset perturbation theory by a Laplace transform approach. *Chem. Phys. Lett.* **1991**, *181*, 319–320.
- [35] Häser, M.; Almlöf, J. Laplace transform techniques in Møller-Plesset perturbation theory. *J. Chem. Phys.* **1992**, *96*, 489–494.
- [36] Häser, M. Møller-Plesset (MP2) perturbation theory for large molecules. *Theor. Chem. Acc.* **1993**, 87, 147–173.
- [37] Wilson, A. K.; Almlöf, J. Møller-Plesset correlation energies in a localized orbital basis using a Laplace transform technique. *Theor. Chem. Acc.* **1997**, *95*, 49–62.
- [38] Ayala, P. Y.; Scuseria, G. E. Linear scaling second-order Moller-Plesset theory in the atomic orbital basis for large molecular systems. *J. Chem. Phys.* **1999**, *110*, 3660–3671.
- [39] Lambrecht, D. S.; Doser, B.; Ochsenfeld, C. Rigorous integral screening for electron correlation methods. *J. Chem. Phys.* **2005**, *123*, 184102.
- [40] Doser, B.; Lambrecht, D. S.; Kussmann, J.; Ochsenfeld, C. Linear-scaling atomic orbitalbased second-order Møller-Plesset perturbation theory by rigorous integral screening criteria. *J. Chem. Phys.* 2009, 130, 064107.
- [41] Hetzer, G.; Pulay, P.; Werner, H.-J. Multipole approximation of distant pair energies in local MP2 calculations. *Chem. Phys. Lett.* **1998**, *290*, 143–149.
- [42] Scuseria, G. E.; Ayala, P. Y. Linear scaling coupled cluster and perturbation theories in the atomic orbital basis. *J. Chem. Phys.* **1999**, *111*, 8330–8343.
- [43] Subotnik, J. E.; Sodt, A.; Head-Gordon, M. A near linear-scaling smooth local coupled cluster algorithm for electronic structure. *J. Chem. Phys.* **2006**, *125*.
- [44] Werner, H.-J.; Knizia, G.; Krause, C.; Schwilk, M.; Dornbach, M. Scalable electron correlation methods I.: PNO-LMP2 with linear scaling in the molecular size and nearinverse-linear scaling in the number of processors. J. Chem. Theory Comput. 2015, 11, 484–507.
- [45] Ma, Q.; Werner, H.-J. Scalable Electron Correlation Methods. 2. Parallel PNO-LMP2-F12 with Near Linear Scaling in the Molecular Size. J. Chem. Theory Comput. 2015, 11, 5291– 5304.
- [46] Menezes, F.; Kats, D.; Werner, H.-J. Local complete active space second-order perturbation theory using pair natural orbitals (PNO-CASPT2). J. Chem. Phys. 2016, 145.
- [47] Schwilk, M.; Ma, Q.; Köppl, C.; Werner, H.-J. Scalable Electron Correlation Methods. 3. Efficient and Accurate Parallel Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD). J. Chem. Theory Comput. 2017, 13, 3650–3675.

- [48] Høyvik, I. M.; Jørgensen, P. Characterization and Generation of Local Occupied and Virtual Hartree-Fock Orbitals. **2016**, *116*, 3306–3327.
- [49] Kleier, D. A.; Halgren, T. A.; Hall, J. H.; Lipscomb, W. N. Localized molecular orbitals for polyatomic molecules. I. a comparison of the Edmiston-Ruedenberg and Boys localization methods. J. Chem. Phys. 1974, 61, 3905–3919.
- [50] Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Pipek-Mezey localization of occupied and virtual orbitals. *J. Comput. Chem.* **2013**, *34*, 1456–1462.
- [51] Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate local approximations to coupledelectron pair approaches: An attempt to revive the pair natural orbital method. *J. Chem. Phys.* 2009, *130*, 114108.
- [52] Liakos, D. G.; Neese, F. Is It Possible to Obtain Coupled Cluster Quality Energies at near Density Functional Theory Cost? Domain-Based Local Pair Natural Orbital Coupled Cluster vs Modern Density Functional Theory. J. Chem. Theory Comput. 2015, 11, 4054–4063.
- [53] Ziółkowski, M.; Jansík, B.; Kjærgaard, T.; Jørgensen, P. Linear scaling coupled cluster method with correlation energy based error control. J. Chem. Phys. **2010**, 133, 014107.
- [54] Eriksen, J. J.; Baudin, P.; Ettenhuber, P.; Kristensen, K.; Kjærgaard, T.; Jørgensen, P. Linear-Scaling Coupled Cluster with Perturbative Triple Excitations: The Divide–Expand–Consolidate CCSD(T) Model. J. Chem. Theory Comput. 2015, 11, 2984– 2993.
- [55] Kendall, R. A.; Früchtl, H. A. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theor. Chem. Acc.* **1997**, *97*, 158–163.
- [56] Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143–152.
- [57] Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
- [58] DiStasio, R. A.; Jung, Y.; Head-Gordon, M. A Resolution-Of-The-Identity Implementation of the Local Triatomics-In-Molecules Model for Second-Order Møller–Plesset Perturbation Theory with Application to Alanine Tetrapeptide Conformational Energies. J. Chem. Theory Comput. 2005, 1, 862–876.
- [59] Neese, F. An Improvement of the Resolution of the Identity Approximation for the Formation of the Coulomb Matrix. *J. Comput. Chem.* **2003**, *24*, 1740–1747.
- [60] Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree–Fock exchange. *Chem. Phys.* **2009**, *356*, 98–109.

- [61] Izsák, R.; Neese, F.; Klopper, W. Robust fitting techniques in the chain of spheres approximation to the Fock exchange: The role of the complementary space. *J. Chem. Phys.* **2013**, *139*, 094111.
- [62] Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013**, *138*, 034106.
- [63] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.
- [64] Pinski, P.; Riplinger, C.; Valeev, E. F.; Neese, F. Sparse maps A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals. *J. Chem. Phys.* 2015, 143, 034108.
- [65] Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* **2016**, *144*.
- [66] Saitow, M.; Becker, U.; Riplinger, C.; Valeev, E. F.; Neese, F. A new near-linear scaling, efficient and accurate, open-shell domain-based local pair natural orbital coupled cluster singles and doubles theory. *J. Chem. Phys.* **2017**, *146*, 164105.
- [67] Anoop, A.; Thiel, W.; Neese, F. A local pair natural orbital coupled cluster study of Rh catalyzed asymmetric olefin hydrogenation. *J. Chem. Theory Comput.* **2010**, *6*, 3137–3144.
- [68] Sparta, M.; Riplinger, C.; Neese, F. Mechanism of olefin asymmetric hydrogenation catalyzed by iridium phosphino-oxazoline: A pair natural orbital coupled cluster study. J. Chem. Theory Comput. 2014, 10, 1099–1108.
- [69] Sparta, M.; Neese, F. Chemical applications carried out by local pair natural orbital based coupled-cluster methods. *Chem. Soc. Rev.* **2014**, *43*, 5032–5041.
- [70] Chan, B.; Kawashima, Y.; Katouda, M.; Nakajima, T.; Hirao, K. From C₆₀ to Infinity: Large-Scale Quantum Chemistry Calculations of the Heats of Formation of Higher Fullerenes. *J. Am. Chem. Soc.* 2016, *138*, 1420–1429.
- [71] Minenkov, Y.; Wang, H.; Wang, Z.; Sarathy, S. M.; Cavallo, L. Heats of Formation of Medium-Sized Organic Compounds from Contemporary Electronic Structure Methods. J. Chem. Theory Comput. 2017, 13, 3537–3560.
- [72] Ochterski, J. W.; Petersson, G. A.; Montgomery Jr., J. A. A complete basis set model chemistry. V. Extensions to six or more heavy atoms. *J. Chem. Phys.* **1996**, *104*, 2598–2619.
- [73] Neese, F.; Hansen, A.; Liakos, D. G. Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis. J. Chem. Phys. 2009, 131, 064103.

- [74] Huntington, L. M.; Hansen, A.; Neese, F.; Nooijen, M. Accurate thermochemistry from a parameterized coupled-cluster singles and doubles model and a local pair natural orbital based implementation for applications to larger systems. *J. Chem. Phys.* **2012**, *136*, 064101.
- [75] Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. Gaussian-1 theory: A general procedure for prediction of molecular energies. J. Chem. Phys. 1989, 90, 5622–5629.
- [76] Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 theory for molecular energies of first- and second-row compounds. *J. Chem. Phys.* **1991**, *94*, 7221–7230.
- [77] Curtiss, L. A.; Carpenter, J. E.; Raghavachari, K.; Pople, J. A. Validity of additivity approximations used in GAUSSIAN-2 theory. *J. Chem. Phys.* **1992**, *96*, 9030–9034.
- [78] Curtiss, L. A.; Raghavachari, K.; Pople, J. A. Gaussian-2 theory using reduced Møller-Plesset orders. J. Chem. Phys. 1993, 98, 1293–1298.
- [79] Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V. A.; Pople, J. A. Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *J. Chem. Phys.* 1998, 109, 7764–7776.
- [80] Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. J. Chem. Phys. 1988, 89, 2193–2218.
- [81] Petersson, G. A.; Tensfeldt, T. G.; Montgomery Jr., J. A. A complete basis set model chemistry. III. The complete basis set-quadratic configuration interaction family of methods. *J. Chem. Phys.* **1991**, *94*, 6091–6101.
- [82] Petersson, G. A.; Al-Laham, M. A. A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *J. Chem. Phys.* **1991**, *94*, 6081–6090.
- [83] Montgomery Jr., J. A.; Michels, H. H.; Francisco, J. S. Ab initio calculation of the heats of formation of CF₃OH and CF₂O. *Chem. Phys. Lett.* **1994**, 220, 391–396.
- [84] Montgomery Jr., J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J. Chem. Phys.* 1999, 110, 2822–2827.
- [85] Martin, J. M. L.; De Oliveira, G. Towards standard methods for benchmark quality ab initio thermochemistry W1 and W2 theory. *J. Chem. Phys.* **1999**, *111*, 1843–1856.
- [86] Montgomery Jr., J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VII. Use of the minimum population localization method. *J. Chem. Phys.* 2000, *112*, 6532–6542.
- [87] DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The correlation consistent composite approach (ccCA): An alternative to the Gaussian-n methods. *J. Chem. Phys.* **2006**, *124*, 114104.
- [88] Wood, G. P. F.; Radom, L.; Petersson, G. A.; Barnes, E. C.; Frisch, M. J.; Montgomery Jr., J. A. A restricted-open-shell complete-basis-set model chemistry. J. Chem. Phys. 2006, 125, 094106.
- [89] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 theory. J. Chem. Phys. 2007, 126, 084108.
- [90] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 theory using reduced order perturbation theory. J. Chem. Phys. 2007, 127, 124105.
- [91] Feller, D.; Dixon, D. A. Predicting the Heats of Formation of Model Hydrocarbons up to Benzene. J. Phys. Chem. A 2000, 104, 3048–3056.
- [92] Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vázquez, J.; Stanton, J. F. HEAT: High accuracy extrapolated ab initio thermochemistry. J. Chem. Phys. 2004, 121, 11599–11613.
- [93] Schuurman, M. S.; Muir, S. R.; Allen, W. D.; Schaefer, H. F. Toward subchemical accuracy in computational thermochemistry: Focal point analysis of the heat of formation of NCO and [H,N,C,O] isomers. J. Chem. Phys. 2004, 120, 11586–11599.
- [94] Daniel Boese, A.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kállay, M.; Gauss, J. W3 theory: Robust computational thermochemistry in the kJ/mol accuracy range. J. Chem. Phys. 2004, 120, 4129–4141.
- [95] Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions. J. Chem. Phys. 2006, 125, 144108.
- [96] Wan, W.; Karton, A. Heat of formation for C_{60} by means of the G4(MP2) thermochemical protocol through reactions in which C_{60} is broken down into corannulene and sumanene. *Chem. Phys. Lett.* **2016**, *643*, 34–38.
- [97] Laury, M. L.; DeYonker, N. J.; Jiang, W.; Wilson, A. K. A pseudopotential-based composite method: The relativistic pseudopotential correlation consistent composite approach for molecules containing 4d transition metals (Y-Cd). J. Chem. Phys. 2011, 135, 214103.
- [98] Jiang, W.; DeYonker, N. J.; Determan, J. J.; Wilson, A. K. Toward accurate theoretical thermochemistry of first row transition metal complexes. J. Phys. Chem. A 2012, 116, 870– 885.
- [99] Laury, M. L.; Wilson, A. K. Examining the heavy p-block with a pseudopotential-based composite method: Atomic and molecular applications of rp-ccCA. J. Chem. Phys. 2012, 137, 1–10.
- [100] Bross, D. H.; Hill, J. G.; Werner, H.-J. J.; Peterson, K. A. Explicitly correlated composite thermochemistry of transition metal species. J. Chem. Phys. 2013, 139, 094302.

- [101] Thanthiriwatte, K. S.; Vasiliu, M.; Battey, S. R.; Lu, Q.; Peterson, K. A.; Andrews, L.; Dixon, D. A. Gas Phase Properties of MX₂ and MX₄ (X = F, Cl) for M = Group 4, Group 14, Cerium, and Thorium. *J. Phys. Chem. A* **2015**, *119*, 5790–5803.
- [102] Peterson, C.; Penchoff, D. A.; Wilson, A. K. Ab initio approaches for the determination of heavy element energetics: Ionization energies of trivalent lanthanides (Ln = La-Eu). J. *Chem. Phys.* 2015, 143, 194109.
- [103] Cox, R. M.; Citir, M.; Armentrout, P. B.; Battey, S. R.; Peterson, K. A. Bond energies of ThO⁺ and ThC⁺ : A guided ion beam and quantum chemical investigation of the reactions of thorium cation with O_2 and CO. J. Chem. Phys. **2016**, 144, 184309.
- [104] Fang, Z.; Both, J.; Li, S.; Yue, S.; Aprà, E.; Keçeli, M.; Wagner, A. F.; Dixon, D. A. Benchmark Calculations of Energetic Properties of Groups 4 and 6 Transition Metal Oxide Nanoclusters Including Comparison to Density Functional Theory. *J. Chem. Theory Comput.* 2016, *12*, 3689–3710.
- [105] Cheng, L.; Gauss, J.; Ruscic, B.; Armentrout, P. B.; Stanton, J. F. Bond Dissociation Energies for Diatomic Molecules Containing 3d Transition Metals: Benchmark Scalar-Relativistic Coupled-Cluster Calculations for 20 Molecules. *J. Chem. Theory Comput.* 2017, 13, 1044–1056.
- [106] Fang, Z.; Vasiliu, M.; Peterson, K. A.; Dixon, D. A. Prediction of Bond Dissociation Energies/Heats of Formation for Diatomic Transition Metal Compounds: CCSD(T) Works. *J. Chem. Theory Comput.* 2017, 13, 1057–1066.
- [107] Vasiliu, M.; Hill, J. G.; Peterson, K. A.; Dixon, D. A. Structures and Heats of Formation of Simple Alkaline Earth Metal Compounds II: Fluorides, Chlorides, Oxides, and Hydroxides for Ba, Sr, and Ra. J. Phys. Chem. A 2018, 122, 316–327.
- [108] Merrick, J. P.; Moran, D.; Radom, L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. J. Phys. Chem. A 2007, 111, 11683–11700.
- [109] Barone, V.; Bloino, J.; Guido, C. A.; Lipparini, F. A fully automated implementation of VPT2 Infrared intensities. *Chem. Phys. Lett.* 2010, 496, 157–161.
- [110] Ramakrishnan, R.; Rauhut, G. Semi-quartic force fields retrieved from multi-mode expansions: Accuracy, scaling behavior, and approximations. J. Chem. Phys. 2015, 142, 154118.
- [111] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. Basis-set convergence of the energy in molecular Hartree–Fock calculations. *Chem. Phys. Lett.* **1999**, *302*, 437–446.
- [112] Martin, J. M. L.; Lee, T. J. The atomization energy and proton affinity of NH₃. An ab initio calibration study. *Chem. Phys. Lett.* **1996**, *258*, 136–143.
- [113] Schwartz, C. Importance of angular correlations between atomic electrons. *Phys. Rev.* **1962**, *126*, 1015–1019.

- [114] Schwartz, C. Methods Comput. Phys.; Academic Press Inc.: New York, NY, 1963; pp 241–266.
- [115] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gn theory. Wiley Interdiscip. Rev. Comput. Mol. Sci. 2011, 1, 810–825.
- [116] Jiang, W.; DeYonker, N. J.; Wilson, A. K. Multireference character for 3d transition-metalcontaining molecules. J. Chem. Theory Comput. 2012, 8, 460–468.
- [117] DeYonker, N. J.; Wilson, B. R.; Pierpont, A. W.; Cundari, T. R.; Wilson, A. K. Towards the intrinsic error of the correlation consistent Composite Approach (ccCA). *Mol. Phys.* 2009, 107, 1107–1121.
- [118] Oyedepo, G. A.; Wilson, A. K. Multireference correlation consistent composite approach [MR-ccCA]: Toward accurate prediction of the energetics of excited and transition state chemistry. J. Phys. Chem. A 2010, 114, 8806–8816.
- [119] Oyedepo, G. A.; Wilson, A. K. Oxidative addition of the C α -C β bond in β -O-4 linkage of lignin to transition metals using a relativistic pseudopotential-based ccCA-ONIOM method. *ChemPhysChem* **2011**, *12*, 3320–3330.
- [120] Nedd, S. A.; DeYonker, N. J.; Wilson, A. K.; Piecuch, P.; Gordon, M. S. Incorporating a completely renormalized coupled cluster approach into a composite method for thermodynamic properties and reaction paths. J. Chem. Phys. 2012, 136, 144109.
- [121] Mahler, A.; Wilson, A. K. Explicitly correlated methods within the ccCA methodology. *J. Chem. Theory Comput.* **2013**, *9*, 1402–1407.
- [122] Riojas, A. G.; Wilson, A. K. Solv-ccCA: Implicit solvation and the correlation consistent composite approach for the determination of pK_a. J. Chem. Theory Comput. 2014, 10, 1500–1510.
- [123] Karton, A.; Martin, J. M. L. Heats of formation of beryllium, boron, aluminum, and silicon re-examined by means of W4 theory. J. Phys. Chem. A. 2007; pp 5936–5944.
- [124] Karton, A.; Martin, J. M. L. Explicitly correlated Wn theory: W1-F12 and W2-F12. J. *Chem. Phys.* **2012**, *136*, 124114.
- [125] Sylvetsky, N.; Peterson, K. A.; Karton, A.; Martin, J. M. L. Toward a W4-F12 approach: Can explicitly correlated and orbital-based ab initio CCSD(T) limits be reconciled? *J. Chem. Phys.* 2016, 144, 214101.
- [126] Bomble, Y. J.; Vázquez, J.; Kállay, M.; Michauk, C.; Szalay, P. G.; Császár, A. G.; Gauss, J.; Stanton, J. F. High-accuracy extrapolated ab initio thermochemistry. II. Minor improvements to the protocol and a vital simplification. *J. Chem. Phys.* 2006, *125*, 064108.
- [127] Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. Highaccuracy extrapolated ab initio thermochemistry. III. Additional improvements and overview. *J. Chem. Phys.* 2008, 128, 114111.

- [128] Feller, D.; Peterson, K. A.; De Jong, W. A.; Dixon, D. A. Performance of coupled cluster theory in thermochemical calculations of small halogenated compounds. *J. Chem. Phys.* 2003, *118*, 3510–3522.
- [129] Feller, D.; Dixon, D. A.; Francisco, J. S. Coupled Cluster Theory Determination of the Heats of Formation of Combustion-Related Compounds: CO, HCO, CO₂, HCO₂, HOCO, HC(O)OH, and HC(O)OOH. J. Phys. Chem. A 2003, 107, 1604–1617.
- [130] East, A. L. L.; Allen, W. D. The heat of formation of NCO. J. Chem. Phys. 1993, 99, 4638–4650.
- [131] Allen, W. D.; East, A. L. L.; Császár, A. G. In Structures and Conformations of Non-Rigid Molecules; Laane, J., Dakkouri, M., Veken, B., Oberhammer, H., Eds.; Springer Netherlands: Dordrecht, 1993; p 343.
- [132] Császár, A. G.; Allen, W. D.; Schaefer III, H. F. In pursuit of the ab initio limit for conformational energy prototypes. J. Chem. Phys. 1998, 108, 9751–9764.
- [133] Császár, A. G.; Tarczay, G.; Leininger, M. L.; Polyansky, O. L.; Tennyson, J.; Allen, W. D. In Spectroscopy from Space; Demaison, J., Sarka, K., Cohen, E. A., Eds.; Springer Netherlands: Dordrecht, 2001.
- [134] Kenny, J. P.; Allen, W. D.; Schaefer III, H. F. Complete basis set limit studies of conventional and R12 correlation methods: The silicon dicarbide (SiC₂) barrier to linearity. J. Chem. Phys. 2003, 118, 7353.
- [135] Gonzales, J. M.; Pak, C.; Cox, R. S.; Allen, W. D.; Schaefer III, H. F.; Császár, A. G.; Tarczay, G. Definitive Ab Initio Studies of Model S_N2 Reactions CH₃X+F (X=F, Cl, CN, OH, SH, NH₂, PH₂). *Chem. - A Eur. J.* **2003**, *9*, 2173–2192.
- [136] Jorgensen, K. R.; Wilson, A. K. Enthalpies of formation for organosulfur compounds: Atomization energy and hypohomodesmotic reaction schemes via ab initio composite methods. *Comput. Theor. Chem.* 2012, 991, 1–12.
- [137] Jorgensen, K. R.; Cadena, M. Theoretical study of bromine halocarbons: Accurate enthalpies of formation. *Comput. Theor. Chem.* 2018, 1141, 66–73.
- [138] Manaa, M. R.; Fried, L. E.; Kuo, I.-F. W. Determination of enthalpies of formation of energetic molecules with composite quantum chemical methods. *Chem. Phys. Lett.* 2016, 648, 31–35.
- [139] Jorgensen, K. R.; Oyedepo, G. A.; Wilson, A. K. Highly energetic nitrogen species: Reliable energetics via the correlation consistent Composite Approach (ccCA). J. Hazard. Mater. 2011, 186, 583–589.
- [140] Alsunaidi, Z. H.; Wilson, A. K. DFT and ab initio composite methods: Investigation of oxygen fluoride species. *Comput. Theor. Chem.* 2016, 1095, 71–82.

- [141] Karton, A.; Daon, S.; Martin, J. M. L. W4-11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem. Phys. Lett.* 2011, *510*, 165–178.
- [142] Simmie, J. M.; Somers, K. P. Benchmarking Compound Methods (CBS-QB3, CBS-APNO, G3, G4, W1BD) against the Active Thermochemical Tables: A Litmus Test for Cost-Effective Molecular Formation Enthalpies. J. Phys. Chem. A 2015, 119, 7235–7246.
- [143] Osmont, A.; Chetehouna, K.; Chaumeix, N.; DeYonker, N. J.; Catoire, L. Thermodynamic data of known volatile organic compounds (VOCs) in Rosmarinus officinalis : Implications for forest fire modeling. *Comput. Theor. Chem.* 2015, 1073, 27–33.
- [144] Ho, D. S.; DeYonker, N. J.; Wilson, A. K.; Cundari, T. R. Accurate enthalpies of formation of alkali and alkaline earth metal oxides and hydroxides: Assessment of the correlation consistent composite approach (ccCA). J. Phys. Chem. A 2006, 110, 9767–9770.
- [145] DeYonker, N. J.; Ho, D. S.; Wilson, A. K.; Cundari, T. R. Computational s-block thermochemistry with the correlation consistent composite approach. J. Phys. Chem. A 2007, 111, 10776–10780.
- [146] DeYonker, N. J.; Mintz, B.; Cundari, T. R.; Wilson, A. K. Application of the correlation consistent composite approach (ccCA) to third-row (Ga-Kr) molecules. J. Chem. Theory Comput. 2008, 4, 328–334.
- [147] Feller, D. Application of systematic sequences of wave functions to the water dimer. *J. Chem. Phys.* **1992**, *96*, 6104–6114.
- [148] Feller, D. The use of systematic sequence of wave functions for estimating the complete basis set, full configuration interaction limit in water. *J. Chem. Phys.* **1993**, *98*, 7059–7071.
- [149] Peterson, K. A.; Woon, D. E.; Dunning Jr., T. H. Benchmark calculations with correlated molecular wave functions. IV. The classical barrier height of the H+H₂→H₂+H reaction. J. *Chem. Phys.* **1994**, *100*, 7410–7415.
- [150] Kutzelnigg, W.; Morgan, J. D. Rates of convergence of the partial-wave expansions of atomic correlation energies. J. Chem. Phys. 1992, 96, 4484–4508.
- [151] Martin, J. M. L. Ab initio total atomization energies of small molecules towards the basis set limit. *Chem. Phys. Lett.* **1996**, 259, 669–678.
- [152] Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. J. Chem. Phys. 1997, 106, 9639–9646.
- [153] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N₂, and H₂O. *Chem. Phys. Lett.* 1998, 286, 243–252.
- [154] Williams, T. G.; DeYonker, N. J.; Ho, B. S.; Wilson, A. K. The correlation Consistent composite Approach: The spin contamination effect on an MP2-based composite methodology. *Chem. Phys. Lett.* 2011, 504, 88–94.

- [155] Douglas, M.; Kroll, N. M. Quantum electrodynamical corrections to the fine structure of helium. Ann. Phys. (N. Y). 1974, 82, 89–155.
- [156] Hess, B. A. Applicability of the no-pair equation with free-particle projection operators to atomic and molecular structure calculations. *Phys. Rev. A* **1985**, *32*, 756–763.
- [157] Hess, B. A. Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators. *Phys. Rev. A* **1986**, *33*, 3742–3748.
- [158] Moore, C. E. *Atomic Energy Levels, Vol. I (Hydrogen through Vanadium)*; Circular of the National Bureau of Standards 467: Washington D.C., 1949.
- [159] Das, S. R.; Williams, T. G.; Drummond, M. L.; Wilson, A. K. A QM/QM multilayer composite methodology: The ONIOM correlation consistent composite approach (ONIOMccCA). J. Phys. Chem. A 2010, 114, 9394–9397.
- [160] Riojas, A. G.; John, J. R.; Williams, T. G.; Wilson, A. K. Proton affinities of deoxyribonucleosides via the ONIOM-ccCA methodology. J. Comput. Chem. 2012, 33, 2590–2601.
- [161] Prascher, B. P.; Lai, J. D.; Wilson, A. K. The resolution of the identity approximation applied to the correlation consistent composite approach. *J. Chem. Phys.* **2009**, *131*, 044130.
- [162] Honig, B.; Karplus, M. Implications of torsional potential of retinal isomers for visual excitation. *Nature* **1971**, *229*, 558–560.
- [163] Pariser, R.; Parr, R. G. A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. II. J. Chem. Phys. **1953**, 21, 767–776.
- [164] Pople, J. A. Electron interaction in unsaturated hydrocarbons. *Trans. Faraday Soc.* **1953**, *49*, 1375.
- [165] Warshel, A.; Karplus, M. Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization. J. Am. Chem. Soc. 1972, 94, 5612–5625.
- [166] Warshel, A.; Karplus, M. Calculation of $\pi\pi^*$ Excited State Conformations and Vibronic Structure of Retinal and Related Molecules. *J. Am. Chem. Soc.* **1974**, *96*, 5677–5689.
- [167] Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J. Mol. Biol. 1976, 103, 227–249.
- [168] Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. Angew. Chemie Int. Ed. 2009, 48, 1198–1229.
- [169] Maseras, F.; Morokuma, K. IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. J. Comput. Chem. 1995, 16, 1170–1179.

- [170] Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. J. Mol. Struct. THEOCHEM 1999, 461-462, 1–21.
- [171] Hopkins, B. W.; Tschumper, G. S. A multicentered approach to integrated QM/QM calculations. Applications to multiply hydrogen bonded systems. J. Comput. Chem. 2003, 24, 1563–1568.
- [172] Humbel, S.; Sieber, S.; Morokuma, K. The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for n-butane conformation and S_N2 reaction: RCl⁺Cl⁻. J. Chem. Phys. **1996**, 105, 1959–1967.
- [173] Karadakov, P. B.; Morokuma, K. ONIOM as an efficient tool for calculating NMR chemical shielding constants in large molecules. *Chem. Phys. Lett.* 2000, 317, 589–596.
- [174] Rega, N.; Iyengar, S. S.; Voth, G. A.; Schlegel, H. B.; Vreven, T.; Frisch, M. J. Hybrid Ab-Initio/Empirical Molecular Dynamics: Combining the ONIOM Scheme with the Atom-Centered Density Matrix Propagation (ADMP) Approach. J. Phys. Chem. B 2004, 108, 4210–4220.
- [175] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels—Alder Reactions and Pt(P(t-Bu)₃)₂ + H₂ Oxidative Addition. J. Phys. Chem. 1996, 100, 19357–19363.
- [176] Vreven, T.; Mennucci, B.; Da Silva, C. O.; Morokuma, K.; Tomasi, J. The ONIOM-PCM method: Combining the hybrid molecular orbital method and the polarizable continuum model for solvation. Application to the geometry and properties of a merocyanine in solution. *J. Chem. Phys.* 2001, 115, 62–72.
- [177] Vreven, T.; Morokuma, K. On the Application of the IMOMO (Integrated Molecular Orbital + Molecular Orbital) Method. *J. Comput. Chem.* **2000**, *21*, 1419–1432.
- [178] Vreven, T.; Morokuma, K. Investigation of the $S_0 \rightarrow S_1$ excitation in bacteriorhodopsin with the ONIOM(MO:MM) hybrid method. *Theor. Chem. Acc.* **2003**, *109*, 125–132.
- [179] Matsubara, T.; Sieber, S.; Morokuma, K. A test of the new "integrated MO + MM" (IMOMM) method for the conformational energy of ethane and n-butane. *Int. J. Quantum Chem.* 1996, 60, 1101–1109.
- [180] Qi, X.-J.; Liu, L.; Fu, Y.; Guo, Q. X. Ab Initio Calculations of pK_a Values of Transition-Metal Hydrides in Acetonitrile. *Organometallics* 2006, 25, 5879–5886.
- [181] Tsai, Y.-C.; Lu, D.-Y.; Lin, Y.-M.; Hwang, J.-K.; Yu, J.-S. K. Structural transformations in dinuclear zinc complexes involving Zn-Zn bonds. *Chem. Commun. (Camb).* 2007, 4125– 4127.

- [182] Ogasawara, M.; Maseras, F.; Gallego-Planas, N.; Kawamura, K.; Ito, K.; Toyota, K.; Streib, W. E.; Komiya, S.; Eisenstein, O.; Caulton, K. G. Competition between Steric and Electronic Control of the Structure in Ru(CO)₂L₂L' Complexes. *Organometallics* 1997, 16, 1979–1993.
- [183] McKee, M. L.; Hill, W. E. ONIOM study of the coordination chemistry of Ag+ with the nitrogen-bridge ligands Ph₂P-NH-PPh₂ and Ph₂P-NCH₃-PPh₂: Ligand chelation versus bridging. J. Phys. Chem. A 2002, 106, 6201–6205.
- [184] Decker, S. A.; Cundari, T. R. Hybrid QM/MM study of propene insertion into the Rh-H bond of HRh(PPh₃)₂(CO)(η^2 -CH₂=CHCH₃): The role of the olefin adduct in determining product selectivity. *J. Organomet. Chem.* **2001**, *635*, 132–141.
- [185] Balcells, D.; Carbó, J. J.; Maseras, F.; Eisenstein, O. Self-consistency versus "best-fit" approaches in understanding the structure of metal nitrosyl complexes. *Organometallics* 2004, 23, 6008–6014.
- [186] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. Phys. Rev. 1964, 136, B864–B871.
- [187] Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- [188] Burke, K. Perspective on density functional theory. J. Chem. Phys. 2012, 136, 150901.
- [189] Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. J. Chem. Phys. 2005, 123, 062201.
- [190] Dirac, P. A. M. Note on Exchange Phenomena in the Thomas Atom. *Math. Proc. Cambridge Philos. Soc.* **1930**, *26*, 376–385.
- [191] Davin, T. J. Computational chemistry of organometallic and inorganic species. Thesis, University of Glasgow, 2010.
- [192] Becke, A. D. Perspective: Fifty years of density-functional theory in chemical physics. J. Chem. Phys. 2014, 140, 18A301.
- [193] Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Many-electron self-interaction error in approximate density functionals. *J. Chem. Phys.* **2006**, *125*, 201102.
- [194] Perdew, J. P.; Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **1981**, *23*, 5048–5079.
- [195] Van Leeuwen, R.; Baerends, E. J. Exchange-correlation potential with correct asymptotic behavior. *Phys. Rev. A* **1994**, *49*, 2421–2431.
- [196] Becke, A. D. A new inhomogeneity parameter in density-functional theory. *J. Chem. Phys.* **1998**, *109*, 2092–2098.

- [197] Zhao, Y.; Truhlar, D. G. Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* **2008**, *41*, 157–167.
- [198] Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. J. Chem. Phys. 2006, 124, 034108.
- [199] Schwabe, T.; Grimme, S. Towards chemical accuracy for the thermodynamics of large molecules: new hybrid density functionals including non-local correlation effects. *Phys. Chem. Chem. Phys.* 2006, 8, 4398.
- [200] Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
- [201] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. 2010, 132, 154104.
- [202] Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-consistent molecular-orbital methods. I. Use of gaussian expansions of slater-type atomic orbitals. *J. Chem. Phys.* **1969**, *51*, 2657–2664.
- [203] Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. J. Chem. Phys. 1972, 56, 2257–2261.
- [204] Raghavachari, K.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. J. Chem. Phys. 1980, 72, 650.
- [205] Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.* **1980**, *102*, 939–947.
- [206] Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. J. Chem. Phys. **1989**, 90, 1007–1023.
- [207] Woon, D. E.; Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. J. Chem. Phys. **1993**, 98, 1358–1371.
- [208] Woon, D. E.; Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. J. Chem. Phys. 1994, 100, 2975–2988.
- [209] Woon, D. E.; Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon. J. Chem. Phys. 1995, 103, 4572–4585.
- [210] Wilson, A. K.; Woon, D. E.; Peterson, K. A.; Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. IX. The atoms gallium through krypton. *J. Chem. Phys.* 1999, *110*, 7667–7676.

- [211] Dunning Jr., T. H.; Peterson, K. A.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. J. Chem. Phys. 2001, 114, 9244–9253.
- [212] De Jong, W. A.; Harrison, R. J.; Dixon, D. A. Parallel Douglas-Kroll energy and gradients in NWChem: Estimating scalar relativistic effects using Douglas-Kroll contracted basis sets. *J. Chem. Phys.* 2001, 114, 48–53.
- [213] Peterson, K. A.; Dunning Jr., T. H. Accurate correlation consistent basis sets for molecular core-valence correlation effects: The second row atoms Al-Ar, and the first row atoms B-Ne revisited. J. Chem. Phys. 2002, 117, 10548–10560.
- [214] Peterson, K. A.; Figgen, D.; Dolg, M.; Stoll, H. Energy-consistent relativistic pseudopotentials and correlation consistent basis sets for the 4d elements Y-Pd. J. Chem. Phys. 2007, 126, 124101.
- [215] Figgen, D.; Peterson, K. A.; Dolg, M.; Stoll, H. Energy-consistent pseudopotentials and correlation consistent basis sets for the 5d elements Hf-Pt. J. Chem. Phys. 2009, 130, 164108.
- [216] Lu, Q.; Peterson, K. A. Correlation consistent basis sets for lanthanides: The atoms La–Lu. J. Chem. Phys. 2016, 145, 054111.
- [217] Hellmann, H. A New Approximation Method in the Problem of Many Electrons. J. Chem. *Phys.* **1935**, *3*, 61–61.
- [218] Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.
- [219] Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, 8, 1057.
- [220] Stoychev, G. L.; Auer, A. A.; Neese, F. Automatic Generation of Auxiliary Basis Sets. J. *Chem. Theory Comput.* **2017**, *13*, 554–562.
- [221] Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1327.
- [222] Born, M. Volumen und Hydratationswärme der Ionen. Zeitschrift für Phys. 1920, 1, 45–48.
- [223] Onsager, L. Electric Moments of Molecules in Liquids. J. Am. Chem. Soc. 1936, 58, 1486– 1493.
- [224] Tomasi, J. Cavity and reaction field: "robust" concepts. Perspective on "Electric moments of molecules in liquids". *Theor. Chem. Acc.* **2000**, *103*, 196–199.
- [225] Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges. J. Chem. Theory Comput. 2007, 3, 2011–2033.

- [226] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. J. Phys. Chem. B 2009, 113, 6378–6396.
- [227] Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. Refinement and parametrization of COSMO-RS. J. Phys. Chem. A 1998, 102, 5074–5085.
- [228] Ho, J.; Coote, M. L. A universal approach for continuum solvent pK_a calculations: Are we there yet? *Theor. Chem. Acc.* **2009**, *125*, 3–21.
- [229] Ho, J. Predicting pK_a in Implicit Solvents: Current Status and Future Directions. *Aust. J. Chem.* **2014**, *67*, 1441.
- [230] Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J. Chem. Soc. Perkins Trans. 2 1993, 799–805.
- [231] Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117–129.
- [232] Miertuš, S.; Tomasi, J. Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes. *Chem. Phys.* **1982**, *65*, 239–245.
- [233] Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3093.
- [234] Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. J. Phys. Chem. A 1998, 102, 1995–2001.
- [235] Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. J. Comput. Chem. 2003, 24, 669–681.
- [236] Hodgson, J. L.; Roskop, L. B.; Gordon, M. S.; Lin, C. Y.; Coote, M. L. Side reactions of nitroxide-mediated polymerization: N-O versus O-C cleavage of alkoxyamines. J. Phys. Chem. A 2010, 114, 10458–10466.
- [237] Qi, X.-J.; Fu, Y.; Liu, L.; Guo, Q. X. Ab initio calculations of thermodynamic hydricities of transition-metal hydrides in acetonitrile. *Organometallics* **2007**, *26*, 4197–4203.
- [238] Mo, S. J.; Vreven, T.; Mennucci, B.; Morokuma, K.; Tomasi, J. Theoretical study of the S_N^2 reaction of $Cl^-(H_2O)^+CH_3Cl$ using our own N-layered integrated molecular orbital and molecular mechanics polarizable continuum model method (ONIOM, PCM). *Theor. Chem.* Acc. 154–161.
- [239] Carter, S.; Bowman, J. M.; Handy, N. C. Extensions and tests of "multimode": A code to obtain accurate vibration/rotation energies of many-mode molecules. *Theor. Chem. Acc.* **1998**, *100*, 191–198.

- [240] Manzhos, S.; Dawes, R.; Carrington, T. Neural network-based approaches for building high dimensional and quantum dynamics-friendly potential energy surfaces. *Int. J. Quantum Chem.* 2015, 115, 1012–1020.
- [241] Kamath, A.; Vargas-Hernández, R. A.; Krems, R. V.; Carrington, T.; Manzhos, S. Neural networks vs Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy. J. Chem. Phys. 2018, 148.
- [242] Wilson Jr, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*, 1st ed.; Dover Publications, Inc.: New York, NY, 1980.
- [243] Bowman, J. M. Self-consistent field energies and wavefunctions for coupled oscillators. J. *Chem. Phys.* **1978**, *68*, 608–610.
- [244] Cohen, M.; Greita, S.; McEarchran, R. Approximate and exact quantum mechanical energies and eigenfunctions for a system of coupled oscillators. *Chem. Phys. Lett.* **1979**, *60*, 445–450.
- [245] Gerber, R. B.; Ratner, M. A. A semiclassical self-consistent field (SC SCF) approximation for eigenvalues of coupled-vibration systems. *Chem. Phys. Lett.* **1979**, *68*, 195–198.
- [246] Bowman, J. M. The Self-Consistent-Field Approach to Polyatomic Vibrations. *Acc. Chem. Res.* **1986**, *19*, 202–208.
- [247] Chaban, G. M.; Jung, J. O.; Benny Gerber, R. Ab initio calculation of anharmonic vibrational states of polyatomic systems: Electronic structure combined with vibrational self-consistent field. J. Chem. Phys. 1999, 111, 1823–1829.
- [248] Roy, T. K.; Gerber, R. B. Vibrational self-consistent field calculations for spectroscopy of biological molecules: New algorithmic developments and applications. *Phys. Chem. Chem. Phys.* 2013, 15, 9468–9492.
- [249] Christiansen, O. Vibrational coupled cluster theory. J. Chem. Phys. 2004, 120, 2149–2159.
- [250] Christoffel, K. M.; Bowman, J. M. Investigations of self-consistent field, scf ci and virtual stateconfiguration interaction vibrational energies for a model three-mode system. *Chem. Phys. Lett.* **1982**, 85, 220–224.
- [251] Pouchan, C.; Aouni, M.; Bégué, D. Ab initio determination of the anharmonic vibrational spectra of P₂O in the region 200–2000 cm⁻¹. *Chem. Phys. Lett.* **2001**, *334*, 352–356.
- [252] Baraille, I.; Larrieu, C.; Dargelos, A.; Chaillet, M. Calculation of non-fundamental IR frequencies and intensities at the anharmonic level. I. The overtone, combination and difference bands of diazomethane, H₂CN₂. *Chem. Phys.* 2001, 273, 91–101.
- [253] Carbonniere, P.; Begue, D.; Pouchan, C. Anharmonic Force Field and Vibrational Spectra of Perfluoromethanimine CF₂NF. *J. Phys. Chem. A* **2002**, *106*, 9290–9293.

- [254] Scribano, Y.; Benoit, D. M. Iterative active-space selection for vibrational configuration interaction calculations using a reduced-coupling VSCF basis. *Chem. Phys. Lett.* 2008, 458, 384–387.
- [255] Huron, B.; Malrieu, J. P.; Rancurel, P. Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions. J. Chem. Phys. 1973, 58, 5745–5759.
- [256] Respondek, I.; Benoit, D. M. Fast degenerate correlation-corrected vibrational self-consistent field calculations of the vibrational spectrum of 4-mercaptopyridine. *J. Chem. Phys.* **2009**, *131*, 054109.
- [257] Scribano, Y.; Lauvergnat, D. M.; Benoit, D. M. Fast vibrational configuration interaction using generalized curvilinear coordinates and self-consistent basis. J. Chem. Phys. 2010, 133.

CHAPTER 3

PREDICTION OF pKa OF LATE TRANSITION METAL HYDRIDES VIA A QM/QM APPROACH

3.1 Introduction

Transition metal (TM) hydrides are important intermediates in many catalytic and stoichiometric processes such as hydrogenation and hydroformylation.^{1–9} As numerous organometallic catalytic reactions include hydride transfers, characterizing metal-ligand binding properties is vital to understanding how these catalysts work. One such thermodynamic property for TM hydrides is the pK_a . Although the pK_a values of a number of TM hydrides have been measured experimentally, experimental characterization of pK_a is not accessible for all TM hydrides. Therefore, with computational approaches, such as density functional theory (DFT), geometries, spectroscopic constants, and energetics, and thermodynamic properties such as pKas, Gibbs free energies, and enthalpies of formation, become an important route to predict various molecular and thermodynamic properties in the absence of experimental measurements.^{10–17}

The development of density functionals has motivated their wide application, and, in 2001, Perdew proposed the Jacob's ladder analogy to classify density functionals into primary rungs that present the hierarchy of density approximations.¹⁸ This is explained in Section 2.3. Essentially, the inherent complexity of functional class increases with higher rungs in Jacob's ladder; however, the accuracy of a functional is not necessarily dependent on its complexity. Therefore, in determining choice of density functional, calibration of density functional approaches with data from experiment or high accuracy wavefunction-based calculations such as CCSD(T) should be done.¹⁹ Though CCSD(T) is often considered as the "gold standard" of quantum chemistry, it is not computationally affordable (memory, disk space, CPU time) for routine calculations of many TM complexes, which often are bound to numerous large ligands.^{20–24}

For the prediction of thermodynamic properties of TM-containing complexes with sterically

hindering ligands, such as for TM hydrides, DFT is often considered, as it is readily used for molecules of increasing size and complexity.²⁵ In a study by Tekarli et al., the gas-phase enthalpies of formation (ΔH_f) of 19 3d TM-containing species were calculated to assess the performance of 44 density functionals paired with cc-pVTZ and cc-pVQZ basis sets.²⁶Among the considered functionals, the B97-1, and PBE1KCIS functionals resulted in the lowest mean absolute deviations (MADs) relative to experiment. A similar study was done by Laury et al. for the ΔH_f 's of 30 4d species, considering the utility of 22 density functionals. Of the functionals considered, B2GP-PLYP and mPW2-PLYP yielded the lowest MAD from experiment.²⁷ Riley and Merz²⁸ examined the performance of 12 functionals with the 6-31G* and TZVP basis sets for the calculation of ΔH_{f} 's of 94 TM species. TPSS1KCIS in combination with the TZVP basis set resulted in the lowest MAD from experiment in their study. Wang et al.'s study of TM atom mediated C_β -O bond cleavage of the β -O-4 linkage of lignin used density functionals to compare the binding, activation, and reaction enthalpies with respect to CR-CCSD(T).²⁹ They found that the property that yielded the lowest MADs from CR-CCSD(T) fluctuated depending on functional choice as well as choice of 3d, 4d, or 5d metal. Overall, the lowest average deviation from CR-CCSD(T) for predicting the reaction energetics was provided by PBE0. These gas phase studies demonstrate that functional choice should be strongly based upon the molecular systems of interest, considering the TM, as well as the number and types of ligands, and the property of interest for TM species.

Since the ligands for bulky TM hydrides predominately consist of main group atoms, approaches that have been useful for main group thermochemistry should be considered in identifying approaches that may be effective for the description of TM hydrides. In a study by Goerigk and Grimme, ³⁰ a thorough benchmark of 47 density functionals from the GMTKN30 database for general main group thermochemistry recommended functionals including the GGA B97-D3 and the meta-GGA, oTPSS-D3. PW6B95 was identified as the most robust hybrid in their study. However, in comparing their results for main group species to the aforementioned TM complexes, ^{26–28} functionals that are optimal for each TM may not perform well for the ligands.

In the development of the MN15-L density functional, Yu et al. ranked 48 density functionals

based on their performance for 33 molecular databases.³¹ This study showed that B97-1, which performed well for the thermochemistry of 3*d* TM-containing compounds,²⁶ ranked 3 out of the 48 chosen functionals for SR-MGM-BE9, which examines single-reference main-group metal bond energies, but ranked 29 for π TC13, which examines thermochemistry of hydrocarbon π systems. The opposite trend is shown for M11-L, which ranked 14 for π TC13 but 45 for SR-MGM-BE9. The rank changes of B97-1 and M11-L for the 2 databases emphasize that some density functionals are good for TM chemistry but poor at describing main group ligands, and vice versa.

As ligand complexity increases, solvation shells of the complex as well as the electronic and steric effects of ligands should be considered alongside the chemically important region with the metal center. However, a single functional may not portray all aspects of increasingly complex systems useful for homogeneous catalysis effectively. Thus, the main goal of this study is to develop a scheme that accounts for an optimal method choice for the metal and an optimal method choice for the ligand.

For systems containing numerous non-hydrogen atoms, the use of cost-effective multilayer fragmentation approaches such as ONIOM, ^{32–36} Molecules-in-Molecules, ³⁷ and the Molecular Tailoring Approach³⁸ can provide a framework for such a combination. However, as the ONIOM method (Section 2.2.5) has been commonly applied to transition metal complexes and homogeneous catalysis, ^{39–41} whereas other fragmentation approaches are often utilized for biomolecules and water clusters, ^{37,42,43} ONIOM is used in this study. However, because of the size of many TM hydrides, it can be costly to use high level theoretical methods (e.g. CCSD(T)) to directly model them, and even within an ONIOM scheme, the size of the model layer can also limit the application of high level theoretical methods in the model layer, making them impractical. Thus, while a combination of a higher level (HL) method and a lower level (LL) method demonstrates a traditional use of ONIOM [i.e., ONIOM(HL:LL)], such a layering scheme can also be utilized to consider the strengths of methods in a metal and non-metal partitioning of a molecule, as is done in this chapter.

As compared with the number of gas phase computational studies on TM species, far fewer

studies have been reported on the solvent effects on TM compounds. Such studies are important, as many TM reactions are carried out in a solvated phase, including TM hydride-mediated catalysis. For the solvated phase, the pK_a exhibits the strongest effects of solvation relative to their gas phase analogs due to the charge separation of the species involved. Previous studies by Liptak and Shields^{44–47} and others^{48–52} have examined the use of both direct and relative thermodynamic schemes for pK_a calculations. These studies show that direct thermodynamic schemes for pK_a s of unknown acids have excellent agreement with experiment with reduced computational cost over relative schemes. Therefore, the direct thermodynamic scheme, shown in Scheme 3.1, will be used for this work.

Implicit solvation models (see Section 2.5) are often utilized for practical computations of bulk TM species.^{53–57} For implicit solvent models, the choice of a cavity model, which defines the shape and size of the cavity occupied by a solute species in the solvent, has been shown to have an impact on the prediction of the pK_a of organic acids using DFT.^{58–60} For instance, in the study of the aqueous solvation free energies of 10 organic species calculated with seven cavities (UAKS, UAHF, UAHF, Bondi, Pauling, UA0, and UFF) using the B3LYP/6-31+G(d) method with the C-PCM solvation model, UAKS and UAHF resulted in the lowest MADs relative to experiment in comparison to the other considered cavity models.⁵⁸ Also, a systematic study of solvation free energy and pK_a values of monoprotic, diprotic, and triprotic acids based on DFT(B3LYP, PBE, BVP86, and M05-2X)/aug-cc-pVTZ methods combined with the C-PCM and SMD solvation models showed that the Pauling cavity in combination with M05-2X resulted in the lowest deviation among the UFF, UAKS, Pauling, and Klamt cavity models.⁶⁰ Though the prediction of pK_a values has been shown to be related to the choice of cavity model, studies showing the utility of density functionals in terms of the choice of cavity models for TM-containing species are limited.⁶¹

In a study by Qi et al.,⁶² using CCSD(T) with an insufficient basis set, such as LANL2DZ+p, to calculate the model layer of TM hydrides was found to fail dramatically in describing TM hydrides, while an improvement of the basis set achieved better results. However, further improving the basis set will make CCSD(T) impractical in the treatment of the model layer. Thus, they tried to use

density functionals to describe the whole systems with a high-level basis set to describe the model layer and low level basis sets to describe the rest of region, which yields much better results than CCSD(T) with a low-level basis set. Therefore, DFT can perform well in calculating properties of TM hydrides and the choice of basis set is more important than the choice of method (e.g., CCSD(T) vs density functionals).

As shown above, ^{26–28,30} TM (model layer) and main group elements (the main component of TM hydrides) can be described well with multiple density functionals. Therefore, instead of using the same density functionals to describe the whole systems, it is worth examining if the combination of different density functionals in ONIOM will provide better description for TM hydrides systems. To assess the appropriateness of density functionals combined with several levels of basis sets within the ONIOM scheme for TM hydrides in solvated phase, comprehensive studies must be carried out where a much wider variety of functionals are considered.

In this chapter, to address the ability of electronic structure methods to describe the pK_as of TM hydrides, density functionals utilized in partnership with basis sets of at least triple- ζ quality are investigated, including ONIOM(DFT:DFT) schemes.⁶³ As well, to consider TM chemistry in solution, the impact of solvent model (SMD, COSMO, and C-PCM) and the degree to which the several cavity models affect the determination of the pK_a values of TM hydrides are analyzed in this study. The influence of the addition of exact exchange and dispersion corrections is considered. As shown in the above examples^{26,27,64} and several other studies,^{65–67} the choice of basis set and the size of the molecules^{28,64} can have an impact on the utility of density functionals; therefore, an understanding of the influence of basis sets. This investigation provides insight about the selection of computational methods for TM hydrides that can be applied to investigate other thermodynamic properties of catalysts for many important chemical reactions, such as hydrogenation and hydroformylation.

3.2 Theoretical Methods

The two layer ONIOM scheme was used with a variety of density functionals and several basis sets to determine pK_a values of Group 10 TM hydrides ([HNi(depe)₂]⁺, [HNi(depp)₂]⁺, [HNi(depe)₂]⁺, [HPd(depe)₂]⁺, [HPd(depe)₂]⁺, [HPd(depe)₂]⁺, [HPd(depe)₂]⁺, [HPt(PNP)₂]⁺). All calculations were performed using the GAUSSIAN 09 software package.⁶⁸ For all considered TM hydrides, geometry optimizations and frequency calculations (using vibrational ZPE scaled by 0.9890)⁶⁹ were performed using B3LYP/cc-pVTZ in both the gas phase and acetonitrile solvent to replicate experimental conditions. Acetonitrile solvent systems were treated using the C-PCM, COSMO, and SMD continuum solvation models.^{53–57} All stationary points were verified to be true minima, with no imaginary frequencies. The thermochemical corrections from B3LYP/cc-pVTZ frequency calculations were added to the single point energies to obtain gas phase and solvation free energies at 298 K.

Subsequently, single-point calculations were performed with the two-layer ONIOM method presented in Section 2.2.5 with Equation 2.38.^{32–36} Since choosing how to partition the molecular systems into layers can have a significant impact upon the calculated energies, several core regions have been considered: (a) the metal atom and four phosphorous atoms; (b) the metal, phosphorous atoms, and the chelate rings; and, (c) all atoms except for the terminal methyl group. The results from these expansions are defined in this study as ONIOM-1, ONIOM-2, and ONIOM-3, respectively, and are shown in Figure 3.1. The ONIOM-1 scheme is primarily used due to computational cost.

To evaluate the impact of the DFT approaches used within ONIOM for the prediction of pK_as of TM hydrides, the following DFT methods were utilized (summarized in Table 3.1), listed by functional class: (a) Generalized Gradient Approximation (GGA): BLYP,^{70,71} PBE,⁷² and B97-D⁷³; (b) meta-GGA (M-GGA): M06L,¹⁶ BB95,⁷⁴,= and TPSS⁷⁵; (c) hybrid GGA (H-GGA): PBE0,^{72,76,77} B3LYP,^{70,71,78} and B3P86^{71,79}; (d) hybrid-meta GGA (HM-GGA): M06,¹⁶ M06HF⁸⁰; and, (e) double hybrid GGA (DH-GGA): B2PLYP⁸¹ based on their utilization for these types of compounds. Additionally, Grimme's empirical dispersion correction (D3)⁸² was added to several density functionals selected from GGA, M-GGA, H-GGA, and HM-GGA functionals,

to evaluate the effect of a dispersion correction on the accuracy of predictions of pK_as of the TM hydrides. To evaluate the impact of the percentage of exact exchange, the percentage of exact exchange for PBE0 was varied from 0% to 80% in intervals of 5% since PBE0 includes no empirical parameters that may affect the utility of DFT; hence, avoiding interference from other empirical parameters.



Figure 3.1: From left to right, the compounds are TM(depe)₂, TM(depp)₂, TM(PNP)₂. (a) The model system (bolded) within the ONIOM-1 QM/QM partitioning scheme for TM hydrides with the TM atom (Ni, Pd, and Pt) and four phosphorous atoms in the layer using the high-level method. (b) ONIOM-2: The QM/QM partitioning scheme for TM hydrides with all the atoms within the chelate rings in the layer using the high-level method. (c) ONIOM-3: The QM/QM partitioning scheme for TM hydrides with all except for the very outside methyl group in the layer using the high-level method.

	Туре	%HF	Exchange/Correlation
BLYP ^{70,71}	GGA ^a	0%	Becke88/Perdew86/Lee-Yang-Parr
PBE^{72}	GGA ^a	0%	Perdew-Burke-Ernzerhof/ Perdew-Burke-Ernzerhof
B97-D ⁷³	GGA ^a	0%	B97-D/B97-D
M06L ¹⁶	M-GGA ^b	0%	M06L/M06L
BB95 ⁷⁴	M-GGA ^b	0%	Becke88/Perdew86/Becke95
TPSS ⁷⁵	M-GGA ^b	0%	Tao-Perdew-Staroverov-Scuseria/Tao-Perdew-
			Staroverov-Scuseria
PBE0 ^{72,76,77}	H-GGA ^c	25%	Perdew-Burke-Ernzerhof/Perdew-Burke-Ernzerhof
B3LYP ^{70,71,78}	H-GGA ^c	20%	Becke88/Perdew86/Lee-Yang-Parr
B3P86 ^{71,79}	H-GGA ^c	20%	Becke88/Perdew86
M06 ¹⁶	HM-GGA ^d	27%	M06/M06
M05-2X ⁸³	HM-GGA ^d	52%	M05-2X/M05-2X
M06-2X ¹⁶	HM-GGA ^d	54%	M06-2X/M06-2X
M06HF ⁸⁰	HM-GGA ^d	100%	M06HF/M06HF
B2PLYP ⁸¹	DH-GGA ^e	50%	Becke88/Perdew86/Lee-Yang-Parr

Table 3.1: Summary of the density functionals utilized.

 a GGA (generalized-gradient approximation)

^bM-GGA (meta GGA)

^cH-GGA (hybrid GGA)

^dHM-GGA (hybrid meta GGA)

^eDH-GGA (double hybrid GGA)

For the lower level within the ONIOM calculations, the relativistic effective core potential (ECP) and valence double- ζ basis set of Hay and Wadt (LANL2DZ)⁸⁴ as well as the Stuttgart/Dresden (SDD)^{85–87} relativistic ECP and valence triple- ζ basis set were considered. For LANL2DZ and SDD, 10, 28, and 60 electrons were frozen for Ni, Pd, and Pt, respectively. For the high-level method, cc-pVDZ, cc-pVTZ, aug-cc-pVDZ, and aug-cc-pVTZ were used.^{88–91} For Ni species, correlation consistent basis sets with the one-particle Douglas-Kroll-Hess Hamiltonian for scalar relativistic effects were applied (e.g., aug-cc-pVTZ-DK) for all atoms.⁹² For Pd and Pt species, the small-core relativistic pseudopotential basis sets (e.g., aug-cc-pVTZ) were used for main group atoms since the pseudopotentials are incompatible with the DKH Hamiltonian and are constructed to account for relativistic effects of the heavy atom.^{90,91,93} In the following sections, the terms DK for Ni and PP

for Pd and Pt are dropped for clarity from the selected basis set notations.

Three implicit solvation models, SMD,⁵³ COSMO,⁵⁴ and C-PCM,^{55–57} were employed to include solvent effects in the single point calculations. The UA0, UAKS, Pauling, Bondi and default cavities (UFF for C-PCM, Klamt for COSMO, and Coulomb-SMD for SMD) were applied.

$$HL_{n}M^{+}_{(solv, 1M)} \xrightarrow{\Delta G_{sol}} H^{+}_{(solv, 1M)} + L_{n}M_{(solv, 1M)}$$

Scheme 3.1: The direct thermodynamic scheme

The direct thermodynamic scheme for calculating pK_as of unknown acids shown in Scheme 3.1 has been used mainly due to its demonstrated utility ${}^{46,48-52,94}$ and was used in this study with the value -4.39 kcal mol⁻¹ for the gas phase free energy of a proton, $\Delta G_{gas}(H^+)$, derived using the Sackur-Tetrode equation. 95 For the value of the experimental solvation phase free energy of the proton in acetonitrile, $\Delta G_{solv}(H^+)$, -260.2 kcal mol⁻¹ has been recommended 96 and was used in this study. Thus, the solvation free energy (ΔG_{sol}) can be calculated using the following equations (Eq. 3.1-3.5).

$$\Delta G_{sol} = \Delta G_{gas} + \Delta \Delta G_{solv} \tag{3.1}$$

$$\Delta G_{gas} = G_{gas}(L_n M) + G_{gas}(H^+) - G_{gas}(HL_n M^+)$$
(3.2)

$$\Delta\Delta G_{solv} = \Delta G_{solv}(L_n M) + \Delta G_{solv}(H^+) - \Delta G_{solv}(H L_n M^+)$$
(3.3)

$$\Delta G_{solv}(L_n M) = E_{solv}(L_n M) - E_{gas}(L_n M)$$
(3.4)

$$\Delta G_{solv}(HL_nM^+) = E_{solv}(HL_nM^+) - E_{gas}(HL_nM^+)$$
(3.5)

The pKa values related to free energies of solvation were calculated as

$$pK_a = \frac{\Delta G_{solv}}{2.303RT} \tag{3.6}$$

All of the calculated gas phase free energies in units atm were converted to molar units and the solvation phase free energies were calculated using $[(E_{soln} + G_{nes}) - E_{gas}]$, as defined in the parametrization of continuum solvent models.^{47,97} An error of 1.36 kcal mol⁻¹ in ΔG_{solv} results in a deviation of 1 pK_a unit. Ho and Coote reported that a direct thermodynamic cycle can be expected to depart from experiment by 3.5 pK_a units.⁹⁸

3.3 Results and Discussion

The considered molecules are grouped based on central TM atoms (Ni, Pd, and Pt) and the ligands (depe, depp, and PNP) in order to evaluate the impact of the selected density functionals, basis sets, cavities, solvation models, and the expansion in size of the high-level region within ONIOM on the calculated pK_as of TM hydrides. Mean absolute deviations (MADs) with respect to experimental data^{99–102} are reported. Since [HPt(depp)₂]⁺ does not have readily available experimental data for pK_a due to the highly reactive nature of Pt complexes, a net equation (Eq. 3.7) of the thermochemical cycle^{103,104} relating hydricities, pK_as , and redox potentials was used to calculate a proposed pK_a based on experimental redox potentials and hydricities. ¹⁰⁰ From Equation 3.7, the proposed pK_a for [HPt(depp)₂]⁺ is 28.3.

$$\Delta G_{H^-} = 1.37(pK_a) + 46.1E^{\circ}(\text{II/0}) + 79.6 \text{ kcal mol}^{-1}$$
(3.7)

None of the considered TM hydrides showed significant structural changes in the gas or solvation phases; therefore, the solvation phase structures obtained with C-PCM were used for the single point calculations based on computational cost.

3.3.1 Utility of DFT in the Real System

The fourteen density functionals (Table 3.1) were chosen for the real layer and PBE, M06-L, B3LYP, and M06 were chosen for the model layer. PBE, M06-L, B3LYP, and M06 were chosen to

showcase the tiers of functional complexity in the model layer. A summary of method and basis set choice for this section is provided in Table 3.2. Using C-PCM for [HNi(depp)₂]⁺ and TPSS for the real layer, the MAD when using PBE, M06-L, B3LYP, and M06 for the model layer was 11.7, 7.4, 6.2, and 9.0 pK_a units, respectively. When using B97-D for the real layer, the MAD when using PBE, M06-L, B3LYP, and M06 for the model layer were 9.7, 5.4, 4.2, and 7.0 pK_a units, respectively. Similarly, using C-PCM for [HPd(depe)₂]⁺, the MAD when using PBE, M06-L, B3LYP, and M06 for the model layer were 10.8, 7.3, 6.6, and 11.9 pK_a units, respectively, using TPSS in the real layer and 8.8, 5.3, 4.6, 9.9 pK_a units, respectively, using B97-D in the real layer. Since the MADs varied significantly based on functional choice in the model layer, the MADs from PBE, M06-L, B3LYP, and M06 are averaged to eliminate bias of functional complexity for the model layer. Therefore, for [HNi(depp)₂]⁺ and [HPd(depe)₂]⁺, the average MAD is 8.6 and 9.2 pK_a units for TPSS, and 6.6 and 7.2 pK_a units for B97-D. Averaging the MADs for the model layers and for the molecule set allowed the choice for the real layer to be compared more readily. Table 3.2: Theoretical methods for the description of real and model systems within the two-layer

ONIOM scheme using C-PCM, COSMO, and SMD for utility of DFT in the real layer.

Method	Model system ^a	Real system ^b
1	PBE/aug-cc-pVTZ	DFT
2	M06L/aug-cc-pVTZ	DFT
3	B3LYP/aug-cc-pVTZ	DFT
4	M06/aug-cc-pVTZ	DFT

^aaug-cc-pVTZ (main group atoms for Pd, Pt species), aug-cc-pVTZ-DK (Ni species), aug-cc-pVTZ-PP (Pd and Pt). ^bDFT functionals are listed in Table 3.1. LANL2DZ is used as the basis set for the real system.

For the molecule set, the average MAD in the pK_a from experiment is provided in Figure 3.2, where the considered density functional approach for the low level of the ONIOM approach has been varied. Among the functionals considered, B97-D performed best with MADs of 5.5, 2.7, and 2.3 pK_a units for C-PCM, COSMO, and SMD, respectively, followed by B3LYP (6.3, 3.4, 2.9 pK_a units), and M06-L (7.2, 4.5, 3.8 pK_a units). Except for B97-D, B3LYP, and M06L, all other GGA, M-GGA and H-GGA functionals performed similarly regarding each solvation model with MAD values of about 7.9, 5.0, and 4.3 pK_a units for C-PCM, COSMO, and SMD, respectively.

The functional with the highest MAD is M06-2X with MAD values of 10.5, 7.7, and 7.1 pK_a units for C-PCM, COSMO, and SMD, respectively. Among the three selected solvation models, SMD provided the best comparison with experimental pK_a data while C-PCM yielded the highest MADs for all fourteen considered density functionals.



Figure 3.2: MADs in pK_a values for the density functionals within low-level methods relative to experiment. All of the results are from calculations with ONIOM(PBE,M06L,B3LYP,M06/aug-cc-pVTZ:DFT/LANL2DZ) scheme. The results of using the four functionals in the model layer are averaged for the molecule set.

It is worth noting that using separate functionals for the core and real layers provided lower MADs than when the same functional is used for both layers within the ONIOM scheme regarding each solvation model. For instance, with SMD, ONIOM (B97-D/ aug-cc-pVTZ : PBE/LANL2DZ), ONIOM (B97-D/ aug-cc-pVTZ : M06L/ LANL2DZ), ONIOM (B97-D/

aug-cc-pVTZ : B3LYP/ LANL2DZ), and ONIOM (B97-D/ aug-cc-pVTZ : M06/ LANL2DZ) yielded MADs lower than ONIOM (PBE/ aug-cc-pVTZ : PBE/ LANL2DZ), ONIOM (M06L/ aug-cc-pVTZ : M06L/ LANL2DZ), ONIOM (B3LYP/ aug-cc-pVTZ : B3LYP/ LANL2DZ), and ONIOM(M06/ aug-cc-pVTZ : M06/ LANL2DZ) by 2.4, 1.6, 1.4, and 4.1 pK_a units, respectively. This shows that a mixed basis set approach may not be advantageous for TM hydride systems.



Figure 3.3: MADs in pK_a values for five types of density functionals, GGA, M-GGA, H-GGA, HM-GGA, and DH-GGA functionals, within low-level methods relative to experiment. All of the results are from calculations with ONIOM(PBE,M06L,B3LYP,M06/aug-cc-pVTZ:DFT/LANL2DZ) scheme. The results of using the four functionals in the model layer are averaged for the molecule set.

The utility of the types of density functionals at modeling pK_as is shown in Figure 3.3. The GGA (7.1, 4.2, and 3.6 pK_a units for C-PCM, COSMO, and SMD, respectively) and H-GGA

(7.3, 4.4, and 3.8 pK_a units for C-PCM, COSMO, and SMD, respectively) functionals produced similar MADs, which were better than all the other types of functionals regardless of solvation method. In contrast, DH-GGAs performed the worst with MADs of 9.8, 6.9, and 6.0 pK_a units for C-PCM, COSMO, and SMD, respectively, which indicates that the addition of a fraction of the PT2 correlation energy is a disadvantage for the description of pK_as of TM hydrides. Compared with HM-GGAs, M-GGA functionals, which do not include exact exchange, yielded lower MADs for all three solvation models. Therefore, exact exchange is not necessary for the description of the real system. COSMO and SMD performed similarly (5.3 and 4.7 pK_a units, respectively) and resulted in MADs ~3 pK_a units lower than that from C-PCM (8.1 pK_a units).

The comparison of functional types of is considered with respect to central TM atoms (Table 3.10) and ligand systems (Table 3.11) employing each of the three solvation models. For the Ni species, the MADs increased with increasing functional complexity, except for DH-GGAs for all three solvation models. For Pd and Pt species, H-GGAs yielded the lowest MADs in comparison to other types of functionals while DH-GGAs always performed the worst for all three solvation models. Moving from Ni to Pt, the MADs of non-local exchange functionals (H-GGA, HM-GGA, and DH-GGA) decrease, which indicates that non-local exchange in functionals can describe TM hydrides with heavier central TM atoms better than those with lighter central TM atoms. Considering the overall MADs of different types of functionals, the increase in MADs upon inclusion of exact exchange is more significant for M-GGA functionals (HM-GGA) than it is for the GGA functionals (H-GGA). As shown in Figure 3.1, the size of the considered ligands increases in the order of depe, depp, and PNP. Similar MAD was found for each type of functional between all three solvation models as the size of the ligand increased.

3.3.2 Utility of DFT in the Model Layer

As seen in the previous section, the fluctuation caused by the choice of the four density functionals in describing the model system of the ONIOM scheme implies that functional choice for both the model and real layer are factors in calculating pK_a values; therefore, the section

focuses on the utility of the density functionals for the model layer while keeping the functionals chosen for the real layer constant. Table 3.3 summarizes the combination of density functionals as ONIOM schemes designed to measure the influence of the fourteen considered density functionals combined with the aug-cc-pVTZ basis set in the description of model layers of the TM hydrides (Figure 3.1a). The real systems (Figure 3.1a) were treated with three density functionals (B97-D, M06L, and B3LYP) paired with the LANL2DZ basis set, which were selected based on their better performance as low-level methods shown in the previous section. The rationale for averaging the MADs from the three selected real system methods is to eliminate bias from the functional chosen for the real layer and gauge the utility of density functionals in the model layer, as done in the previous section for the real layer. The MADs for each high-level method, which are based upon deviations of the calculated pK_a values of the TM hydrides from experimental data for each functional using the C-PCM, COSMO, and SMD, are reported in Figure 3.4. For C-PCM and COSMO, the three bestperforming functionals were B3LYP, M05-2X, and M06-HF, with B3LYP and M06-HF resulting in the lowest average MADs with C-PCM and COSMO, respectively. For SMD, B97-D, TPSS, and M05-2X yielded the same average MAD value of 2.1 pK_a units. Therefore, unlike the consistency for density functionals that were found to perform best in describing the real systems among the solvation models, the utility of density functionals in describing the model layer depended on the selection of the solvation model.



Figure 3.4: MADs in pK_a values for fourteen GGA, M-GGA, H-GGA, HM-GGA, and DH-GGA functionals within high-level methods relative to experiment. All of the results are from calculations with ONIOM(DFT/aug-cc-pVTZ:B97-D,M06L,B3LYP/LANL2DZ) scheme. The MADs in pK_a values for the three functionals in the real layer are averaged for the molecule set.

The most accurate pK_a values were yielded by different density functionals for each solvation model (MAD of 2.0 pK_a units by B3LYP with C-PCM, 1.9 pK_a units by M06-HF with COSMO, and 2.1 pK_a units by B97-D, TPSS, and M05-2X with SMD). PBE resulted in the largest difference from experimental data with MADs of 6.7, 6.5 and 5.5 pK_a units for C-PCM, COSMO, and SMD, respectively. BB95 and M06 also performed considerably worse than other considered functionals (except PBE), which resulted in the same MADs of 6.2 pK_a units for C-PCM and 5.0 pK_a units for SMD, and similar MADs of about 5.6 pK_a units for COSMO.



Figure 3.5: MADs in pK_a values for five types of density functionals, GGA, M-GGA, H-GGA, HM-GGA, and DH-GGA functionals, within high-level methods relative to experiment. All of the results are from calculations with ONIOM(DFT/aug-cc-pVTZ:B97-D,M06L,B3LYP/LANL2DZ) scheme.

The utility of types of density functionals is shown with the three solvation models in Figure 3.5. The H-GGA functionals provided the most comparable pK_a values to the experimental data with MADs of 3.2, 2.8, and 2.5 pK_a units for C-PCM, COSMO, and SMD, respectively, while the DH-GGA functionals resulted in the highest MADs of 5.2, 4.6, and 4.0 pK_a units for C-PCM, COSMO, and SMD, respectively. The large MAD of DH-GGAs infers that the addition of a fraction of the PT2 correlation energy should not be considered for the accurate description of the model layer of TM hydrides. For all three solvation models, GGA and M-GGA functionals yielded larger MADs than H-GGA and HM-GGA, which indicates that inclusion of exact exchange is necessary

to describe the model layer of TM hydrides more appropriately. This lowering of the MADs by including exact exchange in SMD was less obvious then for C-PCM and COSMO.

Method	Model system ^a	Real system ^b
1	DFT/aug-cc-pVTZ	B97-D
2	DFT/aug-cc-pVTZ	M06-L
3	DFT/aug-cc-pVTZ	B3LYP

Table 3.3: Theoretical methods for the description of real and model systems within the two-layer ONIOM scheme using C-PCM, COSMO, and SMD for utility of DFT in the model layer.

^aDFT functionals are listed in Table 3.1. aug-cc-pVTZ (main group atoms for Pd, Pt species), aug-cc-pVTZ-DK (Ni species), aug-cc-pVTZ-PP (Pd and Pt).

^bLANL2DZ is used as the basis set for the real system.

The types of functionals were compared with respect to central TM atoms (Table 3.12) to assess if their ability to describe the model layer was determined by their performance on the description of metal center. The MADs for all types of functionals decrease from lighter to heavier metal for all three solvation models (Table 3.12). For Ni species, the M-GGA functionals yielded the lowest MADs of 3.7, 3.4, and 2.6 pK_a units with C-PCM, COSMO, and SMD, respectively. The H-GGA functionals performed the best for C-PCM and COSMO with MADs of 3.2 and 2.9 pK_a units, respectively. The GGA, M-GGA, and H-GGA functionals resulted in similar MADs of about 2.5 pK_a units with SMD for Pd species. For Pt species, the HM-GGA functionals produced comparable MADs of about 1.8 pK_a units for COSMO and SMD that were lower than for other types of functionals. The DH-GGA functional resulted in the largest MADs for all considered metal species with all three solvation models. The model layer is described better by H-GGA functionals than GGA functionals. Thus, following the same conclusion based on the overall performance of functional type, the reduction in MADs for H-GGA functionals from GGA functionals is more significant for TM hydrides with lighter central TM atoms than for those with heavier central TM atoms.

3.3.3 Impact of Exact Exchange on the Accuracy of DFT

Although there was no systematic trend found between the percentage of exact exchange and the accuracy of Minnesota functionals for the prediction of the pK_as of TM hydrides (Figure 3.4), H-GGA and HM-GGA functionals showed improvement in predicting pK_a values than GGA and M-GGA functionals when applied to the model layer. Therefore, some light might be still shed on the impact of exact exchange by investigating if the implementation of other functionals can be systematically improved as a function of the percentage of exact exchange. PBE0, which has 25% exact exchange included, did improve the accuracy of the local PBE without exact exchange. Additionally, PBE includes no empirical parameters that may affect the utility of DFT. Therefore, using the PBE0 functional to examine the impact of exact exchange on the calculation of pK_as for TM hydrides with density functionals can avoid interference from other empirical parameters. The percentage of exact exchange varied from 0 to 80% in intervals of 5%. The MADs with respect to central TM atoms and size of ligands of TM hydrides were taken into account with the ONIOM(PBE0/aug-cc-pVTZ:B97-D/LANL2DZ) scheme and SMD. B97-D was selected due to its most comparable results to the experimental data and the SMD solvation model was used since it resulted in lower MADs than either C-PCM or COSMO.



Figure 3.6: MADs of PBE0 vs. percentage of exact exchange where (a) the average MAD for each metal center and (b) the average MAD for each ligand. All of the results are from ONIOM(PBE0/aug-cc-pVTZ:B97-D/LANL2DZ) scheme with SMD.

As shown in Figure 3.6, 50% exact exchange was preferred when the ligand (depe, depp, and PNP) is constant and the central atoms changes. For Ni species, the minima all laid at 50%. The MAD curves of Pd and Pt species were significantly flatter than those for the Ni species. For the Pd species, all values between 40 and 80% yielded roughly comparable results with the greatest deviation being 0.6 pK_a units. The Pt species had the minima at 65%. For the overall MADs of the considered species, the minimum can be found at 40% exact exchange. Therefore, the amount of exact exchange needed is dependent on the choice of TM and independent of the ligands.



3.3.4 Impact of Adding Grimme's Empirical Dispersion Correction on the Accuracy of DFT

Figure 3.7: MADs in pK_a values of DFT and DFT-D3 with SMD relative to experiment, with respect to central TM atoms and ligand size of TM hydrides. The results are from calculations involving the ONIOM(DFT(-D3)/aug-cc-pVTZ:B97-D3/LANL2DZ) scheme.

The results of the impact of DFT for the model and real layers indicated that the dispersioncorrected functional, B97-D, was amongst the best functionals in describing both the real and model layers in the QM/QM scheme for TM hydrides with SMD due to having the lowest MADs with respect to experimental pK_a values. Therefore, it is of interest to evaluate the influence of adding the Grimme's empirical (D3) dispersion correction on both LL and HL methods in ONIOM(DFT/augcc-pVTZ:B97-D/LANL2DZ) schemes. The B97-D/LANL2DZ method and basis set combination for the real layer was applied in this section due to its superior performance relative to other methods with SMD. The non-dispersion corrected density functionals, BLYP and PBE from the GGAs, M06L and TPSS from the M-GGAs, PBE0 and B3LYP from the H-GGAs, and M05-2X and M06-2X from the HM-GGAs, were selected from four types of density functionals to describe the model layer of the TM hydrides. To determine the impact of adding the dispersion correction, the overall performance of the functionals with and without the dispersion correction was considered with respect to central TM atoms as well as ligand sizes in the TM hydrides (Figure 3.7). All results are averaged by functional tier in Table 3.4. The values in Table 3.4 are averaged in Figure 3.7 to clearly define the trend when using dispersion-corrected functionals. Although DFT-D3 methods resulted in lower MADs for all considered species, the improvement by adding dispersion correction varies as shown in Figure 3.8 The reductions in the MADs were more significant for the lighter central TM atoms and for TM hydrides with larger sized ligands than for heavier central TM atoms and for TM hydrides with smaller sized ligands. The comparison of different types of density functionals with and without the dispersion correction is shown in Table **??**. For the functionals with non-local exchange functionals, the addition of Grimme's dispersion correction reduced the MADs more significantly than for functionals with local exchange functionals.

Table 3.4: MADs in pK_a values of GGA, M-GGA, H-GGA, and HM-GGA Types of Functionals for Comparison of DFT and DFT-D3 Relative to Experiment with SMD.

	GGA	M-GGA	H-GGA	HM-GGA
DFT	4.1	1.9	2.4	2.8
DFT-D3	2.7	1.3	2.0	2.9


Figure 3.8: MADs of DFT vs. DFT-D3 with SMD for the functionals in the model layer, i.e. ONIOM(DFT(-D3)/aug-cc-pVTZ:B97-D3/LANL2DZ). The MADs are averages of the full molecule set.

3.3.5 Impact on the Choice of Basis Set

It is well-known that the chosen basis set will also affect the accuracy of calculated properties in addition to the selected density functional. Therefore, the influence of the basis set on the accuracy of calculated $pK_{a}s$ was assessed with two double- ζ and two triple- ζ quality correlation consistent basis sets with select density functionals as the HL method, and LANL2DZ and SDD with select density functionals as the LL method. The aug-cc-pVTZ basis set was utilized for the HL methods when comparing the MADs of LANL2DZ and SDD basis sets for LL methods and LANL2DZ was applied for low-level methods for the comparison of the considered correlation consistent basis sets for high-level methods.

The selected density functionals for the high-level method include B97-D, TPSS, B3LYP, and M05-2X since these functionals yield similar MADs of about 2.3 pK_a units and perform better than the other considered density functionals with respect to experimental pK_a values in investigating the impact of functional choice on the model layer. Only B97-D is applied for low-level methods since B97-D yields a lower MAD of 2.3 pK_a units with SMD than other considered functionals in investigating the impact of functional choice on the real layer. All calculations in this section used SMD based on the solvation model's performance in previous sections.

Table 3.5 shows the dependence of the four selected functionals upon the quality of correlation consistent basis set for the high-level methods. B97-D and B3LYP only resulted in a small reduction in MAD of 0.2 pK_a units when the basis set quality was increased from aug-cc-pVDZ to aug-cc-pVTZ while showed a reduction of MADs (more than 0.5 pK_a units) upon improving the basis set from cc-pVDZ to cc-pVTZ.

Table 3.5: MADs in pK _a values relative to experiment for four functionals when changing the basis
set used for the model layer.

	aug-cc-pVDZ ^a	aug-cc-pVTZ ^a	cc-pVDZ ^a	cc-pVTZ ^a
B97-D	3.0	2.8	4.0	2.7
TPSS	1.5	1.5	1.4	3.3
B3LYP	1.1	0.9	1.6	1.1
M05-2X	1.8	1.9	1.9	1.9

^a(aug-)cc-pVnZ-DK was considered for Ni species and (aug-)cc-pVnZ-PP was considered for Pd and Pt species.

As shown in Figure 3.9, the accuracy of the basis set displayed a dependence on the central TM atoms of the TM hydrides, where cc-pVDZ and cc-pVTZ yielded similar pK_a values for Ni and Pt species while cc-pVDZ performed better than cc-pVTZ for Pd. Similarly, aug-cc-pVDZ outperformed aug-cc-pVTZ for Pd species but yielded higher MADs than aug-cc-pVTZ for Pt species. In contrast, the accuracy of the basis sets was not affected by the ligand sizes of the TM hydrides, as both double- ζ and triple- ζ basis sets, with or without the diffuse functions, consistently resulted in similar MADs. Both considered double- and triple- ζ correlation consistent basis sets provided a more accurate description of the model layer of the TM hydrides by including diffuse

functions, except for the Ni species.



Figure 3.9: Mean absolute deviation (MAD) in pK_a values when utilizing different basis sets relative to experiment, with respect to central TM atoms and ligand size of TM hydrides where (a) the cc-pVnZ and aug-cc-pVnZ (n=D,T) are considered for the model layer (HL method) and (b) LANL2DZ and SDD ECPs are considered for the real layer (LL method).

For the low-level methods, SDD performed better than LANL2DZ with respect to central TM atoms and ligand sizes of the TM hydrides, except for Ni species (Figure 3.9). The MADs of both LANL2DZ and SDD decreased as the central TM atoms of TM hydrides becomes heavier.

3.3.6 Impact of Cavity Models on Implicit Solvation Models

The calculated pK_a values were also compared to the experimental data from the viewpoint of the cavities used in computing the C-PCM, COSMO, and SMD reaction fields with the ONIOM(PBE, M06L, B3LYP, and M06/aug-cc-pVTZ:B97-D/LANL2DZ) scheme used in previous sections to eliminate functional bias in choice of the HL method. Five cavity models, Pauling, Bondi, UA0, UAKS, and the default cavity for each solvation model within the GAUSSIAN09 package (UFF for C-PCM, Klamt for COSMO, and SMD-Coulomb for SMD) were applied to determine the effect

of the atomic radii used to build a cavity in the solvent (acetonitrile) on the predicted pK_a values of the TM hydrides. For C-PCM, the Pauling cavity generated the lowest average MAD of 3.4 pK_a units while UA0 resulted in the largest MAD of 5.3 pK_a units for the full molecule set. For both COSMO and SMD, the average MAD of the full molecule set yielded the lowest MADs of 3.0 and 2.3 pK_a units, respectively, with the GAUSSIAN09 default cavity as shown in both Figure 3.10 and Table 3.6, and the highest average MADs with the UA0 cavity with 5.3 and 5.1 pK_a units, respectively.



Figure 3.10: Impact of radii models on (a) C-PCM, (b) COSMO, and (c) SMD. The default cavities for C-PCM, COSMO, and SMD are UFF, Klamt, and SMD-Coulomb, respectively. The average MADs are results from calculation with the ONIOM (PBE, M06L, B3LYP, M06/ aug-cc-pVTZ : B97D/ LANL2DZ) scheme and then categorized by metal and ligand.

			C-PCM		
	Pauling	Bondi	UA0	UAKS	Default
Ni	3.5	3.8	5.8	3.6	5.8
Pd	3.9	4.0	6.4	4.0	4.0
Pt	2.8	2.9	3.7	3.0	2.0
depe	3.8	4.0	5.8	4.5	4.3
depp	3.0	3.2	5.5	3.0	3.5
PNP	3.4	3.6	4.8	3.2	4.0
Overall	3.4	3.6	5.3	3.6	3.9
			COSMO		
	Pauling	Bondi	UA0	UAKS	Default
Ni	3.5	3.8	5.8	3.6	3.1
Pd	3.9	4.0	6.5	4.0	3.9
Pt	2.0	2.1	3.7	2.2	1.9
depe	3.0	3.2	5.8	3.7	3.0
depp	3.0	3.2	5.5	3.0	2.9
PNP	3.4	3.6	4.8	3.2	3.0
Overall	3.1	3.3	5.3	3.3	3.0
			SMD		
	Pauling	Bondi	UA0	UAKS	Default
Ni	3.1	3.5	5.8	3.4	2.6
Pd	3.4	3.5	6.1	3.4	2.8
Pt	1.6	1.8	3.6	1.8	1.4
depe	2.7	2.9	5.6	3.4	2.4
depp	2.5	2.8	5.3	2.6	2.1
PNP	2.9	3.1	4.5	2.7	2.2
Overall	2.7	2.9	5.1	2.9	2.3

Table 3.6: MADs of five cavity models in pK_a values relative to experiment using the ONIOM(PBE, M06-L, B3LYP, and M06/aug-cc-pVTZ:B97-D/LANL2DZ) scheme.

3.3.7 Impact of the Expansion of the Size of Model System

To examine the influence of the size of the model system on the utility of density functionals to predict pK_a values of TM hydrides, four functionals, B97-D, TPSS, B3LYP, and M05-2X were used due to their better agreement with experimental data when used as the HL methods in previous sections. The ONIOM-1, ONIOM-2, and ONIOM-3 models are depicted in Figure 3.1. The ONIOM-1 model used the metal and atoms bound directly to the metal as the high level. The ONIOM-2 model increases the size of the model layer from ONIOM-1 by including the chelating

ring connecting the phosphorous atoms. The ONIOM-3 model increases the size of the model layer in ONIOM-2 by including a methyl group attached to the phosphorous atoms. As shown in Table 3.7, among the four functionals, only B97-D showed improvement when the size of model system was expanded from ONIOM-1 to ONIOM-3, while the MADs of the other three functionals increased. The largest deviation of the MADs between the four functionals are 0.5, 1.8, and 2.7 pK_a units for ONIOM-1, ONIOM-2, and ONIOM-3, respectively. The accuracy of the calculated pK_a values of the TM hydrides showed a larger dependence on the selection of density functionals when a larger sized model system was utilized.

Table 3.7: MADs in pK_a values relative to experiment of three expansions of model system of TM hydrides with SMD.

ONIOM Scheme	ONIOM-1	ONIOM-2	ONIOM-3
B97-D/aug-cc-pVTZ:B97-D/LANL2DZ	1.4	2.8	3.2
TPSS/aug-cc-pVTZ:B97-D/LANL2DZ	1.2	0.9	0.5
B3LYP/aug-cc-pVTZ:B97-D/LANL2DZ	0.9	1.2	1.5
M05-2X/aug-cc-pVTZ:B97-D/LANL2DZ	1.0	1.2	1.3

3.3.8 Comparison of Different Methodologies

Combining the results from all previous sections, the proposed methodology for these systems is B3LYP-D3/aug-cc-pVTZ:B97-D3/SDD (Scheme A). Shown in Figure 3.11, this proposed scheme is compared to four other methodological choices: B97-D3/SDD, B3LYP-D3/SDD, B3LYP/aug-cc-pVTZ:HF/LANL2DZ, and CCSD(T)/aug-cc-pVTZ:B97-D3/SDD, which are Schemes B, C, D, and E, respectively. Schemes B and C outline the use of a single density functional and Schemes D and E outline the use of ab initio methods implemented for both the LL and HL method, respectively. Table 3.8 shows the MADs for each scheme and the average MAD for each scheme presented in Figure 3.11. The performance of each methodology is compared via the average MAD for the molecule set. Scheme A had the lowest MAD of 0.6 pK_a units while Scheme B had the highest MAD of 5.5 pK_a units.



Figure 3.11: Comparison of the experimental and calculated pK_a values via methodological choices represented by their calculated values and the dotted trend lines. The dashed black line denotes the 1:1 correspondence between experiment and calculated pK_a values. Schemes A-E are ONIOM (B3LYP-D3/ aug-cc-pVTZ : B97-D3/ SDD), B97-D3/ SDD, B3LYP-D3/ SDD, ONIOM (B3LYP/ aug-cc-pVTZ : HF/ LANL2DZ), and ONIOM (CCSD(T)/ aug-cc-pVTZ : B97-D3/ SDD).

Table 3.8: Predicted pK_a values for Schemes A-E, which are ONIOM (B3LYP-D3/aug-cc-pVTZ:B97-D3/ SDD), B97-D3/SDD, B3LYP-D3/SDD, ONIOM(B3LYP/aug-cc-pVTZ:HF/LANL2DZ), and ONIOM(CCSD(T)/aug-cc-pVTZ:B97-D3/SDD), respectively.

	Scheme A	Scheme B	Scheme C	Scheme D	Scheme E	Exp
[HNi(depe) ₂] ⁺	23.6	23.3	29.9	20.5	20.6	23.8
[HNi(depp) ₂] ⁺	22.4	22.8	29.8	21.1	19.6	23.3
$[HNi(PNP)_2]^+$	23.6	24.1	20.9	22.7	20.9	22.2
$[HPd(depe)_2]^+$	23.3	26.2	26.3	21.7	18.9	23.2
$[HPd(depp)_2]^+$	24.4	26.6	27.8	22.7	21.4	22.9
$[HPd(PNP)_2]^+$	21.8	24.4	25.5	20.1	19.2	22.1
$[HPt(depe)_2]^+$	30.1	34.1	34.4	28.6	27.8	29.7
$[HPt(depp)_2]^+$	27.9	32.4	32.8	26.8	26.5	28.3 ^a
$[HPt(PNP)_2]^+$	27.8	32.1	32.6	27.0	27.3	27.6
MAD	0.6	5.5	4.4	1.4	2.3	

^aObtained through Equation 3.7.

Schemes B and C were formulated to present how the functionals chosen for Scheme A perform without the use of ONIOM. The average MAD for Scheme B is approximately 1.1 pK_a units higher than for Scheme C (4.4 pK_a units). Both Schemes B and C overestimated the pK_as, thus showing that with the SDD basis set, DFT overestimates the pK_as of these TM hydrides. The decrease in MAD while increasing the complexity of the functional from GGA to H-GGA supports the results from Section 3.3 where exact exchange is necessary for the correct chemical description of the metal center. For Scheme A, which uses a hybrid functional to describe the model layer and a local functional to describe the real system, the quality of the basis set used (aug-cc-pVTZ) at the metal center and the cancellation of inherent DFT errors due to the extrapolative ONIOM method explains why Scheme A has the closest correspondence to experiment. Schemes D and E were chosen to examine wavefunction methods for both the LL and HL method. The average MAD for Scheme D (1.4 pK_a units) is approximately 0.9 pK_a units lower than for Scheme E (2.3 pK_a units). Using wavefunction methods underestimated the pK_as for all molecules examined except for [HNi(PNP)₂]⁺ for Scheme D. In this case, using DFT was advantageous to describe the metal center and directly bound atoms over CCSD(T).

3.4 Conclusions

This study provides insight into density functionals, solvation models, basis sets, cavity models, and model layer size that are needed to examine the chemical properties of TM hydrides. Of the three solvation models considered, the SMD solvation model resulted in lower MADs for predicting pK_a values of TM hydrides than the other two models (COSMO and C-PCM) in comparison to the experimental data. For the high- and low- level methods within the QM/QM ONIOM scheme, B97-D yielded the lowest MADs with B97-D, TPSS, and M05-2X with SMD resulted in lower MADs. The improvement gained including the DFT dispersion correction was more significant for TM hydrides with lighter central TM atoms and bulkier ligands. Therefore, dispersion is recommended for these systems. Generally, the triple- ζ basis sets provided lower MADs than the double- ζ basis sets for the high-level method, while SDD yielded more comparable pK_a values to the experimental data than LANL2DZ for the low-level method. Among the considered cavity models for SMD (Pauling, Bondi, UA0, UAKS, and SMD-Coulomb), the default cavity (SMD-Coulomb), yielded the lowest MADs.

For the selection of ONIOM layers, increasing the number of atoms increases the MAD for all functionals utilized except for B97-D. Thus, the ONIOM-1 scheme (consisting of the metal atom and immediately bound atoms) is recommended. Using *ab initio* methods underestimated the pK_a while the use of a single functional largely overestimated the pK_a. Therefore, the ONIOM scheme (B3LYP-D3/aug-cc-pVTZ:B97-D3/SDD) with SMD can be considered as a computational method to obtain a reliable description of Group 10 TM hydrides, which can serve as a guide for the calibration of bulkier TM hydrides.

APPENDIX

Table 3.9: Summary of the basis sets utilized.

	Ni Species	Pd and Pt Species
Paol system	SDD	SDD
Real system	LANL2DZ	LANL2DZ
Model system	cc-pVDZ-DK	cc-pVDZ-PP
	aug-cc-pVDZ-DK	aug-cc-pVDZ-PP
	cc-pVTZ-DK	cc-pVTZ-PP
	aug-cc-pVTZ-DK	aug-cc-pVTZ-PP

Table 3.10: MADs in pK_a values of GGA, M-GGA, H-GGA, HM-GGA, and DH-GGA functionals within low-level methods with solvation models relative to experiment, with respect to central TM atoms of the TM Hydrides. All of the results are from calculations with ONIOM(B97-D, M06-L, B3LYP, and M06/ aug-cc-pVTZ:DFT/LANL2DZ) scheme.

	Central TM Atom	GGA	M-GGA	H-GGA	HM-GGA	DH-GGA
	Ni	7.2	8.0	8.2	10.7	10.6
C-PCM	Pd	8.0	8.8	7.7	9.7	10.3
	Pt	6.1	6.9	5.9	7.6	8.5
	Ni	4.5	5.5	5.5	8.1	7.8
COSMO	Pd	5.2	6.1	4.9	6.5	7.5
	Pt	3.1	3.8	2.8	4.5	5.2
	Ni	3.8	4.8	4.9	7.5	6.9
SMD	Pd	4.2	5.1	4	5.6	6.4
	Pt	2.9	3.6	2.7	4.5	4.9
Overall-MAD ^a	Ni	5.2	6.1	6.2	8.8	8.4
	Pd	5.8	6.7	5.5	7.3	8.1
	Pt	4.0	4.8	3.8	5.5	6.2

^aAverage results of C-PCM, COSMO, and SMD solvation models

Table 3.11: MADs in pK_a values of GGA, M-GGA, H-GGA, HM-GGA, and DH-GGA functionals within low-level methods with solvation models relative to experiment, with respect to ligands of the TM hydrides. All of the results are from calculations with ONIOM(B97-D,M06-L, B3LYP, and M06/ aug-cc-pVTZ:DFT/LANL2DZ) scheme.

	Ligand	GGA	M-GGA	H-GGA	HM-GGA	DH-GGA
	depe	7.2	7.7	7.4	9.2	10.0
C-PCM	depp	7.1	8.1	7.3	9.3	9.9
	PNP	7.0	7.9	7.1	9.5	9.5
	depe	4.1	4.7	4.2	6.1	6.8
COSMO	depp	4.1	5.2	4.3	6.3	6.8
	PNP	4.5	5.5	4.6	6.6	7.0
	depe	3.8	4.3	3.9	5.9	6.3
SMD	depp	3.5	4.6	3.8	5.9	6.0
	PNP	3.7	4.6	3.8	5.8	5.9
	depe	5.0	5.6	5.2	7.1	7.7
overall-MAD ^a	depp	4.9	6.0	5.1	7.2	7.6
	PNP	5.1	6.0	5.2	7.3	7.5

^aAverage results of C-PCM, COSMO, and SMD

Table 3.12: MADs in pK_a values of GGA, M-GGA, H-GGA, HM-GGA, and DH-GGA functionals within high-level methods with solvation models relative to experiment, with respect to central TM atoms of the TM Hydrides. All of the results are from calculations with ONIOM(DFT/aug-cc-pVTZ:B97-D,M06L, and B3LYP/LANL2DZ) scheme.

	Central TM Atom	GGA	M-GGA	H-GGA	HM-GGA	DH-GGA
	11017100111					
	Ni	7.0	3.7	3.9	4.8	8.9
C-PCM	Pd	3.7	3.7	3.2	4.1	6.2
	Pt	3.0	2.9	2.6	2.8	4.2
	Ni	6.8	3.4	3.5	4.9	7.5
COSMO	Pd	3.7	3.4	2.9	4.1	5.9
	Pt	2.6	2.2	2.1	1.7	3.3
	Ni	5.9	2.6	2.9	5.1	6.7
SMD	Pd	2.7	2.6	2.5	3.5	4.9
	Pt	2.2	2.1	2.2	1.9	3.2
Overall-MAD ^a	Ni	6.6	3.2	3.4	4.9	7.7
	Pd	3.4	3.2	2.9	3.9	5.7
	Pt	2.6	2.4	2.3	2.1	3.6

^aAverage results of C-PCM, COSMO, and SMD solvation models

REFERENCES

REFERENCES

- [1] Wang, W. H.; Muckerman, J. T.; Fujita, E.; Himeda, Y. Mechanistic insight through factors controlling effective hydrogenation of CO₂ catalyzed by bioinspired proton-responsive iridium(III) complexes. *ACS Catal.* **2013**, *3*, 856–860.
- [2] Stewart, M. P.; Ho, M. H.; Wiese, S.; Lindstrom, M. L.; Thogerson, C. E.; Raugei, S.; Bullock, R. M.; Helm, M. L. High catalytic rates for hydrogen production using nickel electrocatalysts with seven-membered cyclic diphosphine ligands containing one pendant amine. J. Am. Chem. Soc. 2013, 135, 6033–6046.
- [3] Liu, T.; Dubois, D. L.; Bullock, R. M. An iron complex with pendent amines as a molecular electrocatalyst for oxidation of hydrogen. *Nat. Chem.* **2013**, *5*, 228–233.
- [4] Luca, O. R.; Blakemore, J. D.; Konezny, S. J.; Praetorius, J. M.; Schmeier, T. J.; Hunsinger, G. B.; Batista, V. S.; Brudvig, G. W.; Hazari, N.; Crabtree, R. H. Organometallic ni pincer Ccomplexes for the electrocatalytic production of hydrogen. *Inorg. Chem.* 2012, 51, 8704–8709.
- [5] Espino, G.; Caballero, A.; Manzano, B. R.; Santos, L.; Pérez-Manrique, M.; Moreno, M.; Jalón, F. A. Experimental and computational evidence for the participation of nonclassical dihydrogen species in proton transfer processes on Ru-Arene complexes with uncoordinated N centers. Efficient catalytic deuterium labeling of H₂ with CD₃OD. *Organometallics* 2012, 31, 3087–3100.
- [6] Crabtree, R. H. *The Organometallic Chemistry of the Transition Metals*, 3rd ed.; Wiley: New York, NY, 1988.
- [7] Bäckvall, J. E. Transition metal hydrides as active intermediates in hydrogen transfer reactions. *J. Organomet. Chem.* **2002**, *652*, 105–111.
- [8] Hoskin, A. J.; Stephan, D. W. Early transition metal hydride complexes: Synthesis and reactivity. *Coord. Chem. Rev.* **2002**, *233-234*, 107–129.
- [9] Andrews, L. Matrix infrared spectra and density functional calculations of transition metal hydrides and dihydrogen complexes. *Chem. Soc. Rev.* **2004**, *33*, 123–132.
- [10] Hyla-Kryspin, I.; Grimme, S. Comprehensive study of the thermochemistry of firstrow transition metal compounds by Spin component scaled MP2 and MP3 methods. *Organometallics* 2004, 23, 5581–5592.
- [11] Gutsev, G. L.; Mochena, M. D.; Jena, P.; Bauschlicher, C. W.; Partridge, H. Periodic table of 3 d-metal dimers and their ions. *J. Chem. Phys.* **2004**, *121*, 6785–6797.
- [12] Yao, C.; Guan, W.; Song, P.; Su, Z. M.; Feng, J. D.; Yan, L. K.; Wu, Z. J. Electronic structures of 5d transition metal monoxides by density functional theory. *Theor. Chem. Acc.* 2007, *117*, 115–122.

- [13] Song, P.; Guan, W.; Yao, C.; Su, Z. M.; Wu, Z. J.; Feng, J. D.; Yan, L. K. Electronic structures of 4d transition metal monoxides by density functional theory. *Theor. Chem. Acc.* 2007, *117*, 407–415.
- [14] Quintal, M. M.; Karton, A.; Iron, M. A.; Daniel Boese, A.; Martin, J. M. L. Benchmark study of DFT functionals for late-transition-metal reactions. J. Phys. Chem. A 2006, 110, 709–716.
- [15] Zhao, Y.; Truhlar, D. G. Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* **2008**, *41*, 157–167.
- [16] 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. *Theor. Chem. Acc.* 2008, *120*, 215–241.
- [17] Cramer, C. J.; Truhlar, D. G. Density functional theory for transition metals and transition metal chemistry. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10757.
- [18] Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. J. Chem. Phys. 2005, 123, 062201.
- [19] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- [20] Czakó, G.; Mátyus, E.; Simmonett, A. C.; Császár, A. G.; Schaefer III, H. F.; Allen, W. D. Anchoring the absolute proton affinity scale. *J. Chem. Theory Comput.* **2008**, *4*, 1220–1229.
- [21] Rahalkar, A. P.; Mishra, B. K.; Ramanathan, V.; Gadre, S. R. "Gold standard" coupled-cluster study of acetylene pentamers and hexamers via molecular tailoring approach. *Theor. Chem. Acc.* 2011, 130, 491–500.
- [22] Liakos, D. G.; Neese, F. Improved correlation energy extrapolation schemes based on local pair natural orbital methods. *J. Phys. Chem. A* **2012**, *116*, 4801–4816.
- [23] Kınal, A.; Piecuch, P. Is the mechanism of the [2+2] cycloaddition of cyclopentyne to ethylene concerted or biradical? A completely renormalized coupled cluster study. J. Phys. Chem. A 2006, 110, 367–378.
- [24] Valeev, E. F.; Daniel Crawford, T. Simple coupled-cluster singles and doubles method with perturbative inclusion of triples and explicitly correlated geminals: The $CCSD(T)_{\overline{R12}}$ Model. *J. Chem. Phys.* **2008**, *128*, 244113.
- [25] Morris, R. H. Estimating the acidity of transition metal hydride and dihydrogen complexes by adding ligand acidity constants. *J. Am. Chem. Soc.* **2014**, *136*, 1948–1959.
- [26] Tekarli, S. M.; Drummond, M. L.; Williams, T. G.; Cundari, T. R.; Wilson, A. K. Performance of density functional theory for 3d transition metal-containing complexes: Utilization of the correlation consistent basis sets. J. Phys. Chem. A 2009, 113, 8607–8614.

- [27] Laury, M. L.; Wilson, A. K. Performance of density functional theory for second row (4d) transition metal thermochemistry. *J. Chem. Theory Comput.* **2013**, *9*, 3939–3946.
- [28] Riley, K. E.; Merz, K. M. Assessment of density functional theory methods for the computation of heats of formation and ionization potentials of systems containing third row transition metals. J. Phys. Chem. A 2007, 111, 6044–6053.
- [29] Wang, J.; Liu, L.; Wilson, A. K. Oxidative Cleavage of the β -O-4 Linkage of Lignin by Transition Metals: Catalytic Properties and the Performance of Density Functionals. *J. Phys. Chem. A* **2016**, *120*, 737–746.
- [30] 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*, 6670.
- [31] Yu, H. S.; He, X.; Truhlar, D. G. MN15-L: A New Local Exchange-Correlation Functional for Kohn-Sham Density Functional Theory with Broad Accuracy for Atoms, Molecules, and Solids. J. Chem. Theory Comput. 2016, 12, 1280–1293.
- [32] Humbel, S.; Sieber, S.; Morokuma, K. The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for n-butane conformation and S_N2 reaction: RCl⁺Cl⁻. J. Chem. Phys. **1996**, 105, 1959–1967.
- [33] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels—Alder Reactions and Pt(P(t-Bu)₃)₂ + H₂ Oxidative Addition. J. Phys. Chem. **1996**, 100, 19357–19363.
- [34] Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. J. Mol. Struct. THEOCHEM 1999, 461-462, 1–21.
- [35] Vreven, T.; Morokuma, K. On the Application of the IMOMO (Integrated Molecular Orbital + Molecular Orbital) Method. *J. Comput. Chem.* **2000**, *21*, 1419–1432.
- [36] Vreven, T.; Mennucci, B.; Da Silva, C. O.; Morokuma, K.; Tomasi, J. The ONIOM-PCM method: Combining the hybrid molecular orbital method and the polarizable continuum model for solvation. Application to the geometry and properties of a merocyanine in solution. *J. Chem. Phys.* 2001, 115, 62–72.
- [37] Mayhall, N. J.; Raghavachari, K. Molecules-in-molecules: An extrapolated fragment-based approach for accurate calculations on large molecules and materials. *J. Chem. Theory Comput.* **2011**, *7*, 1336–1343.
- [38] Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. Molecular tailoring approach for simulation of electrostatic properties. *J. Phys. Chem.* **1994**, *98*, 9165–9169.
- [39] Matsubara, T.; Sieber, S.; Morokuma, K. A test of the new "integrated MO + MM" (IMOMM) method for the conformational energy of ethane and n-butane. *Int. J. Quantum Chem.* **1996**, *60*, 1101–1109.

- [40] Decker, S. A.; Cundari, T. R. Hybrid QM/MM study of propene insertion into the Rh-H bond of HRh(PPh₃)₂(CO)(η^2 -CH₂=CHCH₃): The role of the olefin adduct in determining product selectivity. *J. Organomet. Chem.* **2001**, *635*, 132–141.
- [41] Aguado-Ullate, S.; Saureu, S.; Guasch, L.; Carbó, J. J. Theoretical studies of asymmetric hydroformylation using the Rh-(R,S)-BINAPHOS catalyst Origin of coordination preferences and stereoinduction. *Chem. A Eur. J.* **2012**, *18*, 995–1005.
- [42] Isegawa, M.; Wang, B.; Truhlar, D. G. Electrostatically embedded molecular tailoring approach and validation for peptides. *J. Chem. Theory Comput.* **2013**, *9*, 1381–1393.
- [43] Furtado, J. P.; Rahalkar, A. P.; Shanker, S.; Bandyopadhyay, P.; Gadre, S. R. Facilitating minima search for large water clusters at the MP2 level via molecular tailoring. *J. Phys. Chem. Lett.* **2012**, *3*, 2253–2258.
- [44] Toth, A. M.; Liptak, M. D.; Phillips, D. L.; Shields, G. C. Accurate relative pK_a calculations for carboxylic acids using complete basis set and Gaussian-n models combined with continuum solvation methods. *J. Chem. Phys.* **2001**, *114*, 4595.
- [45] Liptak, M. D.; Shields, G. C. Experimentation with different thermodynamic cycles used for pK_a calculations on carboxylic acids using complete basis set and Gaussian-n models combined with CPCM continuum solvation methods. *Int. J. Quantum Chem.* 2001, 85, 727–741.
- [46] Liptak, M. D.; Shields, G. C. Accurate pK_a calculations for carboxylic acids using Complete Basis Set and Gaussian-n models combined with CPCM continuum solvation methods. J. Am. Chem. Soc. 2001, 123, 7314–7319.
- [47] Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. Absolute pKa determinations for substituted phenols. *J. Am. Chem. Soc.* **2002**, *124*, 6421–6427.
- [48] Topol, I. A.; Tawa, G. J.; Caldwell, R. A.; Eissenstat, M. A.; Burt, S. K. Acidity of organic molecules in the gas phase and in aqueous solvent. J. Phys. Chem. A 2000, 104, 9619–9624.
- [49] Chipman, D. M. Computation of pK_a from dielectric continuum theory. J. Phys. Chem. A 2002, 106, 7413–7422.
- [50] Klicić, J. J.; Friesner, R. A.; Liu, S. Y.; Guida, W. C. Accurate prediction of acidity constants in aqueous solution via density functional theory and self-consistent reaction field methods. *J. Phys. Chem. A* 2002, *106*, 1327–1335.
- [51] Magill, A. M.; Cavell, K. J.; Yates, B. F. Basicity of nucleophilic carbenes in aqueous and nonaqueous solvents Theoretical predictions. *J. Am. Chem. Soc.* **2004**, *126*, 8717–8724.
- [52] Riojas, A. G.; Wilson, A. K. Solv-ccCA: Implicit solvation and the correlation consistent composite approach for the determination of pK_a. *J. Chem. Theory Comput.* **2014**, *10*, 1500–1510.

- [53] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. J. Phys. Chem. B 2009, 113, 6378–6396.
- [54] Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkins Trans.* 2 **1993**, 799–805.
- [55] Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
- [56] Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669–681.
- [57] Andzelm, J.; Kölmel, C.; Klamt, A. Incorporation of solvent effects into density functional calculations of molecular energies and geometries. *J. Chem. Phys.* **1995**, *103*, 9312–9320.
- [58] Takano, Y.; Houk, K. N. Benchmarking the conductor-like polarizable continuum model (CPCM) for aqueous solvation free energies of neutral and ionic organic molecules. J. Chem. Theory Comput. 2005, 1, 70–77.
- [59] Sadlej-Sosnowska, N. Calculation of acidic dissociation constants in water: Solvation free energy terms. Their accuracy and impact. *Theor. Chem. Acc.* **2007**, *118*, 281–293.
- [60] Lee, T. B.; McKee, M. L. Dependence of pK_a on solute cavity for diprotic and triprotic acids. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10258.
- [61] Kovács, G.; Pápai, I. Hydride donor abilities of cationic transition metal hydrides from DFT-PCM calculations. *Organometallics* **2006**, *25*, 820–825.
- [62] Qi, X.-J.; Liu, L.; Fu, Y.; Guo, Q. X. Ab Initio Calculations of pK_a Values of Transition-Metal Hydrides in Acetonitrile. *Organometallics* **2006**, *25*, 5879–5886.
- [63] Djemil, R.; Attoui-Yahia, O.; Khatmi, D. DFT-ONIOM study of the dopamine-β-CD complex: NBO and AIM analysis. *Can. J. Chem.* 2015, 93, 1115–1121.
- [64] Jiang, W.; Laury, M. L.; Powell, M.; Wilson, A. K. Comparative study of single and double hybrid density functionals for the prediction of 3d transition metal thermochemistry. J. Chem. Theory Comput. 2012, 8, 4102–4111.
- [65] Wang, N. X.; Wilson, A. K. Effects of basis set choice upon the atomization energy of the second-row compounds SO₂, CCl, and ClO₂ for B3LYP and B3PW91. *J. Phys. Chem. A* 2003, 107, 6720–6724.
- [66] Wang, N. X.; Wilson, A. K. Density functional theory and the correlation consistent basis sets: The tight d effect on HSO and HOS. *J. Phys. Chem. A* **2005**, *109*, 7187–7196.

- [67] Yockel, S.; Mintz, B.; Wilson, A. K. Accurate energetics of small molecules containing third-row atoms Ga-Kr: A comparison of advanced ab initio and density functional theory. *J. Chem. Phys.* 2004, 121, 60–77.
- [68] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian09 Revision D.01, Gaussian Inc. Wallingford CT 2009.
- [69] DeYonker, N. J.; Wilson, B. R.; Pierpont, A. W.; Cundari, T. R.; Wilson, A. K. Towards the intrinsic error of the correlation consistent Composite Approach (ccCA). *Mol. Phys.* 2009, 107, 1107–1121.
- [70] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- [71] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- [72] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [73] Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
- [74] Becke, A. D. Density-functional thermochemistry. IV. A new dynamical correlation functional and implications for exact-exchange mixing. J. Chem. Phys. 1996, 104, 1040– 1046.
- [75] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- [76] Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew-Burke-Ernzerhof exchangecorrelation functional. J. Chem. Phys. **1999**, 110, 5029–5036.
- [77] Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- [78] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

- [79] Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822–8824.
- [80] Zhao, Y.; Truhlar, D. G. Comparative DFT study of van der Waals complexes: Rare-gas dimers, alkaline-earth dimers, zinc dimer and zinc-rare-gas dimers. J. Phys. Chem. A 2006, 110, 5121–5129.
- [81] Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. J. Chem. Phys. 2006, 124, 034108.
- [82] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. 2010, 132, 154104.
- [83] Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* **2006**, *2*, 364–382.
- [84] Hay, P. J.; Wadt, W. R. Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. J. Chem. Phys. 1985, 82, 270–283.
- [85] Andrae, D.; Häußermann, U.; Dolg, M.; Stoll, H.; Preuß, H. Energy-adjusted ab initio pseudopotentials for the second and third row transition elements. *Theor. Chem. Acc.* 1990, 77, 123–141.
- [86] Dolg, M.; Wedig, U.; Stoll, H.; Preuß, H. Energy-adjusted ab initio pseudopotentials for the first row transition elements. *J. Chem. Phys.* **1987**, *86*, 866–872.
- [87] Igel-Mann, G.; Stoll, H.; Preuß, H. Pseudopotential study of monohydrides and monoxides of main group elements K through Br. *Mol. Phys.* **1988**, *65*, 1329–1336.
- [88] Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- [89] Balabanov, N. B.; Peterson, K. A. Systematically convergent basis sets for transition metals.
 I. All-electron correlation consistent basis sets for the 3d elements Sc-Zn. J. Chem. Phys. 2005, 123, 64107.
- [90] Peterson, K. A.; Figgen, D.; Dolg, M.; Stoll, H. Energy-consistent relativistic pseudopotentials and correlation consistent basis sets for the 4d elements Y-Pd. J. Chem. Phys. 2007, 126, 124101.
- [91] Figgen, D.; Peterson, K. A.; Dolg, M.; Stoll, H. Energy-consistent pseudopotentials and correlation consistent basis sets for the 5d elements Hf-Pt. J. Chem. Phys. 2009, 130, 164108.
- [92] Douglas, M.; Kroll, N. M. Quantum electrodynamical corrections to the fine structure of helium. *Ann. Phys.* (*N. Y*). **1974**, *82*, 89–155.

- [93] Laury, M. L.; DeYonker, N. J.; Jiang, W.; Wilson, A. K. A pseudopotential-based composite method: The relativistic pseudopotential correlation consistent composite approach for molecules containing 4d transition metals (Y-Cd). J. Chem. Phys. 2011, 135, 214103.
- [94] Kallies, B.; Mitzner, R. pK_a Values of Amines in Water from Quantum Mechanical Calculations Using a Polarized Dielectric Continuum Representation of the Solvent. *J. Phys. Chem. B* **1997**, *101*, 2959–2967.
- [95] McQuarrie, D. A. Statistical Mechanics, 1st ed.; University Science Books, 2000.
- [96] Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Supporting Information Single-Ion Solvation Free Energies and the Normal Hydrogen Electrode in Methanol, Acetonitrile, and Dimethylsulfoxide. J. Phys. Chem. B 2006, 111, 1–40.
- [97] Hodgson, J. L.; Roskop, L. B.; Gordon, M. S.; Lin, C. Y.; Coote, M. L. Side reactions of nitroxide-mediated polymerization: N-O versus O-C cleavage of alkoxyamines. J. Phys. Chem. A 2010, 114, 10458–10466.
- [98] Ho, J. Predicting pK_a in Implicit Solvents: Current Status and Future Directions. *Aust. J. Chem.* **2014**, *67*, 1441.
- [99] Berning, D. E.; Noll, B. C.; DuBois, D. L. Relative hydride, proton, and hydrogen atom transfer abilities of [HM(diphosphine)₂]PF₆ complexes (M = Pt, Ni). J. Am. Chem. Soc. 1999, 121, 11432–11447.
- [100] Curtis, C. J.; Miedaner, A.; Ellis, W. W.; DuBois, D. L. Measurement of the hydride donor abilities of [HM(diphosphine)₂]⁺ complexes (M = Ni, Pt) by heterolytic activation of hydrogen. J. Am. Chem. Soc. 2002, 124, 1918–1925.
- [101] Curtis, C. J.; Miedaner, A.; Raebiger, J. W.; DuBois, D. L. Periodic trends in metal hydride donor thermodynamics: Measurement and comparison of the hydride donor abilities of the series HM(PNP)²⁺ (M = Ni, Pd, Pt; PNP = Et₂PCH₂N(Me)CH₂PEt₂). Organometallics 2004, 23, 511–516.
- [102] Raebiger, J. W.; Miedaner, A.; Curtis, C. J.; Miller, S. M.; Anderson, O. P.; DuBois, D. L. Using Ligand Bite Angles to Control the Hydricity of Palladium Diphosphine Complexes. *J. Am. Chem. Soc.* 2004, *126*, 5502–5514.
- [103] Parker, V. D.; Tilset, M. Solution Homolytic Bond Dissociation Energies of Organotransition-Metal Hydrides. J. Am. Chem. Soc. 1989, 111, 6711–6717.
- [104] Parker, V. D.; Handoo, K. L.; Roness, F.; Tilset, M. Electrode Potentials and the Thermodynamics of Isodesmic Reactions. J. Am. Chem. Soc. **1991**, 113, 7493–7498.

CHAPTER 4

UTILIZATION OF THE DOMAIN-BASED LOCAL PAIR NATURAL ORBITAL METHODS WITHIN THE CORRELATION CONSISTENT COMPOSITE APPROACH

4.1 Introduction

Over the years, numerous approaches have been developed to try to reduce the computational cost associated with high-level *ab initio* methods while maintaining similar accuracy. These approaches include but are not limited to *ab initio* composite methods (see Section 2.2.4), $^{1-31}$ and local *ab initio* correlated methods (see Section 2.2.1). $^{32-65}$

The combination of these approaches should, in principle, expand the range of molecules that can be targeted with composite methodologies since composite methods are often limited by molecule size but achieve a high level of accuracy for well-established reliable experiments and local methods reduce the CPU time while reproducing electronic energies analogous to canonical molecular orbital methods. Therefore, the premise of this work is to develop a composite methodology that utilizes local methods to reduce the computational cost while retaining the same level of accuracy as the canonical composite method.

While the correlation consistent Composite Approach (ccCA) results in a reduction in computational cost and is comparable in accuracy relative to its target level of theory for main group species, ³¹ CCSD(T,FC1)/aug-cc-pCV ∞ Z-DK, additional reduction of computational cost is desired to facilitate the description of chemical systems of increasing size. For the methodology, a number of options are available, targeting one or more of the steps of the composite approach. Approaches include RI-ccCA,²⁵ which utilizes the resolution-of-the-identity (RI) approximation for the MP2 steps within ccCA, and ccCA-F12,²⁶ which uses explicitly correlated methods for all steps within ccCA. When using RI-CCSD(T) and CCSD(T)-F12 for RI-ccCA and ccCA-F12, respectively, neither coupled cluster approach reproduced the same energies as CCSD(T), and thus led to a decrease in performance of RI-ccCA and ccCA-F12 relative to ccCA. Another

example is the use of completely renormalized CCSD(T) [CR-CCSD(T)] within ccCA (CR-ccCA), which can be beneficial in situations that might otherwise require multireference wavefunction treatment, such as MR-ccCA, and which resulted in a reduction in the MAD from experiment for open-shell species which was less than that of ccCA.^{23,66}

Another route used to reduce the computational cost of ccCA is via Morokuma's Our own N-layered Integrated molecular Orbitals and molecular Mechanics (ONIOM) framework,⁶⁷ such as in ONIOM-ccCA²⁸ and rp-ccCA-ONIOM,²⁹. While these implementations have been useful, they are not immune to common multilayer method challenges including judicious choice of model layer and method combinations (e.g. QM/QM), as well as reliance on error cancellation to obtain favorable results.⁶⁸

While there have been approaches to reduce the computational cost of ccCA and other composite methods, one of the primary factors in the SCF step is the calculation of four-center two-electron integrals, which formally scales as $N^4/8$. As mentioned earlier, the RI approximation utilized within ccCA can mitigate this computational bottleneck successfully for the MP2 step of ccCA by approximating four-center two-electron integrals as a linear combination of three-center or two-center two-electron integrals through a projection operator using an auxiliary basis set (ABS). The use of an ABS reduces the scaling, and thus cost, of the four-center two-electron integrals from $O(K^4)$ to approximately $O(K^2m)$ where K is the number of basis functions and m is the number of auxiliary basis functions, where $m < K^2$ so that an ABS has enough flexibility to adapt to any Coulomb potential. In practice, different ABS are constructed and used for SCF and correlated integrals.^{25,69}

Alternatives to using RI methods for mitigating the computational bottleneck of four-center two-electron integrals include local methods, which have been extensively developed to localize dynamic correlation. $^{32-65}$ Although canonical orbitals are characteristically delocalized, a localized description of the occupied orbitals is important for dynamic correlation since dynamic correlation for nonmetallic systems is a short-range effect with a dependence on distance of r⁻⁶ like dispersion energy.⁷⁰

The domain-based local pair natural orbital (DLPNO) methods,^{55–58,61} primarily DLPNO-CCSD(T), have been shown to result in reduced computational cost relative to the cost of CCSD(T) for transition metal-based catalysts and larger organic systems such as complex hydrocarbons and fullerenes, and with comparable accuracy.^{71–75} (The original publications and Section 2.2.3 provide more details about DLPNO methods and their development.^{52–54,56–58}). To reduce the costs of DLPNO methods, the DLPNO methods have been paired with the Foster-Boys (FB)^{76,77} and the Pipek-Mezey (PM)⁷⁸ techniques to localize occupied MOs.^{78,79} Both localization schemes have been described in Section 2.2.1. Numerous studies have utilized each of these localization approaches to successfully reduce computational cost. Pipek-Mezey localization has been used to improve the accuracy and efficiency of quantum embedding:⁸⁰ Foster-Boys localization has been paired with an explicitly correlated HF approach for the quantum treatment of protons.⁸¹ Both localization approaches have been used in the development of a linear scaling implementation of the direct random-phase approximation.⁸²

To illustrate, an advantage of methods such as DLPNO is that large energetic differences that can arise from the localization tails generated from orthogonal localized MOS are effectively truncated through integral transformation, enabling the screening of less important contributions to energies.^{83–85}

Localization methods have also been utilized within composite approaches. A previous study by Montgomery et al. utilized the use of the Pipek-Mezey population localization technique based within the CBS-QB3 composite method,^{8,78} yielding a mean absolute deviation (MAD) of 1.10 kcal mol⁻¹ for the G2/97 molecule set for heats of formation, which was comparable to G3 and G3(MP2) with MADs of 0.94 kcal mol⁻¹ and 1.24 kcal mol⁻¹, respectively; thus, demonstrating the utility of localized MOs within composite methods.

As both approaches are widely used, each of the localization schemes will be considered in the incorporation of the DLPNO methods within the ccCA framework. As there have been successes utilizing the DLPNO methods to reproduce RI-MP2 and CCSD(T) correlation energies as well as implementing localization schemes in a composite framework, the goal of this work is to incorporate

the DLPNO methods within the ccCA framework to reduce computational cost with little or no impact upon the accuracy. The accuracy and relative CPU timing for the calculation of enthalpies of formation using DLPNO-ccCA is compared against ccCA and RI-ccCA to elucidate the efficacy of the DLPNO methods.

4.2 Computational Methods

A set of 119 closed shell systems from the G2/97 molecule set, including both first and second row atoms (listed in the Appendix), was used to investigate the enthalpies of formation (ΔH_f).⁸⁶ All calculations were done with the ORCA 4.0 program.^{87,88} Geometry optimizations were done at the B3LYP⁸⁹ level with the cc-pVTZ⁹⁰ basis set. All calculations that include Al-Cl (3p) were done using the recommended version of the correlation consistent basis set, the cc-pV(T+d)Z basis set.⁹¹ Energies were converged to 10⁻⁶ E_h and gradients were converged to 10⁻⁴ E_h/bohr for the geometry optimization. The vibrational ZPE and vibrational contribution to internal energy were scaled by 0.989 at 298.15 K and 1 atm to account for anharmonicity.²⁴ Thermal corrections to enthalpy were calculated at 298.15 K and 1 atm. Experimental spin-orbit corrections for atoms were applied from tables provided by Moore.⁹² The formulation of ccCA is described in section 2.2.4.1 and variants used in this chapter are shown in Table 4.1.

To supplement the experimental ΔH_f from the NIST-JANAF thermochemical tables,⁹³ experimental ΔH_f for LiH, Li₂, LiF, Na₂, and NaCl were obtained based on the work of Cioslowski et al.⁹⁴ Also, as detailed theoretical studies^{95–98} on COF₂, F₂CCF₂, and CH₂CHCl suggest that the experimental ΔH_f were likely in error,⁹⁵ the values for ΔH_f used in this work for COF₂, F₂CCF₂, and CH₂CHCl are -145.6 ± 1.0, -160.8 ± 0.8, and 5.0 ± 1.0 kcal mol⁻¹, respectively, which were obtained via *ab initio* calculations.^{96–99} The values for the atomic enthalpies of formation at 0 K for C and H used in this work were based on the work by Tasi et al.¹⁰⁰ The values for atomic enthalpies of formation for B, Si, and Al atoms were adopted from Karton et al.¹⁰¹ A UHF reference was used for O₃. ΔH_f were calculated using the total atomization approach, which uses open-shell variants for the atoms.

SCF energies were converged to 10^{-8} E_h in all single point energy calculations. The thresholds for DLPNO-MP2 were set to TCutDO = 5.0×10^{-3} , TCutPNO = 10^{-9} , and TCutMKN = 10^{-3} . For DLPNO-CCSD(T) calculations, TCutPairs = 10^{-5} E_h, TCutPNO = 10^{-7} , and TCutMKN = 10^{-4} .¹⁰² These thresholds were established as the TightPNO setting in ORCA.^{88,102} For DLPNO-CCSD(T), the Foster-Boys (FB)^{76,77} and Pipek-Mezey (PM)⁷⁸ localization schemes were used within ORCA to localize the occupied orbitals after the SCF energy was calculated.

Table 4.2 shows a summary of the approximations and the auxiliary basis sets (ABS) used in this work. The AutoAux¹⁰³ feature within ORCA was implemented to generate Li and Na ABS for correlated methods. The basis sets utilized in this work were the cc-pV*n*Z and aug-cc-pV*n*Z basis sets and the cc-pV(n + d)Z and aug-cc-pV(n + d)Z basis sets for Na-Cl, where n = D, T, Q.^{90,104–107} The ABS for coulomb-fitting (RI-J),^{108,109} coulomb-exchange fitting (RI-JK),¹⁰⁹ and correlated methods (RI-C)^{110,111} are denoted as basis/J, basis/JK, and basis/C, respectively, in this work. The implementation of the ABS was done in three schemes. Scheme 1 only utilizes the correlation consistent ABS for correlated methods. Scheme 2 utilizes the correlation consistent ABS for correlated methods and the def2/JK ABS for the SCF energy using the RI-JK approximation. Scheme 3 uses the correlation consistent ABS for correlated methods and either the RIJCOSX¹¹² or RI-JK¹¹³ approximation for the SCF energy with the appropriate def2 ABS. The RIJCOSX approximation is used for RI-MP2 and DLPNO-CCSD(T) (uses the def2/J ABS), and the RI-JK approximation is used for DLPNO-MP2 (uses the def2/JK ABS).^{57,112} The def2 ABS were chosen based on their availability throughout the periodic table. Table 4.1: Summary of the different variants of ccCA utilized in Chapter 4.

	ccCA	RI-ccCA	DLPNO-ccCA
Geometry Optimization	B3LYP/cc-pVTZ	B3LYP/cc-pVTZ	B3LYP/cc-pVTZ
	HF/aug-cc-pVTZ	HF/aug-cc-pVTZ	HF/aug-cc-pVTZ
E.	HF/aug-cc-pVQZ	HF/aug-cc-pVQZ	HF/aug-cc-pVQZ
L _{ref}	HF/aug-cc-pV∞Z	HF/aug-cc-pV∞Z	HF/aug-cc-pV∞Z
	Equation 2.29	Equation 2.29	Equation 2.29
	MP2/aug-cc-pVDZ	RI-MP2/aug-cc-pVDZ	DLPNO-MP2/aug-cc-pVDZ
	MP2/aug-cc-pVTZ	RI-MP2/aug-cc-pVTZ	DLPNO-MP2/aug-cc-pVTZ
MDY/CDS	MP2/aug-cc-pVQZ	RI-MP2/aug-cc-pVQZ	DLPNO-MP2/aug-cc-pVQZ
WIF 2/CDS	MP2/aug-cc-pV∞Z	RI-MP2/aug-cc-pV∞Z	DLPNO-MP2/aug-cc-pV∞Z
	Equations 2.30 - 2.32	Equations 2.30 - 2.32	Equations 2.30 - 2.32
	CCSD(T)/cc-pVTZ	CCSD(T)/cc-pVTZ	DLPNO-CCSD(T) ^a /cc-pVTZ
Δ	-MP2/cc-pVTZ	- RI-MP2/cc-pVTZ	- DLPNO-MP2/cc-pVTZ
$\Lambda C V$	MP2(FC1)/aug-cc-pCVTZ	RI-MP2(FC1)/aug-cc-pCVTZ	DLPNO-MP2(FC1)/aug-cc-pCVTZ
ΔCV	– MP2/aug-cc-pVTZ	– RI-MP2/aug-cc-pVTZ	– DLPNO-MP2/aug-cc-pVTZ
Δρν	MP2/cc-pVTZ-DK	RI-MP2/cc-pVTZ-DK	DLPNO-MP2/cc-pVTZ-DK
$\Delta D \mathbf{K}$	– MP2/cc-pVTZ	- RI-MP2/cc-pVTZ	 – DLPNO-MP2/cc-pVTZ
ΔSO	Experimental atomic values	Experimental atomic values	Experimental atomic values
ZPE	Vibrational ZPE scaled by 0.989	Vibrational ZPE scaled by 0.989	Vibrational ZPE scaled by 0.989

^aThe Pipek-Mezey (PM) and Foster-Boys (FB) localization schemes were considered for orbital localization for DLPNO-CCSD(T) whereas for DLPNO-MP2, only the FB localization was used.

	SCF		Post	t-HF
	RI Approximations	ABS	Methods using RI	ABS
Scheme 1	_	_	RI-MP2 DLPNO-MP2 DLPNO-CCSD(T)	aug-cc-pVnZ/C ^a cc-pVTZ/C aug-cc-pwCVTZ/C
Scheme 2	RI-JK	def2/JK	RI-MP2 DLPNO-MP2 DLPNO-CCSD(T)	aug-cc-pVnZ/C ^a cc-pVTZ/C aug-cc-pwCVTZ/C
Scheme 3	RIJCOSX ^b RI-JK ^c	def2/J def2/JK	RI-MP2 DLPNO-MP2 DLPNO-CCSD(T)	aug-cc-pVnZ/C ^a cc-pVTZ/C aug-cc-pwCVTZ/C
0			DEI NO-CCSD(1)	

Table 4.2: Summary of the approximations, methods, and auxiliary basis sets (ABS) utilized in this work for SCF and post-HF calculations.

 $a_n = D, T, Q.$

^bRIJCOSX is used for calculations involving HF + RI-MP2 and HF + DLPNO-CCSD(T) with the def2/J ABS.

 $^{\rm C}$ RI-JK is used for calculations involving HF + DLPNO-MP2 with the def2/JK ABS.

CPU timing studies were done in serial (single core) on a localDell OptiPlex 390 computer with 16 GB of DDR3 memoryto consider the efficiency of DLPNO-ccCA relative to ccCA and RI-ccCA using ORCA.^{87,88} All energies were calculated without the use of symmetry. The usage of these methods for the molecule set was considered for the first timing study. The use of ABS for the SCF step (Schemes 2 and 3) were timed for DLPNO-ccCA and RI-ccCA. Linear alkanes (C_nH_{2n+2} for n=1-8) were considered in a second CPU timing study to assess the effect of systematically increasing the molecule size on the CPU time and energetics within the ccCA framework. ΔH_f for linear alkanes are computed with the atomization approach and with isodesmic schemes for C_nH_{2n+2} for n = 3-8 since using isodesmic schemes have been shown to reduce the error for computed ΔH_f for linear alkanes.¹¹⁴ The isodesmic schemes are shown in Table 4.6. For RI-ccCA, only the Scheme 1 implementation of ABS was investigated. For DLPNO-ccCA, Schemes 1-3 were considered. The Foster-Boys (FB) localization scheme, the default scheme in ORCA, was used for all DLPNO-ccCA timing calculations.

4.3 **Results and Discussion**

4.3.1 Energetic Properties for the Molecule Set

The ΔH_f 's for each molecule of the set were calculated using ccCA, RI-ccCA, and DLPNO-ccCA. Four schemes for the extrapolation of the reference energy within ccCA, Peterson (P), Schwartz-3 (S3), Schwartz-4 (S4), and Peterson-Schwartz-3 (PS3) extrapolation schemes were considered, and denoted as ccCA-P, ccCA-S3(TQ), ccCA-S4(TQ), and ccCA-PS3(TQ), respectively.^{115–122} These approaches are compared in Table 4.3. The ccCA-PS3(TQ) shows the lowest mean absolute deviation (MAD) for ΔH_f 's of0.94kcal mol⁻¹ and the lowest magnitude for mean signed deviation (MSD) ΔH_f 's of-0.20kcal mol⁻¹ compared to all other extrapolation scheme utilized for RI-ccCA and DLPNO-ccCA in this work. The PS3(TQ) moniker is removed from the name for conciseness.

Table 4.3: Slope, intercept, and R^2 of the calculated and experimental ΔH_f . The mean signed deviation (MSD), mean absolute deviation (MAD), standard deviation (STDEV), and maximum (MAX) deviation for four variants of ccCA based on the Peterson (P), Schwartz-3 (S3), and Schwartz-4 (S4) extrapolation schemes. The P and S3 extrapolated values are averaged for PS3. Triple and quadruple- ζ level basis sets (TQ) were used for all two-point extrapolations. All deviations are in kcal mol⁻¹.

ccCA-P	ccCA-S3(TQ)	ccCA-S4(TQ)	ccCA-PS3(TQ)
1.0006	0.9999	1.0000	1.0003
0.7283	-0.3192	0.7210	0.2046
0.9999	0.9999	0.9999	0.9999
-0.71	0.32	-0.72	-0.20
1.10	1.00	1.11	0.94
1.27	1.14	1.28	1.18
4.10	2.66	4.14	3.38
	ccCA-P 1.0006 0.7283 0.9999 -0.71 1.10 1.27 4.10	ccCA-P ccCA-S3(TQ) 1.0006 0.9999 0.7283 -0.3192 0.9999 0.9999 -0.71 0.32 1.10 1.00 1.27 1.14 4.10 2.66	ccCA-PccCA-S3(TQ)ccCA-S4(TQ)1.00060.99991.00000.7283-0.31920.72100.99990.99990.9999-0.710.32-0.721.101.001.111.271.141.284.102.664.14

The effect of using the Pipek-Mezey and Foster-Boys localization techniques are demonstrated in box plots for DLPNO-MP2/aug-cc-pV ∞ Z electronic energies (Figure 4.1), and DLPNO-CCSD(T)/cc-pVTZ electronic energies (Figures 4.2 and 4.3). When using the def2/JK ABS for the SCF energies in combination with MOs generated with Foster-Boys localization, lower electronic energies were generated than a pairing with MOs generated with Pipek-Mezey localization for DLPNO-MP2/aug-cc-pVnZ (n = D, T, Q). At the complete basis set (CBS) limit, MOs generated with the Pipek-Mezey localization method yielded nearly identical electronic energies to MOs generated with Foster-Boys localization (approximately a 0.006 mE_h difference) for all of the molecule subsets shown in Figure 4.1. However, for DLPNO-CCSD(T), the effect of implementing thresholds such as electron pair screening, domain selection, and PNO generation, on using different localization schemes for the occupied MOs, changed the final DLPNO-CCSD(T) electronic energies within $\pm 2 \text{ mE}_h$ as shown in Figure 4.2, which can affect the total DLPNO-ccCA energies by $\pm 1.4 \text{ kcal mol}^{-1}$.



Figure 4.1: Differences in electronic energies (mE_h) using Pipek-Mezey (PM) and Foster-Boys (FB) localization schemes using the def2/JK ABS within DLPNO-MP2 for complete basis set extrapolation using a combined Peterson-Schwartz-3 extrapolation scheme (PS3(TQ)). Included subsets are based on the presence of certain elements (hydrocarbons, halogenated, chalcogenated, pnictogenated, and Period 3) and electronic features (aromatic, carbonyl, multiple bonds) as well as the full molecule set. Points within the dashed lines represent differences less than 0.1 mE_h. The box plots depict the distribution of data within each subset where the band in the middle represents the median of the data and data points shown as black circles are more than 3 standard deviations from the median.



Figure 4.2: Differences in electronic energies (mE_h) between the Pipek-Mezey (PM) and Foster-Boys (FB) localization methods for all three schemes within DLPNO-CCSD(T) for the same subsets in Figure 4.1. The dashed lines represent differences of less than 0.1 mE_h. The box plots depict the distribution of data within each subset where the band in the middle represents the median of the data and data points shown as black circles are more than 3 standard deviations from the median.



Figure 4.3: Differences between the CCSD(T) electronic energies and the DLPNO-CCSD(T) electronic energies in mE_h using the (a) Pipek-Mezey (PM) and (b) Foster-Boys (FB) localization methods for all three schemes for the subsets of the full molecule set shown in Figure 4.1. The dashed lines represent differences less than 0.1 mE_h. The box plots depict the distribution of data within each subset where the band in the middle represents the median of the data and data points shown as black circles are more than 3 standard deviations from the median.

Figure 4.3 depicts the difference between canonical CCSD(T) electronic energies and DLPNO-CCSD(T) electronic energies generated using localized MOs from the Pipek-Mezey (Figure 4.3a) and Foster-Boys (Figure 4.3b) schemes. As shown in Figure 4.3a, the effect of the DLPNO-CCSD(T) truncation and screening parameters on localized MOs from Pipek-Mezey localization was that DLPNO-CCSD(T) electronic energies were lower than canonical CCSD(T) electronic energies. In Figure 4.3b, using the DLPNO-CCSD(T) truncation and screening parameters on the Foster-Boys localized MOs resulted in higher electronic energies for DLPNO-CCSD(T) relative to CCSD(T) electronic energies, more noticeably for Schemes 2 and 3, since there are more positive outliers in the box plots for the subsets of the full molecule set separated by atom type and functional group. Therefore, depending on which scheme is implemented, the choice of initial localization technique, Foster-Boys or Pipek-Mezey, has a significant effect on the final DLPNO-CCSD(T) electronic energies based on the effect of implementing different thresholds defined within DLPNO-CCSD(T) on localized MOs.

Out of the 119 molecules in this molecule set, only 26 exhibited a negligible difference in the MADs (< 0.01 kcal mol⁻¹) between the two localization schemes using Scheme 1 and ABS (for correlated methods only). These molecules include those with an even charge distribution such as alkanes since the difference in electronic energy between the two localization schemes was within 1 mE_h for the hydrocarbons data subset, as shown in the boxplot for Scheme 1 in Figure 4.2. For Scheme 2 (ABS for correlated methods and RI-JK approximation for SCF), 23 molecules resulted in a negligible difference in the MADs between both localization schemes. Notable cases where deviations decreased more than 0.5 kcal mol⁻¹ when using the Pipek-Mezey localization scheme relative to the Foster-Boys localization scheme include cyclic aromatic systems, halogenated systems, and molecules characterized with triple bonds. This aligns with the known issues with Boys localization for systems with multiple bonds and aromatic ring systems. The largest differences in MAD was 3.31 kcal mol⁻¹ when using the Foster-Boys localization scheme and 1.91 kcal mol⁻¹ when using the Pipek-Mezey localization scheme and 1.91 kcal mol⁻¹ when using the Pipek-Mezey localization scheme and

approximation for SCF (Scheme 2) led to lower electronic energies when using the Pipek-Mezey localization scheme relative to the Foster-Boys localization scheme as shown in Figure 4.2.

When using the RIJCOSX approximation and def2/J ABS (Scheme 3), only 46 molecules yielded a lower MAD for ΔH_f when using the Pipek-Mezey localization scheme and only 22 showed a negligible difference in MAD (< 0.01 kcal mol⁻¹). The largest difference in MAD was 3.16 kcal mol⁻¹ when using Foster-Boys localization and 1.76 kcal mol⁻¹ when using Pipek-Mezey localization for O3 for Scheme 3. With the implementation of ABS, the number of molecules that favored the use of the Pipek-Mezey localization scheme for DLPNO-CCSD(T) decreased from 65 molecules with Scheme 1 to 50 molecules for Scheme 2 and 46 molecules for Scheme 3. This suggests that Pipek-Mezey localization may not be as useful when used with the RI-JK and RIJCOSX approximations for the SCF orbitals even though Pipek-Mezey localization has been proven useful for; (a) cyclic aromatic rings; (b) halogenated compounds; and, (c) molecules such as P₂, which are characterized by triple bonds.

Table 4.4 shows the MSD, MAD, the standard deviation of MADs, and the maximum deviation from experimental ΔH_f for all variations paired with auxiliary basis functions.Overall, DLPNO-ccCA shows good agreement with ccCA for experimental ΔH_f since for Scheme 1 (ABS for correlated methods), DLPNO-ccCA (PM), DLPNO-ccCA (FB), and RI-ccCA lowered the average MAD relative to ccCA by 0.01, 0.01, and 0.04 kcal mol⁻¹, respectively. The maximum deviation from experiment for DLPNO-ccCA(PM), DLPNO-ccCA(FB), and RI-ccCA was2.73 kcal mol⁻¹ for NCCN, 3.77 kcal mol⁻¹ for O₃, and 4.25 kcal mol⁻¹ for O₃ respectively. Using ABS for SCF energies (Scheme 2) lowered the MAD for DLPNO-ccCA (PM), DLPNO-ccCA (FB), and RI-ccCA was2.37, 3.77, and 2.53kcal mol⁻¹, respectively, forSi₂H₆, O₃, and H₂CCHCN. The use of RIJCOSX for SCF calculations for DLPNO-cCCD(T) and RI-MP2 calculations (Scheme 3) lowered the average MAD for DLPNO-ccCA (PM) and DLPNO-ccCA (FB) to 0.92 kcal mol⁻¹. The largest
deviation from experiment for DLPNO-ccCA (PM) and DLPNO-ccCA (FB) was2.69 kcal mol^{-1} for Si₂H₆ and 3.77 kcal mol^{-1} for O₃, respectively. For RI-ccCA, the largest deviation from experiment was 4.52 kcal mol^{-1} for SiCl₄. Therefore, the recommended implementation of ABS for DLPNO-ccCA is the use of RIJCOSX for SCF calculations within DLPNO-CCSD(T) and RI-JK for SCF calculations within DLPNO-MP2 calculations (Scheme 3) in conjunction with Pipek-Mezey localization.

	DLPNO-ccCA (PM)		DI	DLPNO-ccCA (FB)			RI-ccCA		
	Scheme	Scheme	Scheme	Scheme	Scheme	Scheme	Scheme	Scheme	Scheme
	1	2	3	1	2	3	1	2	3
MSD	0.21	0.40	0.52	0.38	0.09	0.21	-0.35	0.09	0.52
MAD	0.95	0.91	0.92	0.98	0.98	0.94	0.92	0.75	1.09
STDEV	1.13	1.03	0.98	1.14	1.21	1.15	1.16	0.93	1.23
MAX	2.73	2.37	2.69	3.76	3.77	3.77	4.25	2.53	4.52

Table 4.4: Mean signed deviation (MSD), mean absolute deviation (MAD), standard deviation (STDEV), and maximum (MAX) deviation for all schemes. All deviations are in kcal mol^{-1} .

4.3.2 CPU Timing

To give insight about the performance of the ccCA variants in terms of time, the total CPU time was measured as the sum of the CPU times of the single point calculations that are included in the Scheme 1 methodology (Table 4.2). Fifteen molecules from the molecule set were selected based on their varying size to demonstrate the performance and potential bottlenecks of the ccCA variants. This subset included CH₃Cl, NF₃, PF₃, H₃CH₂COCH₃, cyclic and linear alkenes, and alkanes (bicyclo[1.1.0]butane, cyclobutane, cyclobutene, isobutene, trans-butane, isobutane, spiropentane), and cyclic aromatic molecules (furan, thiophene, benzene, pyridine). The CPU times were taken as percentages of the total CPU time and show that the bottleneck step is the MP2/aug-cc-pVQZ step. For DLPNO-ccCA, the DLPNO-MP2/aug-cc-pVQZ uses 72.7% of the total CPU time, MP2(FC1)/aug-cc-pCVTZ uses 10.1% of the total CPU time, DLPNO-CCSD(T)/cc-pVTZ uses 5.8% of the total CPU time, and the other steps require 11.4% of the total CPU time as shown in Figure 4.4. This is consistent with RI-ccCA assessed in the same fashion. For Schemes 2 and 3, the use of ABS for the SCF energy, while decreasing the total CPU time, does not change the ratio of CPU time savings significantly.

The CPU timings for the full 119 molecule set are shown in Table 4.5, which displays the mean, largest, and smallest percent CPU time savings for RI-ccCA and DLPNO-ccCA relative to ccCA, and Figure 4.5, which depicts the CPU time savings for DLPNO-ccCA compared to ccCA (Figure 4.5a) and RI-ccCA (Figure 4.5b). In Scheme 1 (ABS for correlated methods), the percent difference in CPU time between DLPNO-ccCA and ccCA averaged 29.1% but 32.8% for RI-ccCA and ccCA, indicating that RI-ccCA is slightly more efficient, overall, than DLPNO-ccCA when using Scheme 1, as shown in Table 4.5. The use of ABS for the SCF energy and correlated methods (Scheme 2) drastically increased the percent CPU time savings to approximately 87.5% and 92.5% for DLPNO-ccCA and RI-ccCA, respectively, relative to use of ABS for correlated methods only (Scheme 1). The changes in percent CPU savings is due to applying the RI approximation to both the SCF and correlation energy calculation energy.



Figure 4.4: CPU time of each individual step within (a) ccCA, (b) RI-ccCA, and (c) DLPNO-ccCA for selected species from the molecule set. The Other category represents the timing of the MP2/aug-cc-pVDZ, MP2/aug-cc-pVTZ, MP2/cc-pVTZ, and MP2/cc-pVTZ-DK calculations as these calculations use a small percentage of the total CPU time. All timing calculations were done with the ORCA software package.

Table 4.5: Percent CPU time savings for the three schemes of ABS implementation within DLPNOccCA and RI-ccCA relative to ccCA. The mean percent difference from ccCA, the most efficient (MAX), and the least efficient (MIN) percent CPU time savings relative to ccCA timings are shown. All timing studies were done with ORCA.

DLPNO-ccCA (FB) (%)			RI-ccCA (%)			
	Scheme	Scheme	Scheme	Scheme	Scheme	Scheme
	1	2	3	1	2	3
MEAN	29.1	87.3	86.5	32.6	92.2	83.6
MAX	40.2	95.3	94.8	38.0	96.0	95.9
MIN	-4.0	27.9	21.0	10.9	34.7	34.7

RI-ccCA, which shows 32.8% CPU time savings relative to ccCA, is slightly more efficient than DLPNO-ccCA, which shows 29.1% CPU time savings relative to ccCA, when using Scheme 1. However, with increasing molecule size and depending on the RI approximation that was used, DLPNO-ccCA is more efficient than RI-ccCA. This is shown especially for Scheme 3, where the use of RIJCOSX for SCF within RI-MP2 and DLPNO-CCSD(T) increased the CPU time for RI-ccCA relative to DLPNO-ccCA. Based on Figure 4.5a and Table 4.5, the RIJCOSX approximation slightly increases the CPU time of DLPNO-ccCA relative to using the RI-JK approximation for the SCF step within DLPNO-CCSD(T), but decreased the MAD when using DLPNO-ccCA to calculate the Δ H_f. The increase in CPU time of RIJCOSX relative to RI-JK is due to the size of the molecules since a threshold exists between the efficiency of RIJCOSX versus RI-JK for molecular size. Therefore, using RI-JK for SCF within DLPNO-MP2 and RIJCOSX for SCF within DLPNO-CCSD(T) for DLPNO-ccCA is recommended for smaller systems.



Figure 4.5: CPU time ratios of DLPNO-ccCA (FB) to (a) ccCA and (b) RI-ccCA. The ratios for Scheme 1 (blue circle), Scheme 2 (black x), and Scheme 3 (green triangle) are shown on a log-log scale. All timing was done with C_1 symmetry enforced and done in ORCA.

4.3.3 Enthalpies and Timing for Linear Alkanes

Deviations from experimental ΔH_f are shown in Table 4.6. Only the Foster-Boys localization scheme was used for linear alkanes since using Pipek-Mezey localization did not significantly change the final DLPNO-CCSD(T) energy. The trend of increasing deviation with increasing number of carbon atoms is consistent with previous ccCA studies and common with many other methods when using the atomization approach.^{5,19,25} DLPNO-ccCA yields a smaller deviation when ccCA overestimates the ΔH_f , as shown for C_nH_{2n+2} for $n \ge 4$ in Table 4.6, and a larger deviation when ccCA underestimates the ΔH_f , as shown for C_nH_{2n+2} for $n \le 3$ in Table 4.6. This is due to the contribution of DLPNO-CCSD(T) as both ccCA and RI-ccCA, which use CCSD(T), follows the same trend for the rate of increase in deviation from experiment when using the atomization approach for ΔH_f . Also, the chosen thresholds for the PNOs allow for noncovalent interactions, which are present in (C_nH_{2n+2} for $n \ge 3$, to be better characterized. The percentage of electron pairs that are screened out of the calculation in the pre-screening process is a potential source of error in the prediction of ΔH_f for smaller molecules.

	Exp ccCA RI-cc		RI-ccCA	DLPNO-ccCA (FB)			
			Scheme	Scheme	Scheme	Scheme	
			1	1	2	3	
CH ₄	$-17.90{\pm}0.10$	-18.26	-18.28	-18.47	-18.45	-18.62	
C_2H_6	-20.03 ± 0.07	-20.72	-20.76	-21.13	-21.08	-20.67	
Calla	25.02 ± 0.12	-25.66	-25.72	-26.21	-26.32	-26.05	
C3118	-23.02±0.12	(-24.64)	(-24.64)	(-24.58)	(-24.76)	(-25.50)	
		20.02	20.00	20.57	20.46	20.22	
$C_4 H_{10}^{b}$	-30.31 ± 0.14	-28.93	-29.00	-29.57	-29.46	-29.22	
04110		(-30.40)	(-30.40)	(-30.45)	(-30.59)	(-30.41)	
		-33 59	-33 68	-34 32	-34 21	-34 30	
$C_5H_{12}^c$	-35.11 ± 0.19	(-34.64)	(-34.64)	(-34.54)	(-34.56)	(-35.48)	
					(,	()	
сиd	20.90 ± 0.10	-38.30	-38.41	-39.15	-39.01	-38.94	
С6Н14	-39.89±0.19	(-39.40)	(-39.40)	(-39.27)	(-39.12)	(-39.73)	
$C_7H_1c^{e}$	-4478 ± 018	-43.05	-43.16	-43.98	-43.81	-43.64	
C/III0	44.70±0.10	(-44.30)	(-44.29)	(-44.19)	(-44.19)	(-44.67)	
		47.92	47.00	49.70	19 60	10 22	
$C_8H_{18}f$	-49.90 ± 0.31	-4/.82	-47.90	-48.79	-48.60	-48.33	
0 10		(-49.22)	(-49.16)	(-49.06)	(-49.07)	(-49.55)	

Table 4.6: Deviations in kcal mol⁻¹ from experimental ΔH_f for linear alkanes (C_nH_{2n+2} 1 \leq n \leq 8) using the atomization approach and using isodesmic approaches (shown in parentheses).

^aIsodesmic Reaction: $2 C_2 H_6 \rightarrow C_3 H_8 + C H_4$

^bIsodesmic Reaction: $C_5H_{12} + C_2H_6 \rightarrow C_4H_{10} + C_3H_8$

 c Isodesmic Reaction: $C_{4}H_{10}$ + $C_{2}H_{6}\ \rightarrow C_{5}H_{12}$ + CH_{4}

 $d_{Isodesmic \ Reaction: \ C_4H_{10} + C_3H_8 \ \rightarrow C_6H_{14} + CH_4}$

^eIsodesmic Reaction: $C_6H_{14} + C_2H_6 \rightarrow C_7H_{16} + CH_4$

^fIsodesmic Reaction: $C_7H_{16} + C_2H_6 \rightarrow C_8H_{18} + CH_4$

The timing results are shown in Table 4.7. Even when using Scheme 1, the time savings associated with increasing the number of carbon atoms in the linear chain is evident for DLPNO-ccCA in comparison to ccCA and RI-ccCA, largely due to DLPNO-CCSD(T). The CPU time savings for RI-ccCA and DLPNO-ccCA for methane, 36.7% and 30.9%, respectively, and for ethane, 37.8% and 36.1%, respectively, show that RI-ccCA is more efficient than DLPNO-ccCA for smaller molecules; however, starting with propane (n = 3), DLPNO-ccCA is more efficient than RI-ccCA with CPU time savings of 39.6% and 38.2%, respectively, and the percent CPU time

saving monotonically increases with increasing carbon atoms for DLPNO-ccCA.

	RI-ccCA		DLPNO-ccCA (FB)	
	Scheme 1	Scheme 1	Scheme 2	Scheme 3
CH4	36.7	30.9	88.3	87.3
C_2H_6	37.8	36.1	93.1	92.5
C ₃ H ₈	38.2	39.6	94.8	94.4
C_4H_{10}	37.1	41.6	95.1	94.8
$C_{5}H_{12}$	35.7	45.8	95.4	95.1
$C_{6}H_{14}$	33.7	50.5	95.7	95.5
C7H16	30.1	56.6	96.2	96.0
C_8H_{18}	35.3	68.7	97.2	97.0

Table 4.7: Percent CPU time savings for RI-ccCA and DLPNO-ccCA (FB) relative to ccCA for linear alkanes (C_nH_{2n+2} 1 \leq n \leq 8). All timing studies were done with ORCA.

For DLPNO-ccCA, Scheme 2 and 3 protocols were implemented to examine further effects of cost savings for larger systems relative to those in the molecule set. When using ABS, the deviations in ΔH_f (Table 4.6) varies depending on which RI approximation was used for the SCF portion of the DLPNO-CCSD(T) calculation. When using RI-JK (Scheme 2), the trend in deviation from experimental ΔH_f remained the same from using ABS for correlated methods only (Scheme 1) but with a slightly higher predicted ΔH_f . Apart from methane, the use of RIJCOSX (Scheme 3) caused the prediction of enthalpy of formation to be lower in magnitude, which caused the deviation for linear alkanes to lie between DLPNO-ccCA when using ABS for correlated methods (Scheme 1) and ccCA results. Isodesmic approaches are used for larger linear alkanes as these have been shown to reduce the error without increasing the cost.¹¹² The isodesmic schemes are shown in Table 4.6. When using the isodesmic approaches, the deviations associated with ΔH_f are reduced by 0.3 to 1.3 kcal mol⁻¹ relative to using the atomization approach for ΔH_f . Regardless of which approach was used, i.e. atomization approach or isodesmic schemes, the calculated ΔH_f generated with all schemes of DLPNO-ccCA yields deviations in agreement with calculated ΔH_f generated with ccCA and RI-ccCA.

As shown in Table 4.7, the percent CPU time savings drastically increases when using ABS for SCF calculations. For methane, the percent CPU time savings increased from 30.9% using ABS

only for correlated methods (Scheme 1) to 88.3% using RI-JK for SCF and ABS for correlated methods (Scheme 2) and 87.3% using RIJCOSX for SCF and ABS for correlated methods (Scheme 3). For octane, the percent CPU time savings increased from 68.7% using Scheme 1 to 97.2% and 97.0% using Scheme 2 and 3, respectively. As with Scheme 1 ABS implementation for DLPNO-ccCA, the percent CPU time savings monotonically increase as the number of carbon atoms increase. Employing RIJCOSX with DLPNO-CCSD(T) caused a slight decrease in percent CPU time savings for all linear alkanes assessed. However, this difference decreases with increasing number of carbons, inferring that RIJCOSX is beneficial for larger molecules.

4.3.4 Applications of DLPNO-ccCA

The proposed DLPNO-ccCA (Pipek-Mezey localization and using RIJCOSX with DLPNO-CCSD(T)) has been applied to the S66 and the L7 (coronene dimer) data sets^{124–127} These datasets target long-range weakly bound systems and are calibrated to CCSD(T)/CBS interaction energies. For S66, examples were picked from the three subcategories of the dataset for presentation, hydrogen-bound molecules (water dimer), dispersion-dominated interactions (stacked uracil dimer), and a combination of both (CH₃NH₂-Peptide, T-shaped benzene dimer). The L7 dataset targets larger noncovalent complexes predominantly exhibiting dispersion interactions. All calculated interaction energies are counterpoise-corrected. Comparing the effectiveness of DLPNO-ccCA interaction energies relative to CCSD(T)/CBS interaction energies provides a computational cost-effective way to have*ab initio* data present for these larger molecular systems and serves as a potential gauge for DFT and other scaling/cost-reduction methods. DLPNO-ccCA calculated interaction energies were compared against the interaction energies generated with CCSD(T)/CBS and an average of MP2/CBS, MP2C/CBS, MP2.5/CBS, SCS-MP2/CBS, SCS(MI)-MP2/CBS given the optimized structures from the original publication of the S66 molecule set.¹²⁶

For all cases except for the uracil stacked dimer, interaction energies calculated with DLPNOccCA yielded smaller deviations from the CCSD(T)/CBS interaction energies than mean MP2/CBS interaction energies. For the T-shaped benzene dimer, the MAD from the reference interaction energy was 0.03 kcal mol⁻¹ and 0.17 kcal mol⁻¹ using DLPNO-ccCA and MP2/CBS, respectively, whereas the MAD from the reference interaction energy for the uracil stacked dimer was 0.34 kcal mol⁻¹ and 0.27 kcal mol⁻¹ using DLPNO-ccCA and MP2/CBS, respectively.

For the coronene dimer, the interaction energy calculated with DLPNO-ccCA deviates from the QCISD/CBS and DFT-D3/def2-QZVP reference interaction energies by 3.62 kcal mol⁻¹ and 6.44 kcal mol⁻¹, respectively. The DFT-D3/def2-QZVP presented in Table4.8is an average of calculated interaction energies with the B3LYP-D3, BLYP-D3, TPSS-D3, PW6D95-D3, and M06-2X-D3 functionals in tandem with the def2/QZVP basis set.¹²⁷ This shows that DLPNO-ccCA has some issues when dealing with complexes that exhibit primarily dispersion-dominated interactions between molecules due to the truncation parameters for screening orbital pairs and triples. This holds since DLPNO-ccCA yields lower MADs than MP2/CBS for CCSD(T)/CBS interaction energies for complexes that exhibited both long-distance hydrogen bonding and dispersion interactions.

Table 4.8: Interactions energies of select examples from the S66 and L7 molecule sets. All interaction energies are in kcal mol^{-1} .

<u>S66</u>	DLPNO-ccCA	CCSD(T)/0	$CBS (\Delta a(TQ)Z)$	MP2/CBS ^a
$(H_2O)_2$	-4.88		-4.86	-4.85 ± 0.17
Uracil dimer (S)	-9.48		-9.82	$-9.56 {\pm} 0.92$
CH ₃ NH ₂ -Peptide	-5.31	-5.42		-5.15 ± 0.38
Benzene dimer (T)	-2.69		-2.58	-3.05 ± 0.38
<u>L7</u>	DLPNO-ccCA	QCISD(T)/CBS	DFT-D3/def2-QZVP ^b	MP2/CBS ^a
Coronene dimer	-27.98	-24.36	$-21.54{\pm}1.28$	-26.32 ± 6.56

^aThe MP2/CBS value presented is the average of the counterpoise-corrected MP2.5/CBS, MP2/CBS, MP2/CBS, SCS(MI)-MP2/CBS, SCS-MP2/CBS interaction energies.^{125,126} FL

^bThe DFT-D3 value presented is an average of the B3LYP-D3, BLYP-D3, TPSS-D3, PW6D95-D3, and M06-2X-D3 functionals used in combination with the def2-QZVP basis set with no counterpoise correction included.¹²⁷ FL

Analyzing the individual components of DLPNO-ccCA for each of the dimers investigated from the S66 and L7 data sets yielded insight into necessary electronic contributions towards calculating interaction energies. As shown in Table 4.9, for the molecules from the S66 data set, the primary contribution towards interaction energies is the inclusion of core-core correlation effects as an additive correction to the DLPNO-MP2/CBS reference energy. While the core-valence correlation effects affected the total interaction energy by less than 0.1 kcal mol⁻¹ for the dimers in S66, core-valence correlation effects increased the total interaction energy by 4.05 kcal mol⁻¹ for the coronene dimer. This shows that core-valence interactions are necessary when considering larger dispersion-dominated molecules.

Table 4.9: Component breakdown of the DLPNO-ccCA calculated interaction energies from the S66 and L7 datasets with counterpoise corrections included. All interaction energies are in kcal mol^{-1} .

	DLPNO-MP2/CBS	ΔCC	ΔCV	ΔDK	DLPNO-ccCA	
<u>S66</u>						
(H ₂ O) ₂	-5.03	0.21	-0.07	0.01	-4.88	
Uracil dimer (S)	-11.03	1.60	-0.06	0.002	-9.49	
CH ₃ NH ₂ -Peptide	-5.53	0.26	-0.04	0.003	-5.31	
Benzene dimer (T)	-3.71	0.88	-0.02	-0.003	-2.85	
<u>L7</u>						
Coronene Dimer	-48.29	16.27	4.05	-0.01	-27.98	
						_

Since DLPNO-ccCA was applied to the coronene dimer, the ΔH_f of coronene can be calculated as well. With DLPNO-ccCA, the computed ΔH_f was 69.3 kcal mol⁻¹ with the atomization approach and the experimental ΔH_f is 70.5±2.7 kcal mol⁻¹.¹²⁸ This shows good agreement with calculating the ΔH_f for larger main group organic species with *ab initio* composite strategies.

4.4 Conclusions

A new formulation of ccCA, DLPNO-ccCA, incorporating the DLPNO methods has been developed and used to determine the ΔH_f for 119 molecules of the first and second row main group from the G2/97 molecule set, a set of 8 linear alkanes, the S66 dataset, and the coronene dimer. The Foster-Boys and Pipek-Mezey localization schemes followed by integral screening have been used to aid in reducing the computational cost of the DLPNO-ccCA approach. It was found that by choice of localization method in one step of the DLPNO-ccCA approach, the DLPNO-CCSD(T) step, can

result in an impact in the MAD from experiment in the enthalpy of formation by as much as 1.4 kcal mol⁻¹, whereas for the MAD in the DLPNO-MP2/aug-cc-pV ∞ Z step, the MAD is impacted by only 0.004 kcal mol⁻¹. For smaller molecules, using localized MOs generated by the Pipek-Mezey localization yielded a lower MAD overall compared to using localized MOs generated by the Foster-Boys localization when ABS are implemented for SCF calculations. Overall, Pipek-Mezey localization of occupied MOs yielded lower MADs for Δ H_f for cyclic aromatic rings, halogenated compounds, and molecules characterized with a triple bond whereas there was no significant numerical difference in the MADs for Δ H_f when using localization techniques for alkanes.

The use of RIJCOSX for the SCF step within DLPNO-CCSD(T) and RI-JK for the SCF step within DLPNO-MP2 is recommended for DLPNO-ccCA paired with Pipek-Mezey localization. DLPNO-ccCA reduces the computational cost compared to ccCA. RI-ccCA tends to save more CPU time (92.5%) for smaller molecules than DLPNO-ccCA when using RI-JK for SCF energies (87.5%). However, DLPNO-ccCA tends to result in more CPU time savings (86.7%) than RI-ccCA (84.3%) with the use of RIJCOSX for RI-MP2 and with increasing molecule size within the molecule set. When using DLPNO-ccCA for methane, ethane and propane, the deviation from experimental ΔH_f increased relative to ccCA, but decreased relative to ccCA for molecules with n ≥ 4 where n is the number of carbon atoms. The percent cost savings in CPU time from utilizing the DLPNO methods for linear alkanes range from approximately 88% to 97% with increasing number of carbon atoms when using ABS for both SCF and correlated methods.

In summary, DLPNO-ccCA reduced the computational cost associated with ccCA by approximately 87% while maintaining an overall MAD of no more than 1 kcal mol⁻¹ from reliable experiment and *ab initio* calculations for ΔH_f of main group complexes. More so than RI-ccCA, DLPNO-ccCA significantly reduces the computational cost of ccCA for the larger molecules in the molecule set, and thus allows access to investigate thermodynamic properties for larger molecules with the same level of accuracy of ccCA.

APPENDIX

APPENDIX

C_4H_6 (cyclobutene)	LiH	Si ₂ H ₆	H ₂ CCO
C_4H_8 (cyclobutane)	$CH_2({}^{1}A_1)$	CH ₃ Cl	C ₄ H ₄ NH
C_4H_8 (isobutene)	CH ₄	H ₃ CSH	H ₃ CCOH
C_4H_{10} (trans-butane)	NH ₃	HOCl	$F_2CCF_2(D_{2h})$
C ₄ H ₁₀ (isobutane)	H_2O	SO_2	H ₃ CCH ₂ OH
C ₅ H ₈ (spiropentane)	HF	BF ₃	H ₃ COCH ₃
C ₆ H ₆ (benzene)	SiH ₂	BCl ₃	(CH3)3N
C ₄ H ₄ S (thiophene)	SiH4	AlF ₃	$(CH_3)_2SO$
H ₂ CSCH ₂ (thiooxirane)	PH ₃	AlCl ₃	H ₃ CCH ₂ SH
$(CH_2)_2O(\text{oxirane})$	SH ₂	CF ₄	H ₃ CSCH ₃
OCHCHO (glyoxal)	HCl	CCl ₄	H ₂ CCHF
NCCN (cyanogen)	Li ₂	OCS $(^{1}\Sigma^{+})$	H ₃ CCH ₂ Cl
H ₂ CCHCl	LiF	CS_2	$(CH_3)_2NH$
$H_2C = CHCN$	C_2H_2	COF_2	H ₃ CCH ₂ NH ₂
H ₃ CONO	C_2H_4	SiF ₄	H ₃ CCOCH ₃
H ₃ CSiH ₃	C_2H_6	SiCl ₄	CH ₃ COOH
CH ₃ CH ₂ CH ₂ Cl	HCN	NNO	CH ₃ CFO
HCOOCH ₃	CO	CINO	CH ₃ C(O)Cl
H_3CCONH_2	H_2CO	NF ₃	HCOOH
H_2CCH_2NH	CH ₃ OH	PF ₃	(CH ₃) ₂ CHOH
C_4H_6 (2-butyne)	N_2	O3	
C_4H_6 (methylene cyclopropane)	H_2NNH_2	F_2O	
C_4H_6 (bicyclo[1.1.0] butane)	O ₂	ClF ₃	
C ₃ H ₄ (propyne)	HOOH	H ₂	
C ₃ H ₄ (allene)	F_2	Cl_2CCCl_2	
C ₃ H ₄ (cyclopropene)	$\operatorname{CO}_2(\mathbf{D}_h)$	CF ₃ CN	
C ₃ H ₆ (propene)	Na ₂	CH_2F_2	
C ₃ H ₆ (cyclopropane)	P ₂	CHF ₃	
C ₃ H ₈ (propane)	Cl ₂	CH_2Cl_2	
C_4H_6 (trans – 1,3 – butadiene)	NaCl	CHCl ₃	
C5H5N (pyridine)	SiO	H ₃ CNH ₂	
C_4H_4O (furan)	CS	H ₃ CCN	
H ₃ CH ₂ COCH ₃	ClF	H ₃ CNO ₂	

Table 4.10: Molecule list used for full set calculations.

Table 4.11: MP2/CBS counterpoise-corrected interaction energies calculated for molecules in the S66 data set used to compare DLPNO-ccCA interaction energies from Reference 126. All interaction energies are in kcal mol⁻¹.

	MP2	MP2.5	MP2C	SCS-MP2	SCS(MI)- MP2
(H ₂ O) ₂	-4.96	-4.93	-4.97	-4.51	-4.91
Uracil dimer (S)	-11.14	-9.47	-9.37	-8.25	-9.56
CH ₃ NH ₂ -Peptide	-5.53	-5.32	-5.41	-4.45	-5.04
Benzene dimer	-3.75	-3.05	-2.96	-2.58	-2.90

Table 4.12: MP2/CBS counterpoise-corrected interaction energies calculated for the coronene dimer used to compare DLPNO-ccCA interaction energies from Reference 127. All interaction energies are in kcal mol^{-1} .

	MP2	MP2.5	MP2C	SCS-MP2	SCS(MI)- MP2
coronene dimer	-38.98	-22.80	-20.88	-27.53	-31.71

Table 4.13: DFT-D3/def2-QZVPP non-counterpoise-corrected interaction energies calculated for the coronene dimer used to compare DLPNO-ccCA interaction energies from Reference 127. All interaction energies are in kcal mol⁻¹.

	B3LYP-D3	BLYP-D3	TPSS-D3	PW6D95-D3	M06-2X-D3
coronene dimer	-23.22	-22.82	-21.19	-19.93	-20.55



Figure 4.6: CPU time for ccCA, RI-ccCA, and all 3 schemes for DLPNO-ccCA for the linear alkanes.



Figure 4.7: Deviations in ΔH_f for ccCA, RI-ccCA, and all 3 schemes for DLPNO-ccCA for the linear alkanes using the isodesmic approach.

REFERENCES

REFERENCES

- [1] Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 theory for molecular energies of first- and second-row compounds. *J. Chem. Phys.* **1991**, *94*, 7221–7230.
- [2] Curtiss, L. A.; Carpenter, J. E.; Raghavachari, K.; Pople, J. A. Validity of additivity approximations used in GAUSSIAN-2 theory. *J. Chem. Phys.* **1992**, *96*, 9030–9034.
- [3] Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-3 theory using density functional geometries and zero-point energies. *J. Chem. Phys.* **1999**, *110*, 7650–7657.
- [4] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 theory. J. Chem. Phys. 2007, 126, 084108.
- [5] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 theory using reduced order perturbation theory. *J. Chem. Phys.* **2007**, *127*, 124105.
- [6] Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vázquez, J.; Stanton, J. F. HEAT: High accuracy extrapolated ab initio thermochemistry. J. Chem. Phys. 2004, 121, 11599–11613.
- [7] Ochterski, J. W.; Petersson, G. A.; Montgomery Jr., J. A. A complete basis set model chemistry. V. Extensions to six or more heavy atoms. *J. Chem. Phys.* **1996**, *104*, 2598–2619.
- [8] Montgomery Jr., J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J. Chem. Phys.* 1999, 110, 2822–2827.
- [9] Montgomery Jr., J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VII. Use of the minimum population localization method. *J. Chem. Phys.* 2000, 112, 6532–6542.
- [10] Feller, D.; Dixon, D. A. Coupled Cluster Theory and Multireference Configuration Interaction Study of FO, F₂O, FO₂, and FOOF. J. Phys. Chem. A **2003**, 107, 9641–9651.
- [11] Feller, D.; Peterson, K. A.; De Jong, W. A.; Dixon, D. A. Performance of coupled cluster theory in thermochemical calculations of small halogenated compounds. *J. Chem. Phys.* 2003, *118*, 3510–3522.
- [12] Feller, D.; Dixon, D. A.; Francisco, J. S. Coupled Cluster Theory Determination of the Heats of Formation of Combustion-Related Compounds: CO, HCO, CO₂, HCO₂, HOCO, HC(O)OH, and HC(O)OOH. J. Phys. Chem. A 2003, 107, 1604–1617.
- [13] Feller, D.; Peterson, K. A.; Dixon, D. A. A survey of factors contributing to accurate theoretical predictions of atomization energies and molecular structures. *J. Chem. Phys.* 2008, 129, 204105.

- [14] Dixon, D. A.; Feller, D.; Peterson, K. A. A Practical Guide to Reliable First Principles Computational Thermochemistry Predictions Across the Periodic Table. *Annu. Rep. Comput. Chem.* 2012, 8, 1–28.
- [15] Martin, J. M. L.; De Oliveira, G. Towards standard methods for benchmark quality ab initio thermochemistry W1 and W2 theory. *J. Chem. Phys.* **1999**, *111*, 1843–1856.
- [16] Daniel Boese, A.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kállay, M.; Gauss, J. W3 theory: Robust computational thermochemistry in the kJ/mol accuracy range. J. Chem. Phys. 2004, 120, 4129–4141.
- [17] Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions. J. Chem. Phys. 2006, 125, 144108.
- [18] Fast, P. L.; Schultz, N. E.; Truhlar, D. G. Multi-coefficient Correlation Method: Comparison of Specific-Range Reaction Parameters to General Parameters for C_nH_xO_y Compounds. J. Phys. Chem. A 2001, 105, 4143–4149.
- [19] DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The correlation consistent composite approach (ccCA): An alternative to the Gaussian-n methods. *J. Chem. Phys.* **2006**, *124*, 114104.
- [20] Jiang, W.; DeYonker, N. J.; Determan, J. J.; Wilson, A. K. Toward accurate theoretical thermochemistry of first row transition metal complexes. J. Phys. Chem. A 2012, 116, 870– 885.
- [21] Laury, M. L.; DeYonker, N. J.; Jiang, W.; Wilson, A. K. A pseudopotential-based composite method: The relativistic pseudopotential correlation consistent composite approach for molecules containing 4d transition metals (Y-Cd). J. Chem. Phys. 2011, 135, 214103.
- [22] Riojas, A. G.; Wilson, A. K. Solv-ccCA: Implicit solvation and the correlation consistent composite approach for the determination of pK_a. J. Chem. Theory Comput. 2014, 10, 1500–1510.
- [23] Oyedepo, G. A.; Wilson, A. K. Multireference correlation consistent composite approach [MR-ccCA]: Toward accurate prediction of the energetics of excited and transition state chemistry. J. Phys. Chem. A 2010, 114, 8806–8816.
- [24] DeYonker, N. J.; Wilson, B. R.; Pierpont, A. W.; Cundari, T. R.; Wilson, A. K. Towards the intrinsic error of the correlation consistent Composite Approach (ccCA). *Mol. Phys.* 2009, 107, 1107–1121.
- [25] Prascher, B. P.; Lai, J. D.; Wilson, A. K. The resolution of the identity approximation applied to the correlation consistent composite approach. *J. Chem. Phys.* **2009**, *131*, 044130.
- [26] Mahler, A.; Wilson, A. K. Explicitly correlated methods within the ccCA methodology. *J. Chem. Theory Comput.* **2013**, *9*, 1402–1407.

- [27] Peterson, C.; Penchoff, D. A.; Wilson, A. K. Prediction of Thermochemical Properties Across the Periodic Table: A Review of the correlation consistent Composite Approach (ccCA) Strategies and Applications. *Annu. Rep. Comput. Chem.* **2016**, *12*, 3–45.
- [28] Das, S. R.; Williams, T. G.; Drummond, M. L.; Wilson, A. K. A QM/QM multilayer composite methodology: The ONIOM correlation consistent composite approach (ONIOMccCA). J. Phys. Chem. A 2010, 114, 9394–9397.
- [29] Oyedepo, G. A.; Wilson, A. K. Oxidative addition of the C α -C β bond in β -O-4 linkage of lignin to transition metals using a relativistic pseudopotential-based ccCA-ONIOM method. *ChemPhysChem* **2011**, *12*, 3320–3330.
- [30] Karton, A.; Daon, S.; Martin, J. M. L. W4-11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem. Phys. Lett.* 2011, *510*, 165–178.
- [31] Weber, R.; Wilson, A. K. Do composite methods achieve their target accuracy? *Comput. Theor. Chem.* **2015**, *1072*, 58–62.
- [32] Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151–154.
- [33] Sæbø, S.; Pulay, P. Local configuration interaction: An efficient approach for larger molecules. *Chem. Phys. Lett.* **1985**, *113*, 13–18.
- [34] Sæbø, S.; Pulay, P. Fourth-order Møller–Plessett perturbation theory in the local correlation treatment. I. Method. J. Chem. Phys. **1987**, *86*, 914–922.
- [35] Sæbø, S.; Pulay, P. The local correlation treatment. II. Implementation and tests. *J. Chem. Phys.* **1988**, 88, 1884–1890.
- [36] Sæbø, S.; Tong, W.; Pulay, P. Efficient elimination of basis set superposition errors by the local correlation method: Accurate ab initio studies of the water dimer. *J. Chem. Phys.* 1993, 98, 2170–2175.
- [37] Sæbø, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, 44, 213–236.
- [38] Schütz, M.; Werner, H.-J. Local perturbative triples correction (T) with linear cost scaling. *Chem. Phys. Lett.* **2000**, *318*, 370–378.
- [39] Schütz, M. Low-order scaling local electron correlation methods. III. Linear scaling local perturbative triples correction (T). *J. Chem. Phys.* **2000**, *113*, 9986–10001.
- [40] Schütz, M.; Werner, H.-J. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J. Chem. Phys.* **2001**, *114*, 661–681.
- [41] Pulay, P.; Sæbø, S. Orbital-invariant formulation and second-order gradient evaluation in Møller-Plesset perturbation theory. *Theor. Chem. Acc.* **1986**, *69*, 357–368.

- [42] Almlöf, J. Elimination of energy denominators in Møller-Plesset perturbation theory by a Laplace transform approach. *Chem. Phys. Lett.* **1991**, *181*, 319–320.
- [43] Häser, M.; Almlöf, J. Laplace transform techniques in Møller-Plesset perturbation theory. *J. Chem. Phys.* **1992**, *96*, 489–494.
- [44] Häser, M. Møller-Plesset (MP2) perturbation theory for large molecules. *Theor. Chem. Acc.* **1993**, 87, 147–173.
- [45] Wilson, A. K.; Almlöf, J. Møller-Plesset correlation energies in a localized orbital basis using a Laplace transform technique. *Theor. Chem. Acc.* **1997**, *95*, 49–62.
- [46] Ayala, P. Y.; Scuseria, G. E. Linear scaling second-order Moller-Plesset theory in the atomic orbital basis for large molecular systems. *J. Chem. Phys.* **1999**, *110*, 3660–3671.
- [47] Lambrecht, D. S.; Doser, B.; Ochsenfeld, C. Rigorous integral screening for electron correlation methods. *J. Chem. Phys.* **2005**, *123*, 184102.
- [48] Doser, B.; Lambrecht, D. S.; Kussmann, J.; Ochsenfeld, C. Linear-scaling atomic orbitalbased second-order Møller-Plesset perturbation theory by rigorous integral screening criteria. *J. Chem. Phys.* 2009, 130, 064107.
- [49] Hetzer, G.; Pulay, P.; Werner, H.-J. Multipole approximation of distant pair energies in local MP2 calculations. *Chem. Phys. Lett.* **1998**, 290, 143–149.
- [50] Scuseria, G. E.; Ayala, P. Y. Linear scaling coupled cluster and perturbation theories in the atomic orbital basis. *J. Chem. Phys.* **1999**, *111*, 8330–8343.
- [51] Subotnik, J. E.; Sodt, A.; Head-Gordon, M. A near linear-scaling smooth local coupled cluster algorithm for electronic structure. *J. Chem. Phys.* **2006**, *125*.
- [52] Neese, F.; Hansen, A.; Liakos, D. G. Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis. *J. Chem. Phys.* **2009**, *131*, 064103.
- [53] Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate local approximations to coupledelectron pair approaches: An attempt to revive the pair natural orbital method. *J. Chem. Phys.* 2009, *130*, 114108.
- [54] Huntington, L. M.; Hansen, A.; Neese, F.; Nooijen, M. Accurate thermochemistry from a parameterized coupled-cluster singles and doubles model and a local pair natural orbital based implementation for applications to larger systems. *J. Chem. Phys.* **2012**, *136*, 064101.
- [55] Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013**, *138*, 034106.
- [56] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.

- [57] Pinski, P.; Riplinger, C.; Valeev, E. F.; Neese, F. Sparse maps A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals. *J. Chem. Phys.* 2015, 143, 034108.
- [58] Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* **2016**, *144*.
- [59] Guo, Y.; Sivalingam, K.; Valeev, E. F.; Neese, F. SparseMaps A systematic infrastructure for reduced-scaling electronic structure methods. III. Linear-scaling multireference domainbased pair natural orbital N-electron valence perturbation theory. J. Chem. Phys. 2016, 144.
- [60] Pavošević, F.; Pinski, P.; Riplinger, C.; Neese, F.; Valeev, E. F. SparseMaps A systematic infrastructure for reduced-scaling electronic structure methods. IV. Linear-scaling second-order explicitly correlated energy with pair natural orbitals. *J. Chem. Phys.* **2016**, *144*.
- [61] Saitow, M.; Becker, U.; Riplinger, C.; Valeev, E. F.; Neese, F. A new near-linear scaling, efficient and accurate, open-shell domain-based local pair natural orbital coupled cluster singles and doubles theory. J. Chem. Phys. 2017, 146, 164105.
- [62] Werner, H.-J.; Knizia, G.; Krause, C.; Schwilk, M.; Dornbach, M. Scalable electron correlation methods I.: PNO-LMP2 with linear scaling in the molecular size and nearinverse-linear scaling in the number of processors. J. Chem. Theory Comput. 2015, 11, 484–507.
- [63] Ma, Q.; Werner, H.-J. Scalable Electron Correlation Methods. 2. Parallel PNO-LMP2-F12 with Near Linear Scaling in the Molecular Size. J. Chem. Theory Comput. 2015, 11, 5291– 5304.
- [64] Menezes, F.; Kats, D.; Werner, H.-J. Local complete active space second-order perturbation theory using pair natural orbitals (PNO-CASPT2). J. Chem. Phys. 2016, 145.
- [65] Schwilk, M.; Ma, Q.; Köppl, C.; Werner, H.-J. Scalable Electron Correlation Methods. 3. Efficient and Accurate Parallel Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD). J. Chem. Theory Comput. 2017, 13, 3650–3675.
- [66] Nedd, S. A.; DeYonker, N. J.; Wilson, A. K.; Piecuch, P.; Gordon, M. S. Incorporating a completely renormalized coupled cluster approach into a composite method for thermodynamic properties and reaction paths. *J. Chem. Phys.* **2012**, *136*, 144109.
- [67] Humbel, S.; Sieber, S.; Morokuma, K. The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for n-butane conformation and S_N2 reaction: RCl⁺Cl⁻. J. Chem. Phys. **1996**, 105, 1959–1967.
- [68] Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643–5677.

- [69] Kendall, R. A.; Früchtl, H. A. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theor. Chem. Acc.* **1997**, *97*, 158–163.
- [70] Schütz, M.; Hetzer, G.; Werner, H.-J. Low-order scaling local electron correlation methods.I. Linear scaling local MP2. J. Chem. Phys. 1999, 111, 5691–5705.
- [71] Anoop, A.; Thiel, W.; Neese, F. A local pair natural orbital coupled cluster study of Rh catalyzed asymmetric olefin hydrogenation. *J. Chem. Theory Comput.* **2010**, *6*, 3137–3144.
- [72] Sparta, M.; Riplinger, C.; Neese, F. Mechanism of olefin asymmetric hydrogenation catalyzed by iridium phosphino-oxazoline: A pair natural orbital coupled cluster study. *J. Chem. Theory Comput.* **2014**, *10*, 1099–1108.
- [73] Sparta, M.; Neese, F. Chemical applications carried out by local pair natural orbital based coupled-cluster methods. *Chem. Soc. Rev.* **2014**, *43*, 5032–5041.
- [74] Chan, B.; Kawashima, Y.; Katouda, M.; Nakajima, T.; Hirao, K. From C₆₀ to Infinity: Large-Scale Quantum Chemistry Calculations of the Heats of Formation of Higher Fullerenes. *J. Am. Chem. Soc.* 2016, *138*, 1420–1429.
- [75] Minenkov, Y.; Wang, H.; Wang, Z.; Sarathy, S. M.; Cavallo, L. Heats of Formation of Medium-Sized Organic Compounds from Contemporary Electronic Structure Methods. J. Chem. Theory Comput. 2017, 13, 3537–3560.
- [76] Boys, S. F. Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Rev. Mod. Phys.* **1960**, *32*, 296–299.
- [77] Foster, J. M.; Boys, S. F. Canonical Configuration Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300–302.
- [78] Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. J. Chem. Phys. 1989, 90, 4916–4926.
- [79] Kleier, D. A.; Halgren, T. A.; Hall, J. H.; Lipscomb, W. N. Localized molecular orbitals for polyatomic molecules. I. a comparison of the Edmiston-Ruedenberg and Boys localization methods. J. Chem. Phys. 1974, 61, 3905–3919.
- [80] Chulhai, D. V.; Goodpaster, J. D. Improved Accuracy and Efficiency in Quantum Embedding through Absolute Localization. *J. Chem. Theory Comput.* **2017**, *13*, 1503–1508.
- [81] Sirjoosingh, A.; Pak, M. V.; Brorsen, K. R.; Hammes-Schiffer, S. Quantum treatment of protons with the reduced explicitly correlated Hartree-Fock approach. J. Chem. Phys. 2015, 142, 214107.
- [82] Kállay, M. Linear-scaling implementation of the direct random-phase approximation. J. *Chem. Phys.* **2015**, *142*, 204105.

- [83] Switkes, E.; Stevens, R. M.; Lipscomb, W. N.; Newton, M. D. Localized bonds in SCF wavefunctions for polyatomic molecules. I. Diborane. J. Chem. Phys. 1969, 51, 2085–2093.
- [84] Levy, M.; Stevens, W. J.; Shull, H.; Hagstrom, S. Transferability of electron pairs between H₂O and H₂O₂. *J. Chem. Phys.* **1974**, *61*, 1844–1856.
- [85] Stoll, H.; Wagenblast, G.; Preuß, H. On the Use of Local Basis-Sets for Localized Molecular-Orbitals. *Theor. Chem. Acc.* 1980, 57, 169–178.
- [86] Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J. Chem. Phys.* 1997, 106, 1063–1079.
- [87] Neese, F. The ORCA program system. 2012; http://doi.wiley.com/10.1002/wcms.81.
- [88] Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1327.
- [89] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [90] Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- [91] Dunning Jr., T. H.; Peterson, K. A.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. *J. Chem. Phys.* 2001, 114, 9244–9253.
- [92] Moore, C. E. *Atomic Energy Levels, Vol. I (Hydrogen through Vanadium)*; Circular of the National Bureau of Standards 467: Washington D.C., 1949.
- [93] Chase Jr, M. W.; Tables, N.-J. T. Data reported in NIST standard reference database 69, June 2005 release: NIST Chemistry WebBook. J. Phys. Chem. Ref. Data, Monograph 1998, 9, 1–1951.
- [94] Cioslowski, J.; Schimeczek, M.; Liu, G.; Stoyanov, V. A set of standard enthalpies of formation for benchmarking, calibration, and parametrization of electronic structure methods. *J. Chem. Phys.* **2000**, *113*, 9377–9389.
- [95] Petersson, G. A.; Malick, D. K.; Wilson, W. G.; Ochterski, J. W.; Montgomery Jr., J. A.; Frisch, M. J. Calibration and comparison of the Gaussian-2, complete basis set, and density functional methods for computational thermochemistry. *J. Chem. Phys.* **1998**, *109*, 10570– 10579.
- [96] Montgomery Jr., J. A.; Michels, H. H.; Francisco, J. S. Ab initio calculation of the heats of formation of CF₃OH and CF₂O. *Chem. Phys. Lett.* **1994**, 220, 391–396.
- [97] Feller, D.; Peterson, K. A.; Dixon, D. A. Ab Initio Coupled Cluster Determination of the Heats of Formation of C₂H₂F₂, C₂F₂, and C₂F₄. *J. Phys. Chem. A* **2011**, *115*, 1440–1451.

- [98] Feller, D.; Peterson, K. A.; Dixon, D. A. Erratum: Ab initio coupled cluster determination of the heats of formation of $C_2H_2F_2$, C_2F_2 , and C_2F_4 (The Journal of Physical Chemistry A (2011) 115 (1440-1451) DOI:10.1021/jp111644h). J. Phys. Chem. A **2011**, 115, 3182.
- [99] Colegrove, B. T.; Thompson, T. B. Ab initio heats of formation for chlorinated hydrocarbons: Allyl chloride, cis- and trans-1-chloropropene, and vinyl chloride. J. Chem. Phys. 1997, 106, 1480–1490.
- [100] Tasi, G.; Izsák, R.; Matisz, G.; Császár, A. G.; Kállay, M.; Ruscic, B.; Stanton, J. F. The origin of systematic error in the standard enthalpies of formation of hydrocarbons computed via atomization schemes. *ChemPhysChem* 2006, 7, 1664–1667.
- [101] Karton, A.; Martin, J. M. L. Heats of formation of beryllium, boron, aluminum, and silicon re-examined by means of W4 theory. J. Phys. Chem. A. 2007; pp 5936–5944.
- [102] Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. Exploring the accuracy limits of local pair natural orbital coupled-cluster theory. *J. Chem. Theory Comput.* 2015, 11, 1525–1539.
- [103] Stoychev, G. L.; Auer, A. A.; Neese, F. Automatic Generation of Auxiliary Basis Sets. J. *Chem. Theory Comput.* **2017**, *13*, 554–562.
- [104] Kendall, R. A.; Dunning Jr., T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- [105] De Jong, W. A.; Harrison, R. J.; Dixon, D. A. Parallel Douglas-Kroll energy and gradients in NWChem: Estimating scalar relativistic effects using Douglas-Kroll contracted basis sets. *J. Chem. Phys.* 2001, 114, 48–53.
- [106] Peterson, K. A.; Dunning Jr., T. H. Accurate correlation consistent basis sets for molecular core-valence correlation effects: The second row atoms Al-Ar, and the first row atoms B-Ne revisited. J. Chem. Phys. 2002, 117, 10548–10560.
- [107] Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning Jr., T. H.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. VII. Valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg. *Theor. Chem. Acc.* 2011, *128*, 69–82.
- [108] Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* **1995**, *240*, 283–290.
- [109] Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 2005, 7, 3297.
- [110] Hättig, C. Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core–valence and quintuple- ζ basis sets for H to Ar and QZVPP basis sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, 7, 59–66.
- [111] Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.

- [112] Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree–Fock exchange. *Chem. Phys.* 2009, 356, 98–109.
- [113] Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.
- [114] Pollack, L.; Windus, T. L.; de Jong, W. A.; Dixon, D. A. Thermodynamic properties of the C5, C6, and C8n-alkanes from ab initio electronic structure theory. J. Phys. Chem. A 2005, 109, 6934–6938.
- [115] Peterson, K. A.; Woon, D. E.; Dunning Jr., T. H. Benchmark calculations with correlated molecular wave functions. IV. The classical barrier height of the H+H₂→H₂+H reaction. J. *Chem. Phys.* **1994**, *100*, 7410–7415.
- [116] Schwartz, C. Importance of angular correlations between atomic electrons. *Phys. Rev.* **1962**, *126*, 1015–1019.
- [117] Schwartz, C. Methods Comput. Phys.; Academic Press Inc.: New York, NY, 1963; pp 241–266.
- [118] Kutzelnigg, W.; Morgan, J. D. Rates of convergence of the partial-wave expansions of atomic correlation energies. J. Chem. Phys. 1992, 96, 4484–4508.
- [119] Martin, J. M. L. Ab initio total atomization energies of small molecules towards the basis set limit. *Chem. Phys. Lett.* **1996**, 259, 669–678.
- [120] Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. J. Chem. Phys. 1997, 106, 9639–9646.
- [121] Martin, J. M. L.; Lee, T. J. The atomization energy and proton affinity of NH₃. An ab initio calibration study. *Chem. Phys. Lett.* **1996**, 258, 136–143.
- [122] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N₂, and H₂O. *Chem. Phys. Lett.* 1998, 286, 243–252.
- [123] Kleier, D. A.; Dixon, D. A.; Lipscomb, W. N. Localized molecular orbitals for polyatomic molecules. *Theor. Chem. Acc.* 1975, 40, 33–45.
- [124] Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
- [125] Janowski, T.; Ford, A. R.; Pulay, P. Accurate correlated calculation of the intermolecular potential surface in the coronene dimer. *Mol. Phys.* 2010, 108, 249–257.
- [126] Řezáč, J., Jan; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. J. Chem. Theory Comput. 2011, 7, 2427–2438.

- [127] Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of quantum chemical methods for large noncovalent complexes. J. Chem. Theory Comput. 2013, 9, 3364–3374.
- [128] Roux, M. V.; Temprado, M.; Chickos, J. S.; Nagano, Y. Critically Evaluated Thermochemical Properties of Polycyclic Aromatic Hydrocarbons. J. Phys. Chem. Ref. Data 2008, 37, 1855– 1996.

CHAPTER 5

COMPUTATIONAL CHEMISTRY CONSIDERATIONS IN CATALYSIS: REGIOSELECTIVITY AND METAL-LIGAND DISSOCIATION

5.1 Introduction

For the prediction of thermodynamic information (i.e., enthalpies, free energies), reaction barriers, HOMO-LUMO gaps, and other fundamental properties, density functional theory (DFT) approaches are commonly used for catalysis. For early main group chemistry (i.e., hydrocarbons), there are many different density functionals can be used, with very little difference in predicted property arising from the choice of functional to describe energetics, and with limited exceptions, as demonstrated by Karton et al.¹ However, for transition metal species, the utility of each functional can vary widely based upon choice of metal, choice of ligand, and property of interest.^{2–8} To illustrate, for a set of ~20 3*d* transition metal species, B3LYP/CEP-31G(d) resulted in errors from experiment from the predicted enthalpies of formation by ~100 kcal mol⁻¹.² However, when the same functional is applied to a different set of the transition metal species —a set that has the smallest reported experimental uncertainties in the enthalpy of formation-- the error is ~6-7 kcal mol⁻¹.^{4,5} So, indeed, extraordinarily large variances can occur depending upon metal and ligand. For catalysis, where there may be interest in understanding the thermochemistry with much smaller errors in energy, this magnitude of error may be of limited utility.

Computational approaches have been designed to improve upon the predictions possible by DFT for transition metal species. With *ab initio* composite approaches like the correlation consistent Composite Approach, ccCA, designed in our group, $^{9-13}$ differences of ~2-3 kcal mol⁻¹, on average, can be achieved in the prediction of enthalpies of formation for 3*d* transition metal species. 12,13 As well, ccCA targeted 4*d* transition metal chemistry by utilizing relativistic pseudopotentials, denoted as rp-ccCA, to model relativistic contributions from core electrons and yielded differences of ~3 kcal mol⁻¹ from experimental enthalpies of formation. 10,13 This is useful, but more costly than

DFT approaches. Strategies have evolved that help to reduce the computational cost, which is the bottleneck in these calculations (e.g., DLPNO-ccCA) while preserving the accuracy or, as a means to provide quantitative energy predictions for transition metal-based catalytic processes.

So far, these comments have focused upon general trends in the prediction of thermodynamic properties of molecular systems. However, a question is, what is the utility of computational approaches for an important industrial process like hydroformylation? More specifically, how useful are computational approaches, particularly new approaches like DLPNO-ccCA, a form of ccCA, for important properties like regioselectivity and metal-ligand binding? And, is the qualitative or quantitative picture impacted by computational method choice?

The mechanism for Rh-based hydroformylation was well-established by Wilkinson in the late 1960s to early 1970s.¹⁴ As the largest volume homogeneous chemical reaction conducted in industry for chemical production, the process converts olefins to aldehydes in a syngas mixture. The advantage of Rh-based hydroformylation as opposed to Co-based hydroformylation is the favorable reaction conditions (ambient temperature and pressure). The efficacy of a catalyst designed for hydroformylation is the ratio of the linear aldehyde to the branched aldehyde (Fig. 5.1), known as the linear-to-branched ratio. In hydroformylation, the formation of the linear aldehyde is favored although there are studies targeting asymmetric hydroformylation, i.e. the production of the branched aldehyde.^{15,16} This is measured through the kinetics of the migratory insertion of the olefin to the catalyst.



Figure 5.1: Hydroformylation reaction converting olefins to linear and branched aldehydes via a Rh catalyst.

Numerous computational studies have targeted modeling the regioselectivity of hydroformylation due to its importance in chemical industry.^{17–26} To account for the size of the catalysts and the limited computing power at the time, earlier computational studies either

substituted PPh₃ ligands with much smaller PH₃ ligands or utilized multilevel computational chemistry methods, such as ONIOM, ^{27,28} to model the bond breaking and formation region with DFT while relegating the sterically bulky ligands to a computationally more affordable method, such as molecular mechanics (MM).^{17–21} While more recent studies also utilize multilevel approaches for hydroformylation, more rigorous *ab initio* methodologies are used to model bond breaking and forming regions and use DFT to model the steric ligands.^{22,23} Other studies have only used DFT to model the olefin insertion step as well as the entire Wilkinson catalytic cycle.^{23–25,29,30} These studies provide insight into potential electronic contributions of the sterically bulky ligands as well as the mechanism by identifying the rate-determining step, which can change based on the type of ligand. Machine learning approaches have recently been developed to screen potential ligands based on their regioselectivity and is a rising trend in computational catalysis.^{26,31}



Figure 5.2: A model of the two reaction pathways for hydroformylation where ΔE_l^{\ddagger} and ΔE_b^{\ddagger} are the reaction barriers for forming the linear and branched product, respectively. The energy difference between the two reaction barriers is denoted as $\Delta \Delta E^{\ddagger}$.

The kinetics of hydroformylation is very sensitive energetically, i.e. the differences in energy between competing pathways ($\Delta\Delta E^{\ddagger}$), illustrated in Fig. 5.2, can be less than 1 kcal mol⁻¹.¹⁷ With such small differences in energy for the competing pathways, calculating the correct linear-to-branched ratio can be difficult to predict with computational methods. For example, as shown in Table 5.1, if $\Delta\Delta E^{\ddagger} = 0$ (reaction barriers are equivalent), the l:b ratio is 50:50. However, lowering the barrier for the linear product by 1 kcal mol⁻¹ ($\Delta\Delta E^{\ddagger} = -1$ kcal mol⁻¹) results in a product ratio

increase to approximately 84:16 and lowering the barrier for the formation of the linear product by an additional 2 kcal mol⁻¹ ($\Delta\Delta E^{\ddagger} = -3$ kcal mol⁻¹) indicates that the reaction highly favors the linear product (100:0 ratio).

Table 5.1: A summary of the effect of $\Delta \Delta E^{\ddagger}$ in kcal mol⁻¹ on the linear-to-branched ratio (l:b) ratio for hydroformylation.

$\Delta\Delta E^{\ddagger}$ (kcal mol ⁻¹)	Linear-to-branched ratio (l:b)
0 to -1	50:50 to 84:16
1 to -2	84:16 to 97:3
-2 to -3	97:3 to 100:0

Basically, considering the challenges mentioned earlier about the prediction of thermochemistry properties for transition metal species, achieving the level of accuracy needed to even predict the correct product distributions seems unsurmountable. Cancellation of errors that can occur from comparing energy differences is helpful, though the errors from experiment are not necessarily the same across a reaction pathway, and, thus, gauging method utility for each problem, considering metal, ligand, and property, is essential. Thus, in this chapter, the impact of method and basis set choice – the route to describe the molecular orbitals – are considered to determine the impact of these choices upon the prediction of linear-to-branched ligand ratio, as well as the ligand dissociation energy.

Another aspect that is important in catalysis is the description of metal-ligand dissociation, as it is a primary step in all homogeneous catalytic reactions, e.g. product dissociation from Rh-catalyzed hydroformylation and solvent interactions with olefin hydrogenation, as well as gas phase ligand dissociation for organometallic reactions targeting C-H activation. Here, to gain understanding about the utility of the *ab initio* composite strategy, DLPNO-ccCA, a cationic (diiimine)(aquo)Pt^{II} complex was examined. This Pt^{II} complex was chosen since Pt^{II} complexes with ligands containing aromatic and aliphatic C-H bonds are involved in the oxidative addition of alkanes and have been a focus in C-H activation studies where the ligand substitution step is rate-determining.^{32–35}

5.2 Computational Methods

5.2.1 Computational methods for hydroformylation

DFT and *ab initio* calculations were done in this study. Several density functionals were utilized, selecting a number of widely used functionals varying in complexity: B3LYP,^{36,37} B3P86,^{36,38} BLYP,^{37,39} BP86,^{38,39} PBE,^{40,41} and PBE0.^{40–42} (It should be noted that while increased complexity often means better property predictions, this is not necessarily guaranteed.) Grimme's dispersion correction with Becke-Johnson dampening (D3BJ) was included for B3LYP and PBE0 to correct for long-range intramolecular interactions.⁴³ The Stuttgart/Dresden basis set, and pseudopotential (SDD) was used for all DFT calculations.^{44,45} Though it is commonly believed that a triple- ζ quality basis set is sufficient for DFT calculations, earlier work has demonstrated that for the predictions of energetic properties of transition metal species, quadruple- ζ level basis sets can have an impact on the energies, and, thus, this level of basis set was considered.^{3,4} As well, this choice of basis set followed earlier work done by Kumar et al.²⁴, and all structures for the DFT and *ab initio* calculations were based on this prior work. DFT calculations in the present work were done with Gaussian16.⁴⁶

Several *ab initio* correlated methods also were used including domain-based pair natural orbital (DLPNO) methods, ^{47–52} DLPNO-MP2 and DLPNO-CCSD(T)), the MP2 and CCSD(T) varieties of the DLPNO approach. The DLPNO approach enables computational cost reductions from typical MP2 and CCSD(T) calculations. And, CCSD(T) is of particular interest, as this method is known for its utility in energy predictions when paired with a high-quality (which typically means large) basis set. The DLPNO calculations were done with the ORCA program suite.⁵³ Calculations were done using Dunning's correlation consistent polarized valence-n- ζ ("zeta") basis sets (aug-cc-pV*nZ*, where n=D (double), T (triple), Q (quadruple)), and considering augmented (aug-cc-pV*nZ*) and augmented core-valence (aug-cc-pCV*nZ*) forms of the sets.^{54,55} For P and Cl, the recommended tight *d* versions of the correlation consistent basis sets, denoted as cc-pV(*n* + *d*)Z, aug-cc-pV(*n* + *d*)Z, and aug-cc-pCV(*n* + *d*)Z were used.⁵⁵ The correlation consistent

pseudopotentials (cc-pVnZ-PP) were used for Rh and Pt atoms.^{56,57} The correlation consistent Composite Approach (ccCA) for 4d transition metals was also considered, ¹⁰ utilizing the DLPNO methods for the composite steps to reduce the computational resources associated with the size of the compound, denoted as DLPNO-rp-ccCA.⁵⁸

To calculate the regioselectivity for hydroformylation, the following equation was used for the linear-to-branched ratio

$$l:b = k_l: k_b = \frac{e^{\left(-\Delta G_l^{\ddagger}/kT\right)}}{e^{\left(-\Delta G_b^{\ddagger}/kT\right)}} = e^{\left(-\Delta \Delta G^{\ddagger}/kT\right)} \approx e^{\left(-\Delta \Delta E^{\ddagger}/kT\right)}$$
(5.1)

where $\Delta\Delta G^{\ddagger}$ is the free energy barrier, $\Delta\Delta G^{\ddagger}$ is the energy difference between the two reaction pathways, k is the Boltzmann constant, and T is the temperature. This equation assumes the olefin insertion step is irreversible. The Rh-catalyst-olefin complex examined with the DLPNO methods is ee–[Rh(H)(CO)(DIPHOS)(propene)] where the bis-phosphine DIPHOS ligand is in the equatorial-equatorial (ee) conformation. The ligands examined with DFT include (PPh₃)₂, and more structurally complex bis-phosphine ligands, TBDCP, DIOP, and DIPHOS. All ligands are attached to a [Rh(H)(CO)] backbone as indicated in the Wilkinson catalytic cycle for Rhbased hydroformylation. Olefins examined with (PPh₃)₂ include pentene, hexene, heptene, octene, decene, dodecene, styrene, and vinyl acetate. Propene is coordinated with all bisphosphine ligands. As the experiments were carried out in toluene, the SMD implicit solvent model⁵⁹ was used to mimic the long-range solvent effects of toluene on the Rh catalyst.

The Rh-catalyst-olefin complex examined with the DLPNO methods is ee-[Rh(H)(CO)(DIPHOS)(propene)] where the bis-phosphine DIPHOS ligand is in the equatorial-equatorial (ee) conformation. The ligands examined with DFT include $(PPh_3)_2$, and more structurally complex bis-phosphine ligands, TBDCP, DIOP, and DIPHOS. All ligands are attached to a [Rh(H)(CO)] backbone as indicated in the Wilkinson catalytic cycle for Rh-based hydroformylation. Olefins examined with $(PPh_3)_2$ include pentene, hexene, heptene, octene, decene, dodecene, styrene, and vinyl acetate. Propene is coordinated with all bisphosphine ligands.



Figure 5.3: Computationally determined 3D structures of ee-[Rh(H)(CO)(DIPHOS) (propene)] catalyst complex (top) and the dissociation reaction of H_2O from the cationic (diimine)(aquo)Pt^{II} complex (bottom).

5.2.2 Computational methods for ligand dissociation

The gas phase ligand dissociation energy was evaluated by the difference between the complex and the respective fragments.

$$\Delta E_{dissoc} = E_{AB} - E_A + E_B \tag{5.2}$$

where E_{AB} is the electronic energy of the complex, E_A is the electronic energy of fragment A, and E_B is the electronic energy of fragment B. Ab initio calculations, in particular, are susceptible to basis set superposition error (BSSE), which can result in overbinding of the ligands.⁶⁰ BSSE can occur when there is an imbalance in basis set size for the species considered in determining an energy difference. The effects have been addressed by conducting calculations on the individual species in the presence of the basis set associated with the other species. A correction for the BSSE was applied to all DLPNO calculations.

To study fundamental organometallic reactions that occur in the gas phase, a cationic (diimine)(aquo)Pt^{II} complex prevalent in C-H activation and oxidative addition of alkanes was chosen (see Fig. 5.3). This molecule was chosen due to computational feasibility based on the molecule size. The calculated zero point energy (ZPE) of the reaction obtained with a frequency calculation at the BP86 level and the PBE0 optimized structures were obtained from Weymuth et

al.⁶¹ This choice of functional for frequency calculations was selected since the ZPE of the reaction did not significantly change with respect to functional choice.⁶¹ PBE0 structures were utilized based on their success for heavier elements. A few density functionals, PBE0, B3LYP, and TPSSh⁶² utilizing cost-saving techniques, i.e the resolution-of-the-identity or RI approximation, were paired with the augmented correlation consistent basis sets and pseudopotentials of triple- and quadruple- ζ level quality (aug-cc-pVnZ, n=T, Q), as well as DLPNO-rp-ccCA to determine ligand dissociation energies. All ligand dissociation calculations were done in the ORCA program suite.

5.3 Results and Discussion

5.3.1 Regioselectivity in hydroformylation

The DFT 1:b ratios for all Rh catalysts are shown in Table 5.2. The corresponding $\Delta\Delta E^{\ddagger}s$ for all DFT results are shown in Table 5.3. The DLPNO results for hydroformylation are shown in Table 5.4 for the l:b ratios, including the l:b determined for calculations that have been corrected for BSSE. With DFT, qualitatively correct 1:b ratios are obtained for most of the examined catalyst-olefin complexes as shown in Table 5.2. However, this largely depends on which type of functional is used. For example, using BLYP, BP86, and PBE generally predicted 1:b ratios that are in disagreement with experiment for (PPh₃)₂ ligands, particularly for hexene, heptene, octene, dodecene, and styrene, which produced lib ratios of 2:98, 26:74, 11:89, 84:16, and 87:13, respectively, for BLYP, and similar ratios for BP86 and PBE (Table 2). With an increase in complexity in the functionals, i.e. B3LYP, B3P86, and PBE0, 1:b ratios of 67:33 47:53 and 76:24 for B3LYP, 71:29, 67:33, 68:32 for B3P86, and 75:25, 73:27, 71:29 for PBE0, were predicted for the conversion of heptane, octene, and dodecane with (PPh₃)₂ ligands, respectively. And for ee-[Rh(H)(CO)(PPh₃)₂(pentene)], the linear product is predicted. However, the inclusion of Grimme's dispersion correction for B3LYP and PBE0 predicted 1:b ratios that predicted the more favorable produce, in agreement with experiment for all examined catalyst-olefin complexes with the exception of ee-[Rh(H)(CO) (PPh₃)₂(decene)] (1:b ratios of 0:100 and 3:97 for B3LYP-D3 and PBE0-D3, respectively) and ee-[Rh(H)(CO)

(DIPHOS)(propene)] (1:b ratios of 17:83 and 11:89 for B3LYP-D3 and PBE0-D3, respectively). Based on the 1:b ratios found, introducing complexity in the functional can but not always improve property prediction.
	BLYP	BP86	PBE	B3LYP	B3P86	PBE0	B3LYP-D3	PBE0-D3	Exp ^a
$L=(PPh_3)_2$									
Pentene	83:17	81:19	79:21	95:5	95:5	95:5	100:0	99:1	95:5
Hexane	2:98	3:97	6:94	10:90	19:81	23:77	100:0	99:1	92:8
Heptane	26:74	28:72	33:67	67:33	71:29	75:25	100:0	99:1	86:14
Octene	11:89	20:80	31:69	47:53	67:33	73:27	100:0	100:0	81:19
Decene	84:16	93:7	89:11	53:47	69:31	64:36	0:100	3:97	74:26
Dodecene	45:55	33:67	41:59	76:24	68:32	71:29	100:0	99:1	87:12
Styrene	87:13	92:8	79:21	83:17	84:16	90:10	0:100	1:99	11:89
Vinyl acetate	0:100	0:100	0:100	0:100	0:100	0:100	0:100	0:100	9:91
L=TBDCP									
Propene	90:10	88:12	89:11	96:4	95:5	96:4	92:8	95:5	92:8
L=DIOP									
Propene	99:1	99:1	100:0	100:0	100:0	100:0	100:0	100:0	90:10
L=ee-DIPHOS									
propene	3:97	3:97	3:97	5:95	5:95	5:95	17:83	11:89	69:31 ^b
	92.17	72.07	71.20	96.14	77.72	76.24	00.17	91.10	60.21b
propene	83:17	15:21	/1:29	80:14	11:25	/0:24	00:12	01:19	09:31

Table 5.2: Comparison of several density functionals to linear-to-branched ratios from experiment for ee-[Rh(H)(CO)(L)(olefin)] complexes.

^aReferences 14,63–69

^bThe data references the ea conformer of DIPHOS.

	BLYP	BP86	PBE	B3LYP	B3P86	PBE0	B3LYP-D3	PBE0-D3	Exp ^a
$L = (PPh_3)_2$									
Pentene	-0.96	-0.88	-0.80	-1.78	-1.74	-1.74	-3.67	-3.05	-1.74
Hexene	2.44	2.07	1.65	1.29	0.87	0.71	-3.93	-2.70	-1.44
Heptene	0.62	0.56	0.41	-0.42	-0.53	-0.66	-3.74	-2.87	-1.07
Octene	1.26	0.81	0.47	0.06	-0.42	-0.60	-4.87	-3.85	-0.85
Decene	-0.98	-1.49	-1.23	-0.08	-0.47	-0.34	3.36	2.05	-0.62
Dodecene	0.11	0.43	0.21	-0.68	-0.43	-0.54	-3.92	-2.91	-1.13
Styrene	-1.10	-1.44	-0.78	-0.94	-1.00	-1.29	5.62	2.87	1.24
Vinyl acetate	6.24	6.55	6.59	6.02	6.35	6.27	5.06	5.42	1.36
L = TBDCP									
Propene	-1.31	-1.16	-1.23	-1.86	-1.71	-1.92	-1.45	-1.70	-1.44
L = DIOP									
Propene	-2.57	-3.12	-3.22	-3.28	-3.76	-4.02	-3.59	-4.22	-1.30
L = ee-DIPHOS									
propene	2.00	2.10	2.08	1.70	1.80	1.75	0.93	1.25	-0.47 ^b
L = ea-DIPHOS									
propene	-0.96	-0.58	-0.54	-1.05	-0.72	-0.69	-1.17	-0.86	-0.47 ^b

Table 5.3: Comparison of the approximate $\Delta \Delta E^{\ddagger}s$ based on the calculated 1:b ratios for ee- [Rh(H)(CO)(L)(olefin)] complexes. Experimental $\Delta \Delta E^{\ddagger}s$ are an approximation of experimental 1:b ratios. All $\Delta \Delta E^{\ddagger}s$ are in kcal mol⁻¹.

^aReferences 14,63–69

^aThe data references the ea conformer of DIPHOS.

For ee-[Rh(H)(CO) (PPh₃)₂(vinyl acetate)], the predicted $\Delta\Delta E^{\ddagger}$ was ~6 kcal mol⁻¹ for each functional considered as shown in Table 3, indicating the branched isomer is favored. Overall, the dispersion-corrected functionals resulted in a lowering of the $\Delta\Delta E^{\ddagger}$ for ee-[Rh(H)(CO) (PPh₃)₂(vinyl acetate)] by approximately 1 kcal mol⁻¹, however, this did not impact the product distribution. Similarly, for ee[Rh(H)(CO)(DIOP)(propene)], the dispersion correction functionals resulted in a lowered the predicted $\Delta\Delta E^{\ddagger}$ by ~0.3 kcal mol⁻¹ and did not impact the product distribution as the predicted $\Delta\Delta E^{\ddagger}$ was ~4 kcal mol⁻¹. However, for ee-[Rh(H)(CO) (PPh₃)₂(styrene)] and ee-[Rh(H)(CO) (PPh₃)₂(decene)], the dispersion-corrected functionals predicted the $\Delta\Delta E^{\ddagger}$ to be ~6 kcal mol⁻¹ and ~3 kcal mol⁻¹ greater than the $\Delta\Delta E^{\ddagger}$ predicted with non-dispersion-corrected functionals. While this change in $\Delta\Delta E^{\ddagger}$ predicted product ratios of 0:100 and 1:99 for B3LYP-D3 and PBE0-D3, respectively, with styrene as the olefin, with decene as the olefin, the predicted product ratios were 0:100 and 3:97 for B3LYP-D3 and PBE0-D3, respectively.

For the DIPHOS ligand, the relative orientation of the Rh-H and Rh-CO bond to the DIPHOS ligand was a major factor in predicted 1:b ratios with DFT. In the ee conformation, all predicted 1:b ratios with DFT predicted the branched product whereas the linear product is predicted for the ea conformation, in qualitative agreement with experiment. This is to be noted for any calculation. The small conformation changes from the ee conformation to the ea conformation led to lowering of the $\Delta\Delta E^{\ddagger}$ by ~2-3 kcal mol⁻¹ for all functionals, changing the product ratio to favor the linear product over the branched ratio. This exhibits the high sensitivity of $\Delta\Delta E^{\ddagger}$, which can greatly affect product formation ratios with changes as small as a few tenths of a kcal mol⁻¹, as exhibited by the $\Delta\Delta E^{\ddagger}$ s of -0.54 and -1.05 kcal mol⁻¹ that yielded product ratios of 71:29 and 86:14 for PBE and B3LYP, respectively. Ergo, based on the observed trends from the DFT calculations, there remains a need to investigate hydroformylation with electron correlation methods.

	l:b	l:b (corrected for BSSE)
DLPNO-MP2/aug-cc-pVDZ-PP	25:75	100:0
DLPNO-MP2/aug-cc-pVTZ-PP	29:71	100:0
DLPNO-MP2/aug-cc-pVQZ-PP	16:84	100:0
DLPNO-MP2/cc-pVTZ-PP	18:82	100:0
DLPNO-CCSD(T)/cc-pVTZ-PP	1:99	100:0
DLPNO-CCSD(T)/aug-cc-pCVDZ-PP	100:0	100:0
DLPNO-CCSD(T,FC1)/aug-cc-pCVDZ-PP	100:0	100:0
DLPNO-rp-ccCA	100:0	100:0
Experimental		69:31

Table 5.4: Results using DLPNO methods to predict the linear-to-branched ratio for ee-[Rh(H)(CO)(DIPHOS)(propene)].

The data references the equatorial-axial (ea) conformer of DIPHOS.

Here, the ee-[Rh(H)(CO)(DIPHOS)(propene)] catalyst-olefin complex is considered, as DFT was unable to address the regioselectivity of this reaction correctly in any case. For the DLPNO methods, the l:b ratio is predicted to favor the branched isomer. When BSSE has been addressed, the linear isomer is favored. For DLPNO-rp-ccCA, the molecular orbital space is well-described, and, thus, accounting for BSSE effects is not necessary, and the l:b ratio predicted with DLPNOrp-ccCA is 100:0. This is primarily due to the interactions between the electrons from core orbitals with electrons in valence orbitals as DLPNO-CCSD(T)/aug-cc-pCVDZ-PP and DLPNO-CCSD(T,FC1)/aug-cc-pCVDZ-PP, which includes sub-valence electron (FC1) excitations within the molecular orbital space, both favored the linear isomer with product ratios of 100:0. The results from implementing the DLPNO methods indicate that electronic effects from including core electrons within the valence basis set are significant in determining $\Delta\Delta E^{\ddagger}$ given the large magnitude relative to other calculated $\Delta \Delta E^{\ddagger}$ s with *ab initio* methods. Even for qualitative predictions, DLPNO-rp-ccCA is useful without having to correct for BSSE. By utilizing a well-described molecular orbital space – DLPNO-rp-ccCA does predict the proper regioselectivity; DFT either does not predict the correct regioselectivity, such as for ee-[Rh(H)(CO) (PPh₃)₂(styrene)] and ee- $[Rh(H)(CO) (PPh_3)_2(hexene)]$, which predicted qualitatively inconsistent product ratios for most of the functionals examined. In addition, the regioselectivity is highly sensitive to functional choice, as the performance is not consistent as the ligand type and olefin changes. However, for the *ab initio* methods considered, simply improving the description of the molecular orbital space, either by BSSE correction or by including sub-valence electrons in the molecular orbital space for interactions yielding qualitative agreement with experiment.

5.3.2 Metal-ligand dissociation in organometallics

The gas phase ligand dissociation energies are shown in Table 5.5 with DLPNO-rp-ccCA compared both experiment and several density functionals utilizing to the resolution-of-the-identity approximation. For gas-phase ligand dissociation, DLPNO-rp-ccCA yields an error of $1.7 \text{ kcal } \text{mol}^{-1}$ relative to experiment. When utilizing the resolution-of-the-identity approximation within DFT calculations, RI-PBE0/aug-cc-pVTZ, RI-B3LYP/aug-cc-pVTZ, and RI-TPSSh yields dissociation energies of 20.7, 20.2, and 19.6 kcal mol^{-1} , respectively. However, increasing the quality of the molecular orbital space, i.e. using aug-cc-pVQZ, increased the error by 0.4, 0.5, and 0.5 kcal mol⁻¹ for RI-PBE0, RI-B3LYP, and RI-TPSSh, causing concern for utilizing DFT with higher quality basis sets. Regardless of functional and basis set choice, the predicted dissociation energy was greater than 5 kcal mol^{-1} lower than the experimental value. With the dispersion correction included for RI-PBE0/aug-cc-pVTZ, the predicted dissociation energy increased to 23.6 kcal mol^{-1} . This would suggest that accounting for dispersion is necessary for DFT predictions of gas-phase properties. DLPNO-rp-ccCA calculations yielded favorable results for ligand dissociation energy in comparison to DFT, but there are factors that can contribute to computationally predicted dissociation energies. For example, as density functionals are primarily used to generate structures for large organometallic complexes, the choice of functional for optimization must be considered. The predicted dissociation energies can change by a few kcal mol^{-1} based on slight structural change (root mean square deviation of ~20 pm) between functionals and by 10's of kcal mol^{-1} for significant structural changes such as ligand reorientation. Also, the basis set choice can affect the quality of predictions as indicated from the lowering of predicted dissociation

energy by increasing basis set quality.

Table 5.5: Comparison of the gas-phase ligand dissociation energy of H_2O from the Pt complex calculated with DLPNO-rp-ccCA and RI-DFT-D3/aug-cc-pVnZ. All energies are in kcal mol⁻¹ and are BSSE-corrected.

RI-PBE0/aug-cc-pVTZ	20.7
RI-B3LYP/aug-cc-pVTZ	20.2
RI-TPSSh/aug-cc-pVTZ	19.6
RI-PBE0/aug-cc-pVQZ	20.3
RI-B3LYP/aug-cc-pVQZ	19.7
RI-TPSSh/aug-cc-pVQZ	19.1
RI-PBE0-D3/aug-cc-pVTZ	23.6
DLPNO-rp-ccCA	24.2
Experiment	25.9 ± 0.7

5.4 Conclusions

Basically, considering the challenges mentioned earlier about the prediction of thermochemistry properties for transition metal species, achieving the level of accuracy needed to even predict the correct product distributions seems unsurmountable. Cancellation of errors that can occur from comparing energy differences is helpful, though the errors from experiment are not necessarily the same across a reaction pathway, and, thus, gauging method utility for each problem, considering metal, ligand, and property, is essential. A typical method choice for the study of transition metal species is DFT. Unfortunately, there is no "magic" functional to use for all problems. While *ab initio* methods like CCSD(T) or composite methods that try to replicate it like ccCA can be quite useful and are more dependable from system to system and, generally, across a reaction pathway, they are more costly, and may require additional measures to ensure quality results are obtained sometimes reaching near saturation of the orbital space (even more costly!). DFT can be very useful, but properly gauging it is important, as illustrated by this study.

REFERENCES

REFERENCES

- Karton, A.; Daon, S.; Martin, J. M. L. W4-11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem. Phys. Lett.* 2011, 510, 165–178.
- [2] Cundari, T. R.; Arturo Ruiz Leza, H.; Grimes, T.; Steyl, G.; Waters, A.; Wilson, A. K. Calculation of the enthalpies of formation for transition metal complexes. *Chem. Phys. Lett.* 2005, 401, 58–61.
- [3] Tekarli, S. M.; Drummond, M. L.; Williams, T. G.; Cundari, T. R.; Wilson, A. K. Performance of density functional theory for 3d transition metal-containing complexes: Utilization of the correlation consistent basis sets. J. Phys. Chem. A 2009, 113, 8607–8614.
- [4] Jiang, W.; Laury, M. L.; Powell, M.; Wilson, A. K. Comparative Study of Single and Double Hybrid Density Functionals for the Prediction of 3d Transition Metal Thermochemistry. J. Chem. Theory Comput. 2012, 8, 4102–4111.
- [5] Laury, M. L.; Wilson, A. K. Performance of density functional theory for second row (4d) transition metal thermochemistry. *J. Chem. Theory Comput.* **2013**, *9*, 3939–3946.
- [6] Determan, J. J.; Poole, K.; Scalmani, G.; Frisch, M. J.; Janesko, B. G.; Wilson, A. K. Comparative Study of Nonhybrid Density Functional Approximations for the Prediction of 3d Transition Metal Thermochemistry. J. Chem. Theory Comput. 2017, 13, 4907–4913.
- [7] Zhao, Y.; Truhlar, D. G. Comparative assessment of density functional methods for 3d transition-metal chemistry. J. Chem. Phys. 2006, 124, 1–7.
- [8] Dohm, S.; Hansen, A.; Steinmetz, M.; Grimme, S.; Checinski, M. P. Comprehensive Thermochemical Benchmark Set of Realistic Closed-Shell Metal Organic Reactions. J. Chem. Theory Comput. 2018, 14, 2596–2608.
- [9] DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The correlation consistent composite approach (ccCA): An alternative to the Gaussian-n methods. *J. Chem. Phys.* **2006**, *124*, 114104.
- [10] Laury, M. L.; DeYonker, N. J.; Jiang, W.; Wilson, A. K. A pseudopotential-based composite method: The relativistic pseudopotential correlation consistent composite approach for molecules containing 4d transition metals (Y-Cd). J. Chem. Phys. 2011, 135, 214103.
- [11] Nedd, S. A.; DeYonker, N. J.; Wilson, A. K.; Piecuch, P.; Gordon, M. S. Incorporating a completely renormalized coupled cluster approach into a composite method for thermodynamic properties and reaction paths. *J. Chem. Phys.* **2012**, *136*, 144109.
- [12] Jiang, W.; DeYonker, N. J.; Wilson, A. K. Multireference character for 3d transition-metalcontaining molecules. J. Chem. Theory Comput. 2012, 8, 460–468.

- [13] Manivasagam, S.; Laury, M. L.; Wilson, A. K. Pseudopotential-Based Correlation Consistent Composite Approach (rp-ccCA) for First- and Second-Row Transition Metal Thermochemistry. J. Phys. Chem. A 2015, 119, 6867–6874.
- [14] Evans, D.; Osborn, J. A.; Wilkinson, G. Hydroformylation of alkenes by use of rhodium complex catalysts. J. Chem. Soc. A Inorganic, Phys. Theor. 1968, 3133.
- [15] Klosin, J.; Landis, C. R. Ligands for practical rhodium-catalyzed asymmetric hydroformylation. Acc. Chem. Res. 2007, 40, 1251–1259.
- [16] Tonks, I. A.; Froese, R. D. J.; Landis, C. R. Very low pressure Rh-catalyzed hydroformylation of styrene with (S,S,S-bisdiazaphos): Regioselectivity inversion and mechanistic insights. *ACS Catal.* 2013, *3*, 2905–2909.
- [17] Kranenburg, M.; van der Burgt, Y. E. M.; Kamer, P. C. J.; van Leeuwen, P. W. N. M.; Goubitz, K.; Fraanje, J. New Diphosphine Ligands Based on Heterocyclic Aromatics Inducing Very High Regioselectivity in Rhodium-Catalyzed Hydroformylation: Effect of the Bite Angle. *Organometallics* **1995**, *14*, 3081–3089.
- [18] Decker, S. A.; Cundari, T. R. Hybrid QM/MM study of propene insertion into the Rh-H bond of HRh(PPh₃)₂(CO)(η^2 -CH₂=CHCH₃): The role of the olefin adduct in determining product selectivity. *J. Organomet. Chem.* **2001**, *635*, 132–141.
- [19] Decker, S. A.; Cundari, T. R. DFT Study of the Ethylene Hydroformylation Catalytic Cycle Employing a HRh(PH₃)₂(CO) Model Catalyst. *Organometallics* 2001, 20, 2827–2841.
- [20] Carbó, J. J.; Maseras, F.; Bo, C.; van Leeuwen, P. W. N. M. Unraveling the Origin of Regioselectivity in Rhodium Diphosphine Catalyzed Hydroformylation. A DFT QM/MM Study. J. Am. Chem. Soc. 2001, 123, 7630–7637.
- [21] Landis, C. R.; Uddin, J. Quantum mechanical modelling of alkene hydroformylation as catalyzed by xantphos-Rh complexesBased on the presentation given at Dalton Discussion No. 4, 10–13th January 2002, Kloster Banz, Germany. J. Chem. Soc. Dalt. Trans. 2002, 729–742.
- [22] Carvajal, M. A.; Kozuch, S.; Shaik, S. Factors Controlling the Selective Hydroformylation of Internal Alkenes to Linear Aldehydes. 1. The Isomerization Step. *Organometallics* 2009, 28, 3656–3665.
- [23] Gellrich, U.; Himmel, D.; Meuwly, M.; Breit, B. Realistic energy surfaces for real-world systems: An IMOMO CCSD(T):DFT scheme for rhodium-catalyzed hydroformylation with the 6-dppon ligand. *Chem. - A Eur. J.* 2013, *19*, 16272–16281.
- [24] Kumar, M.; Chaudhari, R. V.; Subramaniam, B.; Jackson, T. A. Ligand effects on the regioselectivity of rhodium-catalyzed hydroformylation: Density functional calculations illuminate the role of long-range noncovalent interactions. *Organometallics* 2014, 33, 4183– 4191.

- [25] Jacobs, I.; De Bruin, B.; Reek, J. N. Comparison of the full catalytic cycle of hydroformylation mediated by mono- and bis-ligated triphenylphosphine-rhodium complexes by using DFT calculations. *ChemCatChem* 2015, 7, 1708–1718.
- [26] Wodrich, M. D.; Busch, M.; Corminboeuf, C. Expedited Screening of Active and Regioselective Catalysts for the Hydroformylation Reaction. *Helv. Chim. Acta* **2018**, *101*.
- [27] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels–Alder Reactions and Pt(P(*t*-Bu)₃)₂ + H₂ Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357–19363.
- [28] Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. J. Mol. Struct. THEOCHEM 1999, 461-462, 1–21.
- [29] Rush, L. E.; Pringle, P. G.; Harvey, J. N. Computational Kinetics of Cobalt-Catalyzed Alkene Hydroformylation. *Angew. Chemie Int. Ed.* 2014, 53, 8672–8676.
- [30] Kumar, M.; Chaudhari, R. V.; Subramaniam, B.; Jackson, T. A. Importance of Long-Range Noncovalent Interactions in the Regioselectivity of Rhodium-Xantphos-Catalyzed Hydroformylation. *Organometallics* 2015, *34*, 1062–1073.
- [31] Kitchin, J. R. Machine learning in catalysis. *Nat. Catal.* **2018**, *1*, 230–232.
- [32] Labinger, J. A.; Bercaw, J. E. Understanding and exploiting C–H bond activation. *Nature* **2002**, *417*, 507–514.
- [33] Stahl, S. S.; Labinger, J. A.; Bercaw, J. E. Homogeneous Oxidation of Alkanes by Electrophilic Late Transition Metals. *Angew. Chemie Int. Ed.* **1998**, *37*, 2180–2192.
- [34] Hammad, L. A.; Gerdes, G.; Chen, P. Electrospray Ionization Tandem Mass Spectrometric Determination of Ligand Binding Energies in Platinum(II) Complexes. *Organometallics* 2005, 24, 1907–1913.
- [35] Hartwig, J. Organotransition Metal Chemistry: From Bonding to Catalysis; University Science Books: Sausalito, California, 2010; Vol. 2; pp 872–872.
- [36] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. 1993, 98, 5648–5652.
- [37] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- [38] Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822–8824.
- [39] Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

- [40] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [41] Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew-Burke-Ernzerhof exchangecorrelation functional. J. Chem. Phys. 1999, 110, 5029–5036.
- [42] Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. J. Chem. Phys. 1999, 110, 6158–6170.
- [43] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. 2010, 132, 154104.
- [44] Igel-Mann, G.; Stoll, H.; Preuß, H. Pseudopotential study of monohydrides and monoxides of main group elements K through Br. *Mol. Phys.* 1988, 65, 1329–1336.
- [45] Andrae, D.; Häußermann, U.; Dolg, M.; Stoll, H.; Preuß, H. Energy-adjusted ab initio pseudopotentials for the second and third row transition elements. *Theor. Chem. Acc.* 1990, 77, 123–141.
- [46] 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-Young, D.; 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, J. A., Jr.; 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. Gaussian16 [R]evision A.03, Gaussian Inc. Wallingford CT 2016.
- [47] Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013**, *138*, 034106.
- [48] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. J. Chem. Phys. 2013, 139, 134101.
- [49] Pinski, P.; Riplinger, C.; Valeev, E. F.; Neese, F. Sparse maps A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals. J. Chem. Phys. 2015, 143, 034108.
- [50] Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. J. Chem. Phys. 2016, 144.

- [51] Pavošević, F.; Peng, C.; Pinski, P.; Riplinger, C.; Neese, F.; Valeev, E. F. SparseMaps A systematic infrastructure for reduced scaling electronic structure methods. V. Linear scaling explicitly correlated coupled-cluster method with pair natural orbitals. J. Chem. Phys. 2017, 146.
- [52] Saitow, M.; Becker, U.; Riplinger, C.; Valeev, E. F.; Neese, F. A new near-linear scaling, efficient and accurate, open-shell domain-based local pair natural orbital coupled cluster singles and doubles theory. J. Chem. Phys. 2017, 146, 164105.
- [53] Neese, F. Software update: the ORCA program system, version 4.0. Wiley Interdiscip. Rev. Comput. Mol. Sci. 2018, 8, e1327.
- [54] Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. J. Chem. Phys. **1989**, *90*, 1007–1023.
- [55] Dunning Jr., T. H.; Peterson, K. A.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. J. Chem. Phys. 2001, 114, 9244–9253.
- [56] Peterson, K. A.; Figgen, D.; Dolg, M.; Stoll, H. Energy-consistent relativistic pseudopotentials and correlation consistent basis sets for the 4d elements Y-Pd. J. Chem. Phys. 2007, 126, 124101.
- [57] Figgen, D.; Peterson, K. A.; Dolg, M.; Stoll, H. Energy-consistent pseudopotentials and correlation consistent basis sets for the 5d elements Hf-Pt. J. Chem. Phys. 2009, 130, 164108.
- [58] Patel, P.; Wilson, A. K. Utilization of the Domain-Based Local Pair Natural Orbital Methods within the correlation consistent Composite Approach. **2019**, (*Submitted*).
- [59] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. J. Phys. Chem. B 2009, 113, 6378–6396.
- [60] Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.
- [61] Weymuth, T.; Couzijn, E. P.; Chen, P.; Reiher, M. New benchmark set of transition-metal coordination reactions for the assessment of density functionals. J. Chem. Theory Comput. 2014, 10, 3092–3103.
- [62] Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. J. Chem. Phys. 2003, 119, 12129–12137.
- [63] Brown, C. K.; Wilkinson, G. Homogeneous hydroformylation of alkenes with hydridocarbonyltris-(triphenylphosphine)rhodium(I) as catalyst. *J. Chem. Soc. A Inorganic, Phys. Theor.* **1970**, 2753.
- [64] van Leeuwen, P. W.; Clément, N. D.; Tschan, M. J.-L. New processes for the selective production of 1-octene. *Coord. Chem. Rev.* 2011, 255, 1499–1517.

- [65] Deshpande, R. M.; Divekar, S. S.; Gholap, R. V.; Chaudhari, R. V. Enhancement of rate and selectivity in hydroformylation of allyl alcohol through solvent effect. *Ind. Eng. Chem. Res.* 1991, 30, 1389–1390.
- [66] Carlock, J. T. A comparative study of triphenylamine, triphenylphosphine, triphenylarsine, triphenylantimony and triphenylbismuth as ligands in the rhodium-catalyzed hydroformylation of 1-dodecene. *Tetrahedron* **1984**, *40*, 185–187.
- [67] Borole, Y. L.; Chaudhari, R. V. New Route for the Synthesis of Propylene Glycols via Hydroformylation of Vinyl Acetate. *Ind. Eng. Chem. Res.* **2005**, *44*, 9601–9608.
- [68] Casey, C. P.; Whiteker, G. T.; Melville, M. G.; Petrovich, L. M.; Gavney, J. A.; Powell, D. R. Diphosphines with Natural Bite Angles near 120° Increase Selectivity for n-Aldehyde Formation in Rhodium-Catalyzed Hydroformylation. J. Am. Chem. Soc. 1992, 114, 5535– 5543.
- [69] Casey, C. P.; Petrovich, L. M. (Chelating diphosphine)rhodium-Catalyzed Deuterioformylation of 1-Hexene: Control of Regiochemistry by the Kinetic Ratio of Alkylrhodium Species Formed by Hydride Addition to Complexed Alkene. J. Am. Chem. Soc. 1995, 117, 6007–6014.

CHAPTER 6

VIBRATIONAL POTENTIAL ENERGY SURFACES WITH THE CORRELATION CONSISTENT COMPOSITE APPROACH AND DENSITY FUNCTIONAL THEORY

6.1 Introduction

Vibrational spectroscopy is one of the most useful techniques available to science due to its unique window into the structure, dynamical behavior, and bonding properties of molecules.¹ Modeling vibrational interactions is critical to understand infrared absorption, mechanisms, and kinetics of chemical reactions. Calculated frequencies can be utilized to predict thermodynamic properties like the enthalpy of formation and reaction barriers.^{2–6} As well, computational techniques are necessary to substantiate novel experiments where the resolution is insufficient or there are difficulties in isolating the molecule, i.e. diatomics, amino acids, or even short-lived molecules like the transition state of a chemical reaction and radicals.⁷

Computational chemistry techniques can also interpret and assign vibrational features to a specific molecule or type of motion. The increase of computational power in the digital age allows for the development of numerous methods for quantum mechanical modeling within the field, investigating deeper into the electronic structure of many-body systems, such as electronic properties, potential energies, and anharmonic vibrational frequencies.^{8–10} Electronic structure methods have been utilized to investigate vibrational frequencies with potential energy surfaces (PESs) describing dynamical motion, obtaining frequencies that are within several cm⁻¹ of experiment.^{9,11–14} These methods have a caveat, in which the accuracy attained for vibrational frequencies is exchanged for high computational cost, most prominent being disk space and CPU time.^{9,10} This is coupled with the number of grid points needed to generate the requisite PESs, which can easily exceed tens of thousands or even one million.^{11,12} Thus, the combination of electronic structure methods with the vast number of grid points needed for a PES reduces the feasibility of utilizing electronic structure methods for predicting vibrational properties within

several cm⁻¹. This allows for continual development towards creating computational methods with low computational cost while attaining vibrations predictions within several cm⁻¹ from well-established reliable experiments. Efficient processes have been developed to reduce the computational cost while yielding deviations from experimental vibrations by several cm⁻¹ or deviations less than 1 kcal mol⁻¹ for reaction barriers.^{11,15–17}

The correlation consistent Composite Approach $(ccCA)^{18,19}$ is considered, as a route to alleviate the cost of generating potential energy surfaces. ccCA has been utilized to describe PESs. For example, the multireference wavefunction ccCA (MR-ccCA) was utilized to analyze the potential energy curve of the torsional rotation of the carbon-carbon double bond in ethylene, predicting the barrier height of cis-trans isomerism and yielded errors approximately 0.7 kcal mol⁻¹ from experiment.¹⁶ As an alternative to utilizing multireference methods, completely renormalized coupled cluster (CR-CC(2,3)) was implemented within the ccCA formalism, CR-ccCA(2,3).¹⁷ This method utilizes a single reference completely renormalized coupled cluster that can correctly treat reaction pathways such as the thermal pericyclic rearrangement of bicyclo[1.1.0]butane to transbuta-1,3-diene and chemical species, e.g. diradicals, that would normally require multireference methods.

In addition to ccCA, DFT can be used to find properties of these electronic many-body systems at an affordable computational cost relative to the *ab initio* methods utilized for vibrational spectroscopy.²⁰ Density functionals have been largely designed for main group thermochemical properties; however, DFT cannot adequately describe noncovalent interactions, such as π - π stacking or weak hydrogen bonding, crucial in larger polyatomic molecules and could be significant when describing weakly-bound ligands and noncovalent interactions between two molecules.^{21,22}

Anharmonic vibrational frequencies properly characterize vibrational motion more than harmonic frequencies; however, these calculations can be quite expensive.^{23,24} While harmonic frequencies require the second derivative of the potential energy function, anharmonic frequencies require at least the third or fourth derivative to solve.²⁵ Generally, computational methods are

restricted to the harmonic approximation for vibrational frequencies that led to the development of empirical scale factors that can be tailored for high and low frequencies.²⁴ An analysis taking anharmonic effects into account would therefore lead to calculated frequencies that are not perfect overtones and provide a more accurate description of the vibrational behavior.^{26,27}

To account for anharmonicity computationally, computational strategies often include a perturbative correction, such as VPT2, to the potential. However, vibrational self-consistent field (VSCF) theory, which was developed in the late 1970s, fully accounts for anharmonicity by considering the vibrational Schrödinger Equation. Approximations are made which make VSCF theory analogous to Hartree-Fock theory.^{28–32} More recent studies utilizing VSCF theory implement on certain biologically pertinent vibrations for amino acid peptide chains, ^{11,33} which are not typically targeted with the rigorous *ab initio* methods utilized for potential energy surfaces (potentials) of diatomics and small polyatomic molecules like H₂O and formaldehyde. ^{12,34,35} In a study by Roy et al., a VSCF-PT2 approach was utilized with both a B3LYP-D2 potential and a multilevel HF/MP2 potential to characterize anharmonic vibrational motion of an opioid peptide [Ala², Leu⁵]-leucine enkephalin (ALE).¹¹ They found that the B3LYP and multilevel HF/MP2 potential systematically underestimated and overestimated the experimental frequencies for the OH and NH stretching modes for each amino acid, respectively, by a few tens of cm^{-1} . The average of the frequencies compensates for the respective under and overestimation of frequencies and yielded theoretical predictions within 10 cm^{-1} of experiment, which was better than the B3LYP and the multilevel HF/MP2 potentials individually as well as scaled harmonic calculations, thus showing the efficacy of VSCF theory towards predicting anharmonic vibrations for systems as large as a pentapeptide.

In this chapter, the correlation consistent Composite Approach (ccCA) and density functional theory (DFT) have been used to generate potential energy surfaces (PES) for diatomic and small polyatomic molecules to predict structural and vibrational properties such as frequencies and infrared absorbance intensities in tandem with vibrational self-consistent field (VSCF) and post-VSCF theory. Extrapolations schemes for ccCA and functional and basis set choice within DFT

were considered.

This was done to determine the efficacy of each method for each method. The combination of electronic structure methods such as ccCA and DFT with post-VSCF theory aims to reduce the computational cost associated with generating accurate PESs for anharmonic mode-mode couplings as well as calculating contributions from anharmonic corrections to the potential.

6.2 Computational Methods

The PVSCF program^{15,36} was used for vibrational analysis and ORCA 4.0 was used for the electronic structure calculations necessary to generate the potential energy surfaces.^{37,38} The molecule set included 20 molecules: H₂, CO, LiH, N₂, NO⁺, OH, NH, HF, BF, O₂, SiO, H₂O, CO₂, NH₃, C₂H₂, C₂H₄, C₂H₆, cis-3-aminophenol, and trans-3-aminophenol. These were chosen based on the availability of experimental frequencies for these molecules and their presence in the interstellar medium along with the notion of correlating the results with other studies that use more computationally demanding post-HF methods for predicting anharmonic frequencies.²² Experimental vibrational frequencies were obtained from Herzberg, Huber, and Shimanouchi.^{39–41} Equilibrium bond lengths for the diatomic molecules were obtained from CISD/cc-pVTZ calculations and the initial Hessians were generated using RI-B3LYP-D3/aug-cc-pVTZ within the ORCA package. Since the PVSCF program uses the hessian as an initial guess, a Hessian generated with a more approximate method is sufficient for the purpose of this study. For polyatomic molecules, B3LYP/cc-pVTZ geometries and hessians were used in accordance with the ccCA methodology.¹⁹

The potentials were generated via an interpolation of 16 grid points by a multimode expansion using curvilinear coordinates. Potential energy curves (PECs) are generated for diatomics while surfaces that plot the effect of two different vibrational modes concurrently vibrating, or vibrational mode coupling, are generated for polyatomic molecules. The extracted potential energies and dipole moments were then run with the PVSCF program to obtain the anharmonic frequencies and infrared (IR) intensities of each molecule, respectively. For diatomic molecules, a Fourier Grid Hamiltonian approach was used to calculate the single vibrational frequency.^{42,43} For all polyatomic molecules, a vibrational configuration interaction method (VCIPSI-PT2) was used to analyze the effects of vibrational mode coupling.⁴⁴ VCIPSI-PT2 utilizes vibrational configuration interaction with perturbatively selected interactions (VCIPSI), which reduces the computational cost compared to standard VCI approaches while maintaining the same level of accuracy.^{15,44}

6.2.1 DFT Calculations

Density functionals come in numerous flavors based on the number of parameters and the operations performed on the electronic density surface to varying degrees of success. For example, B3LYP^{45,46} is heavily parameterized for main group thermochemistry while TPSS⁴⁷ has no empirical parameters; yet both density functionals are popular and have been reported to yield low mean absolute errors for main group thermochemistry.⁴⁸ TPSS and B3LYP were therefore used as the density functionals in this work.

Dunning's standard and augmented correlation consistent basis sets from double- to quintuple- ζ (cc-pVnZ (VnZ) and aug-cc-pVnZ (aVnZ), n= D, T, Q, 5) were used.⁴⁹ These particular basis sets were built to systematically increase the types of energy contributions and subsequently, these types of functions included in the basis set, which leads to a smooth convergence of energetic properties towards an infinite basis set that would describe all possible space in which electrons exist. The Feller extrapolation scheme was used since this is a three-point extrapolation scheme, which allows the extrapolation function to converge to a limit closer to the experimental values, and uses the exponential form (Equation 2.29). The effects of the Feller extrapolation scheme were examined to provide insight into energies that would be obtained when using a more computationally demanding basis set (such as sextuple- ζ or higher). For polyatomic molecules with more than three atoms, only cc-pVTZ and aug-cc-pVTZ was used.

6.2.2 ccCA Calculations

The implementation of ccCA has been described in Section 2.2.4.1. Standard cartesian, or rectilinear, coordinates were used for diatomics and linear polyatomic molecules (CO₂, C₂H₂). Curvilinear coordinates were used for nonlinear polyatomic molecules (H₂O, aminophenol isomers). SCF energies were converged to 10^{-8} E_h in all single point energy calculations.^{18,19} ccCA electronic energies were used to generate all potential energy curves for singular vibrational motion and surfaces for mode-mode coupling, i.e. simultaneous vibrational motion for two vibrational modes.

For C₂H₄, C₂H₆, and aminophenol isomers, DFT calculations were used to generate all vibrational mode couplings. The number of vibrational mode couplings calculated with ccCA were decreased via a screening threshold to isolate strongly coupled vibrational modes.^{23,50} Selected vibrational modes from coupling maps are provided in the Appendix. The coupling strength is largely independent of the choice of method used to generate the potential energy surface. This is denoted as a FASTVCI approach in this work.

6.3 **Results and Discussion**

The calculated frequencies for diatomics, H₂O, CO₂, and NH₃ are shown in the Appendix. The mean absolute deviation (MAD) was analyzed by basis set, functional, and by number of atoms to note specific trends or certain occurrences within the calculations. For ccCA potentials, utilizing different extrapolation schemes did not significantly affect the predicted vibrational frequency as shown in Table 6.5. Therefore, for conciseness, only the frequencies predicted with ccCA-S4 potentials are presented as ccCA-S4 yielded the lowest errors of all extrapolations schemes utilized.

6.3.1 Diatomics

With DFT potentials at the complete basis set limit, the calculated frequencies yielded a MAD that ranged from 0 to 149 cm⁻¹ depending on the molecule and functional whereas with ccCA potentials, the calculated frequencies yielded a MAD that ranged from 0 to 22 cm⁻¹.

Examining functional choice, frequencies predicted with B3LYP-generated potential energy curves (PECs) had performed well compared to TPSS-generated PECs, with having seven molecules (H₂, BF, HF, NH, OH, O₂, and SiO) that yielded smaller MAD compared to TPSS. This is indicated in Figure 6.1 from TPSS yielding lower MADs for LiH, CO, N₂, and NO⁺ relative to B3LYP. For N₂, the predicted frequency generated with TPSS aligned with those predicted with ccCA-S4. This is plausible, due to B3LYP having parameters fitted. In addition, the deviation values ranged from 33 to 149 cm⁻¹ whereas B3LYP had a larger range of error of 42 to 178 cm⁻¹ when TPSS-generated potentials yielded lower errors for calculated frequencies than B3LYP-generated PECs. The larger range is due to the B3LYP/V ∞ Z PEC for LiH yielding a high error for calculated frequency relative to experiment. Based on the choice of molecule, functional choice had a larger effect on predicted vibrational frequency than basis set choice.



Figure 6.1: Mean absolute deviation (MAD) of vibrational frequencies for diatomics using TPSS/VTZ (blue), B3LYP/VTZ (green), TPSS/aVTZ (purple), B3LYP/aVTZ (red), and ccCA-S4 (black).

When comparing the effects of basis set choice holistically, augmented basis sets tended to produce lower MADs for experimental frequencies in comparison to non-augmented basis sets. This is evident as the MAD decreased from 43 ± 37 and 39 ± 35 for B3LYP/VnZ and B3LYP/aVnZ, respectively, and 57 ± 52 and 55 ± 50 for TPSS/VnZ and TPSS/aVnZ, respectively. This could be due to the extra diffuse function presented in aVnZ basis sets that can better describe a larger internuclear distance for diatomics. From the supplemental tables (Tables 6.6-6.7), using a larger basis set could lead to more accurate predictions as the mean error when using VDZ across all molecules and functionals was 73 ± 10 cm⁻¹ whereas the mean error when using VTZ, VQZ, and V5Z across all molecules and functionals was 44 ± 11 , 42 ± 10 , and 41 ± 10 cm⁻¹, respectively. Although there is an ~30 cm⁻¹ decrease in error between VDZ and the higher ζ -level basis sets,

there is not a consistent lowering of error in calculated frequency with respect to an increase in basis set size as the error in frequency for all DFT/VTZ, DFT/VQZ and DFT/V5Z potentials were statistically not different based on the 25% relative uncertainty in deviations across all molecules and functionals. Therefore, triple- ζ quality basis sets may be useful as a compromise between cost and accuracy for generating potentials describing vibrational motion for polyatomic molecules.

Diatomics that exhibit covalent triple bonds (CO, NO⁺, and N₂) yielded lower deviations from experimental frequencies with TPSS (31, 12, 8 cm⁻¹, respectively) while diatomics with single or double bonds yielded lower deviations from experiment with B3LYP (43, 93, 109 cm⁻¹, respectively). For the diatomics with covalent triple bonds, the parameterization within B3LYP and the increase in electron density between atoms may be the cause of the higher deviations in calculated frequency from experiment obtained with B3LYP-generated PECs opposed to TPSSgenerated PECs. HF, SiO, BF, and OH are molecules for which B3LYP yielded smaller deviations from experiment, and have a larger electron density around the more electronegative atom, which indicates that B3LYP is preferred when calculating polar molecules.

Across all diatomics examined, ccCA yielded a MAD of $9 \pm 7 \text{ cm}^{-1}$ whereas B3LYP/VnZ and TPSS/VnZ yielded MADs of 43 ± 37 and 57 ± 52 , respectively. This is shown in Figure 6.1. This indicates that using potentials generated with ccCA yield lower deviations for predicted frequencies than DFT, with the notable exception of SiO, where using B3LYP regardless of basis set yielded frequencies closer to experiment than ccCA by approximately 20 cm⁻¹. This may be due to the presence of static correlation as the molecule vibrates. With ccCA, diatomics that have a larger difference in mass between the two atoms (OH, BF, HF) tended to yield lower deviations from experiment for calculated frequencies with the exception of NH, which yielded an error of 16 cm⁻¹. PECs for homonuclear diatomics yielded higher deviations with an increase in mass as H₂, N₂, and O₂, yielded errors of 5, 10, and 11 cm⁻¹, respectively. As well, PECs for LiH and NO⁺, heteronuclear diatomics with similar mass between atoms, tended to yield higher errors (8 and 11 cm⁻¹) among the ccCA results although PECs for CO and BF yielded errors of 0 and 2 cm⁻¹, respectively. This would suggest that unlike DFT, there is no consistent trend between molecular weight or atom type present and frequencies generated with ccCA PECs even though PECs at the ccCA level of theory generates lower errors across all diatomics compared to DFT PECs.

The CPU time was measured with a Dell OptiPlex 390 with 16 GB DDR3 memory for a few diatomics to show the approximate cost of generating a PEC of 17 grid points with DFT, ccCA, and CCSD(T,full)/aug-cc-pCV5Z where *full* denotes the inclusion of all electrons for first-row main group diatomics in the correlation space. This is shown in Table 6.1. As expected, B3LYP calculations are more computationally affordable than *ab initio* methods but yielded higher deviations for the molecules used (Table 6.2). Also, generating a PEC at the ccCA level, which aims to model energies at the CCSD(T,full)/aug-cc-pCV ∞ Z-DK level, is more affordable than using CCSD(T,full)/aug-cc-pCV5Z by several hours depending on the molecule size. As shown in Table 6.1 for example, ccCA yielded a percent CPU time savings of 99.16 % for N₂ relative to CCSD(T,full)/aug-cc-pCV5Z. Interestingly, the use of CCSD(T,full)/aug-cc-pCV5Z to generate the PES yielded larger absolute errors (10 ± 3) than ccCA (6 ± 4) for these molecules. The higher absolute errors and large increase in CPU time between CCSD(T,full)/aug-cc-pCV5Z and ccCA suggests that potentials generated with ccCA are more accurate when using a VSCF approach.

	B3LYP/aVTZ	ccCA
H_2	98.81	90.66
LiH	99.89	98.82
CO	99.97	99.01
N_2	99.98	99.16

Table 6.1: Percent CPU Time relative to CCSD(T,full)/aug-cc-pCV5Z to generate all 17 grid points of the PEC for select diatomics.

	B3LYP/aVTZ	ccCA	CCSD(T,full)/aug-cc-pCV5Z	Exp
H ₂	4189	4157	4155	4162.2
LiH	1372	1368	1348	1360
CO	2186	2143	2128	2143
N ₂	2420	2340	2323	2330
$\mathrm{MAD}\pm\mathrm{STD}$	43 ± 29	6 ± 4	10 ± 3	

Table 6.2: Calculated frequencies in cm^{-1} for B3LYP/aug-cc-pVTZ, ccCA, and CCSD(T, full)/aug-cc-pCV5Z for diatomics in Table 6.1.

6.3.2 H₂O, CO₂, NH₃

For DFT potentials, both functional choice and basis set quality affected the accuracy of predicted frequencies and the type of vibration that is observed, i.e. stretching, bending, or inversion. For ccCA, the range in deviation is primarily due to the type of vibration observed. In Figure 6.2, the DFT potentials are obtained at the CBS limit ($V\infty Z$ and $aV\infty Z$) for H₂O and CO₂ since there are only 3 and 4 vibrational normal modes, respectively, which leads to the calculation of 3 and 7 vibrational mode-mode couplings, i.e. the surface describing two vibrations occurring simultaneously. H₂O, CO₂, and NH₃ yielded a larger range of deviations from experimental frequencies than the diatomics, wavering from 2 to 294 cm⁻¹ with DFT potentials and from 2 to 57 cm⁻¹ for ccCA potentials. For NH₃, the vibrational mode-mode coupling potentials were generated at the triple- ζ level based on the cost-to-accuracy ratio for the diatomic molecules, the observation of results from H₂O and CO₂, and the number of vibrational mode-mode coupling potentials consist of 256 grid points.



Figure 6.2: MAD of vibrational frequencies for H₂O, CO₂, and NH₃ using TPSS/V*nZ* (blue), B3LYP/V*nZ* (green), TPSS/aV*nZ* (purple), B3LYP/aV*nZ* (red), and ccCA-S4 (black). For H₂O and CO₂, $n = \infty$. For NH₃, n=T.

In Figure 6.2, the MAD is the average across all frequencies for a particular molecule for each method (B3LYP/vnZ, B3LYP/avnZ, TPSS/vnZ, TPSS/avnZ, and ccCA). For DFT, the choice between avnZ and vnZ altered the curvature of the potentials enough to yield larger variations in the errors for calculated frequencies between the molecules with the exception of H₂O. When avnZ was used in H₂O, the error across all vibrations was the same as $v\infty Z$, regardless of functional choice. For CO₂, using $av\infty Z$ increased the mean error for all frequencies by 13 cm⁻¹ relative to using $v\infty Z$ for TPSS and decreased the mean error by 15 cm⁻¹ for B3LYP. For NH₃, using aVTZ for the potentials decreased the mean error across all frequencies by ~50 cm⁻¹ relative to using VTZ for both TPSS and B3LYP. This indicates that when using DFT to generate potentials for vibrational calculations, augmented basis sets properly characterize both bending and stretching

behavior for small polyatomic species.

Functional choice was a larger factor in terms of general curvature of the potentials as B3LYPgenerated potentials yielded deviations for calculated frequency approximately 20-60 cm⁻¹ lower than TPSS-generated potentials. For H₂O, CO₂, and NH₃, TPSS potentials inadequately described both symmetric and asymmetric stretching modes with errors ranging from 70 to 225 cm⁻¹, whereas B3LYP potentials yielded errors in the range of 3 to 55 cm⁻¹. In comparison, the bending vibrational modes yielded errors ranging from 19 to 42 cm⁻¹ for TPSS potentials. It is plausible that in addition to parameterization for main group species, B3LYP potentials produced results closer to experimental data because B3LYP includes exact exchange. The results infer that B3LYP is preferred in generating potentials for calculating frequencies of polyatomic molecules when utilizing DFT to generate potentials for vibrational motion.

With ccCA, the stretching modes for H₂O yielded deviations within 2 cm⁻¹ of experiment, whereas the calculated bending mode was $\sim 20 \text{ cm}^{-1}$ larger than the experimental frequency. The difference of 20 cm⁻¹ is most likely due to the weak coupling between the bending and stretching modes and may be corrected through coupling all three vibrational modes together simultaneously. For CO₂, ccCA potentials did not properly characterize the vibrational motion of the out-of-plane bending, and stretching normal modes with deviation of 40, 37, and 56 cm⁻¹, respectively. This may be in part due to the use of a standard cartesian coordinate system for displacing the molecule in vibration whereas the deviations are generally lower when using a curvilinear coordinate system as is the case for H₂O. For NH₃, ccCA potentials utilized for VCIPSI-PT2 predicted the inversion barrier to within 10 cm⁻¹ and N-H stretching modes within 15 cm⁻¹. This analysis indicates that when using a curvilinear coordinate system, ccCA potentials yield lower errors for stretching modes opposed to bending modes, such as for H₂O.

6.3.3 Hydrocarbons

This section highlights bonding character and its effect on vibrational potentials generated with DFT and ccCA through analyzing C_2H_2 , C_2H_4 , and C_2H_6 . Potential energy surfaces were

generated with basis set superposition error (BSSE)-corrected energies. With ccCA, only strongly coupled vibrational modes are considered for C_2H_4 and C_2H_6 , which is depicted via coupling maps in the Appendix.



Figure 6.3: MAD of vibrational frequencies for C₂H₂, C₂H₄, and C₂H₆ using TPSS/VTZ (blue), B3LYP/VTZ (green), TPSS/aVTZ (purple), B3LYP/aVTZ (red), and ccCA-S4 (black).

For the hydrocarbons, there is a noticeable improvement among frequencies predicted with DFT potentials. TPSS provided a smaller absolute deviation in ethyne, while B3LYP was preferred for ethene and ethane. The number of C-H bonds largely affected deviations from experimental frequencies. Findings regarding the basis set choice were consistent with other molecules examined thus far in that DFT/aVTZ potentials yielded lower errors for calculated frequencies relative to using DFT/VTZ potentials when using VCIPSI-PT2 to compute the frequencies.

Over all observed frequencies, DFT potentials yield lower MADs from experimental frequencies

than ccCA. For C₂H₂, the difference in magnitude between total MADs of frequencies predicted with DFT and ccCA potentials was approximately 5 cm^{-1} . Yet for C₂H₄ and C₂H₆, this difference increases to 10-15 cm⁻¹. This is primarily due to the large deviations in the C-H stretching vibrations around 3000 cm⁻¹ for ccCA potentials as well as the number of strongly coupled vibrational modes that include a non-IR active symmetric C-H stretching mode used for the FASTVCI approach. This is also consistent with other composite strategies that utilize perturbative anharmonic corrections.⁵¹ This would suggest that for molecules like C₂H₄ and C₂H₆, a softer potential as generated via DFT for the non-IR active symmetric stretches is more characteristic of the vibrational motion.

With ccCA potentials, the C-C stretching mode for C_2H_2 , C_2H_4 , and C_2H_6 yielded errors for predicted frequencies with VCIPSI-PT2 of 17, 21, and 7 cm⁻¹, respectively, all of which are IR inactive. This would suggest that ccCA potentials are more adequate for describing vibrations involving covalent single bonds than double or triple bonds. This is also supported in part by the low deviations observed for H₂O and NH₃ and high deviations for CO₂ observed for stretching modes.

To correct for this discrepancy between frequencies generated with DFT potentials and ccCA potentials, a multilevel approach can be utilized where the coupling elements from ccCA potentials can be added to the frequencies generated via the uncoupled DFT potentials for each vibration. This is denoted as DFT:ccCA in this work. To illustrate this concept for ethene, where the shape of the single mode PECs for both the individual non-IR active C=C and C-H symmetric stretches differ between ccCA and DFT (Figure 6.5), the frequencies are obtained where the mode-mode coupling elements from ccCA potentials are applied to the single mode PECs generated with TPSS/VTZ. As shown in Table 6.3, when using TPSS single mode PECs in tandem with ccCA mode-mode coupling potentials, the error decreases to 24 cm^{-1} from 57 cm⁻¹ when using ccCA single mode and mode-mode coupling potentials, respectively. While the difference between using TPSS and ccCA mode-mode coupling potentials with single mode PECs was only 1 cm⁻¹, ccCA mode-mode coupling potentials lowered the predicted frequency relative to using TPSS mode-

mode coupling potentials, which lowered the deviation as TPSS potentials overestimated the C-H stretching modes (modes 9-12 in Table 6.3). Overall, a multilevel approach may be useful when one method generates a single mode PEC more representative of the vibrational motion indicated by the predicted frequency and for larger polyatomic systems.

Mode	Exp	TPSS:PSS All	TPSS:TPSS FASTVCI	ccCA-S4:ccCA-S4 FASTVCI	TPSS:ccCA-S4 FASTVCI
1	826	821	827	834	827
2	949	976	988	975	991
3	943	963	981	979	977
4	1023	1046		_	
5	1236	1229	1238	1233	1239
6	1342	1354	1363	1360	1363
7	1444	1455	1468	1462	1467
8	1623	1628	1636	1649	1652
9	2989	3030	3050	3143	3043
10	3026	2992	3004	3104	2998
11	3103	3087	3113	3221	3105
12	3106	3118	3145	3247	3132
MAD		18	25	57	24

Table 6.3: VCIPSI-PT2 frequencies using a combination of TPSS and ccCA for single mode and vibrational mode-mode coupling potentials. The use of PECs/PESs is denoted as single:coupled.

Mode 4 did not strongly couple to any other vibrational mode and hence excluded from FASTVCI calculations.

6.3.4 Aminophenol

For cis-3-aminophenol and trans-3-aminophenol, the NH₂ torsion, 318.5 cm⁻¹ and 329 cm⁻¹, respectively, and OH wagging vibrations, 307 cm⁻¹ and 316 cm⁻¹, respectively were examined.⁸ Each chosen vibration was coupled to all 38 other normal modes for DFT calculations. With ccCA, only strongly coupled vibrational modes determined through the DFT calculations were included in vibrational analysis. Coupling maps depicting strongly coupled vibrational modes are included in the Appendix.

Calculated frequencies obtained with ccCA potentials yielded a lower deviation for experimental NH₂ torsion and OH wagging vibrational modes than frequencies obtained with DFT potentials.

This is shown in Table 6.4. B3LYP/aVTZ yields lower deviations than TPSS/aVTZ for both the NH₂ torsion and OH wagging vibrational modes for both cis-3-aminophenol and trans-3-aminophenol. For ccCA potentials the NH₂ torsional mode was better characterized for cis-3-aminophenol with an error of 5.5 cm^{-1} and the OH wagging motion was better characterized for trans-3-aminophenol with an error of 1 cm^{-1} . While the deviations obtained with ccCA potentials are lower than 10 cm^{-1} for trans-3-aminophenol, this approach can be utilized to spectroscopically differentiate between aminophenol isomers that differ by the direction of the OH bond relative to the NH₂ substituent.

Table 6.4: Vibrational frequencies predicted with VCIPSI-PT2 for selected vibrations of cis-3-aminophenol and trans-3-aminophenol.

	Exp	B3LYP/aVTZ	TPSS/aVTZ	ccCA-S4
cis-3-aminophenol				
NH ₂ torsion	318.5	322	345	313
OH wag	307	300	296	275
trans-3-aminophenol				
NH ₂ torsion	329	333	351	322
OH wag	316	330	325	317

When using a multilevel approach for the aminophenol isomers, utilizing ccCA single mode PECs and B3LYP/aVTZ mode-mode coupling potentials (ccCA:B3LYP) for all mode-mode couplings between the NH₂ torsion and all other vibrations as well as between the OH wagging and all other vibrations (75 total mode-mode couplings) yielded lower deviations than if only ccCA mode-mode coupling potentials were used for cis-3-aminophenol for the few strongly coupled modes isolated. The deviations for both the NH₂ torsion and OH wagging decreased by 2 cm⁻¹. For trans-3-aminophenol, the use of this multilevel approach increased the deviation by 2 cm⁻¹. This may be in part due to how the predicted frequencies using B3LYP/aVTZ potentials were higher than those predicted with ccCA potentials.

The computed infrared (IR) spectra uses the frequencies generated via VCIPSI-PT2 for potentials generated with ccCA and DFT. To show the efficacy of the VCIPSI-PT2 predictions, the

generated spectra is compared to harmonic B3LYP/cc-pVTZ frequencies scaled by 1.0066 per the study by Merrick et al.²⁴ The intensities for all frequencies are based on the harmonic calculation. For cis-3- and trans-3-aminophenol, the most intense peaks from harmonic intensities were near 600 cm⁻¹, whereas the most intense experimental peaks were at 307 and 755 cm⁻¹, respectively. The computed spectra shows peaks in the 200-300 cm⁻¹ range indicating torsional motion among the C atoms in the ring. For the NH₂ torsion and OH wagging motions for cis-3-aminophenol, the VCIPSI-PT2 frequencies with ccCA potentials were more closely aligned to experiment than both the scaled harmonic frequencies and VCIPSI-PT2 frequencies with potentials generated with B3LYP/aVTZ. Even with intensities generated via the harmonic frequency calculation, the frequencies obtained via the ccCA potentials with VCIPSI-PT2 yielded a more accurate representation of the spectra than the scaled harmonic frequencies. This would suggest that the computed IR spectra would be more representative of the experimental IR spectra with a full description of the mode-mode couplings.



Figure 6.4: Infrared spectra for cis-3-aminophenol (top) and trans-3-aminophenol (bottom) obtained with VCIPSI-PT2 frequencies with ccCA potentials and B3LYP/cc-pVTZ harmonic frequencies scaled by 1.0066. All intensities are from the harmonic frequency calculations. A Lorentz broadening of 20 cm⁻¹ was applied. The experimental frequencies and relative intensities from Ref 8 are shown for comparison.

6.4 Conclusions

Overall, with ccCA potentials, the mean absolute deviation for calculated frequency from experiment was lower than with DFT potentials. Functional choice had a more significant effect on the predicted frequency than basis set for potentials generated with DFT. For diatomics, TPSS potentials tended to properly characterize molecules that exhibit covalent triple bonds and B3LYP potentials tended to yield lower absolute errors in frequency from experiment for polar molecules. This trend held with the small polyatomics and hydrocarbons. When considering timings, ccCA yielded lower deviations than CCSD(T,full)/aug-cc-pCV5Z with up to 99% CPU time savings for diatomic molecules.

For H_2O , CO_2 , and NH_3 , the predicted frequencies between potentials generated with ccCA and DFT yielded similar errors across all vibrations. DFT predicted the bending behaviors better than ccCA whereas ccCA predicted the stretching behaviors better than DFT. The use of a curvilinear coordinate system yielded lower errors relative to using a standard Cartesian coordinate system as indicated by the deviations observed for H_2O and CO_2 .

For hydrocarbons, DFT characterized the C-H stretching behavior better than ccCA as the errors for DFT potentials were lower than for ccCA potentials for C-H stretching modes. Trends observed for all molecules examined indicate that B3LYP/aVTZ is the more favorable DFT method and basis set combination in terms of generating a PES for vibrational motion when coupled with VCIPSI-PT2 to compute frequencies. A multilevel approach that utilizes the single mode PECs with DFT and the coupled vibrational modes generated with ccCA yields lower frequencies than if only DFT were utilized. This is useful for expanding to larger polyatomic systems and for when one method generates PECs that yield lower deviations than another, as was the case with the hydrocarbons.

For aminophenol, the errors obtained with VCIPSI-PT2 were lower than those for scaled harmonics, indicating the success of utilizing this approach to characterize specific vibrations for polyatomic systems. The FASTVCI approach of only utilizing strongly coupled vibrational modes saves computational resources when generating potentials with electronic structure methods.

B3LYP potentials serve as a good approximation for both the NH₂ torsional mode and OH wagging mode for both the cis-3-aminophenol and trans-3-aminophenol isomers. For cis-3-aminophenol, the ccCA potentials yielded lower deviations for the NH₂ torsion where the opposite is true for trans-3-aminophenol.

Overall, *ab initio* composite strategies and in some cases DFT can be utilized for depicting vibrational behavior of small polyatomic molecules present in the interstellar medium and can be used in tandem with post-VSCF theory as a gauge for predicting anharmonic vibrations without the harmonic frequencies with frequency scaling factors applied and perturbative corrections.

APPENDIX

APPENDIX

	ccCA-P	ccCA-S3	ccCA-S4	ccCA-PS3	Exp ⁴⁰
Diatomics					
H_2	4156	4157	4157	4157	4162
LiH	1372	1370	1368	1371	1360
CO	2143	2144	2143	2144	2143
N_2	2340	2341	2340	2340	2330
BF	1377	1377	1376	1377	3570
HF	3960	3961	3960	3960	2344
NH	3142	3144	3142	3143	1379
NO ⁺	2355	2356	2355	2355	3961
OH	3566	3567	3566	3566	3126
O ₂	1591	1593	1591	1592	1556
SiO	1220	1221	1220	1220	1242
Small Polyato	mics				
	1616	1616	1616	1616	1595
H ₂ O	3660	3661	3660	3660	3657
	3758	3760	3758	3759	3756
	707	708	707	708	667
CO_2	676	676	675	676	667
002	1296	1297	1296	1297	1333
	2292	2293	2292	2293	2349
	939	939	940	939	950
	1613	1613	1613	1613	1626
NH3	1606	1606	1606	1606	1626
	3343	3343	3342	3343	3336
	3455	3456	3455	3455	3443
	3458	3459	3458	3459	3443

Table 6.5: Calculated frequencies of diatomic and small polyatomic molecules in cm^{-1} obtained with ccCA potentials.
	VDZ	VTZ	VQZ	V5Z	V∞Z	Exp ⁴⁰
TPSS/cc-pV	$n\mathbf{Z}$					
H_2	4193	4196	4196	4196	4191	4162
LiH	1333	1359	1357	1360	1366	1360
CO	2100	2111	2113	2112	2111	2143
N_2	2343	2343	2338	2340	2342	2330
BF	1240	1331	1333	1333	1332	1379
HF	3774	3810	3813	3813	3812	3961
NH	3051	3085	3089	3091	3092	3126
NO ⁺	2344	2348	2349	2352	2352	2344
OH	3371	3451	3456	3459	3460	3570
O ₂	1509	1501	1509	1510	1510	1556
SiO	1151	1196	1203	1204	1200	1242
B3LYP/cc-p	$\mathbf{V}n\mathbf{Z}$					
H_2	4148	4190	4187	4188	4189	4162
LiH	1347	1367	1365	1370	1538	1360
CO	2187	2176	2186	2188	2185	2143
N_2	2427	2423	2421	2421	2423	2330
BF	1322	1386	1382	1380	1378	1379
HF	3848	3918	3915	3911	3905	3961
NH	3034	3108	3116	3122	3129	3126
NO ⁺	2463	2459	2452	2453	2453	2344
OH	3473	3545	3551	3554	3555	3570
O ₂	1610	1595	1603	1604	1603	1556
SiO	1194	1234	1247	1247	1240	1242

Table 6.6: Calculated frequencies of diatomics in cm^{-1} obtained with TPSS/cc-pVnZ and B3LYP/cc-pVnZ potentials.

Molecule	aVDZ	aVTZ	aVQZ	aV5Z	aV∞Z	Exp ⁴⁰
TPSS/aug-c	c-pVnZ					
H_2	4181	4195	4196	4196	4195	4162
LiH	1328	1361	1360	1359	1359	1360
CO	2086	2108	2111	2112	2112	2143
N_2	2335	2341	2339	2340	2342	2330
BF	1240	1331	1333	1333	1332	1379
HF	3774	3810	3813	3813	3812	3961
NH	3051	3085	3088	3090	3092	3126
NO ⁺	2344	2348	2349	2352	2352	2344
OH	3424	3453	3458	3459	3460	3570
O ₂	1514	1510	1500	1511	1510	1556
SiO	1156	1203	1201	1205	1205	1242
B3LYP/aug	-cc-p $\mathbf{V}n\mathbf{Z}$					
H_2	4137	4189	4188	4189	4190	4162
LiH	1342	1372	1374	1370	1226	1360
CO	2160	2181	2186	2187	2187	2143
N_2	2416	2420	2421	2422	2423	2330
BF	1282	1375	1379	1380	1379	1379
HF	3880	3902	3907	3908	3961	3961
NH	3092	3118	3122	3123	3124	3126
NO ⁺	2444	2448	2452	2453	2453	2344
OH	3530	3545	3552	3554	3570	3570
O ₂	1612	1605	1594	1605	1604	1556
SiO	1197	1247	1243	1247	1242	1242

Table 6.7: Calculated frequencies of selected diatomics in cm^{-1} with TPSS/aug-cc-pVnZ and B3LYP/aug-cc-pVnZ potentials.

Table 6.8: Calculated vibrational frequencies for H_2O and CO_2 in cm⁻¹ utilizing VCIPSI-PT2 with TPSS/cc-pVnZ potentials.

Molecule	Mode	VDZ	VTZ	VQZ	V5Z	$V\infty Z$	Exp ⁴¹
H ₂ O	1	1614	1615	1614	1614	1614	1595
	2	3515	3540	3546	3547	3548	3657
	3	3612	3630	3635	3637	3638	3756
CO ₂	1	634	641	643	642	642	667
	2	634	641	643	642	642	667
	3	1334	1348	1350	1350	1349	1333
	4	2288	2302	2304	2303	2303	2349

Molecule	Mode	VDZ	VTZ	VQZ	V5Z	$V\infty Z$	Exp ⁴¹
H ₂ O	1	1643	1613	1605	1599	1599	1595
	2	3576	3635	3641	3645	3645	3657
	3	3670	3723	3730	3734	3734	3756
	1	650	667	670	670	669	667
CO	2	650	667	670	670	669	667
CO_2	3	1381	1399	1401	1401	1399	1333
	4	2368	2365	2357	2353	2351	2349

Table 6.9: Calculated vibrational frequencies for H_2O and CO_2 in cm⁻¹ utilizing VCIPSI-PT2 with B3LYP/cc-pVnZ potentials.

Table 6.10: Calculated vibrational frequencies for H_2O and CO_2 in cm⁻¹ utilizing VCIPSI-PT2 with TPSS/aug-cc-pVnZ potentials.

Molecule	Mode	aVDZ	aVTZ	aVQZ	aV5Z	aV∞Z	Exp ⁴¹
H ₂ O	1	1614	1615	1614	1614	1614	1595
	2	3515	3540	3546	3547	3548	3657
	3	3612	3630	3635	3637	3638	3756
	1	591	628	636	636	636	667
CO	2	591	628	636	636	636	667
CO_2	3	1311	1319	1319	1318	1318	1333
	4	2531	2466	2444	2439	2438	2349

Table 6.11: Calculated vibrational frequencies for H_2O and CO_2 in cm⁻¹ utilizing VCIPSI-PT2 with B3LYP/aug-cc-pVnZ potentials.

Molecule	Mode	aVDZ	aVTZ	aVQZ	aV5Z	aV∞Z	Exp ⁴¹
H ₂ O	1	1598	1599	1599	1599	1599	1595
	2	3626	3635	3642	3644	3644	3657
	3	3724	3725	3731	3733	3733	3756
	1	660	669	670	669	669	667
CO	2	660	669	670	669	669	667
CO_2	3	1383	1398	1401	1400	1400	1333
	4	2336	2349	2353	2353	2352	2349



Figure 6.5: Single mode potential energy curves for vibrational modes 8 (left) and 10 (right) of ethene (C=C and C-H symmetric stretches) generated with ccCA (black) and TPSS/VTZ (red).

Table 6.12:	Calculated	vibrational fr	equencies for	NH_3 in cm ⁻¹	utilizing	VCIPSI-PT2	with t	ooth
TPSS and E	B3LYP poter	ntials with the	VTZ and aV	ΓZ basis sets.				

Mode	B3LYP/VTZ	B3LYP/aVTZ	TPSS/VTZ	TPSS/aVTZ	Exp ⁴¹
1	994	886	1039	960	950
2	1632	1576	1659	1598	1626
3	1653	1596	1668	1659	1626
4	3391	3327	3418	3230	3336
5	3474	3441	3619	3357	3443
6	3441	3450	3668	3373	3443



Figure 6.6: Vibrational coupling map for ethene (left) and ethane (right). The vibrational mode couplings shown in black indicate strongly coupled modes that were used for all FASTVCI approaches using ccCA.

Measures to reduce the computational cost includes screening out weakly coupled pair-wise coupling interactions via a threshold established from calculating the coupling strength (Equation 2.66), which can be calculated with only the VSCF potential.^{23,50} By removing non-essential vibrational coupling elements from the potential, a FASTVSCF approach is attained. By utilizing this approach, the computational time to generate all vibrational potential energy surfaces is reduced by approximately a factor of 6 (12 out of 78 coupling modes were calculated with ccCA) for ethene as only the potential energy surfaces for all of the shaded squared were generated with ccCA. For ethene, modes 9-12 characterize the C-H stretching modes (both symmetric and asymmetric). In general, symmetric and asymmetric vibrations are strongly coupled largely due to the effect each type of vibration has on the other.

In contrast to ethene, which exhibited stronger coupling modes for the C-H stretches at approximately 3000 cm^{-1} , ethane only had one strong coupling mode in this region, which is the coupling between C-H symmetric and C-H asymmetric stretches. Other modes that were strongly coupled include the rotational barrier of ethane around the C-C bond (mode 1) and C-C stretching

(mode 9). Including the coupling strength screened out 67 vibrational mode, leaving 11 shown in black in Figure 6.6. This effectively reduced the computational cost of generating the full 2D surface by approximately 93%.



Figure 6.7: Vibrational coupling map for cis-3-aminophenol (left) and trans- 3-aminophenol (right). The vibrational mode couplings shown in black indicate strongly coupled modes that were used for all FASTVCI approaches using ccCA.

For cis-3-aminophenol, only the OH wagging (mode 3) and NH₂ torsion (mode 5) were analyzed. Therefore, all vibrations that coupled to these modes were included. In terms of coupling strength, only 6 of the 78 mode-mode coupling potentials were analyzed with ccCA, again reducing the cost by approximately 92%. For trans-3-aminophenol, only the OH wagging (mode 4) and NH₂ torsion (mode 5) were analyzed. Therefore, all vibrations that coupled to these modes were included. In terms of coupling strength, only 12 of the 78 mode-mode coupling potentials were analyzed with ccCA, reducing the cost of generating the full 2D surface by approximately 84%.

REFERENCES

REFERENCES

- [1] Pretsch, E.; Buhlmann, P.; Badertscher, M. *Structure Determination of Organic Compounds*; 2009; p 443.
- [2] Bron, J. The Importance of Anharmonicity of the Vibrational Excited States in Chemical Kinetics. *Can. J. Chem.* **1975**, *53*, 3069.
- [3] Bakker, J. M.; Aleese, L. M.; Meijer, G.; von Helden, G. Fingerprint IR Spectroscopy to Probe Amino Acid Conformations in the Gas Phase. *Phys. Rev. Lett.* **2003**, *91*, 203003.
- [4] Kawaguchi, K. In *Handb. Vib. Spectrosc.*; Chalmers, J. M., Ed.; John Wiley & Sons, Ltd: Chichester, UK, 2006.
- [5] Barth, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta Bioenerg.* **2007**, *1767*, 1073–1101.
- [6] Almond, M. J.; Jenkins, S. L. *Encycl. Inorg. Bioinorg. Chem.*; John Wiley & Sons, Ltd: Chichester, UK, 2011.
- [7] Reichenbächer, M.; Popp, J. Challenges in Molecular Structure Determination; 2012.
- [8] Yatsyna, V.; Bakker, D. J.; Feifel, R.; Rijs, A. M.; Zhaunerchyk, V. Aminophenol isomers unraveled by conformer-specific far-IR action spectroscopy. *Phys. Chem. Chem. Phys.* 2016, 18, 6275–6283.
- [9] Roy, T. K.; Gerber, R. B. Vibrational self-consistent field calculations for spectroscopy of biological molecules: New algorithmic developments and applications. *Phys. Chem. Chem. Phys.* 2013, 15, 9468–9492.
- [10] Bloino, J.; Baiardi, A.; Biczysko, M. Aiming at an accurate prediction of vibrational and electronic spectra for medium-to-large molecules: An overview. *Int. J. Quantum Chem.* 2016, *116*, 1543–1574.
- [11] Roy, T. K.; Kopysov, V.; Pereverzev, A.; Šebek, J.; Gerber, R. B.; Boyarkin, O. V. Intrinsic structure of pentapeptide Leu-enkephalin: geometry optimization and validation by comparison of VSCF-PT2 calculations with cold ion spectroscopy. *Phys. Chem. Chem. Phys.* 2018,
- [12] Coles, P. A.; Ovsyannikov, R. I.; Polyansky, O. L.; Yurchenko, S. N.; Tennyson, J. Improved potential energy surface and spectral assignments for ammonia in the near-infrared region. J. Quant. Spectrosc. Radiat. Transf. 2018, 219, 199–212.
- [13] Bulik, I. W.; Frisch, M. J.; Vaccaro, P. H. Vibrational self-consistent field theory using optimized curvilinear coordinates. J. Chem. Phys. 2017, 147.

- [14] Knaanie, R.; Šebek, J.; Tsuge, M.; Myllys, N.; Khriachtchev, L.; Räsänen, M.; Albee, B.; Potma, E. O.; Gerber, R. B. Infrared Spectrum of Toluene: Comparison of Anharmonic Isolated-Molecule Calculations and Experiments in Liquid Phase and in a Ne Matrix. *J. Phys. Chem. A* 2016, *120*, 3380–3389.
- [15] Benoit, D. M. PVSCF | A Vibrational Theory Code. 2018; http://pvscf.org.
- [16] Oyedepo, G. A.; Wilson, A. K. Multireference correlation consistent composite approach [MR-ccCA]: Toward accurate prediction of the energetics of excited and transition state chemistry. J. Phys. Chem. A 2010, 114, 8806–8816.
- [17] Nedd, S. A.; DeYonker, N. J.; Wilson, A. K.; Piecuch, P.; Gordon, M. S. Incorporating a completely renormalized coupled cluster approach into a composite method for thermodynamic properties and reaction paths. J. Chem. Phys. 2012, 136, 144109.
- [18] DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The correlation consistent composite approach (ccCA): An alternative to the Gaussian-n methods. J. Chem. Phys. 2006, 124, 114104.
- [19] DeYonker, N. J.; Wilson, B. R.; Pierpont, A. W.; Cundari, T. R.; Wilson, A. K. Towards the intrinsic error of the correlation consistent Composite Approach (ccCA). *Mol. Phys.* 2009, 107, 1107–1121.
- [20] Scott, A. P.; Radom, L. Harmonic Vibrational Frequencies: An Evaluation of Hartree–Fock, Møller–Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. J. Phys. Chem. 1996, 100, 16502–16513.
- [21] Xu, X.; Goddard, W. A. Bonding Properties of the Water Dimer: A Comparative Study of Density Functional Theories. J. Phys. Chem. A 2004, 108, 2305–2313.
- [22] Domin, D.; Benoit, D. M. Assessing Spin-Component-Scaled Second-Order Møller-Plesset Theory Using Anharmonic Frequencies. *ChemPhysChem* 2011, *12*, 3383–3391.
- [23] Respondek, I.; Benoit, D. M. Fast degenerate correlation-corrected vibrational self-consistent field calculations of the vibrational spectrum of 4-mercaptopyridine. J. Chem. Phys. 2009, 131, 054109.
- [24] Merrick, J. P.; Moran, D.; Radom, L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J. Phys. Chem. A* **2007**, *111*, 11683–11700.
- [25] Begue, D.; Carbonniere, P.; Pouchan, C. Calculations of Vibrational Energy Levels by Using a Hybrid ab Initio and DFT Quartic Force Field: Application to Acetonitrile. 2005, 4611–4616.
- [26] Latouche, C.; Palazzetti, F.; Skouteris, D.; Barone, V. High-accuracy vibrational computations for transition-metal complexes including anharmonic corrections: Ferrocene, ruthenocene, and osmocene as test cases. J. Chem. Theory Comput. 2014, 10, 4565–4573.
- [27] Cheng, Q.; Fortenberry, R. C.; DeYonker, N. J. Towards a quantum chemical protocol for the prediction of rovibrational spectroscopic data for transition metal molecules: Exploration of CuCN, CuOH, and CuCCH. J. Chem. Phys. 2017, 147.

- [28] Bowman, J. M. Self-consistent field energies and wavefunctions for coupled oscillators. J. *Chem. Phys.* **1978**, *68*, 608–610.
- [29] Carney, G. D.; Sprandel, L. L.; Kern, C. W. In Adv. Chem. Phys.; Prigogine, I., Rice, S. A., Eds.; John Wiley & Sons, Inc., 1978; Vol. XXXVII; Chapter 6, pp 305–379.
- [30] Gerber, R. B.; Ratner, M. A. A semiclassical self-consistent field (SC SCF) approximation for eigenvalues of coupled-vibration systems. *Chem. Phys. Lett.* **1979**, *68*, 195–198.
- [31] Cohen, M.; Greita, S.; McEarchran, R. Approximate and exact quantum mechanical energies and eigenfunctions for a system of coupled oscillators. *Chem. Phys. Lett.* 1979, 60, 445–450.
- [32] Bowman, J. M. The Self-Consistent-Field Approach to Polyatomic Vibrations. *Acc. Chem. Res.* **1986**, *19*, 202–208.
- [33] Roy, T. K.; Sharma, R.; Gerber, R. B. First-principles anharmonic quantum calculations for peptide spectroscopy: VSCF calculations and comparison with experiments. *Phys. Chem. Chem. Phys.* 2016, 18, 1607–1614.
- [34] Bowman, J. M.; Czakó, G.; Fu, B. High-dimensional ab initio potential energy surfaces for reaction dynamics calculations. *Phys. Chem. Chem. Phys.* 2011, 13, 8094.
- [35] Seager, S. The future of spectroscopic life detection on exoplanets. *Proc. Natl. Acad. Sci.* 2014, 111, 12634–12640.
- [36] Benoit, D. M.; Madebene, B.; Ulusoy, I.; Mancera, L.; Scribano, Y.; Chulkov, S. Towards a scalable and accurate quantum approach for describing vibrations of molecule-metal interfaces. *Beilstein J. Nanotechnol.* **2011**, *2*, 427–447.
- [37] Neese, F. The ORCA program system. 2012; http://doi.wiley.com/10.1002/wcms.81.
- [38] Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1327.
- [39] Herzberg, G. *Electronic Spectra and electronic structure of polyatomic molecules*; Van Nostrand: New York, 1966.
- [40] Huber, K. P.; Herzberg, G. Molecular Spectra and Molecular Structure; Springer US: Boston, MA, 1979.
- [41] Shimanouchi, T. Tables of Molecular Vibrational Frequencies, Consolidated Volume 1; 1972.
- [42] Marston, C. C.; Balint-Kurti, G. G. The Fourier grid Hamiltonian method for bound state eigenvalues and eigenfunctions. *J. Chem. Phys.* **1989**, *91*, 3571–3576.
- [43] Balint-Kurti, G. G.; Ward, C. L.; Clay Marston, C. Two computer programs for solving the Schrödinger equation for bound-state eigenvalues and eigenfunctions using the Fourier grid Hamiltonian method. *Comput. Phys. Commun.* **1991**, 67, 285–292.

- [44] Scribano, Y.; Benoit, D. M. Iterative active-space selection for vibrational configuration interaction calculations using a reduced-coupling VSCF basis. *Chem. Phys. Lett.* 2008, 458, 384–387.
- [45] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- [46] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. J. Chem. *Phys.* **1993**, *98*, 5648–5652.
- [47] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- [48] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. General Performance of Density Functionals General Performance of Density Functionals. 2007, 111, 10439–10452.
- [49] Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. J. Chem. Phys. 1989, 90, 1007–1023.
- [50] Scribano, Y.; Lauvergnat, D. M.; Benoit, D. M. Fast vibrational configuration interaction using generalized curvilinear coordinates and self-consistent basis. J. Chem. Phys. 2010, 133.
- [51] Feller, D.; Peterson, K. A.; Dixon, D. A. The Impact of Larger Basis Sets and Explicitly Correlated Coupled Cluster Theory on the Feller–Peterson–Dixon Composite Method. *Annu. Rep. Comput. Chem.* 2016, 12, 47–48.

CHAPTER 7

CHARGE STABILIZATION OF HIGH POTENTIAL ZINC PORPHYRIN-FULLERENE VIA AXIAL LIGATION OF TETRATHIAFULVALENE

7.1 Introduction

Sustainable production of electricity and fuel using abundant solar photons is one of the most highly researched topics in modern science.^{1–12} Often, the design of light energy harvesting materials follows the concepts developed by Mother Nature in bacterial and green plant photosynthetic systems.^{13–15} The primary photochemical events in natural photosynthesis involves capture capturing and funneling of sun light by a group of well-organized chromophores called 'antenna' systems and promoting electron transfer using the funneled light into the 'reaction center' where a cascade of electron transfer events occurs leading to the generation of long-lived charge separated states. Over the last two to three decades, early photo-events of natural photosynthesis have been mimicked by building donor-acceptor systems to visualize energy and electron transfer or a combination of these two events.^{16–45} One strategy used in building donor-acceptor systems that are capable of producing high-energy charge separated states include choosing donors that are difficult to oxidize and acceptors that are difficult to reduce. Under these conditions, the stored energy in the charge separated state is equivalent to the potential difference between the oxidation and reduction potentials of the donor and acceptor, respectively. However, challenges exist to accomplish this goal where the excited state energy from either the donor or the acceptor may not be sufficient to drive the electron transfer process in an energetically feasible fashion.¹

Collaborators have synthesized a donor-acceptor dyad, (F₁₅P)Zn-C₆₀, capable of generating

¹This chapter is reprinted from Obondi, C. O.; Lim, G. N.; Jang, Y.; Patel, P.; Wilson, A. K.; Poddutoori, P. K.; D'Souza, F. *J.Phys. Chem. C* **2018**, 122, 13636–13647 with permission of the American Chemical Society.

charge separated state carrying an energy of 1.70 eV (see Figure 7.1 for structure of the dyad).⁴⁶ In that study, the electron donor zinc porphyrin was functionalized with meso-pentafluorophenyl substituents that made the zinc porphyrin difficult to oxidize by 0.43 eV compared to simple zinc porphyrins. The singlet excited energy of ${}^{1}(F_{15}P)Zn \cdot (= 2.21 \text{ eV})$ was sufficient to drive the electron transfer process. The lifetime of the charge separated state was persistent for about 50-60 ns. In this chapter, to prolong the lifetime of the charge separated state, collaborators developed supramolecular triads using a hole transporting tetrathiafulvalene, TTF, linked via the well-known metal-ligand axial coordination approach.⁴⁷ Here, the TTF was functionalized with either pyridine or phenylpyridine coordinating ligands. Supramolecular triad formation including binding constants and stoichiometry of the complexes were determined by spectroscopic methods.

7.2 Computational Contributions and Analysis

Computational support provided by the author supplemented the synthesis of these $((F_{15}P)Zn - C_{60})$ supramolecular dyad triads $(C_{60} - (F_{15}P)Zn:Py-TTF)$ and and $C_{60} - (F_{15}P)Zn:Py-phTTF)$ via modeling the molecular electrostatic potential and frontier orbitals. The M06-2X/6-31G* method and basis set combination was based on qualitative modeling indicative of intramolecular charge transfer via the relative location of the HOMO and LUMO.^{48–50} M06-2X was chosen based on its implementation for main group and transition metal thermochemistry.⁴⁸ The Pople-style 6-31G* basis set was used based on its small size and availability for all atoms present in this compound.^{49,50} The molecular electrostatic potential (MEP) for all of complexes were modeled on a scale of strong attractive potential (red) to a strong repulsive potential (blue) with respect to a positive test charge. This provides insight into the potential binding behavior of these systems.

The geometry and electronic structures of the $(F_{15}P)Zn - C_{60}$ dyad, and $C_{60} - (F_{15}P)Zn:Py-TTF$ and $C_{60} - (F_{15}P)Zn:Py-phTTF$ triads were probed using the hybrid-metal Minnesota functional M06-2X with 54% exact exchange and the 6-31G* basis set using Gaussian09.^{48–51} Figure 7.1 depicts molecular electrostatic potential (MEP) maps and frontier HOMO and LUMO for the optimized structures. In the case of the dyads, the frontier HOMO was on the ($F_{15}P$)Zn and LUMO on the C₆₀ making them the donor and acceptor sites, respectively. Interestingly, for the triad, the HOMO was shifted to the TTF site without altering the location of the LUMO, which is attributed to the easier oxidation of TTF over ($F_{15}P$)Zn. HOMO-1 occupied the ($F_{15}P$)Zn for the triads. It may be pointed out here that the density of the dyad was not affected by the addition of TTF ligand except at the porphyrin center where the potential was neutral, which indicates that the central metal is fully coordinated and not likely to bind an additional ligand. The estimated center-tocenter distances between Zn and C₆₀ in the dyad and triads were ~17.5 Å while these distances between Zn and TTF were ~18.0 and ~17.7 Å, respectively, in the case of C₆₀ – ($F_{15}P$)Zn:Py-TTF and C60 – ($F_{15}P$)Zn:Py-phTTF triads.



Figure 7.1: MO6-2X/6-31G* molecular electrostatic potential maps, and the frontier HOMO and LUMO of the optimized structures of (a) $(F_{15}P)Zn-C_{60}$ dyad and (b) $C_{60}-(F_{15}P)Zn$:Py-phTTF triad. The isovalue used for the MO depictions was 0.02 while the density value used was 0.0004.

REFERENCES

REFERENCES

- [1] Connolly, J., Ed. *Photochemical Conversion and Storage of Solar Energy*; Academic Press Inc.: New York, 1981.
- [2] Lewis, N. S.; Nocera, D. G. Powering the planet: Chemical challenges in solar energy utilization. *Proc. Natl. Acad. Sci.* **2006**, *103*, 15729–15735.
- [3] Kamat, P. V. Meeting the Clean Energy Demand: Nanostructure Architectures for Solar Energy Conversion. J. Phys. Chem. C 2007, 111, 2834–2860.
- [4] Armaroli, N.; Balzani, V. The Future of Energy Supply: Challenges and Opportunities. *Angew. Chemie Int. Ed.* **2007**, *46*, 52–66.
- [5] Wasielewski, M. R. Self-Assembly Strategies for Integrating Light Harvesting and Charge Separation in Artificial Photosynthetic Systems. *Acc. Chem. Res.* **2009**, *42*, 1910–1921.
- [6] Gust, D.; Moore, T. A.; Moore, A. L. Solar Fuels via Artificial Photosynthesis. Acc. Chem. Res. 2009, 42, 1890–1898.
- [7] Grätzel, M. Recent Advances in Sensitized Mesoscopic Solar Cells. *Acc. Chem. Res.* **2009**, 42, 1788–1798.
- [8] Young, K. J.; Martini, L. A.; Milot, R. L.; Snoeberger, R. C.; Batista, V. S.; Schmuttenmaer, C. A.; Crabtree, R. H.; Brudvig, G. W. Light-driven water oxidation for solar fuels. *Coord. Chem. Rev.* 2012, 256, 2503–2520.
- [9] Alibabaei, L.; Brennaman, M. K.; Norris, M. R.; Kalanyan, B.; Song, W.; Losego, M. D.; Concepcion, J. J.; Binstead, R. A.; Parsons, G. N.; Meyer, T. J. Solar water splitting in a molecular photoelectrochemical cell. *Proc. Natl. Acad. Sci.* **2013**, *110*, 20008–20013.
- [10] Crabtree, G. W.; Lewis, N. S. Solar energy conversion. *Phys. Today* 2007, 60, 37–42.
- [11] Hammarström, L. Artificial Photosynthesis and Solar Fuels. Acc. Chem. Res. 2009, 42, 1859– 1860.
- [12] Obraztsov, I.; Kutner, W.; D'Souza, F. Evolution of Molecular Design of Porphyrin Chromophores for Photovoltaic Materials of Superior Light-to-Electricity Conversion Efficiency. *Sol. RRL* 2017, *1*, 1600002.
- [13] Cogdell, R., Mullineaux, C., Eds. *Photosynthetic Light Harvesting*; Springer Netherlands: Dordrecht, 2008.
- [14] Green, B. R., Parson, W. W., Eds. *Light-Harvesting Antennas in Photosynthesis*; Advances in Photosynthesis and Respiration; Springer Netherlands: Dordrecht, 2003; Vol. 13.
- [15] Pessarakli, M., Ed. *Handbook of Photosynthesis, Second Edition*, 2nd ed.; Books in Soils, Plants, and the Environment; CRC Press, 2005.

- [16] Balzani, V.; Credi, A.; Venturi, M. Photochemical Conversion of Solar Energy. *ChemSusChem* 2008, 1, 26–58.
- [17] Fukuzumi, S.; Ohkubo, K.; Suenobu, T. Long-Lived Charge Separation and Applications in Artificial Photosynthesis. *Acc. Chem. Res.* **2014**, *47*, 1455–1464.
- [18] Fukuzumi, S.; Ohkubo, K.; D'Souza, F.; Sessler, J. L. Supramolecular electron transfer by anion binding. *Chem. Commun.* 2012, 48, 9801.
- [19] D'Souza, F.; Ito, O. Photosensitized electron transfer processes of nanocarbons applicable to solar cells. *Chem. Soc. Rev.* 2012, 41, 86–96.
- [20] KC, C. B.; D'Souza, F. Design and photochemical study of supramolecular donor-acceptor systems assembled via metal-ligand axial coordination. *Coord. Chem. Rev.* 2016, 322, 104– 141.
- [21] El-Khouly, M. E.; Fukuzumi, S.; D'Souza, F. Photosynthetic Antenna-Reaction Center Mimicry by Using Boron Dipyrromethene Sensitizers. *ChemPhysChem* 2014, 15, 30–47.
- [22] Imahori, H.; Umeyama, T.; Ito, S. Large π -Aromatic Molecules as Potential Sensitizers for Highly Efficient Dye-Sensitized Solar Cells. *Acc. Chem. Res.* **2009**, *42*, 1809–1818.
- [23] Hasobe, T. Supramolecular nanoarchitectures for light energy conversion. *Phys. Chem. Chem. Phys.* **2010**, *12*, 44–57.
- [24] Ulrich, G.; Ziessel, R.; Harriman, A. Die vielseitige Chemie von Bodipy-Fluoreszenzfarbstoffen. *Angew. Chemie* **2008**, *120*, 1202–1219.
- [25] Schwartz, E.; Le Gac, S.; Cornelissen, J. J. L. M.; Nolte, R. J. M.; Rowan, A. E. Macromolecular multi-chromophoric scaffolding. *Chem. Soc. Rev.* 2010, *39*, 1576.
- [26] Guldi, D. M.; Rahman, G. M. A.; Sgobba, V.; Ehli, C. Multifunctional molecular carbon materials—from fullerenes to carbon nanotubes. *Chem. Soc. Rev.* 2006, 35, 471.
- [27] Clifford, J. N.; Accorsi, G.; Cardinali, F.; Nierengarten, J.-F.; Armaroli, N. Photoinduced electron and energy transfer processes in fullerene C₆₀-metal complex hybrid assemblies. *Comptes Rendus Chim.* **2006**, *9*, 1005–1013.
- [28] Martín, N.; Sánchez, L.; Herranz, M. Á.; Illescas, B.; Guldi, D. M. Electronic Communication in Tetrathiafulvalene (TTF)/C₆₀ Systems: Toward Molecular Solar Energy Conversion Materials? Acc. Chem. Res. 2007, 40, 1015–1024.
- [29] Bottari, G.; de la Torre, G.; Guldi, D. M.; Torres, T. Covalent and Noncovalent Phthalocyanine–Carbon Nanostructure Systems: Synthesis, Photoinduced Electron Transfer, and Application to Molecular Photovoltaics. *Chem. Rev.* 2010, *110*, 6768–6816.
- [30] Guldi, D. M.; Sgobba, V. Carbon nanostructures for solar energy conversion schemes. *Chem. Commun.* 2011, 47, 606–610.

- [31] Kato, D.; Sakai, H.; Saegusa, T.; Tkachenko, N. V.; Hasobe, T. Synthesis, Structural and Photophysical Properties of Pentacene Alkanethiolate Monolayer-Protected Gold Nanoclusters and Nanorods: Supramolecular Intercalation and Photoinduced Electron Transfer with C₆₀. J. Phys. Chem. C 2017, 121, 9043–9052.
- [32] Davis, C. M.; Kawashima, Y.; Ohkubo, K.; Lim, J. M.; Kim, D.; Fukuzumi, S.; Sessler, J. L. Photoinduced Electron Transfer from a Tetrathiafulvalene-Calix[4]pyrrole to a Porphyrin Carboxylate within a Supramolecular Ensemble. J. Phys. Chem. C 2014, 118, 13503–13513.
- [33] Voityuk, A. A. Electronic Couplings for Photoinduced Electron Transfer and Excitation Energy Transfer Computed Using Excited States of Noninteracting Molecules. J. Phys. Chem. A 2017, 121, 5414–5419.
- [34] Xu, B.; Wang, C.; Ma, W.; Liu, L.; Xie, Z.; Ma, Y. Photoinduced Electron Transfer in Asymmetrical Perylene Diimide: Understanding the Photophysical Processes of Light-Absorbing Nonfullerene Acceptors. J. Phys. Chem. C 2017, 121, 5498–5502.
- [35] Cai, N.; Takano, Y.; Numata, T.; Inoue, R.; Mori, Y.; Murakami, T.; Imahori, H. Strategy to Attain Remarkably High Photoinduced Charge-Separation Yield of Donor–Acceptor Linked Molecules in Biological Environment via Modulating Their Cationic Moieties. J. Phys. Chem. C 2017, 121, 17457–17465.
- [36] Amati, A.; Cavigli, P.; Kahnt, A.; Indelli, M. T.; Iengo, E. Self-Assembled Ruthenium(II) Porphyrin-Aluminium(III)Porphyrin-Fullerene Triad for Long-Lived Photoinduced Charge Separation. J. Phys. Chem. A 2017, 121, 4242–4252.
- [37] Stangel, C.; Charisiadis, A.; Zervaki, G. E.; Nikolaou, V.; Charalambidis, G.; Kahnt, A.; Rotas, G.; Tagmatarchis, N.; Coutsolelos, A. G. Case Study for Artificial Photosynthesis: Noncovalent Interactions between C₆0-Dipyridyl and Zinc Porphyrin Dimer. *J. Phys. Chem. C* 2017, *121*, 4850–4858.
- [38] Pahk, I.; Kodis, G.; Fleming, G. R.; Moore, T. A.; Moore, A. L.; Gust, D. Artificial Photosynthetic Reaction Center Exhibiting Acid-Responsive Regulation of Photoinduced Charge Separation. J. Phys. Chem. B 2016, 120, 10553–10562.
- [39] Medrano, C. R.; Oviedo, M. B.; Sánchez, C. G. Photoinduced charge-transfer dynamics simulations in noncovalently bonded molecular aggregates. *Phys. Chem. Chem. Phys.* 2016, 18, 14840–14849.
- [40] Pagona, G.; Stergiou, A.; Gobeze, H. B.; Rotas, G.; D'Souza, F.; Tagmatarchis, N. Photoinduced charge separation in an oligophenylenevinylene-based Hamilton-type receptor supramolecularly associating two C₆₀-barbiturate guests. *Phys. Chem. Chem. Phys.* 2016, 18, 811–817.
- [41] Bandi, V.; Gobeze, H. B.; Lakshmi, V.; Ravikanth, M.; D'Souza, F. Vectorial Charge Separation and Selective Triplet-State Formation during Charge Recombination in a Pyrrolyl-Bridged BODIPY–Fullerene Dyad. J. Phys. Chem. C 2015, 119, 8095–8102.

- [42] Manna, A. K.; Balamurugan, D.; Cheung, M. S.; Dunietz, B. D. Unraveling the Mechanism of Photoinduced Charge Transfer in Carotenoid–Porphyrin–C₆₀ Molecular Triad. J. Phys. Chem. Lett. 2015, 6, 1231–1237.
- [43] Obondi, C. O.; Lim, G. N.; D'Souza, F. Triplet–Triplet Excitation Transfer in Palladium Porphyrin–Fullerene and Platinum Porphyrin–Fullerene Dyads. J. Phys. Chem. C 2015, 119, 176–185.
- [44] Bandi, V.; Gobeze, H. B.; Nesterov, V. N.; Karr, P. A.; D'Souza, F. Phenothiazine– azaBODIPY–fullerene supramolecules: syntheses, structural characterization, and photochemical studies. *Phys. Chem. Chem. Phys.* **2014**, *16*, 25537–25547.
- [45] Bandi, V.; Gobeze, H. B.; Karr, P. A.; D'Souza, F. Preferential Through-Space Charge Separation and Charge Recombination in V-Type Configured Porphyrin–azaBODIPY– Fullerene Supramolecular Triads. J. Phys. Chem. C 2014, 118, 18969–18982.
- [46] Lim, G. N.; Obondi, C. O.; D'Souza, F. A High-Energy Charge-Separated State of 1.70 eV from a High-Potential Donor-Acceptor Dyad: A Catalyst for Energy-Demanding Photochemical Reactions. *Angew. Chemie Int. Ed.* 2016, 55, 11517–11521.
- [47] Poddutoori, P. K.; Lim, G. N.; Sandanayaka, A. S. D.; Karr, P. A.; Ito, O.; D'Souza, F.; Pilkington, M.; van der Est, A. Axially assembled photosynthetic reaction center mimics composed of tetrathiafulvalene, aluminum(III) porphyrin and fullerene entities. *Nanoscale* 2015, 7, 12151–12165.
- [48] 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. *Theor. Chem. Acc.* 2008, 120, 215–241.
- [49] Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* 1973, 28, 213–222.
- [50] Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. 6-31G*basis set for atoms K through Zn. J. Chem. Phys. 1998, 109, 1223–1229.
- [51] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian09 Revision D.01, Gaussian Inc. Wallingford CT 2009.

CHAPTER 8

SAMPL6 HOST-GUEST CHALLENGE: BINDING FREE ENERGIES VIA A MULTISTEP APPROACH

8.1 Introduction

Tremendous advances in technological capabilities have enabled computational approaches to be applied to discern a broad range of physical, chemical, and biological phenomena across scales in molecular science. $^{1-6}$ With emphasis on molecular design, computational approaches have found great utility towards innovation in drug discovery. Considering the time and cost of the drug pipeline, from the discovery process to market, in silico biophysical methods serve an important role in expediting and reducing the cost of the discovery process, facilitating the identification, optimization, and refinement of potential drug candidates and providing comprehensive insight into the mechanism of action and structure-property relationships at the atomic level that are ultimately critical to a drug's efficacy. $^{17-12}$

In computational strategies towards structure-based design, an important step is the prediction of probable conformations of a ligand bound to the host. To identify better possible candidate binding modes, they can be ranked via scoring functions and further evaluated via molecular simulation and free energy calculations. From free energy calculations, selectivity profiles may be constructed not only to determine binding affinities but also to provide understanding into how the ligand recognizes its host.

Because of the complexity that occurs in ligand-bound protein systems, relatively smaller representative models such as polymer-based host-guest systems are used to assess free energy methods.^{13–18} Although host structures selected to represent proteins are typically much smaller

¹This chapter is reprinted from Eken, Y.; Patel, P.; Díaz, T.; Jones, M. R.; Wilson, A. K. SAMPL6 Host – Guest Challenge : Binding Free Energies via a Multistep Approach. *J. Comput. Aided. Mol. Des.* **2018**, 32 (10), 1097–1115. with permission of the Springer International Publishing.

than proteins, they are large enough to possess a cavity or binding pocket that allows non-covalent binding of multiple guest molecules. The advantage of using host-guest systems for assessing free energy methods is that they tend to be more rigid and symmetric than proteins, which results in fewer conformations that need to be sampled. ^{19–23} Even in the representation of proteins by more simplistic models, modeling binding free energies for these smaller models is challenging since no clear "best" computational chemistry approach has been identified; efforts are needed to better resolve strategies towards predictions of binding free energies. Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) blind challenges provide a unique platform to validate available methods and stimulate the development of new methods for quantitative predictions. ^{13,16,18,24–26} In these challenges, binding affinities and other physicochemical properties are predicted, using computational models without the benefit of insight from experiment; they are then later compared to unpublished experimental measurements that allow the comparison of different computational prediction methods.

While classical molecular dynamics (MD) methods are commonly used to investigate host-guest interactions, molecular mechanics (MM) force fields result in a limited treatment of effects resulting from polarization, charge transfer, and many body effects which can impact the description of properties such as binding free energies.^{9,27–31} To better account for these effects, quantum mechanical (QM) approaches, which are more costly, are commonly used in drug discovery research,^{9,32} and have been used in previous SAMPL competitions.^{17,33–35} For example, in the SAMPL5 competition for host-guest binding, Caldarau et al.³³ used DFT-D3 and DLPNO-CCSD(T) to predict the binding energies for octa-acid (OA) host-guest systems. In this approach, they used TPSS-D3/def2-SVP optimized structures and host structures are constrained during MD simulations to reduce the flexibility of the host and limit the structural distortions resulting from the repulsion between the negative charge of the ligands and the large negative charge of the OA hosts. This approach yielded binding energies approximately 12.0 kcal mol⁻¹ greater than the experimental binding affinities, with a low correlation coefficient ($\tau^2 \approx 0$), and a statistically insignificant Kendall's rank correlation coefficient ($\tau \leq 0.20$) for all attempts for

the host-guest systems in the SAMPL5 blind challenge due to incorrect representative structures, not sampling enough conformational binding positions for ligands, and thermochemical corrections that yielded up to a 7.2 kcal mol⁻¹ difference depending on the method of choice. This performance demonstrates the limited sampling capabilities of current QM methods compared to MD methods, obtained representative structures, as well as thermodynamic and solvation corrections.

Contrary to this, in the SAMPL4 competition for host-guest binding, Mikulskis et al.³⁵ were successful with both MM- and QM-based approaches for OA hosts with mean absolute deviations (MADs) less than 2.0kcal mol $^{-1}$. Their MM approach, which utilized free energy perturbation (FEP) calculations, yielded MADs of approximately 1.0 kcal mol^{-1} while their QM approaches with DFT-D3 optimized structures yielded MADs of approximately 1.0-2.0 kcal mol⁻¹ depending on the implementation of a solvent in the calculations, i.e. no solvent, implicit solvent, or a combined implicit-explicit solvent. However, the combination of FEP and DFT-D3 did not yield favorable results due to the large difference between the MM and DFT potential energy functions. Sure et al.³⁴ provided another successful attempt at using DFT-D3 for the SAMPL4 competition for host-guest binding of a macrocyclic cucurbit[7]uril host by optimizing the geometry at the TPSS-D3/def2-TZVP level of theory after pre-optimizing possible binding scenarios with the HF-3c semiempirical method. These optimizations were followed by single point calculations using PW6B95-D3/def2-QZVP with the g- and f-functions for non-hydrogen and hydrogen atoms removed, respectively, with the COSMO-RS implicit solvent model, which yielded a MAD of 2.0 \pm 0.5kcal mol⁻¹. These two studies highlight that for the SAMPL4 competition, host-guest structure optimization and higher-level MM-based approaches like FEP can be vital in characterizing correct binding interactions at the QM level.

In this work, efforts in MD and QM methods are combined to predict binding affinities for fourteen ligands to a macrocyclic cucurbit[8]uril host^{19,21,22,36} and eight ligands to two variants of the OA deep-cavity cavitands.^{20,23} Using MD simulations to obtain representative structures, MM-and QM-based methods are utilized to predict binding free energies. Within the QM methods, the

use of a resolution-of-the-identity (RI) approximation designed for larger molecules,³⁷ Grimme's D3 atom-pairwise dispersion corrections with Becke-Johnson damping,³⁸ and truncated correlation consistent basis sets for the hydrogen atoms³⁹ are evaluated to probe how different electronic structure approaches that reduce the computational cost contribute to predicting binding affinities. Insights into what strategies are more favorable for host guest-binding will help to build a framework for predicting host-guest binding affinities using QM approaches.

8.2 Methods

8.2.1 System Preparation and Simulation Protocol

The initial structures for the guest molecules are shown in Figures 8.1 and 8.2, and the three host molecules, cucurbit[8]uril (CB8), octa-acid (OA), and tetramethyl octa-acid (TEMOA), are shown in Figure 8.3. These molecules were issued with the SAMPL6 challenge dataset were used to generate the host-guest systems. The CB8 molecule has no formal charge whereas the octa-acids (OA/TEMOA) have eight deprotonated carboxylic acid groups and thus a formal charge of -8. Even though OA and TEMOA are water-soluble structurally similar deep-cavity cavitands, the TEMOA host has four methyl groups in place of four hydrogen atoms present in the OA host located on the upper rim of the cavitand that enclose the hydrophobic binding pocket.

Initial binding poses of guest molecules binding to the host were generated and refined through a ΔG scoring function.^{40–45} Subsequent molecular dynamics simulations were then carried out in Amber16.7 to relax the host-guest systems in aqueous solution.⁴⁶ An MM-based approach (MMPBSA) was used to calculate the binding free energies at the MM-level, which is a standard level of theory when dealing with drug binding interactions.⁴⁷ This portion was done by co-authors Yiğitcan Eken, Thomas Díaz, and Michael Jones.



Figure 8.1: Guest molecules for the cucurbit[8]uril (CB8) host.



Figure 8.2: Guest molecules for the octa-acid (OA) and tetramethyl octa-acid (TEMOA) hosts.



Figure 8.3: Host molecules: cucurbit[8]uril (CB8), octa-acid (OA), and tetramethyl octa-acid (TEMOA).

8.2.2 Quantum Mechanical Calculations

All quantum mechanical calculations were done by the author. The individual structures generated from the clustering of MD trajectories, shown in Figures 8.4-8.6, for each host-guest complex were used for all quantum chemical calculations. The host and guest molecules were analyzed with the same geometry as from the complex. The thermal corrections for all molecules were calculated at the HF/6-31G(d) level of theory in Gaussian 16 and the vibrational contributions were scaled by 0.8953.⁴⁸ Single point energies were obtained using ORCA 4.0⁴⁹ with the B3PW91 density functional ^{50–52} since B3PW91 has been shown to properly treat long-range covalent interactions. In the treatment of the exact exchange in the functional, the RIJCOSX approximation³⁷ was used with the def2 auxiliary basis set⁵³ to reduce the computational cost associated with the number of atoms in the host-guest complex since the RIJCOSX approximation has been shown to be five times as efficient for molecules of similar size to the host-guest systems. To mimic the aqueous solution, the SMD implicit solvation model⁵⁴ was used with water ($\varepsilon = 78.4$) as the implicit solvent. Grimme's D3 dispersion correction with Becke-Johnson damping was used to investigate long-range covalent interactions as the inclusion of D3 dispersion improves intermolecular interaction energies predicted with DFT.^{34,38,55,56}

The cc-pV nZ^{57} basis sets were used for all single point calculations (see Section 2.2.2 for reasons).^{58–61} Knowing the CBS limit, which removes basis set incompleteness error, the error for the property of interest, i.e. binding free energy, only corresponds to the intrinsic error of the chosen QM method. Therefore, to extrapolate to the Kohn-Sham limit for DFT methods, analogous to the CBS limit for wavefunction-based methods, the cc-pVnZ basis sets were used (n = D, T) with the following extrapolation scheme proposed by Jensen

$$E(l_{max}) = E_{CBS} + A(l_{max} + 1)e^{-B\sqrt{n_s}}$$
(8.1)

where l_{max} is the maximum angular momentum function in the basis set and n_s is the number of s functions in the basis set.⁶² The *B*-parameter was set to 5.5 in agreement with Jensen for use as a two-point extrapolation scheme. Due to the abundance of weak molecular interactions in biomolecules, the calculated binding energies were counterpoise-corrected before the extrapolations were performed on each host, guest, and host-guest complex.^{63,64}

Additional electronic structure modeling techniques were applied to the CB8 host-guest systems to examine the impact of approximations on the binding free energy. Targeting reduction in computational time, the cc-pVnZ basis sets were truncated via the removal of higher angular momentum basis functions for hydrogen atoms. This has been shown to reduce the computational time by approximately 42.9% and 57.8% when removing 1 *d* function from the cc-pVTZ basis set, denoted as cc-pVTZ(-1*d*), and 2 *d* functions and 1 *f* function from the cc-pVQZ basis set, denoted as cc-pVQZ(-1*f*2*d*), respectively, and yielded the results closest to the atomization energies generated with the full basis sets at the complete basis set limit.³⁹

Binding free energies calculated with and without the use of the resolution-of-the-identity (RI) approximation were examined to gauge how the RI approximation, which leads to a reduction in CPU time, affects the accuracy. To characterize the ionic strength of the solution used in experiment, the dielectric constant for the implicit water solvent was also altered from 78.4 for pure water to 76.4 given the concentration of the sodium chloride solution used in the MD simulations and the experimentally determined relation between the concentration of an ionic solution and the dielectric constant.⁶⁵

8.3 Results

The binding free energies submitted as part of the SAMPL6 competition are shown in Tables 8.1-8.3 for CB8, OA, and TEMOA host-guest systems, respectively. For each host-guest complex, statistical measurements were used to gauge the effectiveness of each of the three methods, which are MMPBSA, RI-B3PW91-D3, and RI-B3PW91, in predicting experimental binding free energies. These include the mean absolute error (MAE), the root mean square error (RMSE), Kendall's Tau (τ) rank correlation coefficient, which measures how well a method ranked calculated binding free energies relative to experimental binding free energies where τ values closer to one correspond to increased qualitative accuracy of the prediction, and the correlation coefficient (r^2). To demonstrate

there is no correlation in ranking between the calculated binding free energies and the experimental binding free energies, τ values are compared against τ_{crit} , a cutoff value obtained through a table of critical values generated by Monte Carlo simulations of a τ distribution, which is similar to the normal Z distribution used to reject the null hypothesis.^{66,67}

8.3.1 CB8

The binding free energy predictions for the CB8 host with the three methods submitted were compared to experiment (Table 8.1). The predicted values were significantly more negative than experimental binding free energies with an MAE of 16.69, 33.58, and 15.54 kcal mol⁻¹ for MMPBSA, RI-B3PW91-D3, and RI-B3PW91, respectively.

When the binding affinities of the guests to CB8 are ranked from the lowest to the highest binding affinity, MMPBSA did not correctly rank any of the systems but predicted CB8-G12 to have a stronger binding affinity relative to the other complexes, which correlates to experiment well. RI-B3PW91-D3 correctly ranked CB8-G2 as the tenth strongest bound host-guest complex and predicted that CB8-G12 was more tightly bound relative to the other CB8 host-guest systems. RI-B3PW91 correctly ranked CB8-G6, CB8-G2, CB8-G1, and CB8-G3 as fifth, tenth, eleventh, and fourteenth, respectively, while the remaining systems were ranked incorrectly. Unlike both MMPBSA and RI-B3PW91-D3, RI-B3PW91 predicted CB8-G12 to have a lower binding affinity relative to the other CB8 host-guest systems.

Complex	Exp	MMPBSA	RI-B3PW91-D3	RI-B3PW91
CB8-G0	$\textbf{-6.69} \pm 0.05$	-29.4 ± 0.3	-49.89	6.75
CB8-G1	$\textbf{-7.65}\pm0.04$	-31.5 ± 0.3	-57.22	12.70
CB8-G2	-7.66 ± 0.05	-25.6 ± 0.3	-36.86	10.34
CB8-G3	$\textbf{-6.45} \pm 0.06$	-34.2 ± 0.5	-44.53	26.61
CB8-G4	$\textbf{-7.80}\pm0.04$	-30.8 ± 0.3	-68.09	-11.11
CB8-G5	-8.18 ± 0.05	-18.6 ± 0.3	-35.92	2.39
CB8-G6	$\textbf{-8.34} \pm 0.05$	-19.8 ± 0.2	-31.95	1.26
CB8-G7	-10.00 ± 0.10	-17.6 ± 0.4	-14.90	18.09
CB8-G8	-13.50 ± 0.04	-30.4 ± 0.2	-50.34	4.49
CB8-G9	-8.68 ± 0.08	-19.9 ± 0.5	-37.07	-2.46
CB8-G10	-8.22 ± 0.07	-19.6 ± 0.3	-39.30	0.61
CB8-G11	$\textbf{-7.77}\pm0.05$	-17.5 ± 0.4	-25.75	-1.07
CB8-G12	$\textbf{-9.86} \pm 0.03$	-31.5 ± 0.4	-62.05	15.00
CB8-G13	$\textbf{-7.11}\pm0.03$	-25.4 ± 0.3	-44.04	0.17
MAE		$16.69\pm0.33^{\rm a}$	34.29	14.88
RMSE		$17.80\pm0.76^{\rm b}$	36.99	17.26
au		-0.19	-0.14	0.05
r^2		0.00	0.00	0.00

Table 8.1: Binding free energies for the CB8 host-guest complexes.

The mean absolute error (MAE), root mean square error (RMSE), Kendall's Tau (τ), and r² are shown. These results correspond to those submitted for the competition. ^aThe uncertainty reported for MAE is the average of the absolute uncertainties. ^bThe uncertainty reported for RMSE is the uncertainty of the RMSE with the experimental and calculated uncertainties.







CB8-G2



CB8-G0





CB8-G3



CB8-G5



CB8-G6





CB8-G7

CB8-G8



Figure 8.4: Structures of the CB8 guest molecules inside the binding pocket generated from the clustering analysis.

8.3.2 OA

Complex	Exp	MMPBSA	RI-B3PW91-D3	RI-B3PW91
OA-G0	-5.68 ± 0.03	-12.6 ± 0.2	-41.36	-16.57
OA-G1	-4.65 ± 0.02	-11.6 ± 0.1	-40.67	-17.15
OA-G2	$\textbf{-8.38}\pm0.02$	-18.3 ± 0.2	6.54	44.53
OA-G3	$\textbf{-5.18} \pm 0.02$	-10.0 ± 0.2	-47.94	-17.62
OA-G4	-7.11 ± 0.02	-17.0 ± 0.2	-48.19	-13.49
OA-G5	-4.59 ± 0.02	-9.1 ± 0.2	-38.40	-16.42
OA-G6	-4.97 ± 0.02	-11.3 ± 0.2	-43.19	-23.31
OA-G7	$\textbf{-6.22}\pm0.02$	-11.4 ± 0.1	-47.37	-23.78
MAE		$6.80\pm0.2^{\mathrm{a}}$	35.46 [38.49]	17.86 [12.85]
RMSE		$7.07\pm0.4^{\mathrm{b}}$	36.41 [38.52]	22.51 [13.39]
au		0.64	0.29 [0.71]	-0.21 [0.05]
r^2		0.84	0.44 [0.52]	0.60 [0.03]

Table 8.2: Binding free energies for the OA host-guest complexes.

The mean absolute error (MAE), root mean square error (RMSE), Kendall's Tau (τ), and r^2 are shown. Bracketed values indicate the values after the removal of the statistical outlier (OA-G2). These results correspond to those submitted for the competition. ^aThe uncertainty reported for MAE is the average of the absolute uncertainties.

^bThe uncertainty reported for RMSE is the uncertainty of the RMSE with the experimental and calculated uncertainties.

The three sets of submitted binding free energy predictions for OA are reported in Table 8.2. All values predicted using MMPBSA were significantly more negative than experimental measurements with an MAE of 6.8 ± 0.2 kcal mol⁻¹. When ranking the binding affinities of the guest to the host from lowest to highest binding affinity, MMPBSA correctly placed OA-G2, OA-G4, OA-G6, OA-G5 as first, second, sixth, and eighth, respectively. The other systems were not ranked correctly; OA-G0, OA-G1, OA-G7 and OA-G3 ranked third, fourth, fifth, and seventh, respectively, whereas experimentally ranked fourth, seventh, third, and fifth, respectively.

For RI-B3PW91-D3 and RI-B3PW91, the binding free energy predicted for OA-G2 was determined as a statistical outlier with 99% confidence, visualized in Figure 8.8, using Dixon's Q-Test.⁶⁸ When the statistical outlier (OA-G2) was excluded from the RI-B3PW91-D3 set, the MAE, RMSE, Kendall's Tau (τ), and the correlation coefficient (r^2) increased from 35.46 to 38.39kcal mol⁻¹, 36.41 to 38.52kcal mol⁻¹, 0.29 to 0.71, and 0.44 to 0.52, respectively. When the binding free energy for OA-G2 was excluded from the set of binding free energies obtained

with RI-B3PW91, the MAE, RMSE, and r^2 decreased from 17.87 to 12.85kcal mol⁻¹, 22.51 to 13.39kcal mol⁻¹, and 0.60 to 0.03, respectively, as shown in Table 8.2. In Figure 8.7b, the statistical outlier was removed, which improved and worsened the linear regression model comparing experiment to RI-B3PW91-D3 and RI-B3PW91, respectively. With the exclusion of OA-G2, ranking the binding affinities from lowest to highest, RI-B3PW91-D3 correctly ranked OA-G4, OA-G1, and OA-G5, as first, sixth, and seventh, respectively, while RI-B3PW91 did not correctly ranked any of the systems.



Figure 8.5: Structures of the OA guest molecules inside the binding pocket generated from the clustering analysis.

8.3.3 TEMOA

Complex	Exp	MMPBSA	RI-B3PW91-D3	RI-B3PW91
TEMOA-G0	$\textbf{-6.06} \pm 0.02$	-12.0 ± 0.2	-43.75	-12.80
TEMOA-G1	$\textbf{-5.97} \pm 0.04$	-11.3 ± 0.2	-41.98	-10.18
TEMOA-G2	$\textbf{-6.81} \pm 0.02$	-19.3 ± 0.2	-51.23	-7.22
TEMOA-G3	$\textbf{-5.60}\pm0.04$	-8.3 ± 0.2	-43.56	-15.29
TEMOA-G4	$\textbf{-7.79}\pm0.02$	-19.2 ± 0.3	-51.98	-12.39
TEMOA-G5	-4.16 ± 0.02	$\textbf{-6.1}\pm0.2$	-37.04	-10.66
TEMOA-G6	$\textbf{-5.40}\pm0.03$	-10.4 ± 0.2	-41.05	-16.94
TEMOA-G7	$\textbf{-4.13}\pm0.02$	$\textbf{-6.8}\pm0.3$	-45.98	-10.29
MAE		$5.9\pm0.2^{\mathrm{a}}$	38.83	6.23
RMSE		$7.0\pm0.5^{\mathrm{b}}$	39.03	7.00
au		0.79	0.57	-0.14
\mathbf{r}^2		0.86	0.55	0.00

Table 8.3: Binding free energies for the TEMOA host-guest complexes.

The mean absolute error (MAE), root mean square error (RMSE), Kendall's Tau (τ), and r² are shown. These results correspond to those submitted for the competition.

^aThe uncertainty reported for MAE is the average of the absolute uncertainties.

^bThe uncertainty reported for RMSE is the uncertainty of the RMSE with the experimental and calculated uncertainties.

TEMOA is structurally different from OA because of the substitution of four hydrogens around the portal to the binding pocket of OA with four methyl groups. While the same guests bound to TEMOA and OA with similar binding energies, G7 weakly binds to TEMOA relative to the other guests whereas it binds stronger to OA experimentally. Binding free energy predictions using the submitted methods for the TEMOA host are reported in Table 8.3. Similar to OA, all three methods overestimated the binding free energies relative to experiment. RI-B3PW91-D3 overestimated the binding free energies with an MAE of 38.83kcal mol⁻¹. Of the three methods considered, the MMPBSA method yielded better binding free energies, both quantitatively (MAE of 5.9 ± 0.2 kcal mol⁻¹) and qualitatively ($\tau = 0.79$), than the QM-based calculations. MMPBSA ranked TEMOA-G0 and TEMOA-G1 as the third and fourth strongest bound complexes, respectively. Additionally, MMPBSA predicted that TEMOA-G4 and TEMOA-G2 were the most tightly bound complexes while TEMOA-G7 and TEMOA-G4, TEMOA-G2, and TEMOA-G3 were the first, second, and fifth most tightly bound complexes, respectively. Like MMPBSA, RI-B3PW91-D3 predicted that TEMOA-G5 was a weakly bound host-guest complex relative to the other TEMOA host-guest systems. RI-B3PW91 correctly predicted TEMOA-G0 as the third strongest bound host-guest complex and yielded the lowest deviation from experiment (0.41kcal mol⁻¹) for TEMOA-G2.



Figure 8.6: Structures of the TEMOA guest molecules inside the binding pocket generated from the clustering analysis.

8.3.4 Quantum Mechanical Calculations

The CB8 host-guest systems were used to probe approaches for improving the binding free energy prediction. Specifically, the effects of 1) utilizing truncated correlation consistent basis sets as opposed to standard correlation consistent basis sets; 2) utilizing traditional DFT calculations (neglecting the RI approximation); and 3) modifying the dielectric constant used in the continuum solvation model to reflect the ionic strength of the solution used in experiment were examined.

As shown in Tables 8.1-8.3, for CB8, OA without the statistical outlier (OA-G2), and TEMOA, the MAE, and RMSE increased by approximately 19.4, 25.5, and 32.6 kcal mol⁻¹ when using Grimme's D3 dispersion with RI-B3PW91, respectively, away from experiment. However, when using Grimme's D3 dispersion, the τ value decreases from 0.05 to -0.14 for CB8 but increases from -0.05 to 0.71 when the statistical outlier is removed for OA and increases from -0.14 to 0.57 for TEMOA. This shows the importance of using a dispersion correction for qualitative ranking of binding affinities.

The binding free energies as a result of utilizing truncated basis sets individually and extrapolated to the Kohn-Sham limit with a two-point extrapolation using cc-pVDZ and cc-pVTZ (cc-pV ∞ Z[D,T]) and a three-point extrapolation to the using cc-pVDZ and truncated triple and quadruple correlation consistent basis sets, cc-pVTZ(-1*d*) and cc-pVQZ(-1*f*2*d*), denoted as cc(0,-1,-2), are reported in Tables 8.4 and 8.5, respectively.

		B3PW9	B3PW91-D3 (SMD, ε=78.4)		RI-B3P	W91-D3	(SMD, ε=78.4)	RI-B3P	RI-B3PW91-D3 (SMD, ε=76.4)		
Complex	Evn	ΤZ	Т7	QZ	TZ	Т7	QZ	TZ	Т7	QZ	
Complex	Ехр	(- 1 <i>d</i>)	IZ	(-1f2d)	(- 1 <i>d</i>)	IZ	(-1f2d)	(- 1 <i>d</i>)	1Z	(-1f2d)	
CB8-G0	$\textbf{-6.69} \pm 0.05$	-49.85	-49.91	-49.27	-49.84	-49.89	-49.25	-49.84	-49.82	-36.26	
CB8-G1	-7.65 ± 0.04	-54.54	-57.22	-56.61	-57.21	-57.22	-56.61	-57.21	-57.24	-56.62	
CB8-G2	$\textbf{-7.66} \pm 0.05$	-37.32	-36.86	-36.39	-36.82	-36.86	-36.39	-36.82	-36.87	-36.40	
CB8-G3	$\textbf{-6.45}\pm0.06$	-45.01	-44.54	-44.38	-44.51	-44.53	-44.38	-44.51	-44.55	-44.40	
CB8-G4	$\textbf{-7.80}\pm0.04$	-69.19	-68.10	-67.50	-68.07	-68.09	-67.49	-68.07	-68.12	-67.52	
CB8-G5	$\textbf{-8.18}\pm0.05$	-36.17	-16.10	-35.53	-35.89	-35.92	-35.52	-35.89	-35.95	-35.54	
CB8-G6	$\textbf{-8.34}\pm0.05$	-31.95	-31.96	-31.63	-31.93	-31.95	-31.62	-31.95	-31.97	-31.64	
CB8-G7	$\textbf{-10.00}\pm0.10$	-14.92	-14.95	-12.89	-14.88	-14.90	-12.89	-14.91	-14.92	-12.91	
CB8-G8	-13.50 ± 0.04	-50.61	-27.26	-49.89	-50.30	-50.34	-49.90	-50.30	-50.36	-49.92	
CB8-G9	$\textbf{-8.68} \pm 0.08$	-37.31	-19.22	-36.73	-37.05	-37.07	-36.71	-37.05	-37.09	-36.74	
CB8-G10	-8.22 ± 0.07	-42.27	-15.29	-38.92	-39.28	-39.30	-38.90	-39.28	-39.32	-38.91	
CB8-G11	$\textbf{-7.77}\pm0.05$	-28.63	-10.21	-25.37	-25.74	-25.75	-25.36	-25.74	-25.80	-25.41	
CB8-G12	$\textbf{-9.86} \pm 0.03$	-62.53	-62.08	-61.43	-61.99	-62.05	-61.40	-61.99	-62.07	-61.41	
CB8-G13	$\textbf{-7.11}\pm0.03$	-52.30	-51.72	-50.03	-51.73	-44.04	-50.00	-51.74	-51.75	-50.04	
MAE		35.33	27.68	34.19	34.81	34.29	34.18	34.81	34.85	33.27	
RMSE		37.96	33.79	37.03	37.56	36.99	37.02	37.56	37.6	36.13	
au		-0.14	-0.21	-0.12	-0.12	-0.14	-0.12	-0.12	-0.12	-0.08	
r^2		0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

Table 8.4: Binding free energies for the CB8 complexes in kcal mol^{-1} with schemes involving not using the RI approximation, and changing the dielectric constant of the implicit solvent with the truncated correlation consistent basis sets for hydrogen.

The mean absolute error (MAE), root mean square error (RMSE), Kendall's Tau (τ), and r² are shown.
		B3PW91-D3 (SMD, ε=78.4)		RI-B3PW91-D3 (SMD, ε=78.4)		RI-B3PW91-D3 (SMD, ε=76.4)	
Complex	Exp	cc-pV∞Z [D, T]	cc(0,-1,-2)	cc-pV∞Z [D, T]	cc(0,-1,-2)	$cc-pV\infty Z[D,T]$	cc(0,-1,-2)
CB8-G0	$\textbf{-6.69}\pm0.05$	-49.91	-47.62	-49.89	-47.58	-49.82	-16.15
CB8-G1	$\textbf{-7.65}\pm0.04$	-57.22	-60.08	-57.22	-55.85	-57.24	-55.88
CB8-G2	$\textbf{-7.66} \pm 0.05$	-36.86	-35.25	-36.86	-36.00	-36.87	-36.04
CB8-G3	$\textbf{-6.45} \pm 0.06$	-44.54	-43.50	-44.53	-44.20	-44.55	-44.26
CB8-G4	$\textbf{-7.80}\pm0.04$	-68.10	-64.83	-68.09	-66.23	-68.12	-66.27
CB8-G5	-8.18 ± 0.05	-16.10	-34.89	-35.92	-35.28	-35.95	-35.32
CB8-G6	$\textbf{-8.34}\pm0.05$	-31.96	-31.33	-31.95	-31.32	-31.96	-31.34
CB8-G7	-10.00 ± 0.10	-14.95	-11.27	-14.90	-11.30	-14.92	-11.31
CB8-G8	-13.50 ± 0.04	-27.26	-24.47	-50.34	-49.31	-50.36	-49.38
CB8-G9	-8.68 ± 0.08	-19.22	-36.14	-37.07	-36.50	-37.09	-36.58
CB8-	-8.22 ± 0.07	-15.29	-33.97	-39.3	-38.63	-39.32	-38.67
G10							
CB8-	$\textbf{-7.77}\pm0.05$	-10.21	-20.40	-25.75	-24.88	-25.80	-25.02
G11							
CB8-	$\textbf{-9.86} \pm 0.03$	-62.08	-60.01	-62.05	-60.67	-62.07	-60.70
G12							
CB8-	-7.11 ± 0.03	-51.72	-47.18	-44.04	-37.82	-51.75	-40.12
G13							
MAE		27.68	30.93	34.29	32.69	34.85	30.65
RMSE		33.79	34.71	36.99	35.56	37.6	34.12
τ		-0.21	-0.34	-0.14	-0.12	-0.12	-0.01
r^2		0.07	0.15	0.00	0.00	0.00	0.02

Table 8.5: Binding free energies for the CB8 complexes in kcal mol^{-1} with schemes involving not using the RI approximation, changing the dielectric constant of the implicit solvent, and two options for basis set choice when extrapolating to the Kohn-Sham limit.

These options are cc-pV ∞ Z [D, T], which use cc-pVDZ and cc-pVTZ to extrapolate to the Kohn-Sham limit, and cc(0,-1,-2), which uses cc-pVDZ, cc-pVTZ(-1*d*), and cc-pVQZ(-1*f*2*d*) to extrapolate to the Kohn-Sham limit. The binding energies obtained with RI-B3PW91-D3 (SMD, ε =78.4)/cc-pV ∞ Z [D, T] were submitted. The mean absolute error (MAE), root mean square error (RMSE), Kendall's Tau (τ), and r² are shown.

For the CB8 complexes in Table 8.4, using standard DFT (B3PW91-D3) yielded a MAE of 35.33 kcal mol⁻¹ and 34.19 kcal mol⁻¹ with cc-pVTZ(-1*d*) and cc-pVQZ(-1*f*2*d*), respectively, while RI-DFT (RI-B3PW91-D3) yielded a MAE of 34.81 and 34.18 kcal mol⁻¹ for cc-pVTZ(-1*d*) and cc-pVQZ(-1*f*2*d*), respectively. When changing ε from 78.4 for pure water to 76.4 to account for the ionic strength of the solution (RI-B3PW91-D3 (ε =76.4)), all metrics (MAE, RMSE, τ , and r²) used to gauge the method's predictive qualities for the binding free energies did not significantly change with respect to the binding free energies predicted in pure water (RI-B3PW91-D3 (ε =78.4)).

Table 8.5 shows the predicted binding free energies for B3PW91-D3 (ε =78.4), RI-B3PW91-D3 (ε =78.4), and RI-B3PW91-D3 (ε =76.4) at the Kohn-Sham limit using cc-pV ∞ Z[D,T], a two-point extrapolation using cc-pVDZ and cc-pVTZ, and cc(0,-1,-2), a three-point extrapolation using cc-pVDZ, cc-pVTZ(-1*d*) and cc-pVQZ(-1*f*2*d*) for the CB8 complexes. Using the cc(0,-1,-2) basis set choice for extrapolation, the binding free energies predicted by RI-B3PW91-D3 (ε =78.4) and RI-B3PW91-D3 (ε =76.4) lowered the MAE by approximately 1.6kcal mol⁻¹, and 4.2kcal mol⁻¹, respectively, in regards to using the cc-pV ∞ Z[D,T] scheme.



Figure 8.7: Plots for calculated v. experimental results in kcal mol^{-1} for (a) CB8 (b) OA, and (c) TEMOA for MMPBSA (blue), RI-B3PW91-D3 (black), and RI-B3PW91 (green). The dashed lines in each corresponding color refers to the best fit line where the statistical outlier (OA-G2) is removed for (b) and (c). The dashed gray line is the y = x line.



Figure 8.8: Error plots from experimental results in kcal mol^{-1} for (a) CB8 (b) OA, and (c) TEMOA for MMPBSA (blue), RI-B3PW91-D3 (black), and RI-B3PW91 (green).

8.4 Discussion

8.4.1 Submission Analysis

For the methods submitted to the SAMPL6 competition, using RI-B3PW91-D3 yielded higher τ values for OA and TEMOA than using RI-B3PW91 for predicting binding free energies. Since there are eight guests that are bound to OA and TEMOA, τ_{crit} for α =0.05 is 0.57 for 8 data points. Only MMPBSA correlates with experiment ($|\tau| > \tau_{crit}$), as the τ values are 0.64, 0.29, and -0.21 for MMPBSA, RI-B3PW91-D3, and RI-B3PW91, respectively. However, after removing the statistical outlier, OA-G2, from the dataset, τ increases from 0.29 to 0.71, which implies that RI-B3PW91-D3 also correlates with experiment. As shown in Table 8.2, RI-B3PW91-D3 ranked the binding free energies more correctly than MMPBSA when the outlier is excluded. For TEMOA, both MMPBSA and RI-B3PW91-D3 correlate with experiment with τ values of 0.79 and 0.57, respectively, which are greater than τ_{crit} .

As shown in Figure 8.7a, there is no correlation between experimental and predicted binding free energies for the CB8 host-guest systems. This is supported by $r^2 \approx 0$ and τ values of -0.19, -0.14, 0.12 for MMPBSA, RI-B3PW91-D3, and RI-B3PW91, respectively, which are smaller in magnitude than τ_{crit} for α =0.05 for 14 data points, which is 0.36. This also shows an inconsistency when using Grimme's dispersion correction, which may be due to the abundance of N and O atoms present in the CB8 host and empirical descriptors for those atoms. For all sets of the host-guest systems, RI-B3PW91 had a lower MAE and RMSE than RI-B3PW91-D3 by approximately 19.4-32.6kcal mol⁻¹, but as a tradeoff, resulted in qualitatively better predictions of the binding affinities (Figure 8.8). This implies that using a dispersion correction overbinds the guest to the host but is needed for proper ranking.

To estimate the relative performance of the methods, the mean signed error (MSE) was used to offset the calculated binding free energies. After the removal of MSE from the MMPBSA and RI-B3PW91-D3 predicted binding free energies for OA and TEMOA, the MAE and the RMSE values are recalculated to estimate the performance of methods in relative terms as shown in Table 8.6. This correction improved the MAE and RMSE for MMPBSA by 6.8 and 5.9 kcal mol⁻¹ for OA and TEMOA, respectively. The correction improved RI-B3PW91-D3 MAE and RMSE by 38.39 and 38.83 kcal mol⁻¹ for OA without the OA-G2 outlier and TEMOA, respectively.

Table 8.6: Predicted binding energies for OA and TEMOA using MMPBSA and RI-B3PW91 after the removal of mean signed error (MSE).

	(DA	TEMOA		
	MMPBSA	RI-B3PW91-D3	MMPBSA	RI-B3PW91-D3	
MAE	1.6 ± 0.2^{a}	11.66 [2.81]	$3.0\pm0.2^{\mathrm{a}}$	3.49	
RMSE	$1.9\pm0.4^{ m b}$	17.87 [3.12]	3.7 ± 0.5^{b}	3.95	
au	0.64	0.29 [0.71]	0.79	0.57	
r^2	0.84	0.44 [0.52]	0.86	0.55	

Bracketed values indicate the values after the removal of the statistical outlier (OA-G2). The mean absolute error (MAE) inkcal mol⁻¹, root mean square error (RMSE) inkcal mol⁻¹, Kendall's Tau (τ), and r² are shown.

^aThe uncertainty reported for MAE is the average of the absolute uncertainties.

^bThe uncertainty reported for RMSE is the uncertainty of the RMSE with the experimental and calculated uncertainties.

8.4.2 Impact of Truncated Basis Sets

For the QM calculations, the subset of the CB8 host-guest systems was chosen because the size of these systems is smaller compared to the octa-acid host-guest systems investigated. While using the RI approximation, lowering ε from 78.4 for pure water to 76.4 to account for the ionic strength of the solution increased the MAE by 0.56kcal mol⁻¹. However, altering the dielectric constant from 78.4 to 76.4 to account for the ionic strength of the solution lowered the MAE from 34.85 to 30.65 kcal mol⁻¹ for the three-point extrapolation with truncated triple- ζ and quadruple- ζ correlation consistent basis sets, yet for RI-B3PW91-D3 (ε =78.4), the MAE only decreased from 34.29 to 32.69 kcal mol⁻¹ (Table 8.5). Therefore, factors that can change the dielectric constant should be considered when using implicit solvent models for binding free energy predictions.

The use of the cc(0,-1,-2) basis set scheme lowered the MAE for CB8 complexes by 1.60 kcal mol⁻¹ relative to using cc-pV ∞ Z[D,T] (Table 8.5) for RI-B3PW91-D3 (ε =78.4). In contrast, when using truncated basis sets and standard basis sets for binding free energies (Table 8.4), the MAE decreased by 0.51 kcal mol⁻¹ for the CB8 complexes when using cc-pVTZ as opposed to cc-pVTZ(-1*d*) for RI-B3PW91-D3 (ε =78.4). The MAE decreased by 0.31 kcal mol⁻¹ when

increasing the basis set quality of truncated basis sets for RI-B3PW91-D3 (ε =78.4). Therefore, within the RI approximation, the decrease in MAE when using cc-pVQZ(-1*f*2*d*) highlights the importance of using higher quality basis sets when extrapolating to the Kohn-Sham limit.

For predictions without the RI approximation, the binding free energies determined using B3PW91-D3/cc-pVTZ yielded a decrease in the MAE by 7.65 kcal mol⁻¹ relative to B3PW91-D3/cc-pVTZ(-1*d*) as shown in Table 8.4. This is believed to be a result from including the fourcenter two-electron electron repulsion integrals removed via the RI approximation and the need for additional polarization when describing interactions with hydrogens between the host and the guest. This effect also contributes to the increase of 3.25 kcal mol⁻¹ in the MAE between B3PW91-D3/cc-pV ∞ Z[D,T] and B3PW91-D3/cc(0,-1,-2). However, as shown in Table 8.5, when employing truncated basis sets (cc(0,-1,-2)), binding free energy predictions when using RI-B3PW91-D3 (ε =76.4) are more positive and yield a MAE of 0.28 kcal mol⁻¹ lower than B3PW91-D3 (ε =78.4). This illustrates that within the RI approximation, changing the dielectric constant is as beneficial to predicting binding free energies as utilizing standard DFT, which is more computationally costly than RI-DFT.

For the CB8-G6 host-guest complex, which was one of the smaller systems in the set of hostguest systems, the number of basis functions decreased from 4016 to 3696 with the truncation of 1 d basis function from the cc-pVTZ basis set for hydrogen and decreased from 7640 to 6872 with the truncation of 1 f and 2 d basis functions from the cc-pVQZ basis set for hydrogen. Since DFT scales approximately N^3 to N^5 depending on the complexity of the functional where N is the number of basis functions, truncated basis sets become a practical option for further decreasing the computational cost while improving the quantitative prediction of binding free energies for these host-guest systems as truncating 1 d basis function from cc-pVTZ only affected the binding energy predicted with cc-pVTZ by ≤ 0.06 kcal mol⁻¹ as shown in Table 4 for RI-B3PW91-D3.

8.4.3 Impact of the Extrapolation Scheme *B*-parameter

Another factor that can account for the large deviations between host-guest binding energies is the parameter used to fit Equation 8.1 for two-point extrapolations. The value of 5.5 proposed by Jensen for the *B*-parameter, which was used for atoms and diatomics, caused the extrapolation curve to converge at a very rapid rate and is reflected in the predictions for the CB8 complexes, as the binding affinities in Table 8.1 are identical to those predicted with the cc-pVTZ basis set with the respective method in Table 8.4. Also, when using the three-point extrapolations with truncated basis sets for the CB8 complexes, the *B*-parameter yielded an average value of 0.37 (Table 8.8). Therefore, the value of 0.37 for the *B*-parameter was applied to two-point extrapolations with ccpVDZ and cc-pVTZ to gauge how changing the *B*-parameter affected the extrapolated binding free energies (Table 8.7). The results from using 0.37 as the *B*-parameter in a two-point extrapolation show that the MAE decreased by 0.84 and 0.42 kcal mol⁻¹ for the CB8 and TEMOA complexes, respectively. The MAE did not change for the OA complexes. Setting the *B*-parameter to 0.37 did not change the τ values for CB8 and OA complexes, however, did increase the τ value from 0.57 to 0.71 for TEMOA.

In addition to applying 0.37 for the *B*-parameter to predict binding free energies for all host-guest systems using two-point extrapolations with cc-pVDZ and cc-pVTZ, the value of the *B*-parameter was optimized to the value of 0.12 via minimizing the MAE and was applied (Table 8.7). For the CB8 host-guest systems, shifting the *B*-parameter from 5.5 to 0.12 had a noticeable impact on the MAE, which decreased from 34.29 to 29.84 kcal mol⁻¹ for RI-BWPW91-D3. A similar effect was observed for TEMOA with a decrease in the MAE of 5.07kcal mol⁻¹. There is no notable change in MAE, RMSE, or τ for the OA complexes with the change in the *B*-parameter. Furthermore, τ increases from 0.57 to 0.93 when the *B*-parameter is changed from 5.5 to 0.12 for TEMOA with RI-B3PW91-D3, which provides more evidence that dispersion-corrected functionals should be used for qualitative predictions of binding free energies since $|\tau| > \tau_{crit}$. The observed trends imply that the value of the *B*-parameter should be reoptimized when using Equation 8.1 for macromolecules.

		B=5.5	B=0.37	B=0.12
	MAE	34.29	33.45	29.84
CD 9	RMSE	36.99	36.33	33.34
CD0	au	-0.14	-0.14	-0.03
	r^2	0.00	0.00	0.00
	MAE	35.46 [38.39]	35.46 [38.42]	35.43 [38.74]
	RMSE	36.41 [38.52]	36.43 [38.54]	36.70 [38.86]
0A	au	0.29 [0.71]	0.29 [0.71]	0.29 [0.71]
	r^2	0.44 [0.52]	0.43 [0.52]	0.43 [0.54]
	MAE	38.83	38 41	33.76
	RMSE	39.03	38.60	36.30
TEMOA	τ	0.57	0.71	0.93
	r^2	0.55	0.75	0.58

Table 8.7: Predicted binding energies when using different values for *B* in Equation 8.1 for two-point extrapolations using cc-pVDZ and cc-pVTZ with RI-B3PW91-D3.

Bracketed values indicate the values after the removal of the statistical outlier (OA-G2). The mean absolute error (MAE) inkcal mol⁻¹, root mean square error (RMSE) inkcal mol⁻¹, Kendall's Tau (τ), and r² are shown.

Compared to other submissions employing QM methods in the SAMPL6 Host-Guest binding challenge, our approach yielded quantitatively poorer predictions that may have resulted from the approximations considered in this work. In our approach, only a single conformational state of the guest binding to the host system was considered. Additionally, the representative structures of the individual host-guest systems obtained from clustering the MD trajectories were not optimized with QM methods and is reflected in our model chemistries.

8.4.4 Impact of Representative Geometries

The representative geometries had a notable impact on the binding free energies. For example, the orientation of the substituted cyclohexene ring relative to the OA host might be the potential cause of OA-G2 being a statistical outlier (Figure 8.5). Comparing OA-G2 and TEMOA-G2 in Figures 8.5 and 8.6, where the only difference is the four methyl groups on the host, the structure of the OA-G2 complex has a smaller binding pocket than the TEMOA-G2 complex. While the experimental data suggests that G2 has a stronger binding affinity towards OA than TEMOA, MMPBSA suggests the

opposite. More sampling of representative structures would aid in depicting whether the anomalous binding behavior of OA-G2 correlates with the positive binding free energies predicted with DFT.

Although the only difference between CB8-G6 and CB8-G7 was the expansion of the ring for the guest by one CH2 group, the predicted binding affinities for the CB8-G6 and CB8-G7 complexes differed by approximately 17.0kcal mol⁻¹. This may be due to the binding poses of CB8-G6 and CB8-G7 complexes, as G6 bound in a perpendicular fashion inside the binding pocket relative to the host whereas G7 bound in a parallel fashion inside the binding pocket. This would affect nearby electrostatic interactions and why for B3PW91-D3 (ε =78.4), RI-B3PW91-D3 (ε =78.4), and RI-B3PW91-D3 (ε =76.4), there was a 3.00 kcal mol⁻¹ difference in the change of binding energies between CB8-G6 and CB8-G7 when improving basis set quality via the basis set scheme used for extrapolation (Table 8.5). Ergo, more sampling of chemically relevant structures or enhanced sampling methods can provide a more robust depiction of the host-guest binding environment.

However, these two methods do not correlate to the CB8 binding free energies since the τ values are -0.19 and -0.14 for MMPBSA and RI-B3PW91-D3, respectively. This may result from insufficient sampling as the CB8 guests are larger molecules with higher conformational flexibility. For example, the size of CB8-G4 does not allow the guest to fit entirely into the binding cavity. As a result, most of the CB8-G4 molecule is weakly bound to the host from outside of the binding pocket and only one of the three triethyl amines within the guest can fit into the pocket as shown in Figure 8.4. Each triethyl amine group could bind to the host from inside the binding cavity, which would result in alternative binding conformations and affect the overall binding free energy. To better understand binding free energies of these large structures, more sampling of the different binding modes is needed to generate weighted averages based on the thermodynamic stability of predicted poses.

The results for OA and TEMOA systems illustrate that MMPBSA and RI-B3PW91-D3 methods can be used to qualitatively rank binding energies of small molecules. Among those two methods, MMPBSA is computationally less expensive, but RI-B3PW91-D3 predicted the relative binding affinities better for OA and TEMOA host-guest systems. However, the MAE and the corresponding error plots (Figure 8.8) indicate that both methods overestimated the binding free energies. The MAE reported for the OA and TEMOA complexes state that MMPBSA and RI-B3PW91-D3 predict overbinding by 6.8 and 35.5kcal mol⁻¹, respectively, for OA complexes and 5.9 and 38.8kcal mol⁻¹, respectively, for TEMOA complexes. For all systems, the MMPBSA method was the best approach overall in terms of quantitative predictions.

8.5 Conclusions

When implementing DFT for predicting host-guest binding affinities, the use of Grimme's D3 dispersion correction was essential for qualitatively predicting the binding free energies for the OA and TEMOA systems even though the MAE exceeded 35.0 kcal mol⁻¹ for both the OA and TEMOA systems. When using implicit solvent models, factors that can change the dielectric constant, such as the ionic strength of the solution, are relevant for predicting binding free energies, as lowering the dielectric constant lowered the MAE. While RI-B3PW91-D3 reduced the computational cost relative to B3PW91-D3, B3PW91-D3 yielded a lower MAE. To attain more quantitatively favorable results, using cc-pVQZ(-1f2d) for hydrogen atoms reduces the computational cost relative to using cc-pVQZ while simultaneously providing a better standard for extrapolating to the Kohn-Sham limit than only utilizing cc-pVDZ and cc-pVTZ for extrapolations. Also, truncating 1 d basis function for hydrogen atoms had a very small effect on predicted binding free energies obtained with cc-pVTZ, indicating that truncated basis sets are a viable option to reduce the computational cost while yielding near-identical binding free energies. With the extrapolation scheme utilized, the B-parameter should be revised for macromolecules since reducing the value of the B-parameter from the proposed 5.5 to 0.12 reduced the MAE while providing extrapolated binding energies that were in alignment with those predicted using quadruple- ζ level basis sets. Sampling of different binding poses becomes pertinent for future investigations as binding orientation in the pocket affected the predicted binding free energies by approximately 17.0 kcal mol⁻¹ when using RI-B3PW91-D3 for guests that only differed by one CH₂ group.

All methods presented predict overbinding character for these host-guest systems except for

RI-B3PW91 for CB8 host-guest systems. MMPBSA and RI-B3PW91-D3 worked well at ranking binding affinities for smaller guests regardless of the size of the host. The CB8 guest molecules with a larger van der Waals volume yielded poor prediction of binding free energy due to their higher conformational flexibility, which can complicate predicting binding poses. To better understand binding free energies of these large structures, enhanced sampling methods can be used, and multiple host-guest binding poses can be sampled.

APPENDIX

APPENDIX

Table 8.8: Fitting parameter values obtained when using Jensen's extrapolation scheme for each component in calculating the binding energy (Equation 8.1). The host and guest are counterpoise-corrected before the extrapolation was performed.

Complex	Complex	Host	Guest
CB8-G0	0.37	0.36	0.41
CB8-G1	0.36	0.35	0.37
CB8-G2	0.36	0.36	0.37
CB8-G3	0.36	0.36	0.37
CB8-G4	0.32	0.32	0.34
CB8-G5	0.38	0.38	0.39
CB8-G6	0.39	0.39	0.40
CB8-G7	0.38	0.38	0.40
CB8-G8	0.37	0.37	0.39
CB8-G9	0.39	0.39	0.39
CB8-G10	0.38	0.38	0.38
CB8-G11	0.39	0.39	0.40
CB8-G12	0.36	0.35	0.37
CB8-G13	0.39	0.38	0.40
Average	0.37	0.37	0.38



Figure 8.9: Plots for the correlations calculated after the mean signed errors are removed from the results in Tables 8.1-8.3 versus experimental results in kcal mol^{-1} for (a) OA, and (b) TEMOA for MMPBSA (blue), RI-B3W91-D3 (black). The dashed lines in each corresponding color refers to the best fit line where the statistical outlier (OA-G2) for RI-B3PW91-D3 is removed for (a). The dashed gray line corresponds to the y=x line.

- [1] Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120–127.
- [2] Shan, Y.; Seeliger, M. A.; Eastwood, M. P.; Frank, F.; Xu, H.; Jensen, M. O.; Dror, R. O.; Kuriyan, J.; Shaw, D. E. A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc. Natl. Acad. Sci.* 2009, *106*, 139–144.
- [3] Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* 2013, 497, 643–646.
- [4] Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.* 2015, *31*, 64–74.
- [5] Walkowicz, W. E.; Fernández-Tejada, A.; George, C.; Corzana, F.; Jiménez-Barbero, J.; Ragupathi, G.; Tan, D. S.; Gin, D. Y. Quillaja saponin variants with central glycosidic linkage modifications exhibit distinct conformations and adjuvant activities. *Chem. Sci.* 2016, 7, 2371–2380.
- [6] Hadden, J. A.; Perilla, J. R.; Schlicksup, C. J.; Venkatakrishnan, B.; Zlotnick, A.; Schulten, K. All-atom molecular dynamics of the HBV capsid reveals insights into biological function and cryo-EM resolution limits. *Elife* 2018, 7, e32478.
- [7] García, M. A.; Meurs, E. F.; Esteban, M. The dsRNA protein kinase PKR: Virus and cell control. *Biochimie* 2007, 89, 799–811.
- [8] Tripathi, R. B.; Pande, M.; Garg, G.; Sharma, D. In-silico expectations of pharmaceutical industry to design of new drug molecules. *J. Innov. Pharm. Biol. Sci.* **2016**, *3*, 95–103.
- [9] Ryde, U.; Söderhjelm, P. Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chem. Rev.* **2016**, *116*, 5520–5566.
- [10] Ganesan, A.; Coote, M. L.; Barakat, K. Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discov. Today* 2017, 22, 249–269.
- [11] Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46*, 531–558.
- [12] Huggins, D. J.; Sherman, W.; Tidor, B. Rational Approaches to Improving Selectivity in Drug Design. J. Med. Chem. 2012, 55, 1424–1444.
- [13] Muddana, H. S.; Daniel Varnado, C.; Bielawski, C. W.; Urbach, A. R.; Isaacs, L.; Geballe, M. T.; Gilson, M. K. Blind prediction of host-guest binding affinities: a new SAMPL3 challenge. *J. Comput. Aided. Mol. Des.* **2012**, *26*, 475–487.

- [14] Rogers, K. E.; Ortiz-Sánchez, J. M.; Baron, R.; Fajer, M.; De Oliveira, C. A. F.; McCammon, J. A. On the role of dewetting transitions in host-guest binding free energy calculations. J. Chem. Theory Comput. 2013, 9, 46–53.
- [15] Yang, H.; Yuan, B.; Zhang, X.; Scherman, O. A. Supramolecular chemistry at interfaces: Host-guest interactions for fabricating multifunctional biointerfaces. *Acc. Chem. Res.* 2014, 47, 2106–2115.
- [16] Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K. The SAMPL4 host–guest blind prediction challenge: an overview. J. Comput. Aided. Mol. Des. 2014, 28, 305–317.
- [17] Gallicchio, E.; Chen, H.; Chen, H.; Fitzgerald, M.; Gao, Y.; He, P.; Kalyanikar, M.; Kao, C.; Lu, B.; Niu, Y.; Pethe, M.; Zhu, J.; Levy, R. M. BEDAM binding free energy predictions for the SAMPL4 octa-acid host challenge. *J. Comput. Aided. Mol. Des.* **2015**, *29*, 315–325.
- [18] Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K. Overview of the SAMPL5 host–guest challenge: Are we doing better? J. Comput. Aided. Mol. Des. 2017, 31, 1–19.
- [19] Liu, S.; Ruspic, C.; Mukhopadhyay, P.; Chakrabarti, S.; Zavalij, P. Y.; Isaacs, L. The Cucurbit[n]uril Family: Prime Components for Self-Sorting Systems. J. Am. Chem. Soc. 2005, 127, 15959–15967.
- [20] Gan, H.; Benjamin, C. J.; Gibb, B. C. Nonmonotonic Assembly of a Deep-Cavity Cavitand. J. Am. Chem. Soc. 2011, 133, 4770–4773.
- [21] Biedermann, F.; Scherman, O. A. Cucurbit[8]uril Mediated Donor-Acceptor Ternary Complexes: A Model System for Studying Charge-Transfer Interactions. J. Phys. Chem. B 2012, 116, 2842–2849.
- [22] Vázquez, J.; Remón, P.; Dsouza, R. N.; Lazar, A. I.; Arteaga, J. F.; Nau, W. M.; Pischel, U. A Simple Assay for Quality Binders to Cucurbiturils. *Chem. - A Eur. J.* 2014, 20, 9897–9901.
- [23] Gibb, C. L. D.; Gibb, B. C. Binding of cyclic carboxylates to octa-acid deep-cavity cavitand. J. Comput. Aided. Mol. Des. 2014, 28, 319–325.
- [24] Nicholls, A.; Wlodek, S.; Grant, J. A. The SAMP1 Solvation Challenge: Further Lessons Regarding the Pitfalls of Parametrization *†*. *J. Phys. Chem. B* **2009**, *113*, 4521–4532.
- [25] Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Dill, K. A. Predictions of Hydration Free Energies from All-Atom Molecular Dynamics Simulations *†*. J. Phys. Chem. B 2009, 113, 4533–4537.
- [26] Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. The SAMPL2 blind prediction challenge: introduction and overview. *J. Comput. Aided. Mol. Des.* **2010**, *24*, 259–279.
- [27] Steinmann, C.; Olsson, M. A.; Ryde, U. Relative Ligand-Binding Free Energies Calculated from Multiple Short QM/MM MD Simulations. *J. Chem. Theory Comput.* **2018**, Article ASAP.

- [28] Curutchet, C.; Cupellini, L.; Kongsted, J.; Corni, S.; Frediani, L.; Steindal, A. H.; Guido, C. A.; Scalmani, G.; Mennucci, B. Density-Dependent Formulation of Dispersion-Repulsion Interactions in Hybrid Multiscale Quantum/Molecular Mechanics (QM/MM) Models. J. Chem. Theory Comput. 2018, 14, 1671–1681.
- [29] Sellers, B. D.; James, N. C.; Gobbi, A. A Comparison of Quantum and Molecular Mechanical Methods to Estimate Strain Energy in Druglike Fragments. J. Chem. Inf. Model. 2017, 57, 1265–1275.
- [30] Lu, Y.; Yang, C. Y.; Wang, S. Binding free energy contributions of interfacial waters in HIV-1 protease/inhibitor complexes. J. Am. Chem. Soc. 2006, 128, 11830–11839.
- [31] Bonnet, P.; Bryce, R. A. Molecular dynamics and free energy analysis of neuraminidase ligand interactions. *Protein Sci.* **2004**, *13*, 946–957.
- [32] Kitamura, K.; Tamura, Y.; Ueki, T.; Ogata, K.; Noda, S.; Himeno, R.; Chuman, H. Binding free-energy calculation is a powerful tool for drug optimization: Calculation and measurement of binding free energy for 7-azaindole derivatives to glycogen synthase kinase-3β. J. Chem. Inf. Model. 2014, 54, 1653–1660.
- [33] Caldararu, O.; Olsson, M. A.; Riplinger, C.; Neese, F.; Ryde, U. Binding free energies in the SAMPL5 octa-acid host–guest challenge calculated with DFT-D3 and CCSD(T). J. Comput. Aided. Mol. Des. 2017, 31, 87–106.
- [34] Sure, R.; Antony, J.; Grimme, S. Blind prediction of binding affinities for charged supramolecular host-guest systems: Achievements and shortcomings of DFT-D3. J. Phys. Chem. B 2014, 118, 3431–3440.
- [35] Mikulskis, P.; Cioloboc, D.; Andrejić, M.; Khare, S.; Brorsson, J.; Genheden, S.; Mata, R. A.; Söderhjelm, P.; Ryde, U. Free-energy perturbation and quantum mechanical study of SAMPL4 octa-acid host-guest binding energies. *J. Comput. Aided. Mol. Des.* **2014**, *28*, 375–400.
- [36] Murkli, S.; McNeil, J.; Isaacs, L. CB[8]-guest binding affinities: A blinded dataset for the SAMPL6 challenge. *Supramol. Chem.* **2018**, (Submitted).
- [37] Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree–Fock exchange. *Chem. Phys.* **2009**, *356*, 98–109.
- [38] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- [39] Mintz, B.; Lennox, K. P.; Wilson, A. K. Truncation of the correlation consistent basis sets: An effective approach to the reduction of computational cost? J. Chem. Phys. 2004, 121, 5629–5634.
- [40] Molecular Operating Environment (MOE), 2013.08. Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910. Montreal, QC. 2013.

- [41] Corbeil, C. R.; Williams, C. I.; Labute, P. Variability in docking success rates due to dataset preparation. J. Comput. Aided. Mol. Des. 2012, 26, 775–786.
- [42] Hoffmann, R. An Extended Hückel Theory. I. Hydrocarbons. J. Chem. Phys. 1963, 39, 1397– 1412.
- [43] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 2006, 65, 712–25.
- [44] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. J. Comput. Chem. 2004, 25, 1157–1174.
- [45] Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J. Comput. Chem. 2002, 23, 1623– 1641.
- [46] Case D. A.; Betz R. M.; Cerutti D. S.; Cheatham III, T. E.; Darden T. A.; Duke R. E.; Giese T. J.; Gohlke H.; Goetz A. W.; Homeyer N.; Izadi S.; Janowski P.; Kaus J.; Kovalenko A.; Lee T. S.; LeGrand S.; Li P.; Lin C.; Luchko T.; Luo R.; Madej B.; Mermelstein D.; Merz K. M.; Monard G.; Nguyen H.; Nguyen H. T.; Omelyan I.; Onufriev A.; Roe D. R.; Roitberg A.; Sagui C.; Simmerling C. L.; Botello-Smith W. M.; Swails J.; Walker R. C.; Wang J.; Wolf R. M.; Wu X.; Xiao L.; Kollman P.A. (2016), AMBER 2016, University of California, San Francisco.
- [47] Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py : An Efficient Program for End-State Free Energy Calculations. J. Chem. Theory Comput. 2012, 8, 3314–3321.
- [48] Merrick, J. P.; Moran, D.; Radom, L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. J. Phys. Chem. A 2007, 111, 11683–11700.
- [49] Neese, F. Software update: the ORCA program system, version 4.0. Wiley Interdiscip. Rev. Comput. Mol. Sci. 2018, 8, e1327.
- [50] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. J. Chem. *Phys.* **1993**, *98*, 5648–5652.
- [51] Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* **1992**, *45*, 13244–13249.
- [52] 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*, 6671–6687.
- [53] Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* **1995**, *240*, 283–290.

- [54] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. J. Phys. Chem. B 2009, 113, 6378–6396.
- [55] Goerigk, L.; Grimme, S. A general database for main group thermochemistry, kinetics, and noncovalent interactions - Assessment of common and reparameterized (meta-)GGA density functionals. J. Chem. Theory Comput. 2010,
- [56] 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*, 6670.
- [57] Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. J. Chem. Phys. **1989**, *90*, 1007–1023.
- [58] Feller, D. Application of systematic sequences of wave functions to the water dimer. *J. Chem. Phys.* **1992**, *96*, 6104–6114.
- [59] Martin, J. M. L. Ab initio total atomization energies of small molecules towards the basis set limit. *Chem. Phys. Lett.* **1996**, 259, 669–678.
- [60] Wilson, A. K.; Dunning Jr., T. H. Benchmark calculations with correlated molecular wave functions. X. Comparison with "exact" MP2 calculations on Ne, HF, H₂O, and N₂. J. Chem. Phys. 1997, 106, 8718–8726.
- [61] Feller, D.; Peterson, K. A.; Crawford, T. D. Sources of error in electronic structure calculations on small chemical systems. J. Chem. Phys. 2006, 124, 054107.
- [62] Jensen, F. Polarization consistent basis sets. II. Estimating the Kohn–Sham basis set limit. *J. Chem. Phys.* **2002**, *116*, 7372–7379.
- [63] Faver, J. C.; Zheng, Z.; Merz, K. M. Model for the fast estimation of basis set superposition error in biomolecular systems. *J. Chem. Phys.* **2011**, *135*, 144110.
- [64] Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.
- [65] Gavish, N.; Promislow, K. Dependence of the dielectric constant of electrolyte solutions on ionic concentration: A microfield approach. *Phys. Rev. E* **2016**, *94*, 012611.
- [66] Kendall, M. G. A New Measure of Rank Correlation. *Biometrika* 1938, 30, 81.
- [67] Berry, K. J.; Johnston, J. E.; Zahran, S.; Mielke, P. W. Stuart's tan measure of effect size for ordinal variables: Some methodological considerations. *Behav. Res. Methods* **2009**, *41*, 1144–1148.
- [68] Dean, R. B.; Dixon, W. J. Simplified Statistics for Small Numbers of Observations. *Anal. Chem.* **1951**, *23*, 636–638.

CHAPTER 9

CONCLUDING REMARKS

In this dissertation, several quantum chemical strategies have been shown to be effective for thermodynamic property prediction for main group and transition metal thermochemistry. This includes utilizing density functional methods predicting the pK_as of transition metal hydrides and utilizing *ab initio* composite strategies towards main group thermochemistry, vibrational potential energy surfaces, and organometallic catalysis. Applications include modeling frontier orbitals and predicting host-guest binding interactions with density functional methods as well.

In Chapter 3, a QM/QM scheme utilizing the ONIOM method was used to predict the pK_a of late transition metal hydrides with bidentate phosphine ligands. In predicting the pK_a of TM hydrides via the choice of density functional, *ab initio* method, solvation model, basis set, cavity model, and model layer size within an ONIOM scheme, the optimal scheme is one that utilizes two density functionals, one effective at describing the metal center and immediately bound atoms, and another effective at describing ligands comprised of main group atoms. This was B3LYP-D3/augcc-pVTZ:B97-D3/SDD using the SMD solvation model and default cavity model. Using *ab initio* methods underestimated the pK_a while the use of a single functional largely overestimated the pK_a . In future studies for these systems, the methodology presented can be expanded to sterically bulkier bidentate phosphine ligands for Group 10 hydrides utilized for redox potentials to gauge efficacy. In general studies involving transition metal complexes with sterically bulky ligands, this approach can be utilized to target functional efficacy for transition metal centers and main group ligands independently as different tiers of functionals yielded lower deviations from the experimental pK_a s.

In Chapters 4 and 5, the domain-based local pair natural orbital (DLPNO) methods were utilized within the correlation consistent Composite Approach (ccCA) and applied to main group and transition metal thermochemistry. DLPNO-ccCA yielded lower mean absolute deviations than ccCA for the enthalpies of formation for 119 closed shell main group molecules with a ~87% CPU time reduction relative to ccCA. As DLPNO-ccCA was implemented for linear alkanes up

to octane, the CPU time was reduced by up to 97% for octane relative to ccCA given the lower scaling of DLPNO-CCSD(T) relative to CCSD(T). DLPNO-ccCA was also successfully applied to bioorganic complexes from the S66 dataset and the coronene dimer, which marks one of the largest molecules ever examined with a composite strategy, that exhibit noncovalent interactions.

DLPNO-ccCA can therefore be used to predict thermodynamic properties of organic and bioorganic molecules typically outside the range of *ab initio* composite approaches based on molecule size such as per- and polyfluoroalkyl substances (PFAS). DLPNO-ccCA could also be utilized to examine thermodynamic properties of drug-like molecules, like pK_a or partition coefficients, as these molecules exhibit multiple hydrogen-bonding sites and more aromatic rings. Thus, DLPNO-ccCA increases the applicability of *ab initio* composite strategies for main group species based on the reduction of computer resources and computational cost.

DLPNO-ccCA was implemented to model organometallic catalysis utilizing the variant of ccCA targeting 4*d* transition metal chemistry, rp-ccCA. Denoted as DLPNO-rp-ccCA, this method was successfully applied towards hydroformylation, which is the largest volume homogeneous chemical reaction in industry for chemical production, and gas phase ligand dissociation, which targets modeling metal-ligand interactions with *ab initio* approaches. A continuation of this study would include modeling more metal-ligand interactions prevalent in organometallic catalysis with DLPNO-rp-ccCA and expanding the sample size to include more catalysts utilized in hydroformylation, where the linear isomer is favored, and asymmetric hydroformylation, where the branched isomer is favored. The DLPNO-ccCA variants for transition metals could also be applied towards 3*d* transition metals to further increase the possible molecule space for *ab initio* composite methodologies.

In Chapter 6, ccCA, B3LYP, and TPSS were utilized to generate potential energy surfaces that were then used to predict anharmonic vibrational frequencies with vibrational self-consistent field (VSCF) theory. Overall, with ccCA potentials, the mean absolute deviation for calculated frequency from experiment was lower than with DFT potentials. With DFT-generated potentials, functional choice had a more significant effect on the predicted frequency than basis set choice. A

multilevel approach that utilizes the single mode potential energy curves with DFT and the coupled vibrational modes generated with ccCA yields lower frequencies than if only DFT were utilized, which is useful for expanding to larger polyatomic systems. For aminophenol, the errors obtained with VCIPSI-PT2 were lower than those for scaled harmonics, indicating the success of utilizing this approach to characterize specific vibrations for polyatomic systems.

Future work on this project could include the investigation of astrochemical molecules with unusual binding behavior and utilizing different variants of ccCA, such as completely renormalized ccCA (CR-ccCA(2,3)) to account for bond-breaking behavior occurring in vibrational motion. This approach can also be implemented to metal-carbonyl stretching as well as uncovering vibrational behavior not accounted for by the harmonic approximation. Given the large number of electronic structure calculations involved with generating potential energy surfaces for polyatomic systems, DLPNO-ccCA could also be considered to investigate vibrational phenomena while reducing the CPU time relative to ccCA. As well, multilevel approaches can be utilized to investigate the full anharmonic mode-mode coupling potential energy surfaces to model infrared (IR) spectra that more closely resemble experimental IR spectra than using the harmonic oscillator approximation.

For Chapter 7, calculations were done to complement synthesis of the zinc porphyrin-fullerene supramolecular dyad ($(F_{15}P)Zn-C_{60}$) and the $C_{60}-(F_{15}P)Zn:Py$ -TTF and $C_{60}-(F_{15}P)Zn:Py$ -phTTF triads via modeling the molecular electrostatic potential and frontier orbitals with M06- $2X/6-31G^*$. For the dyads, the frontier HOMO was on the ($F_{15}P$)Zn and LUMO on the C_{60} making them the donor and acceptor sites, respectively. The HOMO was shifted to the tetrathiafulvalene site without altering the location of the LUMO for the triads.¹ For modeling electronic structure and frontier orbitals for supramolecular dyads useful in artificial photosynthesis, time-dependent density functional theory (TDDFT) combined with implicit solvent models can be used to model UV-Vis absorption spectra to verify observed photochemical phenomena for these systems, such as the transition at ~400 nm indicating transitions occurring at the porphyrin.

In Chapter 8, molecular dynamics and DFT methods were used to predict the binding interaction energies of biological host-guest systems for the sixth Statistical Assessment of Modeling Protein and Ligands (SAMPL) blind prediction competition.² Modeling the host-guest systems with RI-B3PW91-D3 predicted qualitative ranking of binding affinity to each of the hosts, exhibited by the Kendall's tau (τ) statistic while predicting binding energies tens of kcal mol⁻¹ away from experimental binding interaction energies. In the future, more orientations could be sampled and binding poses obtained through molecular dynamics simulations could be optimized with density functional theory. As well, different density functionals can be utilized to evaluate the binding interactions of similar systems to provide a gauge for appropriate functionals for hostguest binding. This would provide a linear regression technique that can be implemented to predict binding interaction free energies. In a similar vein, regression-based machine learning approaches can be used with parameters inputted from molecular dynamics and electronic structure calculations, opening a new avenue for thermodynamic property prediction.

- [1] Obondi, C. O.; Lim, G. N.; Jang, Y.; Patel, P.; Wilson, A. K.; Poddutoori, P. K.; D'Souza, F. Charge Stabilization in High-Potential Zinc Porphyrin-Fullerene via Axial Ligation of Tetrathiafulvalene. *J. Phys. Chem. C* **2018**, *122*, 13636–13647.
- [2] Eken, Y.; Patel, P.; Díaz, T.; Jones, M. R.; Wilson, A. K. SAMPL6 host-guest challenge: Binding free energies via a multistep approach. *J. Comput. Aided. Mol. Des.* **2018**, *32*, 1097–1115.