TRUTH AND TRACTABILITY: COMPROMISING BETWEEN ACCURACY AND COMPUTATIONAL COST IN QUANTUM COMPUTATIONAL CHEMISTRY METHODS FOR NONCOVALENT INTERACTIONS AND METAL-SALEN CATALYSIS

A Thesis
Presented to
The Academic Faculty

by

Tait Takatani

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
August 2010

TRUTH AND TRACTABILITY: COMPROMISING BETWEEN ACCURACY AND COMPUTATIONAL COST IN QUANTUM COMPUTATIONAL CHEMISTRY METHODS FOR NONCOVALENT INTERACTIONS AND METAL-SALEN CATALYSIS

Approved by:

Professor C. David Sherrill, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Jean-Luc Brédas
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Kenneth R. Brown
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Peter J. Ludovice
School of Chemical and Biomolecular
Engineering
Georgia Institute of Technology

Professor Thomas M. Orlando
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: 28 June 2010

To my love, my life, my lovely wife, Yue.

Together we can achieve anything.

PREFACE

The advent of the computer has undeniably changed the way people live, work, and play in the modern world. It has become common for the average person to own computers with performances comparable to that of supercomputers built in the 1980's. This increase of computational power has similarly affected the sciences by allowing for robust simulations that support research hypotheses. Remarkably, one can feasibly compute publication quality results by utilizing only personal computers. As such, much of current scientific research includes contributions from theory, experiment, and computation.

While the dramatic progress in the computational sciences shows great promise, many computer simulations are inherently built on approximations of real-world situations. It remains impossible, even for present-day supercomputers, to obtain 100% accuracy for all types of simulations. Therefore, we must understand the limitations of the computational models and be able to assess the magnitude of the errors incurred by the approximations. This becomes more and more imperative with the growing number of computational programs being implemented for scientific research.

Within the field of computational chemistry, there exists a large number of quantum computational methods. In theory, certain methods provide the exact solution to the electronic Schrödinger equation. However, these methods are generally intractable for molecules greater than three atoms. As such, computational chemists apply methods that are built upon many more approximations to drastically reduce the computational costs. This thesis focuses on the performance of those approximate methods when used to describe noncovalent interactions and metal-salen energetics.

In many cases, the approximate methods fail to obtain even qualitatively correct results. As a solution, new trends in empirically scaled quantum computational methods will also be presented. One will discover that reliable results can be obtained for very large systems with tractable computational methods, but only if the proper methods are chosen.

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Dr. C. David Sherrill, for all of his guidance and patience. I began my work with Dr. Sherrill as an undergraduate student in a summer REU program. The first assignment I was given was to explore the potential energy surface of nitrogen oxide. Although it seems simple enough, just vary the distance between the nitrogen and oxygen atoms, it turns out to be one of the most difficult diatomic molecules to properly characterize (resulting in very little literature on the subject). I, disappointingly, was not able to finish the project either. Without being disheartened, I joined the Sherrill group as a graduate student in 2006, and for the past four years Dr. Sherrill has been the best advisor I could ever imagine.

Next, I would like to thank the Sherrill group (past and present members): Dr. John Sears for his guidance in the metal-salen project, the system administrators (Michael Marshall, Steve Arnstein, and Berhane Temelso) for making my life dealing with computers much easier, and Edward Hohenstein for our many collaborations together.

My family has always supported my educational endeavors. I can't thank my parents, Richard and Mae Takatani, enough for all of their help. When I first moved to Atlanta, my parents came with me to ensure that I could move all of my belongings and that I have furniture in my apartment. They also drove from Maryland to Atlanta, on more than one occasion, just to help me move apartments. To my siblings, Sarah and Timothy, I thank you for your financial support during my undergraduate career. And to my sister, Tracie, I will always be grateful that you made sure I had a life outside of research and always a place to go for vacation. I love all of you.

Finally, I could not have done as well without the newest motivation in my life. I

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Computational chemists are concerned about two aspects when choosing between the myriad of theoretical methodologies: the accuracy (the "truth") and the computational cost (the tractability). Among the least expensive methods for the approximation of the electronic Schrödinger equation are the Hartree-Fock (HF), density functional theory (DFT), and second-order Møller-Plesset perturbation theory (MP2) methods. While each of these methods yield excellent results in many cases, the inadequate inclusion of certain types of electron correlation (either high-orders or nondynamical) can produce erroneous results. For example, when utilizing the HF or DFT methods for the computation of noncovalent interactions, both methods fail to predict favorable London dispersion interactions. The MP2 method, on the other hand, dramatically overbinds noncovalent π-π interactions compared to the chemically accurate CCSD(T) method. Likewise, for the computation of transition metal-salen spin-state energy gaps, DFT methods yield large errors (tens of kcal mol^{-1}) compared to the robust CASPT3 method. Both CCSD(T) and CASPT3 methods, however, are limited to the computation of small to medium sized molecules with "modest-sized" basis sets (currently, around 30 atoms for single point computations).

The compromise for the computation of noncovalent interactions often comes from empirically scaling DFT and/or MP2 methods to fit benchmark data sets. The DFT method with an empirically fit dispersion term (DFT-D) often yields semi-quantitative results. The spin-component scaled MP2 (SCS-MP2) method parameterizes the same- and opposite-spin correlation energies to approximate CCSD(T) results and often yields less than 20% error for prototype noncovalent systems. Each

of these methods have the added benefit of being no more computationally expensive than their canonical counterparts. Therefore, the DFT-D and SCS-MP2 methods can be utilized on rather large systems.

There is no simple fix for cases with a large degree of nondynamical correlation (such as transition metal-salen complexes). While testing the new M0-family of meta-GGA DFT functionals on the spin-state energy gaps of transition metal-salen complexes, no DFT method produced reliable results. Therefore each metal-salen complex must be evaluated on a case-by-case basis to determine which methods are the most reliable. For the 'well-behaved' Al(III)-salen complex, all single reference methods produce reliable results because the Al(III)-salen complex contains little to no important nondynamical correlation. The salen-ligand, however, does contain a large degree of π-aromaticity. Therefore, when studying the reaction energy profile for the addition of cyanide to unsaturated imides catalyzed by two Al(Cl)-salen complexes, the BP86-D method was used for geometry optimizations followed by SCS-MP2 single point computations to accurately capture London dispersion forces. Compromising between these two methods, the entire reaction energy profile was tractable with resources found in a personal desktop computer while providing enough 'truth' that explained many experimental observations.

As an introduction to quantum computational methods, chapter one will provide a brief development of the HF, electron correlation, multireference, and DFT methods. Particular attention will be paid to the various approximations these methods are built upon. Chapter two will test the performance of these methods for the description of noncovalent interactions. Also within this chapter, empirically scaled quantum computational methods will be explored. All sections in chapter two are based on previously published work by this author. Section 2.2 provides a pioneering article that investigates the new SCS-MP2 method for a variety of noncovalent interactions.[1] Section 2.3 develops the SCS-CCSD method and provides preliminary results

for noncovalent interactions.[2] Section 2.4 includes part of a feature article (primarily written by C. D. Sherrill) that assesses a myriad of new theoretical methods for the newly obtained complete basis set CCSD(T) noncovalent interaction potential energy curves.[3] Section 2.5 investigates the intricate indole-benzene complex with the SCS-MP2 method.[4] And section 2.6 tests a variety of approximate methods for a set of 22 highly accurate noncovalent interaction energies.[5] Chapter three will focus on various metal-salen complexes. Section 3.2 tests the performance of standard DFT methods in comparison with very robust multireference wavefunction based methods for metal-salen energetics and is based on previously published work.[6] Section 3.3 extends this work on the metal-salen energetics to include the new M0-family of DFT functionals and is based on submitted material.[7] Lastly, section 3.3 investigates the cyanide addition to unsaturated imides reaction as catalyzed by the Al(Cl)-salen complex. This work is currently in preparation for publication.[8] Chapter four provides concluding remarks and important future work.

CHAPTER I

INTRODUCTION

In most cases, the goal of modern quantum chemistry is to utilize the nonrelativistic time-independent Schrödinger equation,

$$\hat{H}\psi = E\psi, \tag{1}$$

where $\hat{H}$ is the molecular Hamiltonian operator and ψ is the wavefunction, to compute chemical reactions and molecular properties. $\hat{H}$ has the form in atomic units,

$$\hat{H} = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{\alpha=1}^{M} \frac{1}{2M_\alpha}\nabla_\alpha^2 - \sum_{i=1}^{N}\sum_{\alpha=1}^{M} \frac{Z_\alpha}{|r_i - d_\alpha|} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{|r_i - r_j|} + \sum_{\alpha=1}^{M}\sum_{\beta>\alpha}^{M} \frac{Z_\alpha Z_\beta}{|d_\alpha - d_\beta|}, \tag{2}$$

where ∇^2 is the Laplacian operator, M_α and Z_α are the mass and charge of nucleus α, respectively, r_i is the position vector of electron i, and d_α is the position vector of nucleus α. The first two terms represent the kinetic energy of the electrons and nuclei, the third term is the Coulomb attraction of the electrons to the nuclei, and the fourth and fifth terms are the electron-electron and nucleus-nucleus Coulomb repulsion terms, respectively. Therefore $\hat{H}$ can be simply re-written as,

$$\hat{H} = T_{elec} + T_{nuc} + U_{elec-nuc} + U_{elec-elec} + U_{nuc-nuc}. \tag{3}$$

Notice that at least one approximation has already been made to ease the complexity of the equation, *i.e.* relativistic effects are neglected. Unfortunately, even with these approximations, this equation remains unsolvable.

At this point, an extremely important additional approximation can be made to the molecular Hamiltonian. An electron has a mass significantly less than any nucleus. Therefore, the electronic motion is much faster than the motion of the

nuclei and to a very good approximation, the motion of the nuclei can be ignored. This is equivalent to holding the nuclei fixed in space for some particular nuclear configuration and allowing electronic motion in the field of the nuclei and is called the Born-Oppenheimer approximation. Mathematically, the third term in equation 2 becomes dependent on only the electronic motion and as a result, the molecular Hamiltonian becomes separable into operators that contain electronic motion and the 'motion' of the nuclei. The separated Hamiltonian operators are as follows:

$$\hat{H}_{nuc} = T_{nuc} + U_{nuc-nuc} \tag{4}$$

and

$$\hat{H}_{elec} = T_{elec} + U_{elec-nuc} + U_{elec-elec} \tag{5}$$

where $\hat{H}_{nuc}$ and $\hat{H}_{elec}$ are the approximate Hamiltonian operators for the 'motion' of the nuclei and the electronic motion, respectively. Within the Born-Oppenheimer approximation, with respect to the electrons, the nuclei are stationary, and so they do not exhibit kinetic energy. Hence, the kinetic energy term is dropped when solving for the electronic motions. The fixed nuclei still contribute Coulomb repulsion between themselves, however; in solving for the electronic motions, the first term of $\hat{H}_{nuc}$ is eliminated and the last term becomes a constant. Now the goal becomes to solve for the electronic energy with $\hat{H}_{elec}$ and is commonly referred to as the many-electron problem. Although the electronic Schrödinger equation is a many-body problem and cannot be solved in closed form, quantum chemists strive to approximate the many-electron problem with the greatest accuracy that can be afforded.

There exists a two-fold hierarchy of accuracy inherent to most current *ab initio* electronic structure approximations (theories based on first principles and not parameterized to fit existing data): one can adjust both the theoretical approximation to the many-electron problem, as well as the underlying one-particle basis set. A basis set in this context refers to a set of functions, each of which represents the motion of

a single electron. There are various standard basis sets of increasing size (and hence increasing quality). There are also various approximate methods for solving the electronic Schrödinger equation for a given basis set. In section 1.1, a brief description of the Hartree-Fock (HF) theory will be made that follows much of Reference 9.[9] The HF method provides a good reference for the more robust post-HF methods. Section 1.2 will discuss the post-HF methods that include higher orders of electron correlation. It is with these methods that the hierarchy is built upon; higher orders of electron correlation generally correspond to greater accuracy. A major drawback of the HF and, sometimes, post-HF methods is the inadequate description of non-dynamical electron correlation which will be explained in section 1.3. Lastly, section 1.4 will address the popular density functional theory (DFT) methods which, unlike the previous methods, are usually not fully derived from first principles ("*ab initio*").

1.1   The Hartree-Fock (HF) Method

The Hartree-Fock (HF) method approximates the electronic Schrödinger equation,

$$\hat{H}_{elec}\psi_{elec} = E_{elec}\psi_{elec}, \tag{6}$$

by considering each electron's motion independent of the instantaneous motion of all other electrons. In other words, $\hat{H}_{elec}$ is separable into parts that only contain the motion of one electron. Through this type of evaluation, the electronic wavefunction can be described as the product of each one-electron function or orbital, ϕ,

$$\psi_{elec} \approx \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\phi_3(\mathbf{r}_3)...\phi_N(\mathbf{r}_N) = \psi_{HP}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3...\mathbf{r}_N) \tag{7}$$

and is referred to as the Hartree product, ψ_{HP}. While this follows our intuitive description of molecules with molecular orbital theory, the wavefunction must also be antisymmetrized.

To enforce antisymmetry, the wavefunction can be written as a generalized Slater

3

determinant,

$$\psi_{elec} \approx \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \chi_3(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \chi_3(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_2) \\ \chi_1(\mathbf{x}_3) & \chi_2(\mathbf{x}_3) & \chi_3(\mathbf{x}_3) & \cdots & \chi_N(\mathbf{x}_3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \chi_3(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix} = |\Psi\rangle, \tag{8}$$

where the coefficient is the normalization factor and $\chi(\mathbf{x})$ are spin orbitals with coordinates $\mathbf{x}$ that depend on the electron position, $\mathbf{r}$, and the generic intrinsic spin of each electron (either α or β). In a simpler shorthand of the Slater determinant, a set of occupied one-electron orbitals can be written as $|\chi_i\chi_j...\chi_k\rangle$ or $|\Psi\rangle$ in bracket notation. Writing the wavefunction in the Slater determinant form not only satisfies the antisymmetry principle, it also implies that electrons are indistinguishable from each other. Furthermore as a result of the antisymmetry principle, electrons with the same spin cannot occupy the same orbital at the same time, which is the Pauli exclusion principle. Inserting the electronic wavefunction approximation into equation 5 and left multiplying by $\langle\Psi|$ yields the approximate electronic energy expression,

$$E_{elec} = \langle\Psi|\hat{H}_{elec}|\Psi\rangle, \tag{9}$$

assuming the wavefunction is normalized.

The electronic Hamiltonian can also be separated into one- and two-electron operators,

$$\hat{H}_{elec} = \sum_i h(i) + \sum_{i<j} v_{ij}, \tag{10}$$

where

$$\sum_i h(i) = T_{elec} + U_{elec-nuc} \tag{11}$$

and

$$\sum_{i<j} v(i,j) = U_{elec-elec}. \tag{12}$$

4

Thus the one-electron operator, $h(i)$, generates the spin orbital energy of an isolated electron in the field the nuclei. The two-electron operator, v_{ij}, introduces the electron-electron Coulomb repulsion and exchange energies. Substituting equation 10 into equation 9 and with some work, the evaluation of the electronic energy in terms of integrals produces the HF energy expression,

$$E_{HF} = \sum_i \langle i|h|i \rangle + \frac{1}{2} \sum_{ij} [ii|jj] - [ij|ji]. \tag{13}$$

The first term is the one-electron integral,

$$\langle i|h|i \rangle = \int d\mathbf{x}_1 \chi_i^*(\mathbf{x}_1) h(\mathbf{r}_1) \chi_j(\mathbf{x}_1), \tag{14}$$

and the second term contains two-electron integrals in Chemists' notation,

$$[ij|kl] = \int d\mathbf{x}_1 d\mathbf{x}_2 \chi_i^*(\mathbf{x}_1) \chi_j(\mathbf{x}_1) \frac{1}{r_{12}} \chi_k^*(\mathbf{x}_2) \chi_l(\mathbf{x}_2). \tag{15}$$

Applying the variational theorem (minimizing E_{HF} by varying the spin orbitals, χ) the HF integro-differential equation is

$$[h(\mathbf{x}_1) + \sum_{j \neq i} \mathcal{J}_j(\mathbf{x}_1) - \sum_{j \neq i} \mathcal{K}_j(\mathbf{x}_1)] \chi_i(\mathbf{x}_1) = \epsilon_i \chi_i(\mathbf{x}_1), \tag{16}$$

where $\mathcal{J}$ and $\mathcal{K}$ are the Coulomb and exchange operators, respectively, and ϵ_i is the energy associated with orbital i. The operators on the left side of the equation, together, are known as the Fock operator, $\mathcal{F}$, and thus the equation becomes,

$$\mathcal{F}(\mathbf{x}_1) \chi_i(\mathbf{x}_1) = \epsilon_i \chi_i(\mathbf{x}_1), \tag{17}$$

otherwise known as the HF equation. Notice that both sides of the equation depend on $\chi_i(\mathbf{x}_1)$ and therefore the equation can be solved iteratively until self-consistency is obtained given an initial set of trial orbitals. This is why the HF method is often referred to as the self-consistent field (SCF) method.

To conclude the discussion of the HF method, there are some important implications of the HF equation and the assumptions it is built upon. The Coulomb operator

has the form,

$$\mathcal{J}_j(\mathbf{x}_1) = \int d\mathbf{x}_2 |\chi_j(\mathbf{x}_2)|^2 r_{12}^{-1}, \tag{18}$$

which is the interaction that electron 1 experiences with the instantaneous positions of electron 2. In other words, electron 1 is considered to be in the mean-field of electron 2 and therefore the HF method is often considered a mean-field approximation. The exchange operator has the form,

$$\mathcal{K}_j(\mathbf{x}_1) = \int d\mathbf{x}_2 \chi_j^*(\mathbf{x}_2) r_{12}^{-1} \chi_i(\mathbf{x}_2), \tag{19}$$

and is a direct consequence of the antisymmetry principle. Therefore, within the HF approximation, electrons of the same-spin provide some amount of electron correlation energy (discussed further section 1.2). Lastly, the preceding derivation assumes that $|\Psi\rangle$ is based on a single Slater determinant. As will be discussed in section 1.3, this leads to an inadequate description of non-dynamical electron correlation.

1.2 Electron Correlation and Post-HF Methods

Generally, the HF method provides an excellent approximation to the electronic energy and generates molecular orbitals which some chemists like to use in understanding and explaining chemistry. However as explained, it is built on an independent particle model which assumes that the electrons experience no effect from the change in positions of other electrons; i.e. they are uncorrelated. Furthermore, since the HF method utilizes the variational theorem, within a given basis, the HF energy (E_{HF}) computed is guaranteed to be a minimum. Then given the exact energy (E_{exact}) in the same basis, the energy due to electron correlation (E_{corr}) is simply defined as

$$E_{corr} = E_{exact} - E_{HF}. \tag{20}$$

Therefore to find E_{corr}, the resulting E_{HF} and its molecular orbitals can be used as a reference point for methods that account for electron correlation. These methods are referred to as post-HF methods and in the following discussion, three types of

post-HF methods will be presented with electron correlation up to double excitations. Higher order correlation will not be presented here for brevity.

1.2.1 Configuration Interaction

The configuration interaction (CI) series of methods is the easiest to conceptually understand. Roughly, the wavefunction approximation for the HF method, $|\Psi\rangle$, is expanded to include combinations of excited state configurations. Formally, the CI wavefunction becomes

$$|\Psi_{CI}\rangle = \sum_I c_I |\Phi_I\rangle = c_o |\Phi_o\rangle + \sum_{ia} c_i^a |\Phi_i^a\rangle + \sum_{\substack{i<j \\ a<b}} c_{ij}^{ab} |\Phi_{ij}^{ab}\rangle + ..., \qquad (21)$$

where $|\Phi_o\rangle$ is the ground state wavefunction, i, j are occupied orbitals, a, b are virtual orbitals, and the generic $|\Phi\rangle$ are linear combination of Slater determinants. Assuming a system with an even number of electrons, Figure 1 shows an example of singly and doubly excited configurations. This expansion can be truncated to include only single

(a)　　　　(b)　　　　(c)
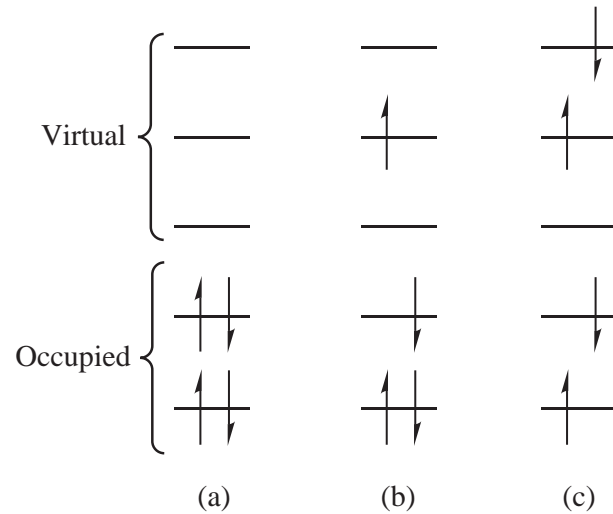
Figure 1: Example configurations of (a) the reference state, (b) a singly excited configuration, and (c) a doubly excited configuration.

and double excitations and expressed in terms of excitation operators creating the CISD wavefunction,

$$|\Psi_{CISD}\rangle = (1 + \hat{C}_1 + \hat{C}_2)|\Phi_o\rangle \qquad (22)$$

where the excitation operators are defined as

$$\hat{C}_1 = \sum_{ia} c_i^a a_a^\dagger a_i \tag{23}$$

and

$$\hat{C}_2 = \frac{1}{4} \sum_{ijab} c_{ij}^{ab} a_a^\dagger a_b^\dagger a_j a_i. \tag{24}$$

The second quantization operators, $a^\dagger$ and a, generate the excited state determinants when acting on the reference wavefunction. For example in Equation 23, the annihilation operator, a_i, removes an electron from an occupied orbital i, and the creation operator, $a_a^\dagger$, places an electron into a virtual orbital a, which is equivalent to a singly excited configuration. The CISD method utilizes the variational theorem to determine the coefficients.

Besides the computational complexity of the CI methods, any truncation suffers from not being size-extensive, *i.e.* not linearly scaling with respect to the number of particles. In a classic example, using the CISD method for two non-interacting hydrogen molecules at an infinite separation does not yield the same results as the sum of two single hydrogen molecules. In regards to the CISD wavefunction, the total CISD wavefunction of two non-interacting hydrogen molecules does not equal the product of the CISD wavefunction of two separate hydrogen molecules. To obtain the correct energies, quadruple excitations are required. Therefore, size-extensivity is important to enforce for rigorous electronic structure methods and other methods are commonly preferred over the CISD method.

1.2.2 Many Body Perturbation Theory

Many body perturbation theory (MBPT) assumes that the many-body problem has been solved to a good approximation and a small perturbation to the approximation will yield an improved result. Since the HF method serves as the reference, or "good approximation", MBPT contributes a perturbative correction due to electron

correlation. Mathematically, the approximation to the electronic Hamiltonian can be written as

$$\hat{H}_{elec} \approx \mathcal{H}_o + \lambda\mathcal{V}, \tag{25}$$

where $\mathcal{H}_o$ is the unperturbed reference Hamiltonian and $\mathcal{V}$ is the electronic correlation perturbation. Expanding the wavefunction, $|\Psi_i\rangle$, and the energy, $\mathcal{E}_i$, in a Taylor series in λ yields,

$$|\Psi_i\rangle = |\Psi_i^{(0)}\rangle + \lambda|\Psi_i^{(1)}\rangle + \lambda^2|\Psi_i^{(2)}\rangle + ... \tag{26}$$

and

$$\mathcal{E}_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + ... \tag{27}$$

The generalized results for the energy expressions up to second-order are

$$E_i^{(0)} = \langle\Psi_i^{(0)}|\mathcal{H}_o|\Psi_i^{(0)}\rangle \tag{28}$$

$$E_i^{(1)} = \langle\Psi_i^{(0)}|\mathcal{V}|\Psi_i^{(0)}\rangle \tag{29}$$

$$E_i^{(2)} = \langle\Psi_i^{(0)}|\mathcal{V}|\Psi_i^{(1)}\rangle, \tag{30}$$

where $|\Psi_i^{(0)}\rangle$ is assumed to be normalized. With some work and defining the operator $\mathcal{V}$ as

$$\mathcal{V} = \sum_{i<j} r_{ij}^{-1} - \mathcal{V}^{HF}, \tag{31}$$

the sum of the zero- and first-order energies in the perturbative expansion exactly equals the HF energy, E_{HF}.

The second-order energy requires a bit more thought; what is the form of the wavefunction $|\Psi_i^{(1)}\rangle$? The zero- and first-order energies only require the use of $|\Psi_i^{(0)}\rangle$, the HF wavefunction. Furthermore, the outcome of utilizing a wavefunction consisting of single excitations will contribute no electron correlation, in part, due to Brillouin's theorem which states that no mixing occurs between the HF ground state and singly excited state wavefunctions. Therefore, because of the two-particle nature of the

9

perturbation, the wavefunction must be then comprised of double excitations and thus the second-order energy has the form,

$$E_o^{(2)} = \frac{1}{4} \sum_{ijab} \frac{|[ai|bj] - [aj|bi]|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}. \tag{32}$$

It is more common to see this equation in the physicist's notation,

$$E_o^{(2)} = \frac{1}{4} \sum_{ijab} \frac{|\langle ab|ij\rangle - \langle ab|ji\rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} = \frac{1}{4} \sum_{ijab} \frac{|\langle ab||ij\rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} = \frac{1}{4} \sum_{ijab} t_{ij}^{ab} \langle ab||ij\rangle, \tag{33}$$

where

$$t_{ij}^{ab} = \frac{\langle ab||ij\rangle}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}. \tag{34}$$

The truncation of the perturbative expansion to the second-order is commonly referred to as the second-order Møller-Plesset perturbation theory (MP2) method. Although the perturbative series of methods are size-extensive, as higher orders of electron correlation are taken into account, the energy is not guaranteed to yield a better approximation for the actual energy. Furthermore due to the denominator, extreme caution should be taken if orbital energies become nearly degenerate. As the sum of the energies of orbitals i and j reach the sum of the energies of orbitals a and b, the denominator proceeds towards zero, and consequently the correlation energy becomes artificially large. This is an indication that either higher orders of correlation are needed, or the reference wavefunction, based on a single determinant, is inadequate to characterize the system.

1.2.3 Coupled-Cluster Theory

Coupled-cluster theory (CC) is perhaps the newest series of electron correlation methods introduced into quantum chemistry that has wide-spread usage. Briefly, the CC wavefunction is defined as an exponential ansatz acting on a reference wavefunction and generally yields a much more robust description of electron correlation. The CC wavefunction has the form,

$$|\Psi_{CC}\rangle = e^{\hat{T}}|\Phi_o\rangle, \tag{35}$$

10

where $\hat{T}$ is the excitation operator that contains all excitations. If $\hat{T}$ includes only single and double excitations, the CC wavefunction becomes,

$$|\Psi_{CCSD}\rangle = e^{\hat{T}_1 + \hat{T}_2}|\Phi_o\rangle, \tag{36}$$

where $\hat{T}_1$ and $\hat{T}_2$ have the same form as the CISD excitation operators, $\hat{C}_1$ and $\hat{C}_2$. The expansion of the exponential ansatz of $|\Psi_{CCSD}\rangle$ yields

$$|\Psi_{CCSD}\rangle = (1 + \hat{T}_1 + \frac{1}{2!}\hat{T}_1^2 + \frac{1}{3!}\hat{T}_1^3 + \hat{T}_2 + \frac{1}{2!}\hat{T}_2^2 + \frac{1}{4!}\hat{T}_1^4 + \hat{T}_2\hat{T}_1 + \frac{1}{2!}\hat{T}_2\hat{T}_1^2)|\Phi_o\rangle. \tag{37}$$

The derivation of the energy expression is somewhat long and there are many excellent articles on this subject.[10–13] For brevity, the correlation energy expression of the coupled-cluster of single and double excitations (CCSD)[10] method is

$$E_{CCSD}^{corr} = \sum_{ia} t_i^a f_i^a + \frac{1}{4}\sum_{ijab} \tau_{ij}^{ab}\langle ij||ab\rangle, \tag{38}$$

where τ_{ij}^{ab} is the CC double excitation amplitude defined as

$$\tau_{ij}^{ab} = t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a, \tag{39}$$

and t_{ij}^{ab} and t_i^a are double and single excitation coefficients, respectively. The first term is based on only single excitations and invoking Brillouin's theorem, it generally results in zero contribution. Notice then that this equation is very similar to the MP2 correlation energy expression.

The CC methods tend to improve on both CI and MBPT methods because they include approximate descriptions of higher-body correlations as products of one- and two-electron correlation terms. Within the CCSD expanded wavefunction, essentially up to four electronic motions are connected by the $\hat{T}_2^2$ and $\hat{T}_1^4$ operations. The CC methods are size-extensive due to the separability of exponential functions. For these reasons CC methods have become the preferred method if a high degree of accuracy is desired; the CCSD method with the perturbative addition of triple excitations, the CCSD(T) method,[14] is considered the 'golden standard' for chemical accuracy when the HF reference is a good approximation to the total electronic wavefunction.[15]

1.3 Dynamical vs. Nondynamical Electron Correlation

The previous discussion on electron correlation only accounted for the energy that arises due to the instantaneous repulsion of electrons, commonly referred to as dynamical electron correlation. Unfortunately, methods that only include dynamical correlation often do not appropriately describe many interesting chemical processes, such as bond breaking reactions, reaction mechanism profiles (transition states and reactive intermediates), and excited electronic states.[16] This is due to the wavefunction dependence on multiple nearly degenerate electron configurations which is created by important nearly isoenergetic orbitals. The energy that is derived from multiple configurations is referred to as nondynamical electron correlation. For this reason, the HF method (based on one electron configuration) cannot describe nondynamical correlation and methods built on the HF reference (MBPT and CC) often yield incorrect results when applied to systems that contain nondynamical correlation to a large degree.

The classic example to demonstrate nondynamical correlation is the bond breaking curve of the H_2 molecule. However, since the metal-ligand complexes, the metal-salens, are under investigation in a subsequent chapter, it is more useful to illustrate nondynamical correlation with electron configuration degeneracies created from partially occupied bonding and anti-bonding d-orbitals. Consider, for example, the square-planar splitting pattern as predicted by the crystal field theory as shown in Figure 2. Now assume a d^6 occupation and that Δ is large enough that a low spin arrangement of electrons is preferred and the lowest three d-orbitals (d_{yz}, d_{xz}, and d_{z^2}) are completely occupied (Figure 3a). This type of system will depend on only one electron configuration and therefore, single reference methods (those based on the HF method) will be sufficient to capture the important electron correlation (in this case, only dynamical correlation). As Δ decreases, the two anti-bonding d-orbitals, d_{xy} and $d_{x^2-y^2}$, become closer in energy to the three lowest d-orbitals. If Δ is small enough,
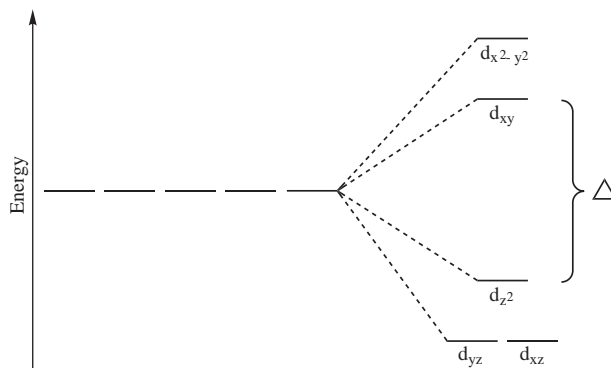
Figure 2: Crystal field splitting pattern of nonbonding d-orbitals for a square planar metal-ligand complex.
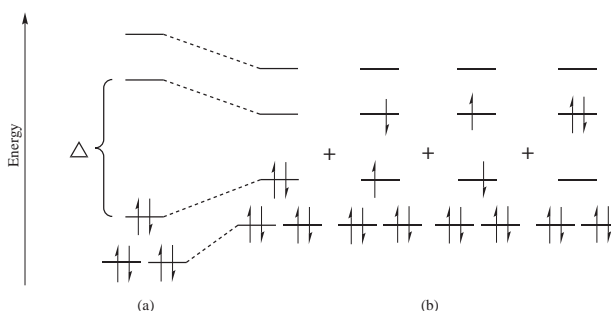
Figure 3: Important electron configurations as Δ decreases from (a) large magnitudes to (b) smaller magnitudes.

the d_{z^2} electrons are able to partially occupy the d_{xy} orbital and the wavefunction may depend on four electron configurations (Figure 3b). This problem increases rapidly if the d_{yz} and d_{xz} electrons are able to partially occupy the $d_{x^2-y^2}$ orbital.

1.3.1 Multiconfiguration Methods

One scheme to successfully account for nondynamical correlation is to compute high enough orders of electron correlation (doubles, triples, quadruples, etc.) until all nearly degenerate configurations are included. The full configuration interaction (FCI) method has been applied to many diatomic molecules which includes all electron configurations and therefore captures all dynamical and nondynamical correlation.[17,18] The FCI method and methods that include high orders of electron correlation, however, are extremely expensive and are feasible only when applied to small

13

molecules.

Another scheme is to enforce the wavefunction dependence on multiple configurations; these types of methods are called multiconfiguration methods. The multi-configuration self-consistent field (MCSCF)[19] method considers a wavefunction built on a linear combination of multiple Slater determinants. The MCSCF wavefunction resembles the truncated CI expansion,

$$|\Psi_{MCSCF}\rangle = \sum_I c_I |\Psi_I\rangle, \tag{40}$$

however, both the expansion coefficients and the orbitals are optimized. Recall that the CI method only allows the optimization of the expansion coefficients and, while the HF method optimizes the orbitals, it is based on only one electron configuration (or determinant). Therefore, the MCSCF method is a combination of the CI and HF methods that accounts for nondynamical correlation.

Although the MCSCF method is certainly useful, it is often unclear where to truncate the expansion series. As with previous electron correlation methods, the computational cost increases rapidly when higher orders of electron correlation are included. The complete-active-space self-consistent field (CASSCF)[20] method allows a FCI procedure within a given active-space (optimally, the important orbitals that are nearly isoenergetic), while simultaneously optimizing all other orbitals. For example, consider again the case of the d^6 square planar complex. If the magnitude of Δ is unknown, then all five d-orbitals could feasibly be included within the active-space. The CASSCF method would compute all dynamical and non-dynamical correlation that arise from the six electrons in the five d-orbitals. Then, the only correlation unaccounted for is the dynamical correlation from all other electrons. To approximate the missing dynamical correlation, perturbative series can be applied to the CASSCF wavefunction. The second-order and third-order perturbation series with the CASSCF wavefunction are called the CASPT2[21,22] and CASPT3[23] methods, respectively.

1.4 Density Functional Theory (DFT)

The major goal of density functional theory (DFT)[24,25] is to approximate the nonrelativistic electronic Hamiltonian, $\hat{H}_{elec}$. The difference between the traditional (those based on the HF method) and DFT methods is that DFT methods depend on the electron density, $n(r)$, rather than the many-electron wavefunction; i.e. the many-electron problem is reduced to the computation of non-interacting electrons in an external potential, $v(r)$, added to the effect of interacting electrons as a functionalized form that varies with respect to $n(r)$.[26] The ability to utilize $n(r)$ to determine $\hat{H}_{elec}$ is built on Hohenberg and Kohn's lemma that the ground state density uniquely defines $v(r)$, and that the number of electrons can be extracted from $n(r)$ through integration. Minimizing the energy as a functional of $n(r)$, Kohn and Sham derived the equation

$$[-\frac{1}{2}\nabla^2 + v(r) + \int \frac{n(r')}{|r - r'|}dr' + v_{xc}(r)]\phi_j(r) = \epsilon_j\phi_j(r), \qquad (41)$$

to compute DFT orbitals, $\phi_j(r)$, and orbital energies, ϵ_j. In terms of the independent particle model, the first, second, and third terms on the left hand side are simply T_{elec}, $U_{elec-nuc}$, and $U_{elec-elec}$, respectively (refer to equation 5). The so-called exchange-correlation term, $v_{xc}(r)$, separates DFT from HF theory, and has the form

$$v_{xc}(r)] = \delta E_{xc}[n(r)]/\delta n(r), \qquad (42)$$

where E_{xc} is the exchange-correlation energy with respect to $n(r)$.

In principle, DFT is an exact theory if the exact $E_{xc}[n]$ is utilized. However in practice, the exact form of $E_{xc}[n]$ is not known. Furthermore, it is difficult to even approximate the functional because $E_{xc}[n]$ must correct for T_{elec} and the correlation energy, which are approximated due to the independent particle model, and for electron self-interaction and exchange energies, which are neglected by utilizing electron densities.[26] Among the earliest DFT functionals, the local density approximation

(LDA) provided a simple functional form of $E_{xc}[n]$,

$$E_{xc}^{LDA} = \int \epsilon_{xc}(n(r))n(r)dr, \qquad (43)$$

where ϵ_{xc} is the exchange-correlation energy per particle of a uniform interacting electron gas of density n. The DFT orbitals within the LDA formalism provides orbitals very similar to HF orbitals. DFT was generally improved by including the gradient of $n(r)$ into the functional form for $E_{xc}[n]$ and is referred to as the generalized gradient approximation (GGA),

$$E_{xc}^{GGA} = \int f(n(r), |\nabla n(r)|)dr, \qquad (44)$$

where $f(n, |\nabla_n|)$ is any function of the electron density and its gradient. In 1993, Becke mixed the GGA functional form with the exchange energy obtained from the HF method to create the highly successful B3LYP functional.[27] These types of DFT functionals that include some amount of HF exchange energy are called hybrid functionals. Lastly, of recent development, the kinetic-energy density has been included into the functional form of $E_{xc}[n]$ and shows great promise towards improving the accuracy of DFT.[28–31] While there exists no scheme to systematically improve DFT (unlike post-HF methods), the utility of DFT methods is undeniable due to their computational tractability on large systems.

CHAPTER II

NONCOVALENT INTERACTIONS

Noncovalent interactions exhibit significant influence on the stability of molecular structures and affect many chemical processes such as self-assembly and drug binding. The structures of DNA, RNA, and proteins can be largely attributed to several noncovalent interaction types including π-π,[32–34] C-H/π,[35] and sulfur-π,[36] among others. These interactions also contribute to the intercalation of drugs, crystal structures, and many host-guest complexes.[37–42] There is no doubt that the accurate characterization of noncovalent interactions would be of great fundamental importance for areas ranging from biology to chemistry to materials science.

Much research has been conducted on the simplest relevant aromatic model system of π-π interactions, the benzene dimer.[3,43–52] Unfortunately, HF and many popular DFT methods incorrectly predict unfavorable (repulsive) noncovalent interaction energies. Therefore, even a qualitative evaluation of intermolecular interaction energies require the inclusion of electron correlation. MP2 is among the least expensive methods that account for electron correlation and is therefore attractive to use for larger systems. However, MP2 usually overestimates intermolecular binding energies for π-π systems. CC methods provide a more robust treatment of electron correlation. Although the CCSD method has been shown to underestimate binding energies by about as much as MP2 overestimates them for several dimer systems,[53] the CCSD(T) method[14] provides very accurate results. Indeed, estimates of CCSD(T) energies with augmented triple-zeta basis sets or larger are considered to be of benchmark quality. Unfortunately CCSD(T) formally scales as $O(N^7)$, where N is proportional to the size of the molecule. Computational costs are increased further when the

17

counterpoise (CP) correction is employed to reduce the basis set superposition error (BSSE), which can be significant unless extremely large basis sets are used.

The following research studies will be presented chronologically. The first study, presented in section 2.2, is one of the pioneering publications that thoroughly explored the newly developed spin-component scaled MP2 (SCS-MP2)[54] method for the computation of noncovalent interactions.[1] Inspired by the SCS-MP2 method and its much improved accuracy over the canonical MP2 method, the next study developed the spin-component scaled CCSD (SCS-CCSD) method as covered in section 2.3.[2] Within the years 2006-2009, numerous approximate methods based on empirical scalings of quantum chemistry methods were published for noncovalent interactions. In response, a review article primarily written by C. D. Sherrill, provided the most accurate potential energy curves (PECs) for three structures of the benzene dimer (sandwich, T-shaped, and parallel-displaced), and the methane-benzene and H_2S-benzene complexes, to-date.[3] Section 2.4 provides the portion of that publication which examines the errors incurred by various approximations to the CCSD(T) method. Since the SCS-MP2 method yields accurate results (normally to within 10% error) and is relatively inexpensive, an extensive study was performed on the indole-benzene complex.[4] This work generated some of the first 3-dimensional PECs for noncovalent interactions and is presented in section 2.5. Lastly in section 2.6, as a 'culmination' of the previous research a basis set consistent evaluation of the S22 test set[55] for noncovalent interactions was performed and is, arguably, the most accurate set of noncovalent interaction energies to-date.[5] In the next section (section 2.1), a generalized theoretical methodology for the computation of noncovalent interactions is given. If the methods for a specific study utilized different procedures, they will be given in a specialized methodology section.

2.1 Generalized Theoretical Methodology

The MOLPRO *ab initio* suite of quantum chemistry programs[56] was primarily used for the computation of interaction energies and PECs. For the benzene dimer (sandwich, T-shaped, and parallel-displaced), and the methane-benzene and H_2S-benzene complexes, the benzene monomer geometries were set to those suggested by Gauss and Stanton, R(C-C)=1.3915 Å and R(C-H)=1.0800 Å (Figure 4).[57] For the methane
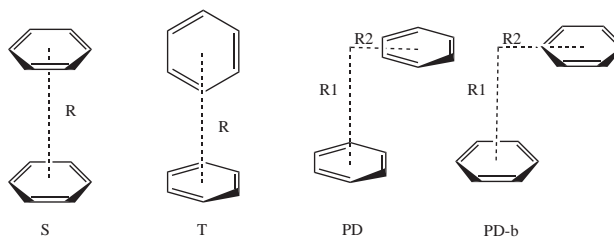
Figure 4: Orientation for the sandwich, T-shaped, and two parallel displaced benzene dimer configurations.

dimer (Figure 5c) and the methane-benzene complex (Figure 5a), methane geometries optimized at the MP2/cc-pVDZ level of theory by Ringer *et al*[58] were used. And for the H_2S-benzene complex (Figure 5b), experimental values[59] for the H_2S monomer, R(S-H)=1.3356 Å and θ(H-S-H)=92.12º were used. Monomer geometries were kept
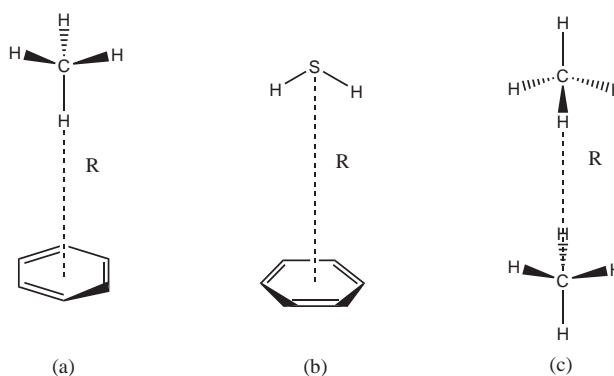
Figure 5: Orientations for (a) methane-benzene, (b) H2S-benzene, and (c) methane dimer.

frozen as the intermolecular distances were varied.

The resolution-of-the-identity (RI) approximation speeds up the computation of

the two-electron integrals by introducing an auxiliary basis set and expressing products four-index integrals as products of three-index integrals and two-index electron repulsion integrals:[60]

$$(\mu\nu|\rho\sigma) \approx \sum_{PQ} (\mu\nu|P)(P|Q)^{-1}(Q|\rho\sigma). \tag{45}$$

The two-index quantity $(P|Q)$ is also an electron repulsion integral, not an overlap:

$$(P|Q) = \int P(\mathbf{r}_1) \frac{1}{r_{12}} Q(\mathbf{r}_2) d^3\mathbf{r}_1 d^3\mathbf{r}_2. \tag{46}$$

This RI approach is closely related to the pseudospectral approximation,[61,62] which uses a grid instead of atom-centered functions for the auxiliary basis. This way of expressing the two-electron integrals can be helpful for more effective factorizations of Hartree-Fock,[60,63] MP2,[64,65] and other kinds of quantum chemistry computations. In recent years, the RI approximation has also been called density fitting (DF).[65]

Interaction energy computations applied the augmented correlation-consistent basis sets aug-cc-pVXZ (X=D,T,Q) of Dunning and co-workers.[66,67] For MP2-type methods, density fitting was used to reduce the computational cost of the HF (DF-HF) procedure and the subsequent MP2 (DF-MP2) computation.[65,68] For HF, density fitting employed the JK auxiliary[60] basis sets, and for the MP2 correlation energy, density fitting employed the auxiliary MP2-Fitting[69] basis sets. Density fitting appears to introduce only very small errors for interaction energies.[3] Core orbitals were constrained to remain doubly-occupied in all correlated computations. All interaction energies were CP corrected using the method of Boys and Bernardi to reduce the BSSE effects.[70]

MP2 interaction energies were extrapolated to the complete basis set (CBS) limit, by adding a two-point extrapolation[71] of the aug-cc-pVTZ and aug-cc-pVQZ correlation energies to the aug-cc-pVQZ reference energy. Estimates of the CBS limit of the CCSD(T) and SCS-CCSD[2] interaction energies were obtained via an additive approximation which combines MP2/CBS results with the difference between

20

coupled-cluster (CC) and MP2 methods evaluated with a smaller basis set where,

$$E_{\text{CC}}^{\text{CBS}} \approx E_{\text{MP2}}^{\text{CBS}} + \Delta CC, \tag{47}$$

and ΔCC is defined as

$$\Delta CC = E_{\text{CC}}^{\text{small-basis}} - E_{\text{MP2}}^{\text{small-basis}}. \tag{48}$$

When computing the estimated CCSD(T)/CBS interaction energy, this scheme is commonly referred to as the ΔCCSD(T) correction and appears to be very good for the computation of noncovalent interactions.[3]

Spin-component scaled (SCS) methods utilized various parameters for the same- and opposite-spin contributions to the correlation energy and are given in Table 1. In the cases where the SCS(MI)-MP2 method was utilized, the interaction energies

Table 1: Scaling Parameters for Various Spin-Component Scaled Methods

	$c_{SS}{}^a$	$c_{OS}{}^b$
SCS-MP2[c]	0.33	1.20
SCSN-MP2[d]	1.76	0.00
SCS(MI)-MP2[e]	1.29	0.40
SCS-CCSD[f]	1.13	1.27

[a]Same-spin scaling coefficient. [b]Opposite-spin scaling coefficient. [c]Spin-component scaled MP2 method.[54] [d]Spin-component scaled MP2 method for nucleic acids.[72] [e]Spin-component scaled for intermolecular interactions.[73] [f]Spin-component scaled CCSD.[2]

were extrapolated to the CBS limit with the cc-pVTZ and cc-pVQZ basis sets.[73]

Local correlation methods based on the use of localized molecular orbitals, as introduced by Pulay and Saebø,[74–76] have proven to be highly effective for the approximation of correlation energies.[65,74–82] The subsequent localization procedures followed that of Hill *et al.*[83] The Pipek-Mezey[84] localization method was applied with the Newton-Raphson algorithm to ensure convergence. The contributions from the most diffuse basis functions were eliminated by setting the corresponding overlap matrix elements used in the Pipek-Mezey procedure to zero. A completeness criterion of 0.985 for the orbital domain selection was applied as described by Boughton and

Pulay.[85] The π orbital domains of benzene were merged and only atoms with Löwdin charges greater than 0.2 for hydrogen and 0.01 for all other atoms were included in the domain. Furthermore, local methods have been shown to incur little BSSE and therefore were not normally CP corrected.[86]

2.2 Performance of Spin-Component-Scaled Møller-Plesset Theory (SCS-MP2) for Potential Energy Curves of Noncovalent Interactions

2.2.1 Introduction

This section is based on previously published work by Takatani and Sherrill.[1]

Recent publications have explored the use of new and cheaper methods to obtain accurate interaction energies. Grimme originally proposed the SCS-MP2 method,[54] which exploits the ability to separately scale the singlet and triplet electron pair contributions to the correlation energy. Using 51 reaction energies, the scaling factors were optimized to 6/5 and 1/3 for the singlet and the triplet pairs, respectively. Results for π-π interactions showed only a 10-20% binding energy underestimation compared to CCSD(T) with costs no greater than conventional MP2. A related approach, the scaled opposite-spin MP2 (SOS-MP2) approach of Head-Gordon and co-workers,[87] takes advantage of increased computational efficiencies that occur when the triplet pairs are neglected [the singlet pairs can be evaluated without any of the usual O(N^5) work using the resolution of the identity and a Laplace transform approach]. While the SOS-MP2 method works well in many situations,[87] it does not appear to perform as well as SCS-MP2 for noncovalent interactions.[88]

Using the benzene dimer as a test case, additional work on variants of the SCS-MP2 method have sought to further reduce computational costs.[83] DF approximations for electron repulsion integrals can significantly decrease the computation time while introducing only small, controllable errors.[65,68] Local electron correlation methods provide a more favorable computational scaling by eliminating many unimportant,

long-distance contributions to the energy.[89,90] Additionally, local MP2 methods have significantly reduced BSSE and a counterpoise correction is typically not required.[86] Furthermore, Schütz *et al.* described a localized DF-MP2 (DF-LMP2) procedure that required computational resources that scale only $O(N)$ with respect to molecular size.[91] Such methods have the potential to reduce the computational cost for large systems immensely. This DF-SCS-LMP2 approach, without CP corrections, was shown to match CCSD(T) to within 0.1-0.2 kcal mol^{-1} for three orientations of the benzene dimer interaction energies.[83] However, this initial study did not investigate whether these results were representative of other types of noncovalent interactions.

In a subsequent study, Hill and Platts recently re-optimized the scaling factors with a set of ten stacked nucleic acid base pairs (SCSN-MP2).[72] The optimized parameters were 0 and 1.76 for the singlet and triplet scaling factors, respectively. Exploring 22 single point noncovalent interaction energies, DF-SCSN-LMP2 had a mean absolute deviation (MAD) of 0.27 kcal mol^{-1} and DF-SCS-LMP2 had a MAD of 1.26 kcal mol^{-1}.

Here the SCS-MP2 and SCSN-MP2 methods, and their density-fitted and local correlation variants are tested, for a wider variety of noncovalent interactions: namely, π-π, C-H/π, sulfur-π, and alkane-alkane interactions as exemplified by the benzene dimer, the methane-benzene complex, the H$_2$S-benzene complex, and the methane dimer. Comparisons are made to estimated CCSD(T)/CBS PECs which should be close to the *ab initio* limit. New CCSD(T)/CBS curves are presented here for the H$_2$S-benzene complex and the methane dimer. PECs of these systems are of interest because in more complex systems such as proteins, individual noncovalent interactions may not be able to achieve their optimal geometries due to steric constraints. As this article was being finalized, a related work by Antony and Grimme was published[88] which examines the SCS-MP2 approach with the original and nucleic acid (SCSN) parameters, as well as the related SOS-MP2 approach, for a wide array of van der

Waals complexes. That work considered potential energy minima, in contrast to the potential curves considered here. In addition, the effect of local correlation approximations are also considered, which can be problematic in aug-cc-pVDZ basis sets for the systems considered, as demonstrated below.

2.2.2   Specialized Theoretical Methodology

In their testing of the DF-SCS-MP2 and DF-SCS-LMP2 methods for the benzene dimer, Hill, Platts, and Werner[83] utilized non-augmented basis sets for hydrogen and augmented basis sets for carbon. In this study, most results employed augmented basis sets for all atoms. For double-ζ basis sets, the effects of augmented versus non-augmented basis sets for hydrogen were compared. Local correlation approximations did not lead to a significant reduction in computational cost for the small systems considered here, perhaps because of increased overhead as well as the loss of point-group symmetry in the local orbitals.

For the estimation of the CCSD(T)/CBS PECs, the basis sets for the ΔCCSD(T) correction varied for each complex. The smaller basis was aug-cc-pVDZ for the benzene dimer and the methane-benzene complex and aug-cc-pVTZ for the H_2S-benzene complex. For the methane dimer, CCSD(T)/CBS approximations were obtained by adding a two-point extrapolation of the aug-cc-pVTZ and aug-cc-pVQZ CCSD(T) correlation energies to the reference aug-cc-pVQZ reference energy. The CCSD(T)/CBS curves for the H_2S-benzene complex and the methane dimer are presented for the first time here, and for other systems they have been obtained in the course of other studies.[3,58,92]

For the methane dimer, dispersion, induction, exchange-repulsion, and electrostatic contributions to the intermolecular interaction were determined using symmetry-adapted perturbation theory through the second order (SAPT2)[93]. A single point

computation using the aug-cc-pVDZ basis was performed at the CCSD(T)/aug-cc-pVQZ optimal inter-monomer distance with the SAPT2006 program.[94,95]

2.2.3 Results and Discussion

Figures 6a-b show the CBS extrapolated PECs for the methane-benzene complex and the methane dimer, respectively. Unlike some approximate electronic structure meth-
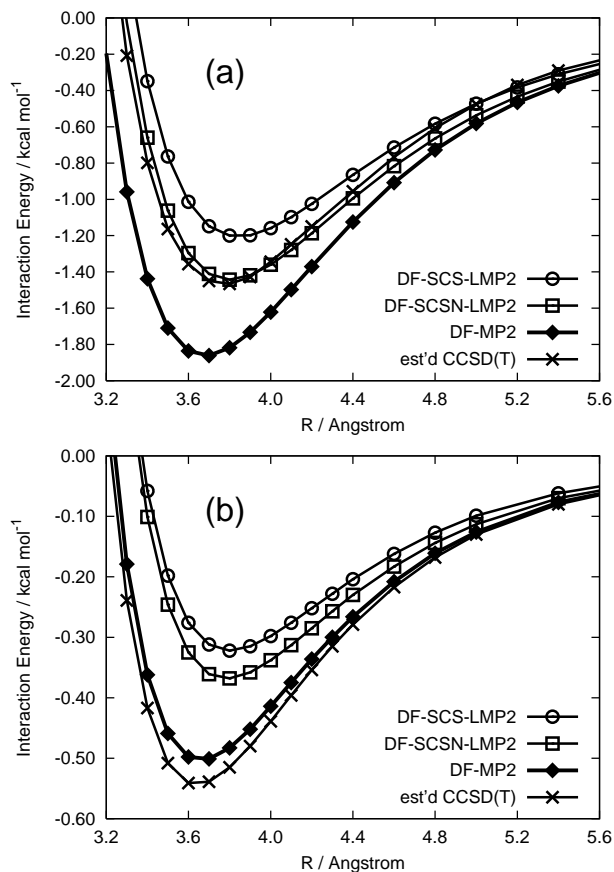
Figure 6: Potential energy curves at the CBS limit for (a) the methane-benzene system and (b) the methane dimer

ods (including many density functional approaches and semiempirical methods), the SCS methods and their variants are always qualitatively correct for these systems. For the methane-benzene complex, the SCS-LMP2 and SCSN-LMP2 methods offer striking improvements over the MP2 method for the reproduction of the CCSD(T)/CBS PEC. This trend is seen for all systems except the methane dimer. For the methane

dimer, the MP2 method nearly reproduces the CCSD(T)/CBS curve. The SCS-LMP2 and SCSN-LMP2 methods underestimate the binding energy far more than MP2 and in this case, conventional MP2 is superior to the SCS-type methods. On the other hand, the error in the SCS(N) energies remains below about 0.3 kcal mol^{-1}.

Figures 7a-e show the CBS extrapolated deviations from the estimated CCSD(T) PECs for the sandwich benzene dimer, T-shaped benzene dimer, methane-benzene complex, H_2S-benzene complex, and methane dimer, respectively. The most significant deviation observed around the equilibrium intermolecular distances is for the methane dimer, where SCS-type methods give deviations of 0.22-0.26 kcal mol^{-1}. While a small difference on an absolute scale, this corresponds to a large relative error (40-50%). However, in the CBS limit all SCS-type methods are fairly accurate, with errors on the order of a few tenths of one kcal mol^{-1}. Except for at short intermolecular distances, the local approximation has very little effect of the CBS limit. For all the systems except for the methane dimer, the SCS-MP2 and SCS-LMP2 deviations tend to diverge with decreasing intermolecular distances, most notably in the sandwich benzene dimer and the H_2S-benzene complex. The SCSN-MP2 and SCSN-LMP2 deviations seems to diverge with decreasing intermolecular distances in the T-shaped benzene dimer and the H_2S-benzene complex, albeit not as significantly as the SCS-MP2 and SCS-LMP2 methods. The SCSN-(L)MP2 method yields deeper potential curves than SCS-(L)MP2 and is considerably closer to CCSD(T) for the T-shaped benzene dimer and for the methane-benzene complex. The sandwich benzene dimer, which is typically the most difficult for approximate electronic structure methods, is well described by all SCS-type methods considered. For the T-shaped benzene dimer and the H_2S-benzene complex, SCSN-(L)MP2 overestimates and SCS-(L)MP2 underestimates interaction energies compared to CCSD(T). Except for the methane dimer or for short intermolecular distances in the sandwich benzene dimer and the
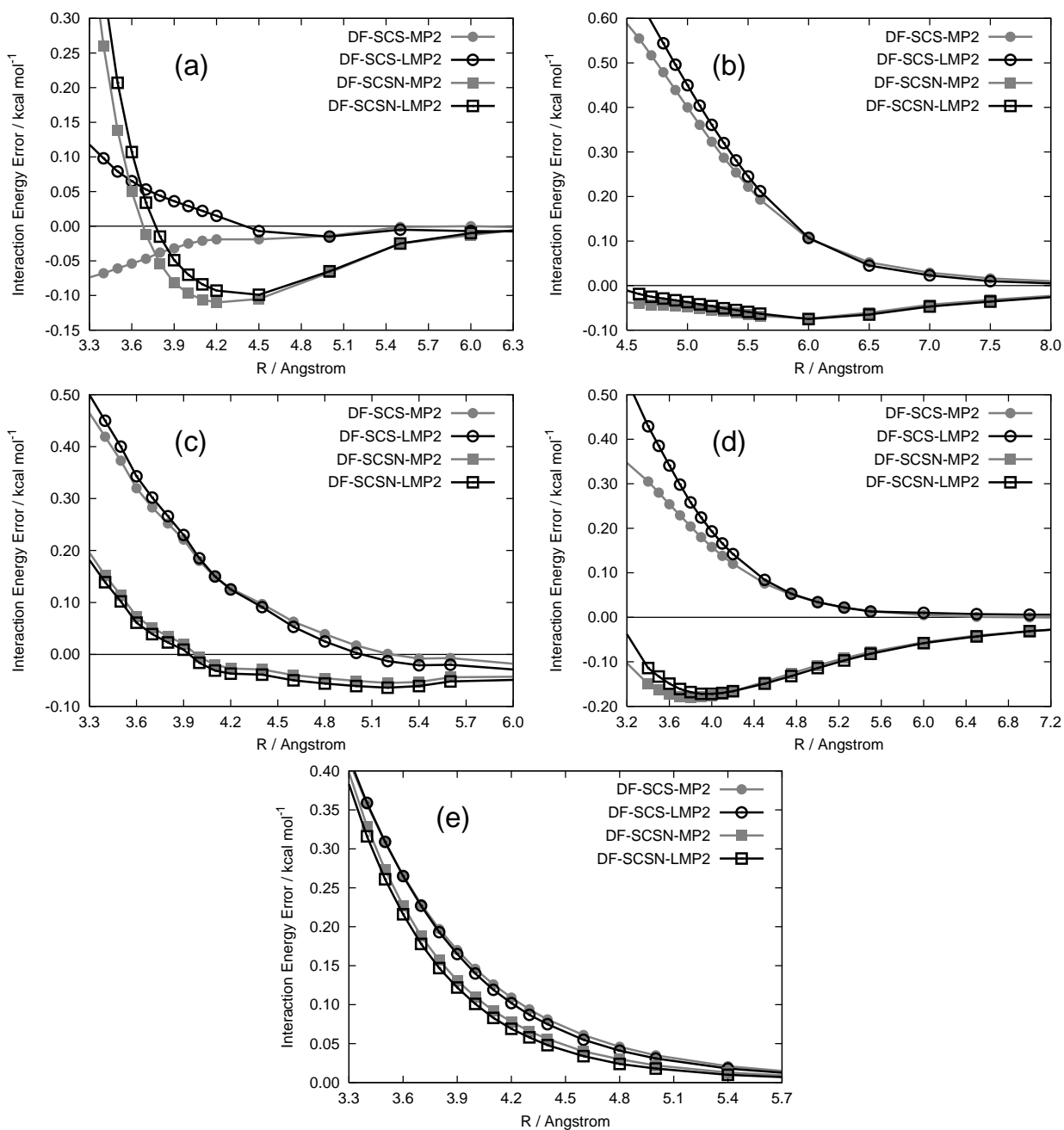
Figure 7: Potential energy errors with respect to the estimated CCSD(T)/CBS curve for (a) the sandwich benzene dimer, (b) the T-shaped benzene dimer, (c) the methane-benzene complex, (d) the H_2S-benzene complex, and (e) the methane dimer.

methane-benzene complex, SCSN-(L)MP2 provides stronger intermolecular interactions than CCSD(T). Although the performance of SCSN-(L)MP2 is better than that of SCS-(L)MP2 on average, at large intermolecular separations for all systems but

the methane dimer, the SCS-(L)MP2 curves approach the benchmark CCSD(T) curve more rapidly than does SCSN-(L)MP2, which tends to be slightly overbound.

Table 2 presents the benchmark CCSD(T) interaction energies and optimum intermolecular distances for each complex considered using each basis set. Table 3 presents

Table 2: Estimated CCSD(T) Interaction Energies and Optimal Intermolecular Distances. All Energies in kcal mol^{-1}.

Basis X	Sandwich Benzene		T-Shaped Benzene		Methane-Benzene[a]		H$_2$S-Benzene		Methane-Methane	
	ΔE	R_e/Å	ΔE	R_e/Å	ΔE	R_e/Å	ΔE	R_e/Å	ΔE	R_e/Å
D	-1.386	4.0	-2.312	5.1	-1.214	3.9	-2.346	3.9	-0.429	3.8
T	-1.640	3.9	-2.593	5.0	-1.399	3.8	-2.640	3.8	-0.505	3.7
Q	-1.786	3.9	-2.809	5.0	-1.439	3.8	-2.754	3.8	-0.525	3.7
CBS	-1.833	3.9	-2.871	4.9	-1.466	3.8	-2.843	3.8	-0.541	3.6

[a]Energies from Ref[58].

the MP2/SCS-type interaction energies at the corresponding CCSD(T)/aug-cc-pVXZ intermolecular distances. Conventional MP2 overestimates the binding energy for all cases but the methane dimer, with relative errors ranging from 8-90% depending on the basis set. By contrast, except for the methane dimer, the SCS-type methods provide much better agreement with CCSD(T), with errors typically 0.5 kcal mol^{-1} or less (or within about 20%). However, larger errors are observed when the local correlation approximation is used in conjunction with the aug-cc-pVDZ basis set; SCS-LMP2 and SCSN-LMP2 provide a much stronger interaction with this basis than their SCS-MP2 and SCSN-MP2 counterparts, and the errors approach 1 kcal mol^{-1} in some cases (in the case of methane dimer, by chance this local correlation error in the aug-cc-pVDZ basis nearly cancels the error of the SCS methods). For larger basis sets, the local approximation is much better. The large errors seen for the aug-cc-pVDZ basis set will be discussed in more detail below. For the rest of this section, SCS methods and their variants using only triple-ζ quality basis sets or better will be discussed. With these basis sets, SCS-(L)MP2 tends to underestimate the intermolecular interaction strength compared to CCSD(T), while SCSN-(L)MP2 tends to overestimate the interaction strength. Overall, excepting the methane dimer,

Table 3: Interaction Energies and Errors vs Estimated CCSD(T) for Various MP2 Methods Using aug-cc-pVXZ Basis Sets.[a,b] All Energies in kcal mol^{-1}.

Basis X	Sandwich Benzene ΔE	Error	T-Shaped Benzene ΔE	Error	Methane-Benzene ΔE	Error	H$_2$S-Benzene ΔE	Error	Methane-Methane ΔE	Error
SCS-MP2										
D	-1.516	-0.130	-2.104	0.208	-0.987	0.227	-2.193	0.153	-0.248	0.181
T	-1.758	-0.118	-2.329	0.264	-1.118	0.281	-2.431	0.210	-0.275	0.230
Q	-1.824	-0.038	-2.410	0.399	-1.177	0.262	-2.551	0.203	-0.297	0.228
CBS	-1.865	-0.032	-2.431	0.439	-1.214	0.252	-2.639	0.204	-0.276	0.266
SCS-LMP2										
D	-2.086	-0.700	-2.660	-0.348	-1.325	-0.111	-2.273	0.073	-0.427	0.003
T	-1.809	-0.169	-2.371	0.222	-1.147	0.252	-2.348	0.292	-0.269	0.236
Q	-1.752	0.034	-2.336	0.473	-1.148	0.291	-2.453	0.301	-0.292	0.233
CBS	-1.797	0.036	-2.374	0.496	-1.200	0.266	-2.584	0.258	-0.276	0.265
SCSN-MP2										
D	-1.552	-0.166	-2.644	-0.332	-1.253	-0.039	-2.730	-0.384	-0.296	0.133
T	-1.827	-0.187	-2.853	-0.260	-1.378	0.021	-2.934	-0.294	-0.325	0.181
Q	-1.882	-0.096	-2.890	-0.081	-1.412	0.019	-2.985	-0.231	-0.341	0.184
CBS	-1.915	-0.082	-2.917	-0.046	-1.431	0.035	-3.022	-0.180	-0.315	0.227
SCSN-LMP2										
D	-2.114	-0.728	-3.271	-0.959	-1.612	-0.398	-2.873	-0.527	-0.424	0.005
T	-1.839	-0.199	-2.883	-0.290	-1.404	-0.005	-2.892	-0.252	-0.312	0.194
Q	-1.814	-0.028	-2.834	-0.025	-1.397	0.042	-2.929	-0.175	-0.338	0.187
CBS	-1.882	-0.049	-2.904	-0.033	-1.443	0.023	-3.011	-0.168	-0.325	0.216
MP2										
D	-2.619	-1.233	-2.969	-0.657	-1.488	-0.274	-2.989	-0.643	-0.397	0.032
T	-3.047	-1.407	-3.359	-0.766	-1.718	-0.319	-3.408	-0.768	-0.462	0.043
Q	-3.123	-1.337	-3.433	-0.624	-1.779	-0.340	-3.529	-0.775	-0.486	0.039
CBS	-3.170	-1.337	-3.599	-0.728	-1.818	-0.352	-3.618	-0.776	-0.498	0.043

[a]Intermolecular energies and errors are evaluated at the estimated CCSD(T) optimized geometries for the given basis set (see Table 1). [b]Density fitting approximations applied to all methods except for CCSD(T).

the SCS-type methods are within 20% of CCSD(T). SCSN-LMP2 tends to give the best results for the complexes considered, with errors of 0.3 kcal mol^{-1} or less for basis sets of aug-cc-pVTZ quality or better.

Upon increasing the size of the basis set, the comparison to CCSD(T) binding energies in the same basis generally improves for the SCSN-MP2 and SCSN-LMP2 methods. By contrast, the errors for canonical MP2 versus CCSD(T) are nearly constant with respect to basis set. The SCS-MP2 and SCS-LMP2 methods have a more erratic behavior with respect to basis set, with errors sometimes decreasing but other times increasing as the basis set is expanded.

Having examined the performance of the SCS-type methods compared to CCSD(T) for each basis set, now consider the basis set dependence of these methods and how they compare to the CCSD(T)/CBS limit, which should be within a few tenths of one kcal mol^{-1} of the *ab initio* limit for these systems. Figure 8 presents the potential curves for the methane-benzene system with the SCSN-LMP2 method and the aug-cc-pVDZ through aug-cc-pVQZ basis sets. As previously noted, local approxi-
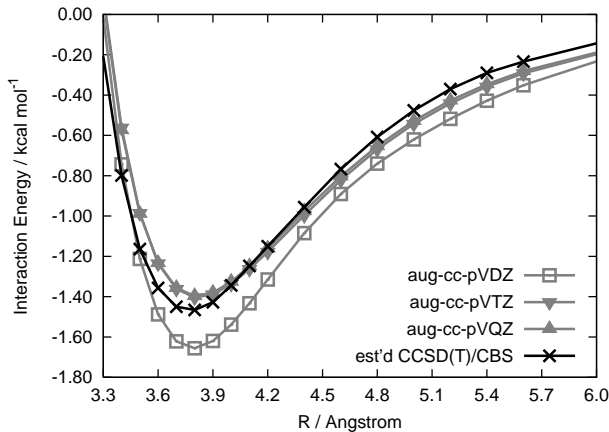
Figure 8: DF-SCSN-LMP2 potential energy curves for the methane-benzene complex.

mations with the aug-cc-pVDZ basis set lead to significant overbinding compared to CCSD(T)/CBS. However, the aug-cc-pVTZ and aug-cc-pVQZ curves are nearly coincident for this system, indicating rapid convergence with respect to basis set once the

aug-cc-pVTZ basis is used. The SCSN-LMP2 curves lie very close to CCSD(T)/CBS, being slightly above near equilibrium and slightly below at larger intermolecular separations.

Table 4 presents errors for the SCS-type and MP2 methods for each basis set as compared to the benchmark estimated CCSD(T)/CBS interaction energies from Table 2. Both because of the small size of the aug-cc-pVDZ basis and also because

Table 4: MP2-type/aug-cc-pVXZ Methods in Comparison to Estimated CCSD(T)/CBS Energies.[a,b] All Energies in kcal mol^{-1}.

Basis X	Sandwich Benzene ΔE	Error	T-Shaped Benzene ΔE	Error	Methane-Benzene ΔE	Error	H$_2$S-Benzene ΔE	Error	Methane-Methane ΔE	Error
SCS-MP2										
D	-1.527	0.306	-1.933	0.938	-0.940	0.526	-2.137	0.706	-0.155	0.386
T	-1.758	0.075	-2.262	0.609	-1.118	0.348	-2.431	0.412	-0.231	0.310
Q	-1.824	0.009	-2.360	0.511	-1.177	0.289	-2.551	0.292	-0.259	0.283
CBS	-1.865	-0.032	-2.431	0.439	-1.214	0.252	-2.639	0.204	-0.276	0.266
SCS-LMP2										
D	-2.123	-0.299	-2.542	0.329	-1.321	0.145	-2.231	0.611	-0.384	0.158
T	-1.809	0.024	-2.302	0.569	-1.147	0.319	-2.348	0.495	-0.221	0.321
Q	-1.752	0.081	-2.276	0.595	-1.148	0.318	-2.453	0.389	-0.250	0.291
CBS	-1.797	0.036	-2.374	0.496	-1.200	0.266	-2.584	0.258	-0.276	0.265
SCSN-MP2										
D	-1.558	0.275	-2.609	0.261	-1.239	0.227	-2.737	0.106	-0.216	0.326
T	-1.827	0.006	-2.840	0.031	-1.378	0.088	-2.934	-0.092	-0.283	0.258
Q	-1.882	-0.049	-2.885	-0.014	-1.412	0.054	-2.985	-0.142	-0.303	0.238
CBS	-1.915	-0.082	-2.917	-0.046	-1.431	0.035	-3.022	-0.180	-0.315	0.227
SCSN-LMP2										
D	-2.151	-0.318	-3.294	-0.424	-1.645	-0.179	-2.907	-0.064	-0.384	0.157
T	-1.839	-0.006	-2.866	0.004	-1.404	0.062	-2.892	-0.095	-0.266	0.275
Q	-1.814	0.019	-2.820	0.050	-1.397	0.069	-2.929	-0.087	-0.298	0.243
CBS	-1.882	-0.049	-2.904	-0.033	-1.443	0.023	-3.011	-0.168	-0.325	0.216
MP2										
D	-2.782	-0.949	-3.065	-0.194	-1.514	-0.048	-3.054	-0.211	-0.356	0.186
T	-3.047	-1.214	-3.441	-0.570	-1.718	-0.252	-3.408	-0.566	-0.449	0.092
Q	-3.123	-1.290	-3.533	-0.662	-1.779	-0.313	-3.529	-0.686	-0.479	0.062
CBS	-3.170	-1.337	-3.599	-0.728	-1.818	-0.352	-3.618	-0.776	-0.498	0.043

[a]Intermolecular energies and errors are evaluated at estimated CCSD(T)/CBS optimized geometries. [b]Density fitting approximations applied to all methods except for CCSD(T).

SCSN parameters were optimized using aug-cc-pVTZ basis sets, the aug-cc-pVDZ basis usually yields the largest errors for SCSN-(L)MP2 vs. CCSD(T)/CBS results. SCSN-MP2/aug-cc-pVDZ underbinds in all cases, with errors up to 16% (methanebenzene) and 60% (methane dimer), while SCSN-LMP2/aug-cc-pVDZ overbinds in all

cases except for the methane dimer, the largest relative errors being for the sandwich benzene dimer (17%) and the methane dimer (29%). On the other hand, for aug-cc-pVTZ basis sets or better, errors for SCSN-(L)MP2 becomes 6% or less and tends to overbind in the CBS limit for cases other than the methane dimer. In contrast to SCSN-(L)MP2, SCS-(L)MP2 tends to remain underbound even at the CBS limit, and errors vs. CCSD(T)/CBS generally decrease with increasing basis set. Significant differences in the quality of the SCS methods with respect to basis sets was also observed by Szabados[96] when comparing Feenburg's scaling parameters to Grimme's scaling parameters.

With respect to the optimal intermolecular distance, SCS-type methods also tend to outperform the MP2 method as shown in Table 5. Whereas the MP2 method un-

Table 5: Optimal Intermolecular Distances in Angstroms for Various Methods Evaluated at the CBS Limit.[a]

	CCSD(T)	SCS-MP2	SCS-LMP2	SCSN-MP2	SCSN-LMP2	MP2
Sandwich Benzene	3.9	3.9	3.9	3.9	3.9	3.7
T-Shaped Benzene	4.9	5.0	5.0	4.9	5.0	4.8
Methane-Benzene	3.8	3.8	3.8	3.8	3.8	3.7
H_2S-Benzene	3.8	3.8	3.8	3.8	3.8	3.7
Methane-Methane	3.6	3.8	3.8	3.8	3.8	3.7

[a]Density fitting approximations applied to all methods except for CCSD(T).

derestimates the optimal intermolecular distances for all systems except the methane dimer, SCS-type methods tend to reproduce the CCSD(T) minima. For the methane dimer, both the MP2 and the SCS-type methods overestimate the CCSD(T) optimal distances, with SCS-type methods producing larger distances than the MP2 method.

With the excellent performance of the SCSN-(L)MP2 method established for all systems except perhaps for the methane dimer, a brief discussion on the particulars of this method is warranted. As described above, the SCSN parameters have been optimized for the intermolecular interactions of nucleic acids. Specifically, the nucleic acid interactions contain of a set of ten nucleic acid-base pair stacked homo- and

hetero-dimers as optimized by Šponer *et al.*[97] All of these dimers consist of aromatic-π intermolecular interactions thus explaining the superior results obtained by the SCSN method for the systems containing the aromatic benzene ring. However, since the singlet pair contribution has been eliminated from the MP2 correlation energy, it is quite unphysical and is possibly only applicable for the computation of intermolecular interaction energies. The SCS method, on the other hand, still correlates both singlet and triplet pair contributions to the MP2 correlation energy and therefore has a wider applicability as a quantum chemical method.

The local correlation approximation has a significant effect on interaction energies when using the aug-cc-pVDZ basis set. Figure 9a shows the behavior of the SCS-type methods and their local counterparts for the methane-benzene system using this basis. In contrast to results seen for larger basis sets, for the aug-cc-pVDZ basis the local correlation approximation causes a significant shift in the PECs, with the local methods increasing the strength of the interaction by up to several tenths of one kcal mol^{-1}.

In an attempt to reduce the size of these shifts due to the local correlation approximation in the aug-cc-pVDZ basis set, the counterpoise correction (even though local correlation is already supposed to eliminate most of the leading terms contributing to the BSSE) was also considered.[86] Figure 9b shows the PECs for the methane-benzene complex when the counterpoise (CP) correction is made to the local methods. The CP correction shifts the curves up significantly and leads to underbinding. Comparing to Figure 9a, agreement with CCSD(T)/aug-cc-pVDZ becomes better for SCSN-LMP2 but worse for SCS-LMP2 when the CP correction is applied.

The large effect of the CP correction even on local methods when using a double-ζ quality basis was previously noted by Hill *et al.*[83] in their study of the benzene dimer. However, those researchers did not notice as large a shift due to the local approximation indicated in this work. One difference in the two studies is that Hill *et*
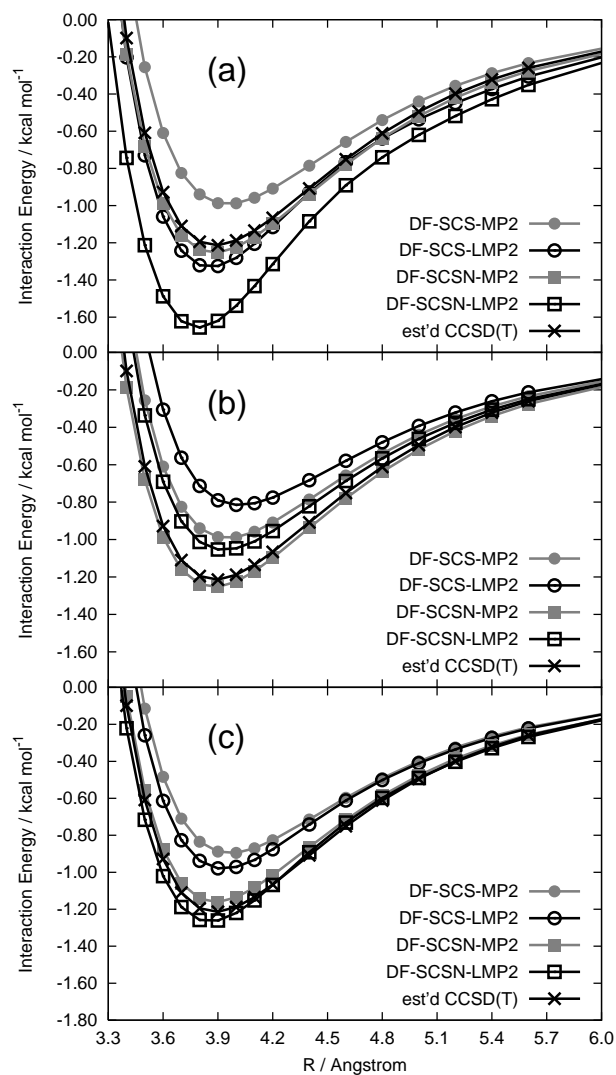
Figure 9: Potential energy curves for the methane-benzene complex using (a) the aug-cc-pVDZ basis, (b) the aug-cc-pVDZ basis and CP corrections for all methods, and (c) the aug-cc-pVDZ basis neglecting diffuse functions for hydrogen for MP2 methods and CP corrections only for nonlocal methods.

al. did not include diffuse functions on hydrogens (which is denoted here as aug-cc-pVDZ*). Figure 9c shows the PECs for the methane-benzene system with this aug-cc-pVDZ* basis. While the local (not CP corrected) energies still tend to overbind with respect to CP-corrected energies without the local approximation, they are much more similar to each other.

The other complexes generally show the same behavior, as shown in Table 6,

which presents interaction energies and errors versus CCSD(T)/aug-cc-pVDZ for the SCS-type methods using the aug-cc-pVDZ and aug-cc-pVDZ* basis sets. Except for

Table 6: Comparison of SCS and SCSN Methods with Various Double-ζ Basis Sets to CCSD(T)/aug-cc-pVDZ Energies.[a–d]. All Energies in kcal mol^{-1}.

| | Sandwich Benzene | | T-Shaped Benzene | | Methane-Benzene | | H$_2$S-Benzene | | Methane-Methane | |
Basis	ΔE	Error	ΔE	Error	ΔE	Error	ΔE	Error	ΔE	Error
SCS-MP2										
aug-cc-pVDZ	-1.516	-0.130	-2.104	0.208	-0.987	0.227	-2.193	0.153	-0.248	0.181
aug-cc-pVDZ*	-1.469	-0.089	-2.067	0.245	-0.888	0.325	-2.077	0.269	-0.168	0.262
SCS-LMP2										
aug-cc-pVDZ	-2.086	-0.700	-2.660	-0.348	-1.325	-0.111	-2.273	0.073	-0.427	0.003
aug-cc-pVDZ*	-1.592	-0.206	-2.220	0.092	-0.979	0.235	-2.076	0.270	-0.265	0.165
aug-cc-pVDZ(CP)	-1.313	0.073	-1.734	0.578	-0.790	0.424	-1.799	0.547	-0.209	0.220
SCSN-MP2										
aug-cc-pVDZ	-1.552	-0.166	-2.644	-0.332	-1.253	-0.039	-2.730	-0.384	-0.296	0.133
aug-cc-pVDZ*	-1.502	-0.122	-2.627	-0.315	-1.161	0.053	-2.628	-0.306	-0.213	0.216
SCSN-LMP2										
aug-cc-pVDZ	-2.114	-0.728	-3.271	-0.959	-1.612	-0.398	-2.873	-0.527	-0.424	0.005
aug-cc-pVDZ*	-1.634	-0.248	-2.824	-0.512	-1.261	-0.047	-2.652	-0.282	-0.291	0.139
aug-cc-pVDZ(CP)	-1.376	0.010	-2.284	0.028	-1.053	0.161	-2.450	-0.103	-0.250	0.180

[a]Intermolecular energies and errors are evaluated at CCSD(T)/aug-cc-pVDZ optimized geometries. [b]Density fitting approximations applied to all methods except for CCSD(T). [c]aug-cc-pVDZ* indicates cc-pVDZ for hydrogen and aug-cc-pVDZ for all other atoms. [d](CP) indicates counterpoise corrected energies for the local methods.

the H$_2$S-benzene complex, the local approximation leads to stronger intermolecular attraction. For the aug-cc-pVDZ basis, this shift can be as large as 0.6 kcal mol^{-1}. For the aug-cc-pVDZ* basis, the shifts become smaller and are typically 0.1-0.2 kcal mol^{-1}. Comparing to CCSD(T) within the aug-cc-pVDZ basis set, the best overall method appears to be SCSN-LMP2/aug-cc-pVDZ(CP) with errors up to 0.16 kcal mol^{-1} (13%) for H$_2$S-benzene or 0.18 kcal mol^{-1} (42%) for methane dimer. However this approach forfeits the computational advantage of neglecting the CP correction, which is one of the usual advantages of applying local correlation methods. For large systems where only double-ζ quality basis sets are feasible, it would appear that the best approach may be to apply SCSN-MP2/aug-cc-pVDZ (with CP correction); however, if local correlation methods are required to reduce computational costs, then results should be counterpoise corrected or else diffuse functions should be neglected

for hydrogen.

Throughout this discussion, it should be clear that poorer results for the methane dimer cast a shadow on the otherwise stellar performance of the SCS methods. While the errors observed for the methane dimer remain small on an absolute scale, they are large on a relative scale. The recent publication by Antony and Grimme[88] also notes less impressive performance for n-alkane dimers, with relative errors for binding energies remaining in the 30-40% range as one progresses from methane dimer to pentane dimer.

This raises the immediate question of whether one can determine ahead of time whether or not SCS methods are likely to be highly accurate for a given weakly-interacting system, beyond the already-apparent distinction between alkane dimers and other systems. A symmetry-adapted perturbation theory[95] (SAPT2/aug-cc-pVDZ) analysis was performed on the methane dimer at the CCSD(T)/aug-cc-pVQZ optimal inter-monomer distance and found, as expected, that this is a dispersion-dominated complex, with energetic contributions of -0.20 (electrostatic), 0.67 (exchange), -0.03 (induction), and -0.87 kcal mol^{-1} (dispersion), where exchange-induction and exchange-dispersion cross terms have been treated as induction and dispersion, respectively. Here the ratio of dispersion to electrostatic energies is much larger than for the T-shaped benzene dimer,[51] the H_2S-benzene complex,[92] or the methane-benzene complex.[58] However, near its optimal inter-monomer separation, the sandwich benzene dimer also has a much larger dispersion contribution than electrostatic contribution,[51] and in that case, SCS methods work well. The dispersion interaction typically has about the same size but opposite sign compared to the exchange-repulsion term; if the sum of these two contributions are considered, they approximately cancel for the the H_2S-benzene and methane-benzene complexes and for the T-shaped benzene dimer, leaving the electrostatic effect as the dominant remaining term. For the sandwich benzene dimer, the sum of exchange and dispersion remains

attractive by approximately 0.5 kcal mol^{-1} but is overtaken by a 1 kcal mol^{-1} attractive electrostatic interaction. Only in the case of the methane dimer does the sum of exchange and dispersion remain as large as the electrostatic component (both at about 0.2 kcal mol^{-1}). Whether such considerations might be diagnostic of the performance of SCS methods for other types of noncovalent interactions remains an open question.

2.2.4 Conclusions

For basis sets of augmented triple-ζ quality or better, the SCS-MP2, SCS-LMP2, SCSN-MP2, and SCSN-LMP2 methods can reliably reproduce the PECs from the more robust CCSD(T) method for the sandwich and T-shaped benzene dimers and for the methane-benzene and H$_2$S-benzene complexes. Results for the methane dimer are less accurate, although the absolute errors remain on the order of a few tenths of one kcal mol^{-1} across the PEC. The SCSN parameters appear to work best on average, and for the aug-cc-pVTZ basis set or larger, interaction energies are within 6% of the benchmark CCSD(T)/CBS values for all cases but the methane dimer [with an error of 0.2-0.3 kcal mol^{-1}, or 40-50% at the CCSD(T) inter-molecular distance]. In general, the SCS-type methods performed significantly better than standard MP2 for the cases considered here. With the exception of alkane dimers, the SCS methods appear to provide a significant improvement over conventional MP2 for PECs of weakly interacting systems. Except for the methane dimer, in the complete-basis-set limit the SCS-(L)MP2 potential curves tend to lie slightly above, and SCSN-(L)MP2 curves slightly below, the benchmark CCSD(T)/CBS curves, with SCSN-(L)MP2 being a little overbound at large intermolecular separations.

Results with the aug-cc-pVDZ basis are less accurate, and in this basis set, local correlation approximations lead to significant increases in binding. For the systems considered, errors for SCS-type LMP2 methods using this basis can be significantly

reduced by applying the counterpoise correction or by neglecting diffuse functions on hydrogen atoms.

2.3 Improvement of the CCSD method Via Scaling Same- and Opposite-Spin Components of the Double Excitation Correlation Energy

2.3.1 Introduction

This section is based on previously published work by Takatani, Hohenstein, and Sherrill. [2]

The CCSD method has received little attention compared to the MP2-type methods and the CCSD(T) method for applications to intermolecular interactions, being computationally more expensive than the MP2 method and less accurate than the CCSD(T) method. Indeed, in comparison to the CCSD(T) method, it appears to underbind by about as much as the MP2 method overbinds. [53] On the other hand, the CCSD method does offer a more robust treatment of electron correlation than the MP2 method while being less expensive than the CCSD(T) method. Moreover, just as for the MP2 method, the double excitation terms in the CCSD expression can be decomposed into same- and opposite-spin components. Since the reference Hartree-Fock energy already has some same-spin electron correlation contribution (the same argument used for SCS-MP2), this unbalanced description also carries into the CCSD method (albeit to a smaller degree than in the MP2 method). Therefore by separating the spin-components and scaling them accordingly, one should expect greater accuracy from a spin-component scaled CCSD (SCS-CCSD) method than the SCS-MP2 method.

This study explores the separation and scaling of the CCSD method. The parameters will be determined by a procedure similar to that of Grimme, [54] fitting to a set of reaction energies computed with the accurate CCSD(T) method. To correct for the unbalanced description of the correlation energy of different spin components,

the opposite-spin scaling factor is expected to be larger than the same-spin scaling factor. Finally, full PECs will be presented for the sandwich benzene dimer and the methane dimer interaction energies. The SCS-CCSD method is expected to outperform SCS-MP2 for both systems even though neither dimer is included in the set of reaction energies used to determine the scaling parameters.

2.3.2 Specialized Theoretical Methodology

Assuming a Hartree-Fock reference determinant, the closed shell MP2 correlation energy summed over unique pairs of spatial orbitals is

$$E_{MP2} = \sum_{ijab} \frac{(ia|jb)}{\Delta_{ij}^{ab}} [2(ia|jb) - (ib|ja)] \tag{49}$$

where

$$\Delta_{ij}^{ab} = \epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j, \tag{50}$$

and the indices i, j and a, b correspond to occupied and virtual orbitals, respectively. ϵ_i denotes the energy of the orbital i and $(ia|jb)$ is the two-electron integrals over spatial orbitals in the usual notation.[9] The MP2 energy can be further decomposed to

$$E_{MP2} = \sum_{ijab} \frac{(ia|jb)^2 + (ia|jb)[(ia|jb) - (ib|ja)]}{\Delta_{ij}^{ab}}, \tag{51}$$

where the same-spin component is

$$E_{SS-MP2} = \sum_{ijab} \frac{(ia|jb)[(ia|jb) - (ib|ja)]}{\Delta_{ij}^{ab}} \tag{52}$$

and the opposite-spin component is

$$E_{OS-MP2} = \sum_{ijab} \frac{(ia|jb)^2}{\Delta_{ij}^{ab}}. \tag{53}$$

The closed shell CCSD correlation energy summed over unique pairs of spatial orbitals is[10]

$$E_{CCSD} = \sum_{ijab} \tau_{ij}^{ab} [2(ia|jb) - (jb|ia)], \tag{54}$$

where

$$\tau_{ij}^{ab} = t_{ij}^{ab} + t_i^a t_j^b - t_i^a t_j^b. \tag{55}$$

t_{ij}^{ab} and $t_i^a t_j^b$ can be combined because they both describe an overall two-electron excitation. The expression looks similar to the MP2 equation and can be split to give the same-spin component

$$E_{SS-CCSD} = \sum_{ijab} \tau_{ij}^{ab}[(ia|jb) - (ib|ja)] \tag{56}$$

and the opposite-spin component

$$E_{OS-CCSD} = \sum_{ijab} \tau_{ij}^{ab}(ia|jb). \tag{57}$$

The SCS-CCSD energy can then be expressed as

$$E_{SCS-CCSD} = \sum_{ia} f_{ia} t_i^a + p_{SS} E_{SS-CCSD} + p_{OS} E_{OS-CCSD}, \tag{58}$$

where p_{SS} and p_{OS} are the scaling parameters for the same- and opposite-spin contributions, respectively. Scaling the energy contribution of the first term (products of single excitation amplitudes with Fock matrix elements) was not attempted.

The MOLPRO package of *ab initio* programs[56] was used to obtain all geometries for the 48 reaction energy test set. Excluding three reaction energies, this test set was identical to the test set used to parameterize the SCS-MP2 method.[54] No zero-point energy or entropy corrections were made. DFT optimizations, employing the B3LYP functional,[98,99] were performed with the correlation-consistent triple-ζ (cc-pVTZ) basis set of Dunning *et al.*[66] For later comparison, single point B3LYP energies were computed for the optimized geometries with the cc-pVQZ basis set.

The PSI3 suite of *ab initio* programs[100] was used to split the CCSD method into its respective same- and opposite-spin components. Single point CCSD and CCSD(T) energies with the cc-pVQZ basis set were computed at the converged B3LYP geometries. Using the CCSD(T)/cc-pVQZ reaction energies as a reference, the parameters

of the spin-component scaled CCSD (SCS-CCSD) method were optimized by minimizing the mean absolute deviation and were determined to be 1.13 and 1.27 for the same- and opposite-spin scaling factors, respectively.

The estimated CCSD(T), CCSD, and SCS-CCSD aug-cc-pVTZ PECs for the sandwich benzene dimer were obtained by an additive approximation combining MP2/aug-cc-pVTZ results with the difference between the coupled-cluster methods and the MP2 method evaluated at the aug-cc-pVDZ basis set.[101] CBS PECs for the methane dimer were obtained by adding a two-point extrapolation[71] of the aug-cc-pVTZ and aug-cc-pVQZ CCSD(T) correlation energies to the reference Hartree-Fock aug-cc-pVQZ energy. The CCSD(T)/CBS curves were obtained in separate studies.[1,3]

2.3.3   Results and Discussion

Table 7 shows the errors of the B3LYP, MP2, SCS-MP2, CCSD, and SCS-CCSD methods compared to 48 CCSD(T)/cc-pVQZ reaction energies. Minimizing the mean absolute deviation (MAD), the scaling factors of SCS-CCSD were determined to be 1.13 and 1.27 for the same- and opposite-spin factors, respectively. With respect to the MAD of each method, the MP2 method performs similarly to the B3LYP method (3.0 and 2.6 kcal mol^{-1}, respectively), and the SCS-MP2 method performs similarly to the CCSD method (1.8 and 1.9 kcal mol^{-1}, respectively). Both the SCS-MP2 and CCSD methods outperform the MP2 and B3LYP methods. The SCS-CCSD method clearly outperforms all the methods considered here with a MAD of only 1.1 kcal mol^{-1}.

By minimizing the MAD rather than the least squares deviation, the SCS-CCSD method incurs a greater maximum deviation than the SCS-MP2 method. However, most large SCS-CCSD errors correspond to even larger CCSD errors. This is best exemplified by the reaction of O_3 and H_2. The CCSD method gives an error of -14.1

Table 7: CCSD(T) reaction energies and errors for other methods compared to CCSD(T) all evaluated with the cc-pVQZ basis set. All energies and errors in kcal mol^{-1}

Reaction	CCSD(T)	MP2	SCS-MP2	CCSD	SCS-CCSD	B3LYP
$F_2 + H_2 \rightarrow 2HF$	-134.1	-7.6	-2.6	-3.0	0.0	4.9
$F_2O + H_2 \rightarrow F_2 + H_2O$	-68.4	-3.3	-2.0	-2.7	-1.2	5.5
$O_3 + 3H_2 \rightarrow 3H_2O$	-223.4	1.9	3.4	-14.1	-5.4	5.5
$H_2O_2 + H_2 \rightarrow 2H_2O$	-86.3	-4.2	-1.6	-1.9	-0.3	3.6
$CO + H_2 \rightarrow CH_2O$	-4.6	-0.6	0.8	-0.1	-1.1	-3.0
$CO_2 + 3H_2 \rightarrow CH_4 + H_2O$	-63.6	-1.9	2.2	-1.6	-2.4	-0.5
$N_2 + 3H_2 \rightarrow 2NH_3$	-38.0	0.2	4.5	-1.6	-1.7	-2.0
$^1CH_2 + H_2 \rightarrow CH_4$	-128.8	-7.1	-2.0	1.0	-0.4	-1.3
$N_2O + H_2 \rightarrow N_2 + H_2O$	-80.7	4.7	1.4	-5.7	-2.8	8.0
$HNO_2 + 3H_2 \rightarrow 2H_2O + NH_3$	-121.6	-4.2	-1.5	-5.7	-3.3	5.7
$C_2H_2 + H_2 \rightarrow C_2H_4$	-49.4	2.0	2.5	-0.8	-0.4	-0.2
$CH_2=C=O + 2H_2 \rightarrow CH_2O + CH_4$	-43.1	2.0	0.8	-2.2	-1.8	3.2
$BH_3 + 3HF \rightarrow BF_3 + 3H_2$	-94.3	-0.7	0.4	1.3	1.4	-1.9
$HCOOH \rightarrow CO_2 + H_2$	2.0	-2.6	-3.0	1.3	1.7	-1.8
$CO + H_2O \rightarrow CO_2 + H_2$	-6.7	-4.0	-1.9	2.4	1.9	-6.2
$C_2H_2 + HF \rightarrow C_2H_3F$	-27.3	3.0	2.8	-0.1	0.3	-3.1
$HCN + H_2O \rightarrow CO + NH_3$	-12.6	2.9	1.7	-0.6	1.0	0.1
$HCN + H_2O \rightarrow HCONH_2$	-21.8	0.7	2.6	0.3	1.1	-4.9
$HCONH_2 + H_2 \rightarrow HCOOH + NH_3$	0.4	0.8	0.3	0.1	0.1	0.5
$HCN + NH_3 \rightarrow N_2 + CH_4$	-38.2	0.9	-0.7	-0.6	0.4	1.5
$CO + CH_4 \rightarrow CH_3CHO$	3.4	-1.4	1.1	0.7	-0.7	-2.3
$O_3 + CH_4 \rightarrow 2H_2O + CO$	-159.7	3.7	1.2	-12.5	-3.0	6.0
$N_2 + F_2 \rightarrow N_2F_2$	17.3	1.2	4.7	1.8	1.2	-5.9
$BH_3 + 2F_2 \rightarrow BF + 3HF$	-248.1	-11.5	-2.3	-4.9	0.4	11.8
$2BH_3 \rightarrow B_2H_6$	-43.3	-1.4	3.7	2.8	-0.7	3.7
$2BH_3 \rightarrow B_2H_6$	-43.3	-1.4	3.7	2.8	-0.7	3.7
$2\,^1CH_2 \rightarrow C_2H_4$	-199.4	-13.1	-4.0	3.4	0.0	-2.6
$CH_3ONO \rightarrow CH_3NO_2$	-3.2	-4.9	-2.7	0.8	0.6	-0.7
$CH_2=C \rightarrow C_2H_2$	-44.8	-7.7	-4.7	1.4	-0.3	1.3
Allene $\rightarrow$ Propyne	-1.3	-3.3	-2.8	-0.3	-0.3	3.2
Cyclopropene $\rightarrow$ Propyne	-23.4	-0.3	-1.1	-0.3	0.3	-0.4
Oxirane $\rightarrow$ CH$_3$CHO	-26.5	1.0	0.1	-0.2	0.7	-1.5
Vinylalcohol $\rightarrow$ CH$_3$CHO	-10.9	-0.4	-1.1	-0.2	0.2	-0.1
Cyclobutene $\rightarrow$ 1,3-Butadiene	-11.3	2.1	0.8	0.3	1.1	-4.0
$C_2H_4 + \,^1CH_2 \rightarrow C_3H_6$	-115.1	-8.1	-2.1	1.8	-0.2	0.4
$C_2H_2 + C_2H_4$ Cyclobutene	-32.5	-2.0	1.4	0.4	-0.7	4.8
$3C_2H_2 \rightarrow$ Benzene	-153.3	-5.5	4.5	2.1	2.5	2.0
$HCN \rightarrow CNH(TS)^a$	47.7	4.3	4.5	0.4	0.8	-0.2
Cyclobutene $\rightarrow$ 1,3-Butadiene(TS)^a	35.0	-0.8	1.4	3.3	1.9	-1.8
$^3CH_2 \rightarrow \,^1CH_2$	9.7	4.6	-1.7	0.9	-4.6	0.8
$HF + H^+ \rightarrow H_2F^+$	-122.9	0.9	0.0	-0.2	-0.4	0.8
$H_2O + H^+ \rightarrow H_3O^+$	-173.0	1.2	-0.1	-0.7	-0.3	0.5
$NH_3 + H^+ \rightarrow H_4N^+$	-212.9	1.1	-0.5	-0.9	-0.4	0.7
$F^- + H^+ \rightarrow HF$	-390.4	2.3	0.9	-0.2	-0.3	2.7
$OH^- + H^+ \rightarrow H_2O$	-412.5	2.5	0.8	-0.8	-0.3	1.8
$NH_2^- + H^+ \rightarrow NH_3$	-426.1	1.9	0.0	-1.3	-0.6	1.1
$2NH_3 \rightarrow (NH_3)_2$	-3.4	-0.1	0.3	0.3	0.0	0.6
$2H_2O \rightarrow (H_2O)_2$	-5.4	-0.1	0.4	0.3	0.1	0.1
$2HF \rightarrow (HF)_2$	-4.9	0.0	0.4	0.3	0.1	-0.2
MAD^b		3.0	1.8	1.9	1.1	2.6
RMS^c		4.1	2.3	3.3	1.6	3.6
MAX^d		13.1	4.7	14.1	5.4	11.8

^{a}Transition states ^{b}Mean absolute deviation ^{c}Root mean squared deviation ^{d}Maximum absolute deviation

kcal mol^{-1} while the SCS-CCSD method corrects this significantly to give a deviation of -5.4 kcal mol^{-1}. If one were to minimize the least squares deviation, the errors from the two reactions with O_3 would be weighted too heavily to yield very accurate results for the other reaction energies. One possible pitfall of the SCS-CCSD method is that excitation energies can be inaccurate, as shown by the relatively large deviation (-4.6 kcal mol^{-1}) for the singlet-triplet gap of CH_2. Without additional data, it is not clear if this error is representative for energy gaps between different electronic states in general.

PECs for the sandwich benzene dimer with the aug-cc-pVTZ basis set are shown in Figure 10. The MP2 energies are overbound (not shown in the figure for clarity), while

Figure 10: Potential energy curves with the aug-cc-pVTZ basis set for the sandwich benzene dimer. Coupled-cluster methods estimated via an addition of the difference between MP2 and coupled-cluster correlation energies with the aug-cc-pVDZ basis set to the MP2/aug-cc-pVTZ energies. Energies are counterpoise-corrected.
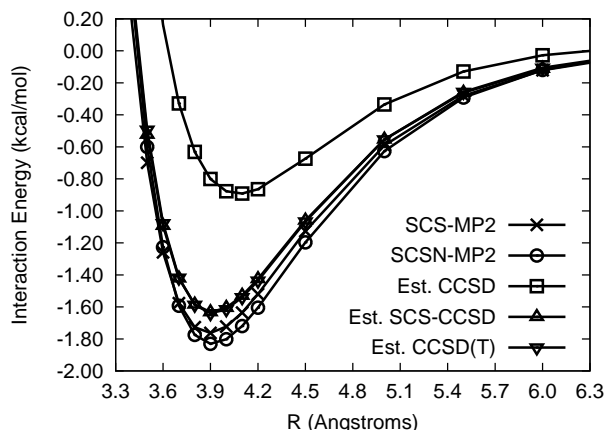
the CCSD method only recovers about half the interaction energy of the CCSD(T) method at the minimum. Both the SCS-MP2 and the SCSN-MP2 energies are within a couple tenths of one kcal mol^{-1} of the CCSD(T) minimum energy. The SCS-CCSD method performs extremely well by reproducing the CCSD(T) energies almost exactly.

PECs for the methane dimer extrapolated to the CBS limit are shown in Figure 11. In this case the MP2 energies are good approximations to the CCSD(T) energies,

Figure 11: Potential energy curves at the CBS limit for the methane dimer. Energies are counterpoise-corrected.
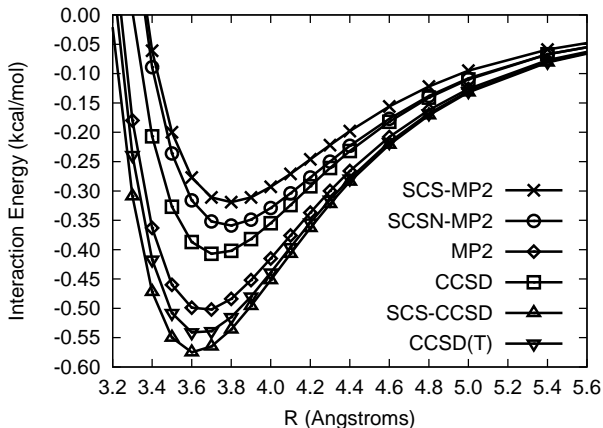
being within a few hundredths of one kcal mol^{-1}. On the other hand, the SCS-MP2 and SCSN-MP2 energies are underbound by a couple tenths of one kcal mol^{-1} at the CCSD(T) minimum. The CCSD energies are underbound by a little more than one tenth of one kcal mol^{-1} at the CCSD(T) minimum, and for this system the CCSD method outperforms both the SCS-MP2 and SCSN-MP2 methods. The SCS-CCSD method corrects the CCSD energies to overbind by only a few hundredths of one kcal mol^{-1}, about as much as the MP2 method underbinds.

Table 8 presents the interaction energies for each method at their respective minima and the optimal inter-monomer distances for the sandwich benzene dimer and the methane dimer. For the methane dimer, the MP2 method has an error of only 0.04

Table 8: Correlation methods in comparison to estimated CCSD(T) method. Energies in kcal mol^{-1} and distances in Angstroms.

	Sandwich Benzene[a,b]		Methane-Methane[c]	
	ΔE	R_e	ΔE	R_e
MP2	-3.25	3.7	-0.50	3.7
SCS-MP2	-1.76	3.9	-0.32	3.8
SCSN-MP2	-1.83	3.9	-0.36	3.8
CCSD	-0.89	4.1	-0.41	3.7
SCS-CCSD	-1.63	3.9	-0.57	3.6
CCSD(T)	-1.64	3.9	-0.54	3.6

[a]Evaluated with the aug-cc-pVTZ basis set. [b]Coupled-cluster methods estimated via an addition of the difference between MP2 and coupled-cluster correlation energies with the aug-cc-pVDZ basis set to the MP2/aug-cc-pVTZ energies. [c]Extrapolated to the CBS limit.

44

kcal mol^{-1}; on the other hand, the MP2 method dramatically overbinds the sandwich benzene dimer, with an absolute error of 1.61 kcal mol^{-1}. The CCSD method underbinds both the sandwich benzene dimer and the methane dimer with errors of 0.75 kcal mol^{-1} and 0.13 kcal mol^{-1}, respectively. The SCS-MP2 and SCSN-MP2 absolute errors are within a couple tenths of one kcal mol^{-1} for both systems. However, while a couple tenths of one kcal mol^{-1} error corresponds to a small relative error for the sandwich benzene dimer ($\sim$10%), for the methane dimer the relative error reaches 30-40%. The SCS-CCSD errors are significantly smaller than those for any of the other methods considered, with errors of 0.01 and -0.03 kcal mol^{-1} for the sandwich benzene dimer and methane dimer, respectively, compared to the CCSD(T) benchmarks. The SCS-CCSD method also predicts the optimal inter-monomer distance more consistently than any of the other methods considered. The MP2 and CCSD methods fail to obtain correct intermolecular distances for either system, with errors of 0.2 Å and 0.1 Å for the sandwich benzene dimer and the methane dimer, respectively. The SCS-MP2 type methods obtain the correct distance for the sandwich benzene dimer, but deviate by 0.2 Å for the methane dimer. Only the SCS-CCSD method predicts the correct optimal distances for both systems.

The main drawback of the SCS-CCSD method is the computational cost associated with the CCSD method. Because the SCS-CCSD method still scales less favorably than MP2-type methods [$O(\mathrm{N}^6)$ vs $O(\mathrm{N}^5)$], the MP2-type methods have a greater range of applicability for larger systems. However, work on coupled-cluster methods using density-fitting (or resolution of the identity) approximations has reduced the necessary disk storage[102] and computational cost, and local approximations have also helped speed up the CCSD method significantly.[103–105] Schütz and Manby improved the linear scaling of the local DF-CCSD method by localizing the auxiliary fitting functions (LDF-LCCSD).[106] They report that the LDF-LCCSD method can be as much as 100 times faster than the CCSD method. These new approaches to

coupled-cluster methods make the SCS-CCSD method considerably more practical for computations on large systems and the results are more reliable than those from the SCS-MP2 method.

2.3.4 Conclusions

The overall double excitation contribution to the CCSD method has been split into same- and opposite-spin components. The components have been parameterized for a set of 48 reaction energies by minimizing the mean average deviation in comparison to CCSD(T) energies. The scaling parameters are 1.13 and 1.27 for the same- and opposite-spin components, respectively. This new SCS-CCSD method outperforms the B3LYP, MP2, SCS-MP2, and CCSD methods with a MAD of only 1.1 kcal mol^{-1} for the reaction energies in the test set, and only deviates substantially from the CCSD(T)/cc-pVQZ benchmarks for reactions involving O_3 and the singlet-triplet gap of CH_2. The SCS-CCSD method also outperforms the SCS-MP2 type methods in describing the intermolecular interactions for the sandwich benzene dimer and methane dimer for both interaction energies and optimal inter-monomer distances. A previous study has computed CCSD energies for molecules as large as $(Gly)_{16}$,[106] and thus the SCS-CCSD method is feasible for systems with at least 65 heavy atoms. The density fitting and local CCSD approximation schemes are expected to allow for SCS-CCSD quality energies of even larger systems.

2.4 An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H_2S

2.4.1 Introduction

This section is based on previously published work by Sherrill, Takatani, and Hohenstein.[3]

In this work, the efficiency and reliability of several theoretical models and approximations as applied to nonbonded interactions are reviewed. A number of quantum-mechanical models, including DFT approaches, empirically-corrected DFT, and SCS-MP2[54,72,73] and SCS-CCSD[2] are considered. Approximations such as local correlation, dual-basis techniques, and RI/DF are also considered. These approximations are evaluated by comparison to benchmarks near the *ab initio* limit for several prototypes of noncovalent interactions. Given the demonstrated level of interest in such benchmark data, significantly larger computations are performed in order to improve on previously-published benchmark potential curves for the benzene dimer,[51] H_2S-benzene,[92,107] and CH_4-benzene.[58]

2.4.2 Specialized Theoretical Methodology

In generating earlier set of potential curves for the benzene dimer,[51] CCSD(T) energies were evaluated in a basis called aug-cc-pVDZ*, which is aug-cc-pVDZ for carbon and cc-pVDZ for hydrogen. Following Tschumper and co-workers,[108] in this work it will referred to as a "heavy-aug-cc-pVDZ" basis, abbreviated haDZ, meaning that only the heavy (non-hydrogen) atoms are augmented by diffuse functions. (Such a basis is also sometimes denoted as aug'-cc-pVDZ).

The performance of the various methods and approximations was assessed by comparison to very accurate benchmark values which estimate the CCSD(T) binding energies at the CBS limit. For the smallest test system, the methane dimer, the CCSD(T)/CBS limit value is most easily estimated by the two-point extrapolation scheme of Halkier et al[71] using the CCSD(T) energies in the aug-cc-pVTZ and aug-cc-pVQZ basis sets. This was the procedure used for the entire PEC for the methane dimer.

For larger benchmark systems, explicit CCSD(T)/aug-cc-pVQZ computations are much more difficult, and the ΔCCSD(T) approximation is required. In the past,

it was argued that augmented double-ζ basis sets seem large enough to converge ΔCCSD(T).[51] However, recent work by Janowski and Pulay suggests that additional improvements in the basis can change the value of ΔCCSD(T) somewhat.[109] The basis used to obtain ΔCCSD(T) is noted in parenthesis after the CCSD(T)/CBS designation. Thus, CCSD(T)/CBS(ΔaDZ) indicates that ΔCCSD(T) is evaluated in an aug-cc-pVDZ basis. Similarly, CCSD(T)/CBS(Δha(DT)Z) indicates that the ΔCCSD(T) correction was evaluated using a haDZ/haTZ extrapolation. In practice, the ΔCCSD(T) term did not change much whether a (h)aTZ basis or a (h)aDZ/(h)aTZ extrapolation was used. For the benzene dimer, complete PECs were computed at the CCSD(T)/CBS(Δha(DT)Z) level. For the methane-benzene and H_2S-benzene complexes, complete PECs were computed at the CCSD(T)/CBS(Δa(DT)Z) level.

All DFT computations were performed with Q-Chem 3.1[110] and MOLPRO[56]. Energy computations were performed with B3LYP,[111] PBE,[112] M05-2X,[29] and M06-2X[30,113] functionals along with Dunning's aug-cc-pVTZ basis set.[67] Computations using the M05-2X and M06-2X functionals utilized a Lebedev grid with 302 angular points for each of the 100 radial points included (100,302). This grid was chosen to avoid artifacts due to numerical integration for noncovalent interactions, particularly when using meta-GGA functionals.[114,115] In the case of methane dimer, an even larger, (200,1202), grid was necessary to completely remove these artifacts. Empirical dispersion corrections were added to B3LYP and PBE to improve their performance for treating noncovalent complexes; the dispersion corrections of Grimme were employed,[116,117] as implemented in Q-Chem 3.2. The empirical dispersion terms were scaled by 0.75 and 1.05 for PBE and B3LYP, respectively, as recommended by Grimme.[117] The dispersion corrected functionals will be referred to as PBE-D and B3LYP-D for the remainder of this work. Counterpoise correction was not applied to the DFT computations. The dual-basis (DB) SCF approximation,[118] also

48

implemented in Q-Chem 3.1,[110] employed corresponding basis subsets for the self-consistent field iterations.

2.4.3 Results and Discussion

2.4.3.1 Density Fitting, Local Correlation, and Dual-Basis Approximations

The RI/DF, local, and DB approximations are meant to speed up conventional *ab initio* electronic structure methods while introducing negligible errors. Tables 9 and 10 show the effectiveness for RI/DF and local correlation approximations as observed in the MOLPRO[56] and Q-Chem[110] packages. Here, errors and computational times are compared against a canonical MP2/aug-cc-pVTZ computation for the sandwich benzene dimer (828 basis functions) using a standard 3.2GHz Intel Woodcrest EM64T Linux workstation. For MOLPRO, the canonical computation requires 23.0 hours including the counterpoise correction computation. Spatial symmetry was not used for these tests (which would reduce the time to 12.7 hours), because the density fitting and local correlation approximations implemented in MOLPRO do not use spatial symmetry; this is unlikely to be a drawback in studies of large molecules, which are unlikely to have much, if any, symmetry to exploit. Density fitting the MP2 correlation energy (DF-MP2/HF) incurs essentially no error and reduces the computational time to 19 hours. Localizing the MP2 correlation energy (LMP2/HF) leads to a slight increase in computational time (to 28.7 hours), probably because this test case is too small for the benefits of local correlation to make up for the additional overhead costs. However, it is encouraging to note that the LMP2/HF result incurs an error of only 0.015 kcal mol^{-1}. Such an error is quite small and should not be of any practical concern for most studies. Using local correlation in conjunction with density fitting for MP2 reduces the computational cost compared to DF-MP2/HF by 1.5 hours.

Density fitting the Hartree-Fock procedure only (MP2/DF-HF) requires about one quarter of the time of the standard MP2 method and incurs an error of 0.006

Table 9: Effectiveness of Density Fitting and Local Correlation Approximations for the Sandwich Benzene Dimer. Errors versus MP2/aug-cc-pVTZ in kcal mol^{-1} and Time in Hours.[a]

MP2		SCF		
DF	LOC[b]	DF	Error	Time
			0.000	23.0
X			-0.001	19.2
	X		0.015	28.7
		X	0.006	6.2
X	X		0.011	17.7
X		X	0.005	1.7
X	X	X	0.013	3.2

[a]Evaluated at an intermolecular separation of 3.8 Å, using the cc-pVTZ JK and MP2-fitting basis sets as the auxiliary basis sets for the Hartree-Fock and MP2 procedures, respectively. Timings using the MOLPRO 2006.1 program[56] obtained on a 3.2 GHz Intel Woodcrest EM64T workstation. An 'X' denotes the use of the given approximation. All computations evaluated without use of spatial symmetry. [b]Not counterpoise corrected.

Table 10: Effectiveness of Resolution-of-the-Identity (RI) MP2 and Dual-Basis Self-Consistent-Field (SCF) Approximations for the Sandwich Benzene Dimer. Errors versus MP2/aug-cc-pVTZ in kcal mol^{-1} and Time in Hours.[a]

MP2	SCF		
RI	DB	Error	Time
		0.000	24.3
X		-0.001	12.2
	X	0.012	16.6
X	X	0.010	2.8

[a]Evaluated at an intermolecular separation of 3.8 Å, using the cc-pVTZ MP2-fitting basis set as the auxiliary basis set for the MP2 procedure, and the default Q-Chem basis subset for the dual-basis SCF procedure. Timings using the Q-Chem 3.1 program[110] obtained on a 3.2 GHz Intel Woodcrest EM64T workstation. An 'X' denotes the use of the given approximation. All computations evaluated without use of spatial symmetry.

kcal mol^{-1}. This signifies that the time consuming part for this test case is the SCF iterations. The DF-MP2/DF-HF method dramatically reduces the computational time to under 2 hours with an error of only 0.005 kcal mol^{-1}. Given this result, density fitting both the MP2 and HF parts of the computation is recommended for routine or even benchmark-quality computations. Adding local correlation (DF-LMP2/DF-HF) increases the computational time somewhat, probably because of overhead which is not recovered for such a small molecular system.

As noted above, the SCF procedure can be one of the major time-consuming steps in an MP2 computation. Another technique for reducing this cost is the DB-SCF procedure of Head-Gordon and co-workers,[118] which is implemented in the Q-Chem program.[110] Table 10 examines the RI and DB-SCF approximations. In Q-Chem, the reference MP2/aug-cc-pVTZ computation on the sandwich benzene dimer requires 24.3 hours to complete using the same hardware used above (again, spatial symmetry is not utilized in these comparisons). Applying the resolution-of-the-identity (RI-MP2) approximation to the MP2 method (equivalent to density fitting) requires about half the time of the standard MP2 method and incurs an extremely small error of -0.001 kcal mol^{-1}. The MP2 method combined with the dual-basis SCF (MP2/DB-SCF) approximation decreases the time to 16.6 hours, and incurs a somewhat greater (but still negligible) error of 0.012 kcal mol^{-1}. The most efficient approximation tested in Q-Chem is the combined RI-MP2 and DB-SCF procedures (RI-MP2/DB-SCF). The RI-MP2/DB-SCF method has an error of 0.01 kcal mol^{-1} with a computational time of 2.8 hours. It would be interesting to also apply the RI procedure to the SCF iterations in combination with the dual-basis and RI-MP2 approximations to further decrease computational costs.

Figures 12-14 depict for the benzene dimer test cases (a) the convergence of the (explicit or estimated) CCSD(T) binding energies with respect to basis set; (b) the errors in interaction energies for spin-component-scaled methods; and (c) the errors in various DFT approximations. In this section, the focus will be on the performance of the SCS methods.

Figure 12: (a) CCSD(T) PECs for the sandwich benzene dimer. Interaction energy errors for (b) SCS-type approximations and (c) DFT methods compared to the CCSD(T)/CBS(Δha(DT)Z) interaction energies.
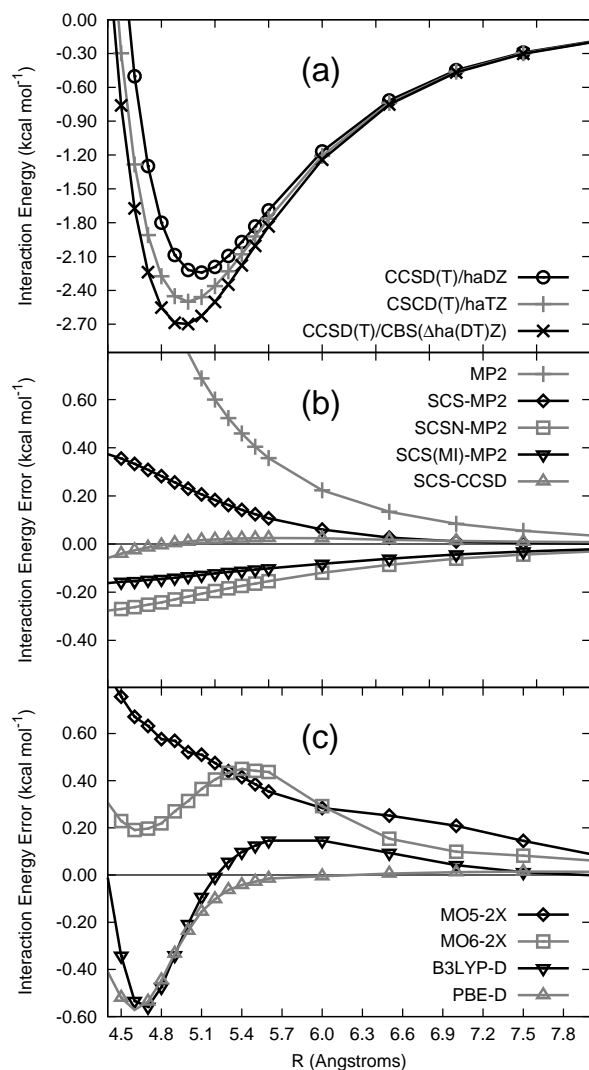
Figure 13: (a) CCSD(T) PECs for the T-shaped benzene dimer. Interaction energy errors for (b) SCS-type approximations and (c) DFT methods compared to the CCSD(T)/CBS(Δha(DT)Z) interaction energies.

Of all the SCS or DFT methods considered in this work, the only one which uniformly exhibits a high degree of accuracy is the SCS-CCSD method (it is, not surprisingly, also the most computationally expensive, having the same cost as conventional CCSD). The magnitude of the interaction energy error is below 0.2 kcal mol^{-1} for any of the points considered along the benzene dimer PECs; it only approaches this value at short intermolecular distances, where electron correlation effects are largest (and where most other approximate methods also have increasing errors). The SCS- and

53
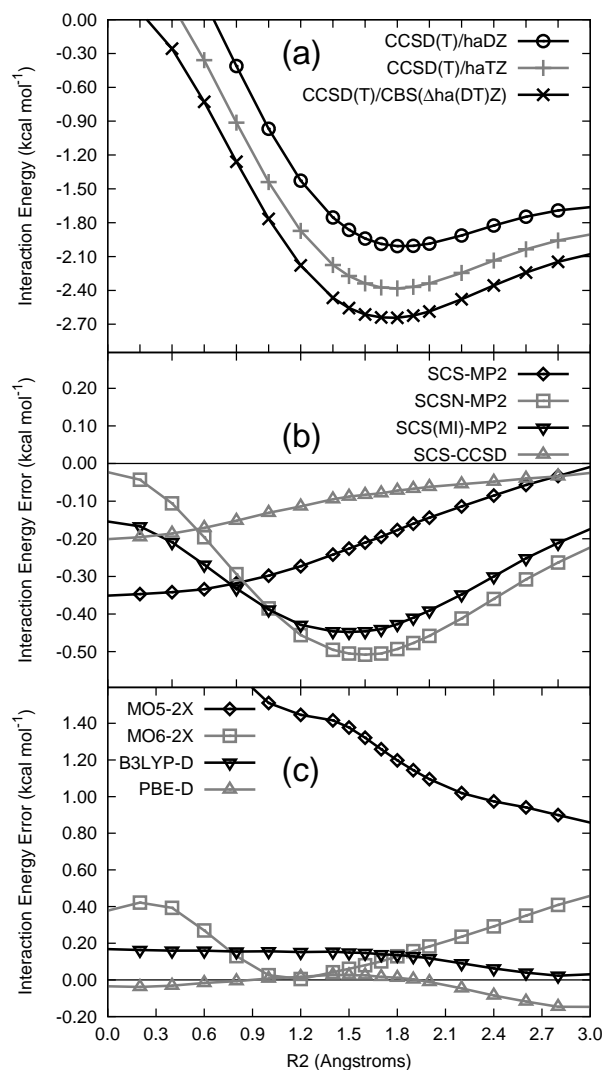
Figure 14: (a) CCSD(T) PECs for the PD benzene dimer at an R1 displacement of 3.4 Å. Interaction energy errors for (b) SCS-type approximations and (c) DFT methods compared to the CCSD(T)/CBS(Δha(DT)Z) interaction energies.

SCSN-MP2 methods do not perform nearly as well, having errors of one to several tenths of one kcal mol^{-1}. While the SCS-MP2 error is generally a monotonic function of intermolecular separation in Figures 12-14, the SCSN-MP2 error is more erratic and can increase or decrease with distance. The SCS(MI)-MP2 error curves have a similar shape to those from SCSN-MP2, but the magnitude of the error is generally smaller.

Figures 15 and 16 show results of the SCS and DFT approximations for the

methane-benzene and H$_2$S-benzene complexes. Again, the SCS-CCSD method does extremely well, with errors of about 0.1 kcal mol^{-1} or less across the PECs. The SCS(MI)-MP2 errors again mimic the SCSN-MP2 errors, but they are generally somewhat better. The SCS-MP2 method exhibits errors of a couple tenths of one kcal mol^{-1} at equilibrium distances for both complexes.
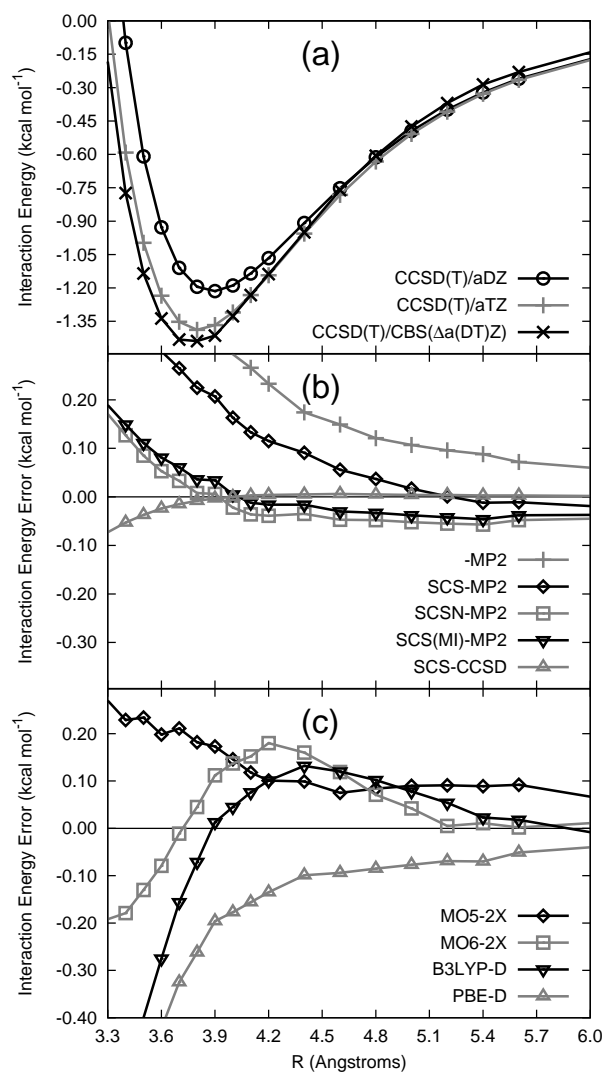
Figure 15: (a) CCSD(T) PECs for the methane-benzene complex. Interaction energy errors for (b) SCS-type approximations and (c) DFT methods compared to the CCSD(T)/CBS(Δa(DT)Z) interaction energies.

Interaction energy errors for the methane dimer exhibit a different behavior than the other noncovalent complexes, as shown Figure 17. In this case, the unaltered
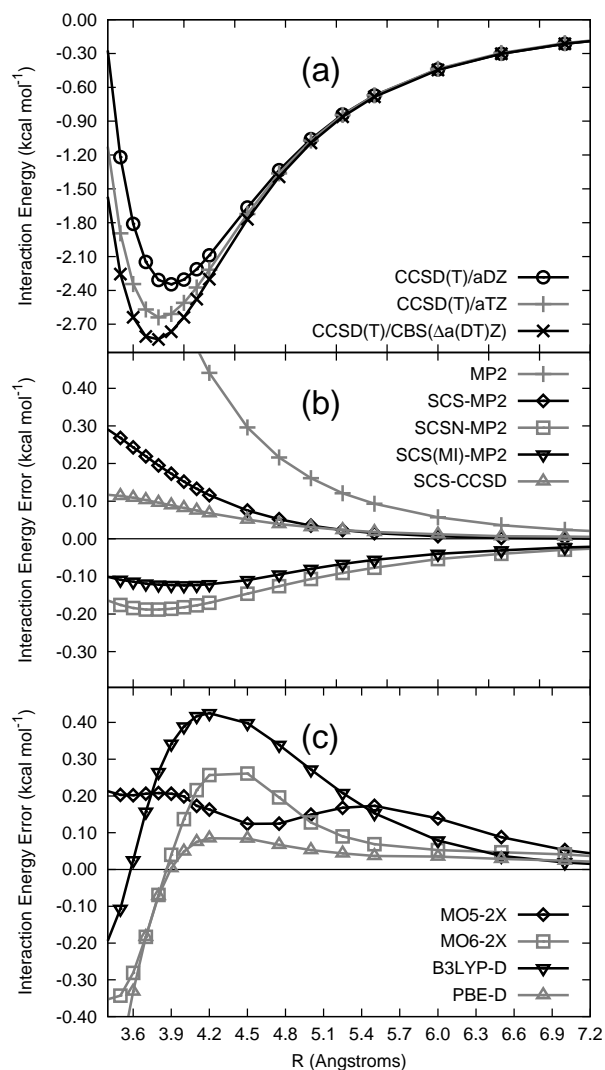
Figure 16: (a) CCSD(T) PECs for the H$_2$S-benzene complex. Interaction energy errors for (b) SCS-type approximations and (c) DFT methods compared to the CCSD(T)/CBS(Δa(DT)Z) interaction energies.

MP2 method reproduces the benchmark PEC to less than four hundredths of one kcal mol^{-1}. Because MP2 typically overbinds van der Waals complexes, the SCS methods typically scale the interaction energy to make it smaller in magnitude; unfortunately, because MP2 is actually correct in this case, the scaling overcorrects. Thus, all scaled MP2 values are underbound up to about 3.9 Å (particularly SCS-MP2). Around the equilibrium distance, the scaled MP2 methods all underestimate the binding energy by 0.2 kcal mol^{-1}. The SCS-CCSD method reliably incurs errors of less than 0.1 kcal
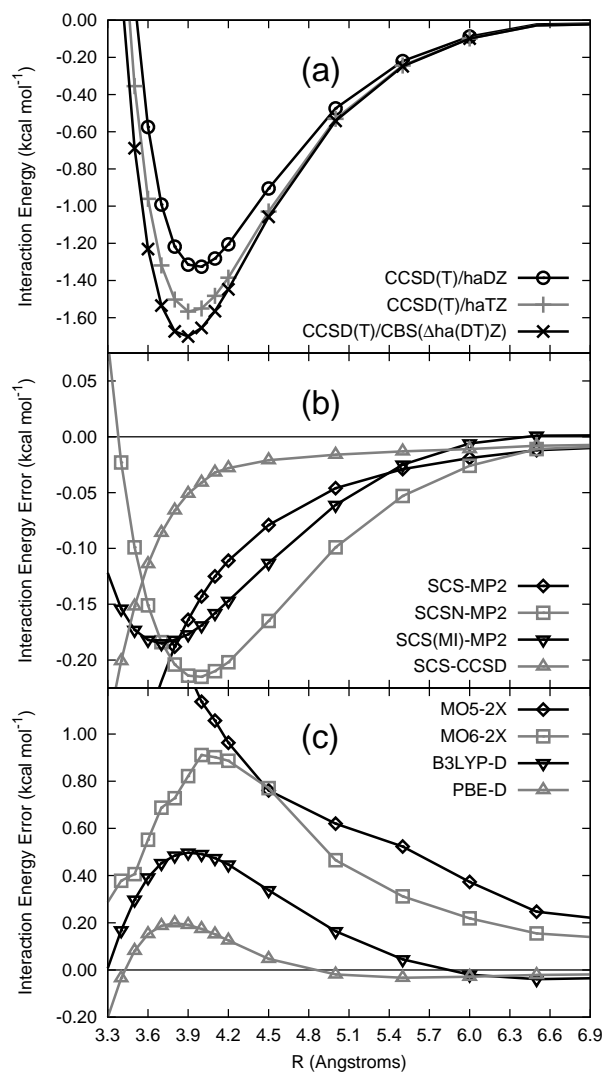
mol^{-1} across the PEC.
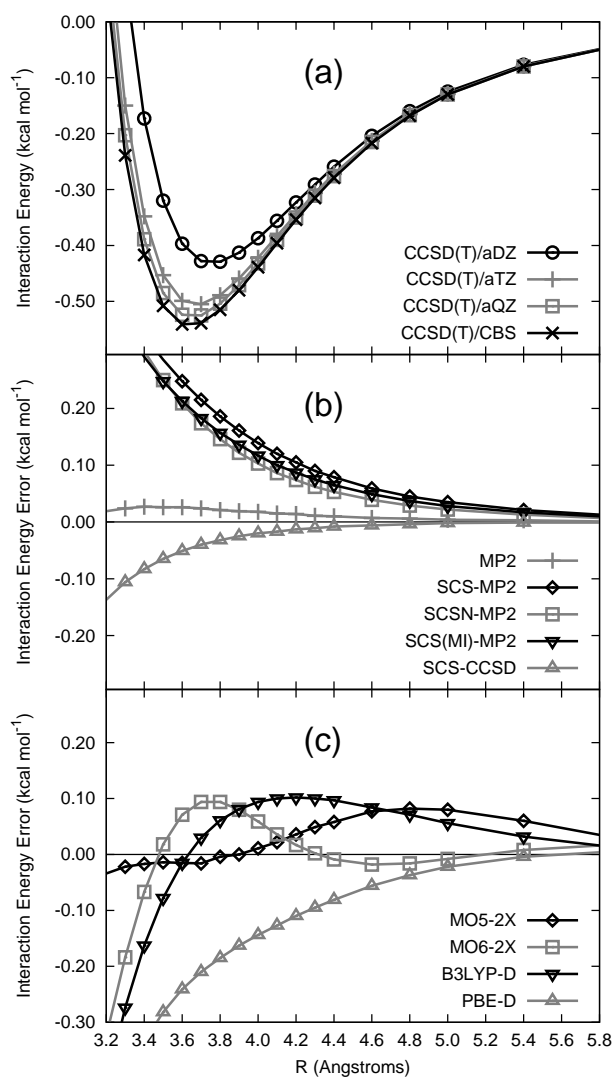
Figure 17: (a) CCSD(T) PECs for the methane dimer. Interaction energy errors for (b) SCS- and MP2-type approximations and (c) DFT methods compared to the CCSD(T)/CBS interaction energies.

Table 11 shows the interaction energies and optimum equilibrium distances for the various methods utilized for the PECs. For all systems, the SCS-CCSD method reproduces CCSD(T) interaction energies reliably with errors of less than a tenth of one kcal mol^{-1}, and it obtains the proper intermolecular separations. The scaled MP2 methods generally yield intermolecular distances correct within 0.1 Å, except

Table 11: Interaction Energies (kcal mol^{-1}) at Optimum Intermolecular Separations (Å, in Parentheses) for Prototype Nonbonded Complexes.

Method	S	T	PD[a]	CH$_4$-benzene	H$_2$S-benzene	CH$_4$-CH$_4$
CCSD(T)[b]	-1.70 (3.9)	-2.70 (5.0)	-2.71 (3.5,1.7)	-1.47 (3.8)	-2.83 (3.8)	-0.54 (3.6)
SCS-CCSD[c]	-1.75 (3.9)	-2.69 (5.0)	-2.77 (3.5,1.7)	-1.48 (3.8)	-2.74 (3.8)	-0.58 (3.6)
SCS-MP2[d]	-1.87 (3.9)	-2.47 (5.0)	-2.87 (3.5,1.6)	-1.21 (3.8)	-2.64 (3.8)	-0.32 (3.8)
SCSN-MP2[d]	-1.92 (3.9)	-2.92 (4.9)	-3.17 (3.5,1.7)	-1.43 (3.8)	-3.02 (3.8)	-0.36 (3.8)
SCS(MI)-MP2[e]	-1.88 (3.9)	-2.83 (5.0)	-3.10 (3.5,1.7)	-1.40 (3.8)	-2.96 (3.8)	-0.35 (3.8)
M05-2X[f]	-0.52 (4.0)	-2.18 (5.0)	-1.57 (3.5,2.0)	-1.26 (3.8)	-2.63 (3.8)	-0.54 (3.7)
M06-2X[f]	-0.95 (3.8)	-2.42 (4.9)	-2.54 (3.4,1.7)	-1.44 (3.7)	-2.99 (3.7)	-0.47 (3.5)
B3LYP-D[f]	-1.20 (3.9)	-3.03 (4.9)	-2.51 (3.4,1.8)	-1.61 (3.6)	-2.65 (3.7)	-0.56 (3.5)
PBE-D[f]	-1.51 (3.9)	-3.02 (4.9)	-2.63 (3.4,1.8)	-1.76 (3.6)	-2.99 (3.7)	-0.77 (3.5)

[a]Optimized distances given as (R1,R2) in Å. [b]Estimated CCSD(T)/CBS(Δha(DT)Z) results for benzene dimer. CCSD(T)/CBS results using aTZ/aQZ extrapolation for other cases. [c]Estimated SCS-CCSD/CBS(ΔaDZ) results for benzene dimer. SCS-CCSD/CBS results using aTZ/aQZ extrapolation for other cases. [d]aTZ/aQZ extrapolation to the CBS limit. [e]TZ/QZ extrapolation to the CBS limit. [f]Evaluated with the aug-cc-pVTZ basis.

for the methane dimer, which exhibits an error of 0.2 Å. Interestingly, the SCS(MI)-MP2 method seems to be the most reliable amongst the MP2-type methods for the interaction energies even though non-augmented basis sets were used for this method. This is likely because the SCS(MI)-MP2 method was fit to the S22 test set[55] (which contains 4 of the 6 complexes tested here) with non-augmented basis sets. It is therefore recommended that the SCS-CCSD method be used when computationally feasible. Otherwise, the scaled MP2 approaches appear similar for these test cases, but the performance of the SCS(MI)-MP2 method is perhaps somewhat better, and it is also more computationally efficient because it has been parameterized with the smaller non-augmented basis sets.

2.4.3.3 Evaluation of Density Functional Approximations

The lower panel of Figure 17 shows the performance of four DFT methods for the methane dimer. The methane dimer is the smallest of the six test cases considered and the most dominated by dispersion interactions. PBE and B3LYP cannot accurately

describe binding in this dimer without correction for dispersion interactions; however, the addition of empirical dispersion terms should provide reasonable results. Unfortunately, PBE-D does not perform well for the methane dimer, overbinding at all geometries considered, particularly around equilibrium. B3LYP-D improves over PBE-D near the equilibrium geometry; at larger separations, however, PBE-D performs somewhat better. On an absolute scale, the errors for both PBE-D and B3LYP-D are modest (0.3 kcal mol^{-1} or less except at very short intermolecular separations), but they are significant compared to a total CCSD(T)/CBS binding energy of only 0.54 kcal mol^{-1}.

M05-2X and M06-2X appear to implicitly model the interaction energy of the methane dimer without the addition of empirical terms. These functionals tend to overestimate the binding of the methane dimer at short intermonomer separations but they perform well at the equilibrium geometry. M05-2X significantly underbinds at larger intermonomer separations; at distances larger the 5 Å, the error between M05-2X and CCSD(T) begins to behave like R^{-6}. M06-2X incurs almost no error from 4.6 to 5.8 Å, although it exhibits some underbinding at larger distances.

The methane-benzene complex is the next largest of the six included in this study, this complex includes important electrostatic interactions as well as dispersion interactions. The behavior of the four DFT methods methane-benzene (see Figure 15(c)) is similar to that for the methane dimer; this is perhaps not surprising because both complexes contain only carbon and hydrogen atoms, and the same three types of pair potentials are included in the dispersion correction. PBE-D again overestimates the attractive interactions in the complex, incurring significant errors at small intermonomer separations. B3LYP-D performs well around the equilibrium geometry, but underestimates the strength of the attractive interactions within the complex at larger separations. As in methane dimer, M06-2X performs well for configurations near equilibrium but overestimates binding at short intermolecular separations.

Compared to M06-2X, the error curve for M05-2X is flatter for methane-benzene and methane dimer, but it is shifted up to somewhat larger errors for methane-benzene and underestimates binding at all separations.

Compared to the methane dimer and methane-benzene, the H_2S-benzene complex contains stronger electrostatic contributions.[58,92] Thus one might expect this case to be easier for DFT methods than the previous two considered. Figure 16(c) displays the errors in the interaction energies. Here, PBE-D performs well; for the equilibrium geometry and separations beyond equilibrium, PBE-D incurs errors of less than 0.10 kcal mol^{-1}. B3LYP-D underestimates the strength of the interaction and does not perform as well as PBE-D except at intermonomer distances shorter than the equilibrium distance. M05-2X performs similarly for this complex as it does for methane-benzene, consistently underestimating the attractive interactions. In this case, M05-2X incurs errors 0.2 kcal mol^{-1} or less across the PEC. M06-2X appears to capture a larger portion of the dispersion interactions than does M05-2X, as it overestimates the attractive interactions around equilibrium and at shorter separations. M06-2X also outperforms M05-2X at distances larger than 5 Å. Again, the error curve for M06-2X is somewhat more erratic than that for M05-2X with respect to intermolecular separation.

The benzene dimer has already been used as a challenging test case to assess the performance of new density functional approximations,[113,119–127] because its potential energy surface is not reproduced by standard functionals.[114] The performance of the DFT approaches considered here for three configurations of the benzene dimer can be seen in Figures 12, 13 and 14. As was seen for the three smaller complexes, PBE-D tends to predict more attractive interaction energies than does B3LYP-D. Of the three configurations, the DFT-D methods perform best for the parallel-displaced case. Here, PBE-D and B3LYP-D incur absolute errors of less than 0.20 kcal mol^{-1} across the PEC. M05-2X and M06-2X do not perform nearly as well for the benzene dimer

as they did for the three smaller complexes. M05-2X performs particularly poorly for the sandwich and parallel-displaced configurations, where it underestimates the strength of the interaction by over 1.00 kcal mol^{-1} at the equilibrium geometries. In the sandwich and T-shaped configurations, the error grows at larger intermolecular separations and then decreases as the magnitude of the interaction energy decreases. In the sandwich configuration, after a separation of 4.5 Å the error between M06-2X and CCSD(T) appears to be dominated by neglected dispersion interactions. For the T-shaped configuration, this occurs at about 5.5 Å. Thus the benzene dimer is a more challenging system for M05-2X and M06-2X than the three smaller complexes examined in this work. Somewhat better performance was reported for M05-2X for the parallel-displaced and T-shaped configurations in the work of Zhao and Truhlar, [127] who used the 6-311+G(2df,2p) basis set. That basis is somewhat smaller than aug-cc-pVTZ, leading to larger basis set superposition errors which partially cancel out the underbinding exhibited by in M05-2X (this is consistent with the significantly decreased binding observed by Zhao and Truhlar when they applied the counterpoise correction).

Thus the M05-2X and M06-2X functionals appear to implicitly capture some of the effects of "medium-range" dispersion interactions. For the smaller systems (methane dimer, methane-benzene and H$_2$S-benzene), M05-2X and M06-2X work well, performing as well or better than PBE-D and B3LYP-D. However, for the slightly larger benzene dimer test case, the performance of these methods is not as good; M05-2X significantly underbinds at shorter intermolecular distances (although this could be partially compensated for by using a smaller basis set), and both M05-2X and M06-2X fail to capture the attraction due to dispersion at large intermolecular distances.

2.4.4   Conclusion

This work evaluated several methods including spin-component-scaled MP2 and CCSD, density functional theory empirically corrected for dispersion interactions (DFT-D), and the meta-generalized-gradient approximation functionals M05-2X and M06-2X. Generally speaking, these approaches provided reasonable results for the prototype test cases (errors of a few tenths of one kcal mol^{-1} across the potential energy curves), although typically the spin-component-scaled wavefunction techniques performed better than the density functional approximations. M05-2X significantly underbound the benzene dimer when used with the fairly large aug-cc-pVTZ basis without counterpoise correction (this error could be partially cancelled by using smaller basis sets which exhibit larger basis set superposition errors). The performance of M06-2X and the DFT-D methods is roughly comparable, although DFT-D tends to be more reliable at large intermolecular distances. Intermolecular distances are well-predicted by the methods considered, although DFT-D tended to slightly underestimate these distances by about 0.1 Å. M05-2X and M06-2X are more sensitive to the integration grid than standard density functionals.[114,115]

Several general approximations for electronic structure theory were also tested in the context of nonbonded interactions, including RI/DF approximations, local correlation approximations for MP2, and the dual-basis approach of Head-Gordon and co-workers.[110,118]. The errors introduced by these approximations are negligible for the cases considered here, and they can lead to significant reductions in the computational cost — up to a factor of 7 when density fitting is used for both the Hartree-Fock and MP2 steps in an MP2/aug-cc-pVTZ computation for the sandwich benzene dimer, even when the canonical computation is allowed to use spatial symmetry (which the density fitting computations do not).

2.5 Accurately Characterizing the π-π Interaction Energies of Indole-Benzene Complexes

2.5.1 Introduction

This section is based on previously published work by Geng, Takatani, Hohenstein, and Sherrill.[4]

Among the most prevalent noncovalent interactions in biological complexes are π-π interactions, which arise from favorable electrostatic and London dispersion forces between the aromatic subunits.[128] These interactions have been shown to stabilize the vertical base-base stacking of DNA and RNA, and the structures of proteins.[32,34,129] The individual π-π interactions, however, are not always at their equilibrium separation due to effects such as steric constraints. Therefore, it is useful to obtain potential energy curves to accurately characterize how noncovalent interaction energies vary with the distance and orientation of the aromatic rings.

Phenylalanine, tyrosine, and tryptophan are the three neutral amino acids with constituent aromatic moieties; the side chains are commonly modeled with benzene, phenol, and indole, respectively. Interactions involving tryptophan stabilize the tertiary structure of biological macromolecules through both hydrogen bonding and π interactions, via the N-H group and the π-cloud of the indole, respectively.[130] Specifically regarding tryptophan-phenylalanine interactions, [NiFe] hydrogenase[131] is an enzyme that is directly involved in the metabolism of molecular hydrogen. This enzyme contains both the T-shaped (the plane of the benzene perpendicular to the plane of indole) and the parallel displaced configurations of the tryptophan-phenylalanine noncovalent complex as shown in Figure 18. Despite the importance of the indole-benzene model system, there have been few theoretical studies because the indole-benzene complex is somewhat large for very high-accuracy computations.

Similar methodologies as those used to examine the energetics for the benzene dimer[3,51] are applied here to study the indole-benzene complex. Because indole is
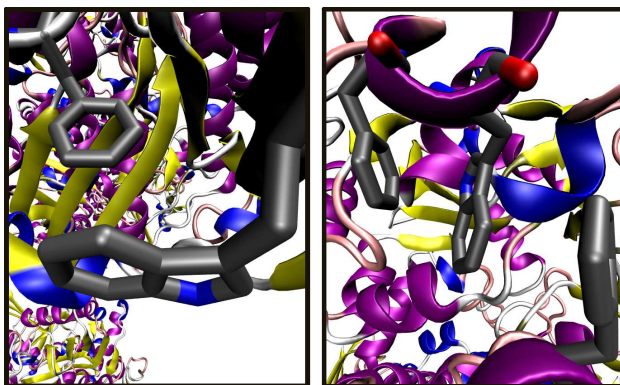
Figure 18: NiFe hydrogenase showing T-shaped (left) and parallel-displaced (right) type configurations of phenylalanine and tryptophan.

larger than the benzene ring, and because the lower symmetry means more configurations are possible, the indole-benzene complex is more challenging to study than simpler systems like the benzene dimer. Hence, there are only limited prior high-level studies available for this complex.[132] Accurate estimates of the CCSD(T)/CBS limit for binding energies of a large number of indole-benzene geometries are provided here. This data should be useful in benchmarking and parameterizing new theoretical approximations. Energy differences among different configurations are analyzed in light of energy component analysis using symmetry-adapted perturbation theory (SAPT)[95]. Detailed 3-D potential energy surfaces (displacements along the x, y, and z axes) are presented for the parallel-displaced configuration; to date, only very limited 3D potential data is available for any π-π interactions,[133] and this is the first such data available for a system as complex as indole-benzene. The detailed exploration of geometrical configurations, potential energy surfaces, and energy component analysis, as well as comparisons to the benzene dimer and the pyridine dimer, provides deeper insight into the nature of π-π interactions.

2.5.2 Specialized Theoretical Methodology

The MOLPRO package of *ab initio* programs[56] was used to compute all indole-benzene interaction energies. The benzene monomer geometry was set to those suggested by Gauss and Stanton, R(C-C)=1.3915 Å and R(C-H)=1.0800 Å.[57] The indole monomer geometry was optimized at the MP2/aug-cc-pVDZ level. Monomer geometries were kept frozen as the intermolecular distances were varied. Three "axial" T-shaped configurations are shown in Figure 19 and seven "equatorial" T-shaped configurations are shown in Figure 20 (where the plane of the indole ring is defined as the 'equator'). The displacements of the parallel configuration are depicted in Figure 21. To rationalize the interaction energy variance between the T-shaped
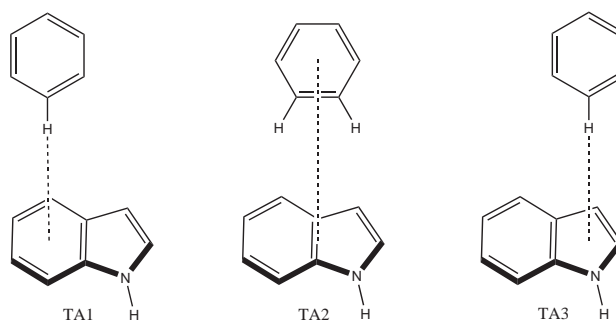
Figure 19: Three axial T-shaped configurations of the indole-benzene complex.

configurations, electrostatic potentials (ESPs) were mapped and natural population analyses were completed for indole and benzene at the B3LYP/6-31G* level of theory utilizing the Spartan[134] and Jaguar[135] programs, respectively. At selected minima, the configurations were extrapolated to the CBS limit utilizing aug-cc-pVDZ for the ΔCCSD(T) correction.

The interaction energy of selected T-shaped configurations of the indole-benzene complex was decomposed using symmetry-adapted perturbation theory (SAPT).[93,95] Due to the computational expense of SAPT, a truncation of SAPT which computes terms to second order with respect to intermonomer electron correlation and neglects intramonomer electron correlation, denoted SAPT0, was utilized. The performance of
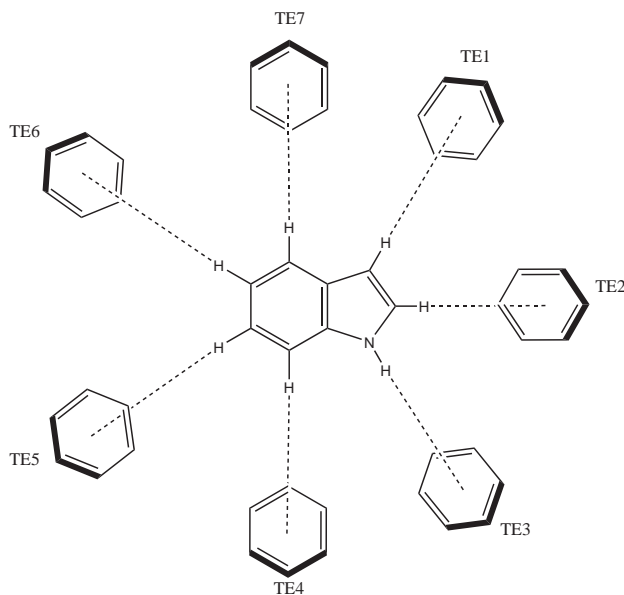
Figure 20: Seven equatorial T-shaped configurations of the indole-benzene complex.
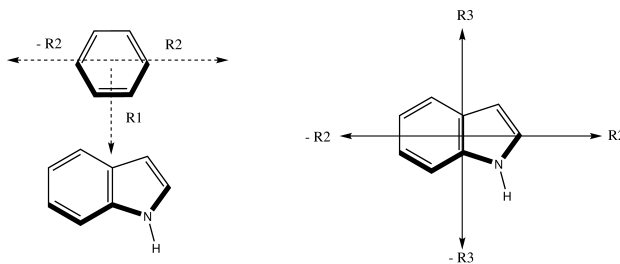
Figure 21: Displacements for the sandwich indole-benzene complex.

this method should be comparable to MP2. The individual SAPT terms were grouped into electrostatic, exchange, induction, and dispersion terms, where the individual components are defined as in Ref.[95].

$$E_{electrostatic} = E_{elst}^{(10)} \tag{59}$$

$$E_{exchange} = E_{exch}^{(10)} \tag{60}$$

$$E_{induction} = E_{ind,resp}^{(20)} + E_{exch-ind,resp}^{(20)} + \delta E_{ind,resp}^{(HF)} \tag{61}$$

$$E_{dispersion} = E_{disp}^{(20)} + E_{exch-disp}^{(20)} \tag{62}$$

The $\delta E_{ind,resp}^{(HF)}$ term, which recovers the Hartree-Fock interaction energy, is computed as follows.

66

$$\delta E_{ind,resp}^{(HF)} = E_{int}^{HF} - \left(E_{elst}^{(10)} + E_{exch}^{(10)} + E_{ind,resp}^{(20)} + E_{exch-ind,resp}^{(20)} \right), \tag{63}$$

This term describes the induced multipole-induced multipole interactions that are neglected in the SAPT based description of induction. For ease of comparison of energy components between similar dimer configurations, for all axial T-shaped configurations SAPT results were evaluated for the same distance between the center of benzene and the plane of indole (4.9 Å); likewise, for all equatorial T-shaped configurations, a fixed distance of 2.48 Å was used for the distance between the center of benzene to the closest hydrogen in indole. The SAPT0 computations were performed with a truncated aug-cc-pVDZ basis set denoted aug-cc-pVDZ$'$. This basis set removes all diffuse functions from hydrogen atoms and diffuse d functions from non-hydrogen atoms. Fortuitous error cancellation occurs for π-π interactions when this basis set is paired with MP2-like methods.[101] The two-electron integrals necessary for SAPT0 were computed with the density fitting (DF) approximation. A truncated version of the aug-cc-pVDZ-RI fitting basis set[136] was designed to match the aug-cc-pVDZ$'$ orbital basis. The diffuse functions were removed from the fitting basis for hydrogen atoms and diffuse f functions were removed for non-hydrogen atoms. All SAPT computations were performed with a locally modified version of PSI 3.4[100] using a new DF-SAPT implementation to be reported elsewhere.[137]

2.5.3 Results and Discussion

2.5.3.1 T-Shaped Configurations

Figure 22 shows the MP2 and SCS-MP2 interaction energies for the three axial T-shaped configurations at their respective equilibrium intermolecular separations. The left- and right-most configurations of Figure 22 (TA1 and TA3) feature a hydrogen on benzene pointing directly towards the center of the 6- and 5-member rings,
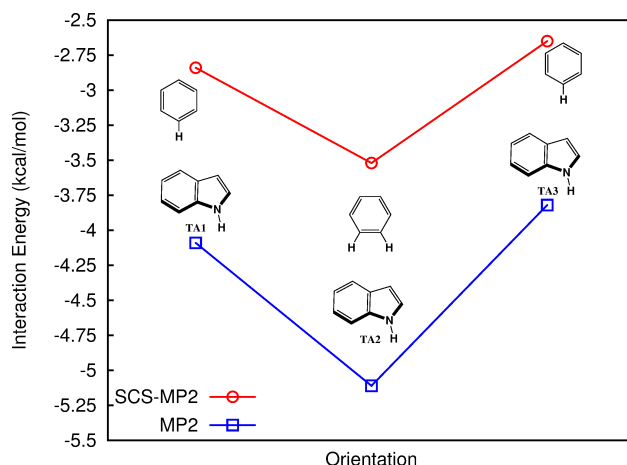
Figure 22: MP2 and SCS-MP2/aug-cc-pVDZ interaction energies (kcal mol^{-1}) for the three axial T-shaped configurations.

respectively, where the planes of indole and benzene are perpendicular. In the center configuration (TA2), however, two hydrogens point toward the indole (one toward each ring) with the center of benzene aligned with the mid-point of the shared bond of the two indole rings. Configuration TA2 is the most favored with MP2 and SCS-MP2 interaction energies of -5.11 and -3.52 kcal mol^{-1}, respectively, while configuration TA3 is the least favored with MP2 and SCS-MP2 interaction energies of -3.82 and -2.65 kcal mol^{-1}, respectively. Configuration TA1 is only slightly more bound than configuration TA3 with an MP2 interaction energy of -4.09 kcal mol^{-1} and an SCS-MP2 interaction energy of -2.84 kcal mol^{-1}. Consistent with previous studies, the MP2 interaction energies are larger in magnitude than the SCS-MP2 interaction energies for all three configurations; however the qualitative trend of both methods is the same. The SCS-MP2 results are expected to be more reliable[1,88] because of the known tendency of MP2 to overbind complexes involving π interactions.

The trend observed in Figure 22 can be explained by examining the electrostatic potentials (ESPs) of benzene and indole as shown in Figure 23 and the decomposition of the interaction energies with SAPT as shown in Table 12. Favorable interactions occur for all configurations due to the interaction between the positive hydrogens
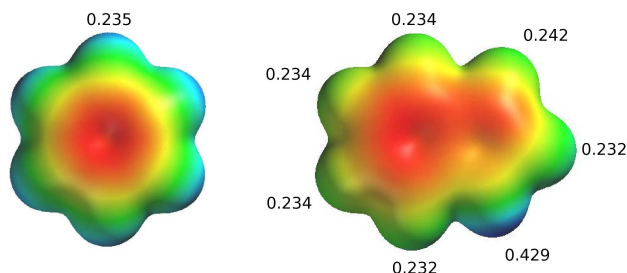
Figure 23: B3LYP/6-31G* electrostatic potential mapping for benzene (left) and indole (right), where blue is positive and red is negative. Numbers are the hydrogen atomic charges.

Table 12: Physical components (kcal mol^{-1}) of the total interaction energy determined using SAPT0 for the axial T-shaped configurations.[a,b]

	elst	exch	ind	disp	net dispersion[c]	Total
TA1	-1.94	3.13	-0.57	-4.32	-1.19	-3.70
TA2	-2.19	3.25	-0.60	-4.95	-1.71	-4.49
TA3	-2.32	4.27	-0.68	-4.83	-0.56	-3.55
TE1	-1.71	3.77	-0.55	-4.28	-0.50	-2.77
TE2	-2.31	3.48	-0.72	-3.74	-0.26	-3.29
TE3	-3.31	3.09	-1.01	-4.16	-1.07	-5.40
TE4	-2.18	4.05	-0.76	-4.63	-0.57	-3.46
TE5	-1.66	3.83	-0.53	-4.19	-0.36	-2.55
TE6	-1.71	3.74	-0.54	-4.15	-0.41	-2.66
TE7	-1.86	4.02	-0.58	-4.70	-0.68	-3.12
PD(MP2)	-3.24	10.49	-1.22	-11.00	-0.51	-4.96
PD(SCS)	-2.99	6.37	-0.91	-7.58	-1.21	-5.11

[a]Computed using the aug-cc-pVDZ' basis set. [b]For TA configurations, distances from the center of benzene to the plane of indole are 4.9 Å. For TE configurations, distances from the center of benzene to each respective hydrogen of indole are 2.48 Å. For PD configurations, minima are (R1=3.4, R2=0.1, R3=-1.0 Å, MP2/aug-cc-pVDZ) and (R1=3.4, R2=1.3, R3=-1.8 Å, SCS-MP2/aug-cc-pVDZ). [c]Net dispersion is the sum of the exchange and dispersion components.

of benzene and the negative π-cloud of indole. Furthermore, these interactions are enhanced via favorable dispersion interactions between the π-clouds of indole and benzene. Considering the ESP of indole, one might expect configuration TA1 to be more favorable than TA3 due to stronger electrostatic interactions; the ESP above the 6-membered ring appears more negative than above the 5-membered ring. However, the SAPT analysis of Table 12 indicates that TA3 has an electrostatic attraction about 0.4 kcal mol^{-1} stronger than that in TA1. On the other hand, net dispersion (the difference between dispersion and exchange)[138] is more favorable (by 0.6 kcal mol^{-1}) for TA1, in which the benzene is above the larger 6-membered ring. Overall,

69

the interaction energies of TA1 and TA3 are very similar, with TA1 being slightly favored in the MP2, SCS-MP2, or SAPT computations.

The center configuration, TA2, is more favorable than TA1 or TA3 because it has two C-H/π contacts and allows the benzene ring to interact with the π-clouds of both the 5- and 6-member rings of indole. A similar trend was observed for the methane-indole complex in a study by Ringer *et al.*[58] For that system, a decomposition of the interaction energy through SAPT suggested that the configuration in which two hydrogens of methane point toward the rings of indole is stabilized primarily by increased electrostatic attraction relative to the configurations with one interacting hydrogen. However, in the case of the indole-benzene complex, the electrostatic interaction of configuration TA2 strengthens compared to that in configuration TA1, but it does not become as strong as that in TA3. Instead, the increased net dispersion from the interaction of the benzene ring with both π-clouds is the primary component that stabilizes configuration TA2 over configurations TA1 and TA3; the net dispersion of configuration TA2 is more favorable than configurations TA1 and TA3 by 0.52 kcal mol^{-1} and 1.15 kcal mol^{-1}, respectively.

Figure 24 shows the MP2 and SCS-MP2 interaction energies for the seven equatorial T-shaped configurations at their respective equilibrium intermolecular separations. For each configuration, the hydrogen of the indole points directly at the center of the benzene ring, where the planes of indole and benzene are perpendicular. The MP2 and SCS-MP2 global minima occur at configuration TE3 with interaction energies of -6.09 and -4.65 kcal mol^{-1}, respectively. Configurations TE5 and TE6 yield similar interaction energies to each other and are the least favored positions, with energies of around -2.9 and -1.9 kcal mol^{-1} as computed by the MP2 and SCS-MP2 methods, respectively. Configuration TE1 is slightly more favorable than configurations TE5 and TE6 by about -0.2 kcal mol^{-1} for both methods, and configuration TE2 is more favored than TE1 by around -0.3 kcal mol^{-1} for both methods. The
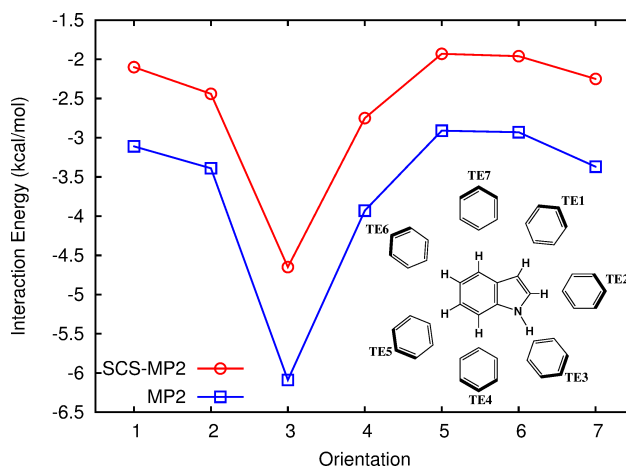
Figure 24: MP2 and SCS-MP2/aug-cc-pVDZ interaction energies (kcal mol^{-1}) for the seven equatorial T-shaped configurations.

MP2 method predicts that configuration TE7 is nearly isoenergetic to configuration TE2, whereas the SCS-MP2 method favors configuration TE2 over TE7 by -0.2 kcal mol^{-1}. Configuration TE4 is more favored than all other configurations except for TE3, with MP2 and SCS-MP2 interaction energies of -3.93 and -2.75 kcal mol^{-1}, respectively. Again, there exist large quantitative differences between the MP2 and SCS-MP2 interaction energies, but the trend is similar.

The equatorial T-shaped indole-benzene interaction energy trend can be partially rationalized by examining the ESPs and hydrogen atomic charges in Figure 23, and more completely by the components of SAPT in Table 12. The hydrogen connected to the nitrogen of the 5-member indole ring exhibits the largest positive charge by far. Therefore, the most favorable interaction is TE3, where the positive hydrogen interacts with the negative benzene π-cloud. The SAPT analysis confirms that this configuration has by far the most stabilizing electrostatic interaction. The partial charges of the other hydrogens in indole are all fairly similar, and therefore one might expect that configurations besides TE3 would have similar electrostatic contributions. Instead, a significant stabilization (0.3 - 0.7 kcal mol^{-1}) of configurations TE2 and TE4 relative to the remaining configurations is seen. The hydrogens pointed at

71

benzene in these configurations are those hydrogens which are on either side of the most positive hydrogen, and it is believed that the extra electrostatic stabilization of these configurations is due to longer-range attractions between the π-cloud of benzene and that most positive hydrogen. Note that the configurations with the largest electrostatic stabilization also have the largest induction contributions; apparently the partial charge on the N–H hydrogen is large enough to produce a significant polarization of the benzene. Also note that configurations TE1, TE3, TE4, and TE7 are those with the largest net dispersion interaction; in the other configurations, the benzene is much closer to one ring than the other, and so interactions with the second ring are greatly diminished.

2.5.3.2   Parallel Configurations

Figures 25 and 26 show the 3-D PECs for the parallel configurations at vertical displacements (R1) of 3.4, 3.6, and 3.8 Å computed by the MP2 and SCS-MP2 methods, respectively, with the aug-cc-pVDZ basis set.   The parallel configurations are stabilized by π-π dispersion interactions between the benzene ring and the two rings of indole, as well as favorable electrostatic effects due to charge interpenetration. Each PEC is constructed by holding the vertical distance R1 constant and varying the horizontal distances R2 and R3. MP2 and SCS-MP2 agree that the most favorable vertical separation of those considered is R1=3.4 Å. As shown by the top views of Figure 27, for this value of R1, the MP2 method locates a minimum at R2=0.1, R3=-1.0 Å. This minimum corresponds to the center of the benzene ring laying over the shared bond of indole and shifted towards the nitrogen. On the other hand, at this same vertical separation, the SCS-MP2 minimum places the center of the benzene ring almost directly over the nitrogen of indole with R2 and R3 values of 1.3 and -1.8 Å, respectively. The SAPT investigation on these two structures, shown in Table 12, suggests that the SCS-MP2 minimum is preferred over the MP2 minimum (according
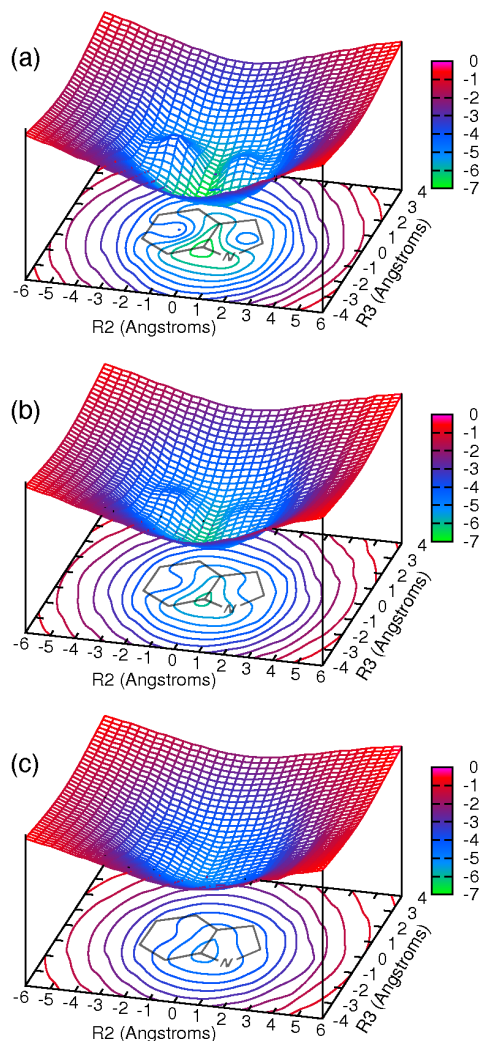
Figure 25: MP2/aug-cc-pVDZ 3-Dimensional potential energy surfaces (kcal mol^{-1}) for the parallel displaced configurations of the indole-benzene complex at (a) 3.4 Å, (b) 3.6 Å, and (c) 3.8 Å.

to SAPT) due to a dramatic decrease in the exchange contribution, which leads to a more favorable net dispersion interaction. The total SAPT interaction energies, utilizing the truncated basis set, indicates that the SCS-MP2 minimum is slightly more favored than the MP2 minimum. Further analysis by coupled-cluster methods are presented below to locate the true parallel configuration minimum.
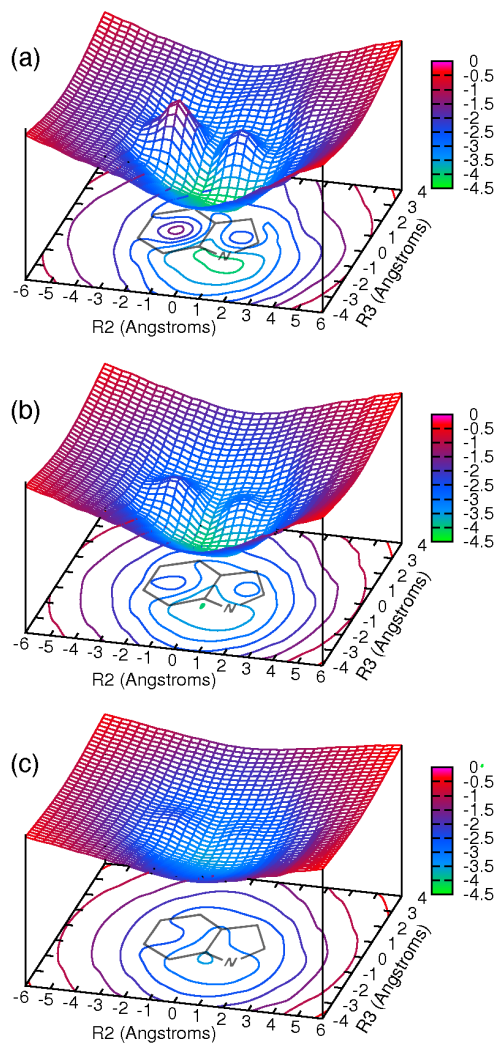
Figure 26: SCS-MP2/aug-cc-pVDZ 3-Dimensional potential energy surfaces (kcal mol^{-1}) for the parallel displaced configurations of the indole-benzene complex at (a) 3.4 Å, (b) 3.6 Å, and (c) 3.8 Å.
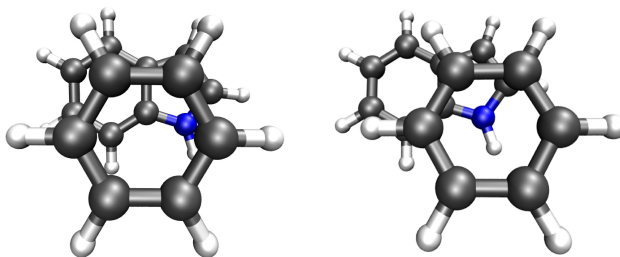
Figure 27: Representation of the minimum parallel displaced geometries predicted by the MP2 (left) and SCS-MP2 (right) methods with both vertical distances at 3.4 Å.

2.5.3.3 Coupled-Cluster Results

Due to the computational difficulty of obtaining CCSD(T) results for complexes of this size, only the two most favorable geometries from each of the axial and equatorial T-shaped configurations were further considered using the CCSD(T) method. For the parallel displaced (PD) configurations, CCSD(T) interaction energies of the minima predicted by both the MP2 and SCS-MP2 methods were computed. Results are shown in Table 13. Estimated CCSD(T)/CBS interaction energies indicate that the

Table 13: Interaction energies (kcal mol^{-1}) for selected minima with various levels of theory.[a,b]

	aug-cc-pVDZ		CBS[c]		Est'd CBS[d]	
	MP2	SCS-MP2	MP2	SCS-MP2	SCS-CCSD	CCSD(T)
TA1[e]	-3.83	-2.57	-4.31	-3.01	-3.27	-3.34
TA2[f]	-5.02	-3.52	-5.56	-4.02	-4.30	-4.39
TE3[g]	-6.08	-4.64	-6.80	-5.31	-5.47	-5.62
TE4[h]	-3.93	-2.71	-4.41	-3.15	-3.39	-3.46
PD MP2[i]	-6.81	-4.04	-7.59	-4.71	-4.35	-4.46
PD SCS[j]	-6.20	-4.22	-6.68	-4.67	-4.53	-4.64

[a]Density fitting approximations applied to all methods except for coupled-cluster methods. [b]All T-shaped configurations computed at the SCS-MP2/CBS minima. [c]Two-point extrapolation of the aug-cc-pVTZ and aug-cc-pVQZ correlation energies to the aug-cc-pVQZ reference energy. [d]Complete basis-set estimated via an addition of the difference between MP2 and coupled-cluster correlation energies with the aug-cc-pVDZ basis set to the DF-MP2/CBS energies. [e]Left axial T-shaped configuration (R=5.0 Å). [f]Center axial T-shaped configuration (R=4.8 Å). [g]Equatorial T-shaped orientation 3 (R=4.5 Å); global minimum. [h]Equatorial T-shaped orientation 4 (R=5.0 Å). [i]Parallel displaced minimum (R1=3.4 Å, R2=0.1 Å, R3=-1.0 Å) as predicted at the MP2/aug-cc-pVDZ level. [j]Parallel displaced minimum (R1=3.4 Å, R2=1.3 Å, R3=-1.8 Å) as predicted at the SCS-MP2/aug-cc-pVDZ level.

global minimum, among the configurations considered, occurs at configuration TE3, favored by one kcal mol^{-1} or more compared to any other configuration reported here. This confirms the preference for configuration TE3 reported in an earlier study[132] of a few of the geometries considered here, using similar methodologies but smaller basis sets. CCSD(T) computations agree with SCS-MP2/aug-cc-pVDZ results for the PD minimum; the SCS-MP2 PD configuration is more favorable than the MP2 PD configuration by almost -0.2 kcal mol^{-1}. Configuration TA2 lies only around 0.2 kcal mol^{-1} higher than the PD minima.

Of the methods considered here, those that predict the correct trend in interaction energies are the estimated SCS-CCSD/CBS and SCS-MP2/aug-cc-pVDZ levels of theory. The SCS-MP2/CBS interaction energies of the indole-benzene complex are normally within 10% of CCSD(T) results; however, the CBS-extrapolated SCS-MP2 results predict the PD MP2 configuration to be more favorable than the PD SCS configuration (although by only -0.04 kcal mol^{-1}). The PD configurations are also overstabilized compared to the T-shaped configurations at the SCS-MP2/CBS level of theory; the CCSD(T) method favors configuration TE3 over the PD structures by about one kcal mol^{-1}, whereas the extrapolated SCS-MP2 method only predicts a difference of about 0.6 kcal mol^{-1} between the TE3 and PD configurations. The SCS-CCSD method, on the other hand, only incurs around 2% error compared to CCSD(T) results for all configurations considered here; interaction energy differences are also well represented. The MP2 method greatly overstabilizes the PD configurations, which leads to a failure in obtaining the correct global minimum.

2.5.4 Conclusions

Although the MP2 and SCS-MP2 methods with the aug-cc-pVDZ basis set significantly differ for interaction energies of the indole-benzene complex (on average, by about 1-1.5 kcal mol^{-1} at selected minima), the trends for the axial and equatorial T-shaped configurations are similar. SAPT0 total interaction energies with the truncated basis set (aug-cc-pVDZ$'$) are similar in magnitude to the CCSD(T) results. The TA2 configuration is predicted as the most favorable axial T-shaped configuration, which is analogous to the methane-indole complex studied previously.[58] However the underlying increase in the dispersion interaction is the primary stabilizing force for this configuration relative to the other axial T-shaped configurations, not the electrostatic contribution as it was in the indole-methane case. Furthermore, both methods predict the N–H/π interaction as the minimum T-shaped configuration due to the

highly favorable electrostatic interaction with the strongly polarized N–H bond. This is in agreement with previous work on the indole-benzene complex[132] and with this study's high-quality CCSD(T)/CBS estimates.

For the parallel configurations, however, the MP2 and SCS-MP2 methods predict different minima. At the vertical distance of 3.4 Å, the MP2 minimum situates the benzene over the shared bond of the indole rings (slightly displaced towards the nitrogen side), whereas the benzene in the SCS-MP2 minimum is directly over the N–H bond. SAPT results indicate that the SCS-MP2 minimum is preferred due to a more favorable net dispersion interaction. As the vertical distance is increased, the MP2 and SCS-MP2 minima appear more similar, where the benzene is located over the shared bond of the indole rings.

Highly accurate CCSD(T)/CBS estimates reinforce the conclusion that the global minimum is the TE3 configuration, with an interaction energy of -5.62 kcal mol^{-1}. CCSD(T) results also support optimizations by the SCS-MP2 method rather than the MP2 method for the parallel configuration; the PD SCS structure is more favorable than the PD MP2 structure by 0.18 kcal mol^{-1}. Moreover, the MP2 method fails to predict the correct global minimum, favoring the parallel configuration over the N-H/π T-shaped configuration. Among the approximate methods considered here, the SCS-CCSD method appears to be the most reliable with errors of only around 2%, whereas the SCS-MP2 CBS extrapolated results are normally within 10% of the CCSD(T)/CBS estimates. The CCSD(T)/CBS benchmark results reported here should be useful in further benchmarking of new methods for nonbonded interactions.

Several conclusions concerning molecular recognition can be drawn from these results in comparison with the previous studies on the benzene dimer and the pyridine-benzene complex.[101,138] Aromatic N-H/π interactions are much stronger than aromatic C-H/π interactions. Using the same CCSD(T)/CBS estimation procedure as that used here [the ΔCCSD(T) correction being computed in the aug-cc-pVDZ basis,

abbreviated as CCSD(T)/CBS(ΔaDZ)], the T-shaped benzene dimer has an interaction energy of -2.73 kcal mol^{-1} and the most favorable T-shaped pyridine-benzene interaction is -3.18 kcal mol^{-1}, whereas the indole-benzene N-H/π interaction energy is -5.62 kcal mol^{-1}. The extension of aromaticity (i.e., the addition of a second aromatic ring in indole compared to benzene) can enhance aromatic C–H/π interactions. Configuration TA1 exhibits an interaction energy 0.6 kcal mol^{-1} more favorable than the T-shaped benzene dimer. The extension of aromaticity and the inclusion of the nitrogen heteroatom also increase π-π interactions. The PD SCS configuration of the indole-benzene complex is around 1.8 kcal mol^{-1} more favored than the PD benzene dimer. Unlike the pyridine-benzene complex, the benzene of the indole-benzene complex prefers to be situated around the nitrogen, as exemplified by the preference of the PD SCS and the TE3 configurations over all others considered. This is due to the interaction of the highly positive hydrogen of the N-H bond in indole with the negative π-cloud of benzene. In pyridine, there is no N-H bond and the potential around the nitrogen is highly negative. Lastly, like the PD and T-shaped benzene dimer, the interaction energies of the PD MP2 and TA2 configurations are roughly equivalent, and therefore a more 'V-shaped' configuration is possible.[139] This structure for the indole-benzene complex is provided in the S22 test set where the plane of benzene is slightly tilted from the plane of indole.[55] However, the N–H/π T-shaped indole-benzene configuration is still reported to be more stable.

2.6 Basis Set Consistent Revision of the S22 Test Set of Noncovalent Interaction Energies

This section is based on previously published work by Takatani et al.[5]

Excellent work by Hobza and co-workers has provided several key databases of CCSD(T) noncovalent interaction energies.[55,97,140,141] Among the most popular is the S22 test set, which contains 22 dimers of various types (H-bonded, dispersion dominated and mixed) and system sizes (from water dimer to adenine-thymine

complexes).[55] It has become increasingly common to benchmark existing and new methods against the S22 test set;[72,142–149] some methods are even being parameterized or adjusted based on reducing the error compared to the S22 interaction energies.[73,150,151]

Despite the high accuracy of the initial S22 CCSD(T) interaction energies, the basis sets used vary between complexes.[55] As has become customary in such studies, estimates of the CCSD(T) CBS limit were obtained using the ΔCCSD(T) approximation, where the ΔCCSD(T) term was evaluated in a small or medium-sized basis. No diffuse functions were used for the CCSD(T) computations in the S22 set except for the well documented T-shaped and parallel displaced benzene dimers.[51,101] The MP2 interaction energies were extrapolated with the cc-pVQZ and cc-pV5Z basis sets, or the cc-pVTZ and cc-pVQZ basis sets, depending on the size of the complex (diffuse functions were included only in a few cases). The inconsistent use of basis sets (driven by necessity in the 2006 computations) has been noted previously by Marchetti and Werner while testing explicitly correlated wave function methods for the computation of intermolecular interactions.[144] Furthermore, for the larger complexes such as conformations of adenine-thymine and indole-benzene, only double-ζ quality basis sets, without diffuse functions, were used for the CCSD(T) correction. Recently, potential energy curves for 20 complexes of the S22 test set were computed at the CCSD(T)/CBS limit, in which the ΔCCSD(T) terms were consistently evaluated with the aug-cc-pVDZ basis set.[152]

Unfortunately, as pointed out by Janowski and Pulay,[109] double-ζ quality basis sets are not always sufficient to converge the coupled-cluster correction. Marchetti and Werner have worked to increase the basis sets for CCSD(T) computations of the S22 complexes.[145] However, they neglected the diffuse functions on the hydrogens and only computed the interaction energies for the smaller complexes of the S22 test set. The CCSD(T)/aug-cc-pVTZ interaction energies have been obtained for the stacked

methyl adenine-methyl thymine complex[153] and two uracil dimer conformations.[154] Both studies suggested that the interaction energies for the larger complexes of the S22 test set could change as much as a few to several tenths of one kcal mol^{-1} when evaluating the ΔCCSD(T) term with the larger aug-cc-pVTZ basis set. CCSD(T) computations on the benzene dimer using full or truncated aug-cc-pVQZ basis sets also demonstrate that the ΔCCSD(T) term can change by around one tenth of one kcal mol^{-1} from the aug-cc-pVDZ values.[3,48,109]

This study consistently extrapolates all interaction energies of the S22 test set to the CCSD(T)/CBS limit. The MOLPRO package of *ab initio* programs[56] was used to compute most interaction energies. The largest coupled-cluster computations used NWChem 5.1[155,156] on the Cray XT4 "Jaguar" at Oak Ridge National Laboratory. Computations employed the aug-cc-pVXZ (X=D,T,Q) basis sets of Dunning *et al.*[66,67] and were counterpoise corrected.[70] MP2/CBS energies were computed by adding a two-point extrapolation[71] of the aug-cc-pVTZ and aug-cc-pVQZ correlation energies to the aug-cc-pVQZ reference self-consistent field energy, denoted as (aT-aQ). To these values, a coupled-cluster correction, ΔCCSD(T), evaluated in a smaller basis was added (see below). For smaller systems, a direct extrapolation of CCSD(T) correlation energies was used to obtain CCSD(T)/CBS(aT-aQ) interaction energies. The largest changes to the SCF component of the interaction energies when increasing from the aug-cc-pVTZ to the aug-cc-pVQZ basis sets were observed for the formic acid and formamide dimers, both by 0.09 kcal mol^{-1}. When increasing further to the aug-cc-pV5Z basis set, the SCF interaction energies for both of these dimers only changed by 0.01 kcal mol^{-1}. Similar results were previously obtained utilizing truncated versions of the basis sets by Marchetti and Werner, where the SCF/aug-cc-pVQZ′ errors were no greater than 0.02 kcal mol^{-1} compared to the SCF/aug-cc-pV5Z′ results for the entire S22 test set.[144] Thus, the SCF component of the interaction energies seems essentially converged with the aug-cc-pVQZ basis

set.

Regarding the MP2/CBS estimates, it is not obvious which values are more reliable, the present (aT-aQ) extrapolations, or the (aQ'-a5') extrapolations of Marchetti and Werner.[144] The latter study went up to quintuple-ζ basis sets (compared to quadruple-ζ in the present work), but it neglected the diffuse functions on hydrogen atoms, which is included here. However, there is a high level of agreement between this study's MP2/CBS estimates and those of Marchetti and Werner (the largest difference is 0.03 kcal mol^{-1}).

The coupled-cluster correction, ΔCCSD(T), was evaluated with the aug-cc-pVDZ and aug-cc-pVTZ basis sets; these results are referred to as the ΔaDZ and ΔaTZ corrections, respectively. A more reliable ΔCCSD(T) correction was estimated as the difference between the two-point extrapolated CCSD(T) and MP2 correlation energies using the aug-cc-pVDZ and aug-cc-pVTZ basis sets; this is referred to as the Δa(DT)Z correction. All ΔCCSD(T) terms were added to the MP2/CBS(aT-aQ) energies. By comparison to larger-basis CCSD(T) computations, the quality of the Δa(DT)Z correction scheme has been shown to be highly accurate for the benzene dimer, methane-benzene, and H$_2$S-benzene complexes.[3] The MP2, SCS-MP2,[54] SCS-CCSD,[2] SCS(MI)-MP2,[73] and B2PLYP-D[151] methods have been re-evaluated against the new CCSD(T)/CBS(Δa(DT)Z) results.

The original S22 geometries were adopted. Interaction energies and geometries would change slightly upon further optimization (*e.g.*, interaction energies would change by -0.01 and -0.08 kcal mol^{-1} for the T-shaped and parallel-displaced benzene dimers, respectively).[3] Some of the dimers which were originally optimized without diffuse functions might change somewhat more than this. However, as the primary value of the S22 test set has been in validation of theoretical binding energies rather than in comparisons to experiment, any reasonable set of geometries will suffice, and using the original geometries makes it easier to compare prior work to this study's

revised benchmark data.

Table 14 shows all CCSD(T)/CBS interaction energy approximation schemes, along with fully extrapolated CCSD(T)/CBS(aT-aQ) interaction energies for the smallest complexes and the original S22 data. Comparing the fully extrapolated

Table 14: CCSD(T) Basis Set Consistent Interaction Energies in kcal mol^{-1} for the S22 Test Set

Complex	Orig.[a]	CBS(ΔaDZ)[b]	CBS(ΔaTZ)[c]	CBS(Δa(DT)Z)[d]	CBS[e]
H-Bonded Complexes					
$(NH_3)_2$	-3.17	-3.10	-3.15	-3.17	-3.15
$(H_2O)_2$	-5.02	-4.92	-4.99	-5.02	-5.07
Formic Acid Dimer	*-18.61*	-18.46	-18.70	-18.80	-18.81
Formamide Dimer	*-15.96*	-15.84	-16.03	-16.12	-16.11
Uracil Dimer	-20.65	-20.42	-20.61	-20.69	—
2-Pyridoxine·2-Aminopyridine	*-16.71*	-16.70	-16.91	-17.00	—
Adenine·Thymine WC	*-16.37*	-16.43	-16.65	-16.74	—
Dispersion Dominated Complexes					
$(CH_4)_2$	-0.53	-0.53	-0.53	-0.53	-0.53
$(C_2H_4)_2$	-1.51	-1.48	-1.49	-1.50	-1.48
Benzene·CH$_4$	-1.50	-1.47	-1.45	-1.45	—
PD Benzene Dimer	*-2.73*	-2.73	-2.65	-2.62	—
Pyrazine Dimer	*-4.42*	-4.33	-4.24	-4.20	—
Uracil Dimer	*-10.12*	-9.83	-9.77	-9.74	—
Stacked Indole·Benzene	*-5.22*	-4.62	-4.60	-4.59	—
Stacked Adenine·Thymine	*-12.23*	-11.82	-11.71	-11.66	—
Mixed Complexes					
Ethene·Ethine	-1.53	-1.50	-1.51	-1.51	-1.50
Benzene·H$_2$O	-3.28	-3.27	-3.28	-3.29	—
Benzene·NH$_3$	-2.35	-2.33	-2.32	-2.32	—
Benzene·HCN	-4.46	-4.52	-4.54	-4.55	—
T-shaped Benzene Dimer	-2.74	-2.75	-2.72	-2.71	—
T-shaped Indole Benzene	*-5.73*	-5.63	-5.63	-5.62	—
Phenol Dimer	-7.05	-7.04	-7.08	-7.09	—

[a]Taken from Ref[55]. Italicized energies denote the significant changes (> 0.1 kcal mol^{-1}) compared to the CBS(Δa(DT)Z) results. [b]Estimated via an addition of the difference between MP2 and CCSD(T) correlation energies with the aug-cc-pVDZ basis set to the MP2/CBS(aT-aQ) energies. [c]Estimated via an addition of the difference between MP2 and CCSD(T) correlation energies with the aug-cc-pVTZ basis set to the MP2/CBS(aT-aQ) energies. [d]Estimated via an addition of the difference between MP2/CBS(aD-aT) and CCSD(T)/CBS(aD-aT) correlation energies to the MP2/CBS(aT-aQ) energies. [e]Two-point extrapolation of the CCSD(T) correlation energies with the aug-cc-pVTZ and aug-cc-pVQZ.

CCSD(T)/CBS(aT-aQ) interaction energies to the approximate extrapolations, all schemes give similar errors for the smallest complexes (ammonia, water, methane, ethene, and ethene·ethine dimers), with the ΔaDZ scheme being slightly less accurate for the hydrogen bonded cases. The Δa(DT)Z scheme, however, yields superior results for the interaction energies of the formic acid and formamide dimers;

at the ΔaTZ level, the interaction energy errors are around one tenth of one kcal mol^{-1}, whereas at the Δa(DT)Z level the errors are within one hundredth of one kcal mol^{-1}. Note that the interaction energies were nearly converged at the ΔaDZ level for all the mixed complexes, with differences of less than 0.04 kcal mol^{-1} when comparing to the ΔaTZ interaction energies. It is, therefore, reasonable to assume that the Δa(DT)Z interaction energies for the mixed complexes must also result in, at most, a few hundredths of one kcal mol^{-1} error compared to the fully extrapolated CCSD(T)/CBS(aT-aQ) interaction energies. Furthermore, compared with the CCSD(T)/CBS(aT$'$-aQ$'$) interaction energies (with truncated basis sets) for several of the test cases published by Marchetti and Werner,[145] and the benzene·CH$_4$ dimer and three orientations of the benzene dimer in previous work,[3] the Δa(DT)Z correction also incurs only a few hundredths of one kcal mol^{-1} error. Having established the effectiveness of the counterpoise corrected CCSD(T)/CBS(Δa(DT)Z) approximation scheme, these values will be used as the benchmarks in the subsequent discussion.

In general, the severity of the ΔaDZ approximation scheme errors increases as the system sizes increase for the hydrogen bonded and dispersion dominated complexes, whereas the ΔaTZ errors remain relatively constant throughout. For the hydrogen bonded complexes, the ΔaDZ approximation underestimates the interaction energy anywhere from 0.07-0.34 kcal mol^{-1}, with the worst cases being for the formic acid dimer and adenine-thymine complex. For the dispersion dominated complexes, the ΔaDZ approximation overestimates the interaction energy by as much as 0.16 kcal mol^{-1}, with the worst case being for adenine-thymine complex. The ΔaTZ absolute errors are less than 0.1 and 0.05 kcal mol^{-1} for the hydrogen bonded and dispersion dominated complexes, respectively. For the mixed complexes, both approximation schemes perform rather well with absolute errors of less than 0.05 and 0.01 kcal mol^{-1} for the ΔaDZ and ΔaTZ schemes, respectively.

The ΔaDZ scheme yields larger absolute errors with increasing system size, and

the original S22 data is also less accurate for the interaction energies of the larger complexes. This is exemplified by the formic acid dimer and adenine-thymine (Watson-Crick geometry) interaction energy errors; the original S22 data outperforms the ΔaDZ approximation for the formic acid dimer, but is less accurate for adenine-thymine. For the original S22 data, ΔTZ and ΔDZ CCSD(T) corrections (without diffuse functions) were computed for the formic acid dimer and adenine-thymine, respectively. The use of different basis sets leads to the inconsistent reliability of the original S22 data, and the neglect of diffuse functions yields absolute errors greater than the ΔaDZ approximation errors for the larger complexes. Overall, the result of consistently utilizing the Δa(DT)Z CCSD(T) correction changes 10 of the 22 interaction energies by more than 0.1 kcal mol^{-1} from the original S22 data as indicated in Table 14.

Table 15 presents the mean absolute deviation (MAD), mean deviation (MD), and root mean squared deviation (RMS) for the MP2, SCS-MP2, SCS-CCSD, SCS(MI)-MP2, and B2PLYP-D methods, as well as the original S22 data, compared to the S22A benchmark energies. Overall, the MP2 and SCS-MP2 methods perform similarly, with MADs of 0.88 and 0.80 kcal mol^{-1}, respectively. The MP2 method outperforms the SCS-MP2 method for hydrogen bonded complexes, whereas the SCS-MP2 method outperforms the MP2 method for both dispersion and mixed complexes. The overall performance of the SCS-CCSD and SCS(MI)-MP2 methods is somewhat similar, despite the parameterization of the SCS(MI)-MP2 method for S22 test set. The SCS(MI)-MP2 method incurs similar errors for the hydrogen-bonded and dispersion dominated complexes with a slightly better performance for the mixed complexes, leading to the overall 0.28 kcal mol^{-1} MAD. The SCS-CCSD method outperforms all other methods tested here for the mixed complexes, but underestimates most interaction energies to yield an overall MD of -0.23 and a MAD of 0.24 kcal mol^{-1}. While this performance is quite good, a re-parameterization of the SCS-CCSD method for

Table 15: Interaction Energy Error Statistics in kcal mol^{-1} for the S22A[a] Test Set with Various Methods

Test Set	Orig.[b]	MP2[c]	SCS-MP2[c]	SCS-CCSD[d]	SCS(MI)-MP2[e]	B2PLYP-D[f]
H-Bonded Complexes						
MAD[g]	0.15	0.24	1.54	0.40	0.30	0.14
MD[h]	-0.15	-0.14	-1.54	-0.40	-0.20	0.08
RMS[i]	0.20	0.27	1.69	0.47	0.36	0.17
Dispersion Dominated Complexes						
MAD[g]	0.25	1.69	0.55	0.23	0.37	0.12
MD[h]	0.25	1.68	-0.32	-0.22	0.19	0.02
RMS[i]	0.34	2.15	0.69	0.31	0.43	0.17
Mixed Complexes						
MAD[g]	0.05	0.61	0.37	0.08	0.17	0.11
MD[h]	0.01	0.61	-0.37	-0.08	0.06	0.07
RMS[i]	0.06	0.72	0.41	0.09	0.20	0.14
Overall						
MAD[g]	0.15	0.88	0.80	0.24	0.28	0.12
MD[h]	0.04	0.76	-0.72	-0.23	0.02	0.06
RMS[i]	0.45	1.05	0.96	0.54	0.58	0.40

[a]Using CCSD(T)/CBS(Δa(DT)Z) results. [b]Original S22 test set from Ref[55]. [c]CBS(aT-aQ) interaction energies. [d]CBS(Δa(DT)Z) interaction energies. [e]Energies evaluated with the two-point extrapolation of the correlation energies with the cc-pVTZ and cc-pVQZ basis sets taken from Ref[73]. [f]Energies counter-poise corrected evaluated with the aug-cc-pVTZ basis taken from Ref[151]. [g]Mean absolute deviation. [h]Mean deviation. [i]Root mean squared deviation.

the computation of noncovalent interactions would likely increase its accuracy and more easily justify the expense of performing CCSD computations. The most accurate method tested is the B2PLYP-D method with an overall MAD of 0.12 kcal mol^{-1}; this method has been parameterized specifically for the original S22 interaction energies. Interestingly, increasing the basis set results in much greater errors for the dispersion dominated complexes than for the H-bonded and mixed complexes when comparing the original S22 data to the revised data. This is perhaps not surprising in that polarizability (to which dispersion is related) is known to be difficult to converge with respect to basis set.

Given the very small errors exhibited by SCS-CCSD and B2PLYP-D, the usefulness of the new benchmark energies becomes apparent. The original S22 data incurs an overall MAD of 0.15 kcal mol^{-1} compared to the CCSD(T)/CBS(Δa(DT)Z) interaction energies; the overall MAD of the B2PLYP-D method is less than that of the original S22 data. Furthermore, if the original S22 data are used as the benchmarks,

the MADs for the hydrogen bonded and dispersion dominated complexes of the SCS-CCSD method would be 0.25 and 0.54 kcal mol^{-1}, respectively, as opposed to 0.40 and 0.23 kcal mol^{-1}, respectively. Thus the original S22 data incorrectly suggests that the SCS-CCSD method approximates hydrogen bonded complexes more accurately than dispersion dominated complexes.

A small test set of noncovalent interactions, such as the S22 test set, is indeed useful for benchmarking and parameterization purposes. However, approximate methods have now reached an accuracy where higher-quality benchmark data (using consistent basis sets across the set of dimers) is necessary to allow for further improvements. Overall, the B2PLYP-D method outperforms all other approximate methods tested here, while the SCS-CCSD method outperforms all other methods for mixed complexes. The revised benchmark energies provided here, along with recent potential energy curves for small dimers[3] are recommended for the future assessment and parameterization of methods for noncovalent interactions.

CHAPTER III

METAL-SALEN CATALYSTS

The salen [bis(salicylaldehydo)ethylenediamine] and salen-type ligands with various metal centers (shown in Figure 28) comprise a highly active class of synthetic complexes.[157–160] Depending on the metal center (normally 3d- and 4d-metals), these
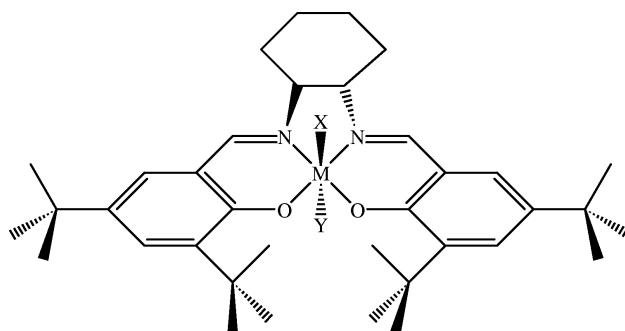
Figure 28: The metal-salen complex, where M is a metal-center, and X and Y are various axial ligands.

complexes have been utilized for numerous asymmetric catalysis applications.[161–163] Research on the immobilization of the metal-salen complexes have shown that different schemes can greatly effect the catalytic activity.[164] The molecular supports, furthermore, may have a significant influence on the electronic structure of the metal-salen complexes.[165] Understanding the underlying catalytic reaction mechanisms via electronic structure computations could immensely aid past and ongoing efforts to tune the reactivity and selectivity of metal-salen catalysts.

Previous theoretical studies on metal-salen systems are largely limited to the Mn-salen complex[166–182] while very few have explored different metal centers with multireference methods.[183–185] Density functional theory (DFT) methods[24,25] have seen much success in its applications to organic chemistry and has become attractive for coordination chemistry as well.[30,186–190] However, currently popular density functional

87

approximations can be less reliable for systems containing transition metals when several low-lying electronic states are present.[181,191–195] Furthermore, there exist benchmark studies of DFT computations on transition-metal ligand complexes,[196–198] but few reach the size and complexity of the metal-salen systems. Of significant importance, the application of two common DFT methods on the olefin epoxidation mechanism with the Mn-salen catalyst yields large qualitative discrepancies.[170]

Transition metal complexes can feature strong dynamical and nondynamical electron correlation effects, which makes them difficult to describe theoretically. One possible solution is to include higher-order excitations in the treatment of electron correlation. Unfortunately, such approaches become computationally intractable for larger molecules. Another approach is to treat the system with multireference methods.[20,199,200] The difficulty of utilizing this approach is complicated by the selection of appropriate active spaces and the sufficient treatment of dynamical electron correlation. Work in on the d^0 and d^2 metal-salens[184,185] has outlined a procedure for the proper treatment of metal-salen complexes with minimal active space multireference methods and has shown that DFT methods, in general, fail to properly describe the singlet, triplet, and quintet energy gaps for a series metal-salen complexes.

As in the previous chapter, the following section contains a generalized theoretical methodology for the computation of metal-salen energetics. Specialized theoretical methodology sub-sections specify the unique methods utilized within each study. Sections 3.2 and 3.3 extend the systematic study on the transition metal-salen systems to the d^6 and d^4 metal centers, respectively, as initiated by Sears and Sherrill.[184,185] While both studies provide benchmark multireference computations and test B3LYP and BP86 performances, section 3.2 studies the smaller model versus the larger model salen ligands[1] and section 3.3 studies the performance of new meta-GGA DFT functionals.[7] Section 3.4 presents the only mechanistic study on a 'well-behaved' metal-salen catalyst, the Al-Salen.[8] As will become apparent, only by utilizing many of the

approximate single reference methods to accurately capture noncovalent interactions can one obtain a reliable reaction energy profile.

3.1 Generalized Theoretical Methodology

Figure 29 shows the two model systems for the metal-salens, as well as the full salen ligand. Generally, no axial ligands were used in this study in an attempt to focus on
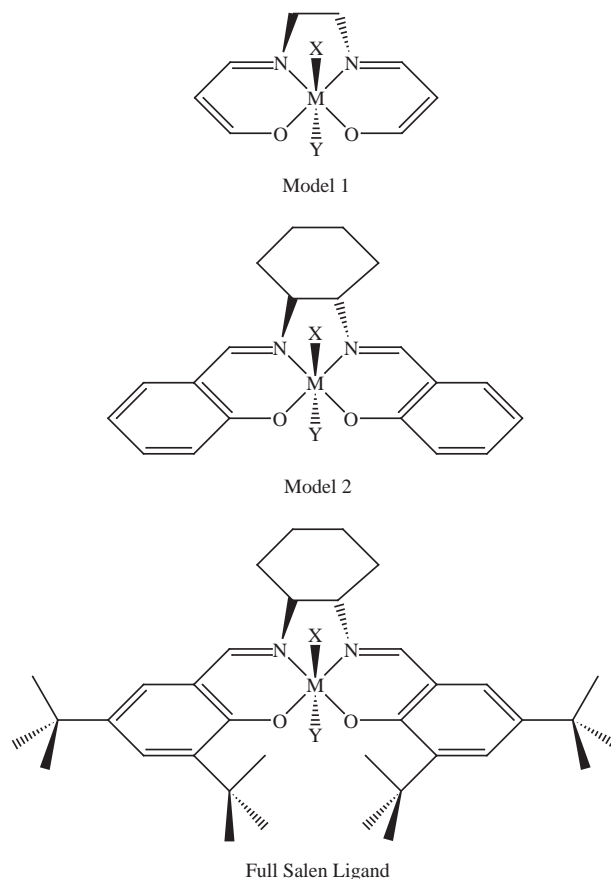
Figure 29: Two model systems for the metal-salen catalysts and the full salen ligand.

the nature of the interaction between the metal center and only the salen ligand.

All complete active-space (CAS) type methods were performed with the MOL-PRO 2006.1 package of *ab initio* programs.[56] The converged BP86 geometries were used as starting points for the complete active-space self-consistent field (CASSCF)[20] optimizations of the model 1 complexes. Starting orbitals were generated using the

configuration of singles and doubles (CISD) method with the STO-3G[201–203] basis sets. Large CAS-configuration interaction (CAS-CI)[200] computations were performed with the CISD natural orbital guesses. The active spaces were chosen from an examination of the most important orbitals in the CAS-CI vector. From this starting information, the final CASSCF wavefunctions were determined using the LANL2DZ* basis.

CAS second-order and third-order perturbation theory (CASPT2[21,22] and CASPT3[23]) methods were performed at the converged CASSCF geometries (RMS gradient 10^{-3} au) for the model 1 systems. Due to the limitation on the number of correlated orbitals that can be included in the CASPT3 program, the lowest 27 molecular orbitals were frozen throughout the CASPT3 computations. Estimates of the full CASPT3 energies were obtained by correcting the internally contracted CASPT2 energies with the difference between CASPT3 and CASPT2 with the 27 lowest orbitals frozen. All computations employed a restricted-core approximation with a small-core, 1s2s2p3s and 1s2s2p3s3p3d4s for the 3d and 4d transition metals, respectively. The leading determinants from the CASSCF CI expansions indicated the degree of multireference character for each complex.

Optimized model 1 CASSCF geometries were compared to optimized model 1 DFT geometries and the least root mean squared deviation (LRMSD) values were computed with the Visual Molecular Dynamics (VMD) program.[204] Molecular orbital isosurfaces were generated using MOLEKEL with contour values of 0.05.[205]

3.2 Assessing the Performance of Density Functional Theory for the Electronic Structure of Metal-Salens: The d^6-Metals

3.2.1 Introduction

This section is based on previously published work by Takatani, Sears, and Sherrill.[6]

This work extends the systematic studies on the accuracy of DFT methods to

the d^6 metal-salen complexes. In particular, the Fe(II)-, Co(III)-, Ni(IV)-, Ru(II)-, Rh(III)-, and Pd(IV)-metal salen complexes are considered. Optimized geometries and relative energies obtained using various density functional approximations are benchmarked against high-level multireference *ab initio* methods. This work improves upon prior studies of d^0 and d^2 metal-salen complexes by considering both the smaller model salen ligand used previously, as well as a larger model system. As before, this work is not intended to study the chemical properties of the metal-salen systems *per se*, but to test the reliability of DFT methods for use in such studies. The d^6-metal salen series does, however, contain important complexes utilized for catalytic transformations. The Ru(II)-salen has been shown to catalyze the cyclopropanation of olefins and the Co(III)-salen has been used to catalyze the hydrolytic kinetic resolution (HKR) reaction.[206–209] Furthermore, the efficiency of supported Co(III)-salen for HKR reactions has also been extensively studied.[164,165,210–212] The results presented here serve to guide future theoretical investigations on these highly important series of metal-salen catalysts.

3.2.2 Specialized Theoretical Methodology

The metal centers, M, studied were Fe(II), Co(III), Ni(IV), Ru(II), Rh(III), and Pd(IV). All DFT computations were performed with the Jaguar 5.5 suite of programs.[135] The combination of Becke's 1988[213] exchange functional with Perdew's 1986[214] correlation functional (BP86) and the combination of Becke's three parameter hybrid functional[27] with the correlation functional of Lee, Yang, and Parr[99] (the hybrid functional, B3LYP) DFT methods were utilized with the Los Alamos basis sets and corresponding effective core potentials of Hay and Wadt (LANL2DZ) for all transition-metal atoms[215] and the 6-31G* basis sets for all other atoms[216] (LANL2DZ*). The pseudospectral DFT implementation[217] with a fine grid, as found in Jaguar 5.5, was used to completely optimize the geometries [root mean square

(RMS) gradient 10^{-3} au] for the lowest singlet, triplet, quintet states of each functional for each model system. A restricted open shell reference was used in DFT computations of all triplet and quintet states. Frequency computations were performed to ensure the structures corresponded to potential energy minima.

Similar to model 1 geometry comparisons, the LMRSDs were calculated between the optimized model 1 B3LYP geometries and the optimized model 2 B3LYP geometries, where only the heavy atoms of the model 1 systems and the corresponding atoms of the model 2 systems were compared. CASSCF single-point computations were performed on the Fe(II)-, Co(III)-, Ru(II)-, and Rh(III)-salens with the optimized B3LYP model 2 geometries.

3.2.3 Results and Discussion

Table 16 provides the relative energies for the singlet, triplet, and quintet states for the model 1 and 2 Fe(II)-, Co(III)-, Ru(II)-, and Rh(III)-salens, and the model 1 Ni(IV)- and Pd(IV)-salens. States that are nearly isoenergetic with the lowest states provided in Table 16 and explicit geometry analyses are not presented here for the sake of brevity.

Figure 30 shows a representation of the active space orbitals used in the CAS-type computations. The computed active space orbitals match chemical intuition in that the highest lying occupied and partially occupied molecular orbitals are the five non/anti-bonding d-orbitals of the metal center. Previous work, furthermore, outlines the importance of the two C-C-C (three-center-two-electron) π bonds referred to as $R\pi_1$ and $R\pi_2$ hereafter. In Figure 30, both $R\pi_1$ and $R\pi_2$ are represented by the $R\pi$ orbital plot. These $R\pi$ orbitals are included only for the Ni(IV)- and Pd(IV)-salens leading to active spaces consisting of 10 electrons in 7 orbitals. This is due to the increased positive charge on the metal center compared to other metal-salens considered, which lowers the energy of the d orbitals relative to that of the $R\pi$ orbitals.

Table 16: Relative Energies (kcal mol^{-1}) for the Low-Lying Electronic States of the Model 1 and Model 2 (in Parentheses) d^6-Metal Salens Computed at Various Levels of Theory.

	State	CASPT3[a]	CASPT2[a]	CASSCF	B3LYP	BP86
Fe(II)	1^1A	56.17	61.42	69.72 (79.86)	28.78 (28.00)	16.60 (15.48)
	1^3A	21.90	25.43	36.30 (41.06)	0.00 (1.33)	0.00 (0.00)
	1^5A	0.00	0.00	0.00 (0.00)	3.91 (0.00)	15.55 (11.82)
Co(III)	1^1A	33.54	41.07	46.56 (43.79)	12.36 (13.08)	0.48 (1.06)
	1^3A	0.00	1.76	9.93 (10.97)	0.00 (0.00)	0.00 (0.00)
	1^5A	21.43	0.00	0.00 (0.00)	15.39 (11.93)	24.78 (22.27)
Ni(IV)	1^1A	1.13	0.81	0.61	0.00	0.00
	1^3A	0.41	0.25	0.00	20.88	8.97
	1^5A	0.00	0.00	0.06	8.24	39.40
Ru(II)	1^1A	22.61	22.44	26.06 (32.63)	10.18 (8.09)	2.48 (15.48)
	1^3A	0.00	0.00	0.00 (2.85)	0.00 (0.00)	0.00 (0.00)
	1^5A	14.69	10.09	3.33 (0.00)	37.62 (34.43)	46.77 (36.95)
Rh(III)	1^1A	17.42	11.75	22.34 (24.49)	0.00 (0.00)	0.00 (0.00)
	1^3A	0.00	0.00	0.00 (0.00)	2.91 (2.79)	8.69 (6.98)
	1^5A	63.11	24.45	35.96 (29.68)	49.03 (45.74)	58.48 (54.26)
Pd(IV)	1^1A	0.00	0.00	0.00	4.84	0.00
	1^3A	0.41	0.47	0.14	0.00	6.53
	1^5A	29.08	26.07	13.24	43.35	51.57

[a]Relative energies computed at the CASSCF optimized geometries.
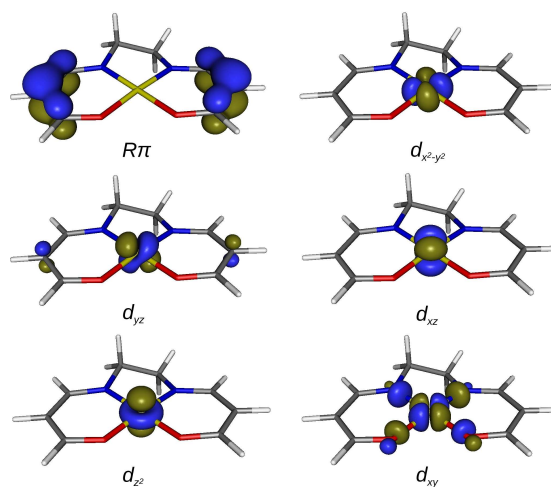

Figure 30: Representation of the active space orbitals for the d^6 metal-salen catalysts.

For the Fe(II)-, Co(III)-, Ru(II)-, and Rh(III)-salens, the $R\pi_1$ and $R\pi_2$ orbitals remain doubly occupied leading to active spaces consisting of 6 electrons in 5 orbitals.

3.2.3.1 Fe(II)- and Ru(II)-Salens

For the neutral Fe(II)- and Ru(II)-salens, the active spaces chosen were comprised of the five d-orbitals. CAS-type computations for the Fe(II)-salen predict a quintet ground state followed by a triplet state and then a singlet state, 22 and 56 kcal mol^{-1} higher in energy relative to the quintet state, respectively, for the CASPT3 method. Both DFT methods predict a triplet ground state followed by a quintet state. For the Ru(II)-salen, all methods predict a triplet ground state. The energetic orderings, however, are not consistent. The CASPT3 method indicates the quintet state to be 15 kcal mol^{-1} above the triplet state, and a singlet state 8 kcal mol^{-1} higher than the quintet state. The B3LYP method predicts the singlet state to follow the triplet state with a relative energy of 10 kcal mol^{-1} and then the quintet state 28 kcal mol^{-1} higher than the singlet state. The BP86 relative energies show the same trend as the B3LYP relative energies.

The discrepancy between the DFT and CAS-type methods can be rationalized by examining the leading determinants from CASSCF computations (a representative example is shown in Table 17). The small coefficients of the most important determinants indicate the high degree of multireference character of the Fe(II)- and Ru(II)-salen wavefunctions. For example, the largest determinant in the Fe(II)-salen singlet state CI wavefunction is only 0.663. By contrast, wavefunctions dominated by a single reference typically have a leading CI coefficient of about 0.9 or larger. The CI wavefunctions are consistent with near degeneracies of the $d_{x^2-y^2}$, d_{xz}, d_{z^2}, and d_{yz} orbitals. The d_{xy} orbital is higher in energy, however, with a small energy gap in the Fe(II)-salen (which favors the high-spin quintet state) and a somewhat larger energy gap in the Ru(II)-salen (which favors the triplet state). The orbital degeneracies make the Fe(II)- and Ru(II)-salens difficult to describe with the B3LYP and BP86 DFT methods. This strong multireference character continues throughout the d^6 metal-salen series.

Table 17: Leading Determinants in the Natural Orbital Basis from SA-CASSCF Calculations on the Low-Lying Electronic States of Fe(II)- and Ru(II)-Salens

	State	Determinant	Coeff
Fe(II)	1^1A	$(d_{x^2-y^2})^2 (d_{xz})^2 (d_{z^2})\beta (d_{yz})\alpha$	0.663
		$(d_{x^2-y^2})^2 (d_{xz})^2 (d_{z^2})\alpha (d_{yz})\beta$	-0.663
		$(d_{x^2-y^2})^2 (d_{xz})^2 (d_{z^2})^2$	-0.165
	1^3A	$(d_{x^2-y^2})^2 (d_{xz})\alpha (d_{z^2})^2 (d_{yz})\alpha$	0.976
		$(d_{x^2-y^2})^2 (d_{xz})\alpha (d_{yz})\alpha (d_{xy})^2$	-0.116
	1^5A	$(d_{x^2-y^2})\alpha (d_{xz})\alpha (d_{z^2})^2 (d_{yz})\alpha (d_{xy})\alpha$	0.986
		$(d_{x^2-y^2})^2 (d_{xz})\alpha (d_{z^2})\alpha (d_{yz})\alpha (d_{xy})\alpha$	-0.164
Ru(II)	1^1A	$(d_{x^2-y^2})^2 (d_{xz})^2 (d_{z^2})^2$	0.795
		$(d_{x^2-y^2})^2 (d_{xz})^2 (d_{yz})^2$	-0.334
		$(d_{x^2-y^2})^2 (d_{z^2})^2 (d_{yz})^2$	-0.268
		$(d_{x^2-y^2})\beta (d_{xz})^2 (d_{z^2})\alpha (d_{yz})^2$	-0.254
		$(d_{x^2-y^2})\alpha (d_{xz})^2 (d_{z^2})\beta (d_{yz})^2$	0.254
	1^3A	$(d_{x^2-y^2})^2 (d_{xz})^2 (d_{z^2})\alpha (d_{yz})\alpha$	0.925
		$(d_{x^2-y^2})\alpha (d_{xz})^2 (d_{z^2})^2 (d_{yz})\alpha$	0.366
	1^5A	$(d_{x^2-y^2})\alpha (d_{xz})\alpha (d_{z^2})^2 (d_{yz})\alpha (d_{xy})\alpha$	0.741
		$(d_{x^2-y^2})^2 (d_{xz})\alpha (d_{z^2})\alpha (d_{yz})\alpha (d_{xy})\alpha$	-0.624
		$(d_{x^2-y^2})\alpha (d_{xz})\alpha (d_{z^2})\alpha (d_{yz})^2 (d_{xy})\alpha$	0.248

3.2.3.2   Co(III)- and Rh(III)-Salens

As with the Fe(II)- and Ru(II)-salens, the active space chosen for the Co(III)- and Rh(III)-salens were comprised of the five d-orbitals. Interestingly for the Co(III)-salen, the CASPT3 and DFT methods predict a triplet ground state, and the CASPT2 and CASSCF methods predict a quintet ground state. As the higher-order perturbations are added to the CASSCF computations, the triplet and quintet states become nearly degenerate, as predicted by the CASPT2 method, and eventually switch order as in the CASPT3 computations. The corrections, however, do not seem to be convergent. The difference in relative energies for the CASSCF triplet and quintet states is $\sim$10 kcal mol^{-1}. This difference changes to $\sim$2 kcal mol^{-1} and then $\sim$21 kcal mol^{-1} for the CASPT2 and CASPT3 triplet and quintet states, respectively. This trend makes it difficult to assess the performance of DFT methods because, in this case, the CAS-type computations do not appear to have converged.

3.2.3.3 Ni(IV)- and Pd(IV)-Salens

The charge of the metals in Ni(IV)-salen and Pd(IV)-salen pull electron density towards the center. Consequently, the active spaces for the Ni(IV)- and Pd(IV)-salens were increased to include both Rπ orbitals as well as the five d-orbitals, whereas in the previous metal-salens the Rπ orbitals always remained doubly occupied. CAS-type relative energies for the Ni(IV)-salen suggest that the lowest singlet, triplet, and quintet states are nearly degenerate. Both DFT methods predict an apparent singlet ground state; however, the ordering of the other states is in disagreement. The B3LYP method predicts a quintet state to follow the singlet state with a relative energy of 8 kcal mol^{-1} and the triplet state 13 kcal mol^{-1} higher than the quintet state. The BP86 method, on the other hand, predicts the triplet state to follow the singlet state with a relative energy of 9 kcal mol^{-1} and the quintet state 30 kcal mol^{-1} higher than the triplet state. For the Pd(IV)-salen, CAS-type computations predict nearly degenerate singlet and triplet states with a high lying quintet state (29 kcal mol^{-1} above the ground state according to the CASPT3 method). Both DFT methods, in this case, are in agreement with the CAS-type methods, but with overestimations of the quintet energies (43 and 52 kcal mol^{-1} for the B3LYP and BP86 methods, respectively). This could either be due to the overstabilization of the low spin states or the understabilization of the quintet states.

3.2.3.4 Geometry Analyses

Although the energetic ordering of states does not appear reliable as computed by DFT methods, DFT optimized geometries are in good agreement with CASSCF optimized geometries; errors are mostly on the order of 0.1 Å LMRSD. Figure 31 overlays the DFT geometries with the CASSCF geometries for the singlet and quintet states of Co(III)-salen. The worst agreement between the DFT and CASSCF methods for geometries is observed for the quintet state of the Co(III)-salen, where there is a
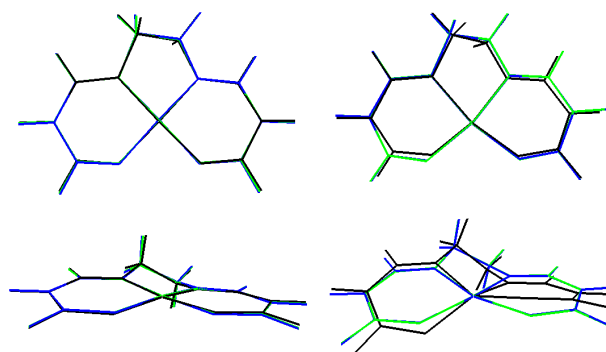
Figure 31: Overlay of the CASSCF (black), B3LYP (green), and BP86 (blue) optimized geometries for the 1^1A (left) and 1^5A (right) states of Co(III)-salen

deviation of 0.3 Å LRMSD. LMRSDs over 0.2 Å for DFT methods normally appear when comparing the quintet states of the metal-salens [specifically the Ru(II)-, Co(III)-, and Pd(IV)-salens]. The salen ligands in the quintet state CASSCF geometries twist out of planarity to a greater degree than the lower spin state geometries as depicted in Figure 31. DFT methods do not predict this out of plane twisting for the quintet states. Uncharacteristically, DFT singlet state optimized geometries for the Ni(IV)-salen incur greater deviations ($\sim$0.24 Å LMRSDs) than the higher spin state geometries compared to CASSCF optimized geometries. For all other metal-salens considered, the largest deviations are between the quintet geometries. Nevertheless, the DFT and CASSCF optimized geometries are fairly similar for all systems considered.

3.2.3.5   Larger Model Salens

Due to the reliable performance of DFT optimizations, CASSCF computations for the model 2 salens were performed at B3LYP optimized geometries. The relative energies for the model 2 Fe(II)-, Co(III)-, Ru(II)-, and Rh(III)-salens are shown in parentheses in Table 16. All methods predict the same energetic trends for the model 2 systems as for the model 1 systems. The similar trends indicate that there are no dramatic changes in the electronic structure as the larger model systems are used. This result is

expected since the most important orbitals are the d-orbitals of the metal center and it is unlikely that they will mix to a significant degree when moving from the model 1 to the model 2 system. For the model 2 Ni(IV)- and Pd(IV)-salens, difficulties in keeping the appropriate orbitals within the active spaces for the CASSCF computations were encountered. It remains unknown whether the differences between the Rπ orbitals of the model 1 and model 2 systems change the energetic ordering of spin states and the energy gaps between them.

The B3LYP optimized geometry LMRSDs between the model 1 and 2 systems are presented in Table 18. For all triplet geometries considered, the model 1 systems

Table 18: LRMSD (Å) in Molecular Geometries^a for the Optimized 1^1A, 1^3A, and 1^5A States for the B3LYP Model 1 Compared to Model 2 Systems

	State	LRMSD
Fe(II)	1^1A	0.23
	1^3A	0.05
	1^5A	0.19
Co(III)	1^1A	0.20
	1^3A	0.06
	1^5A	0.14
Ru(II)	1^1A	0.33
	1^3A	0.06
	1^5A	0.24
Rh(III)	1^1A	0.08
	1^3A	0.06
	1^5A	0.10

^{a}Comparisons based on the heavy atoms of the model 1 system.

are very similar to the model 2 systems with 0.06 Å LMRSDs or less. The optimized Rh(III)-salen model 1 geometries reproduce the the model 2 geometries for the singlet, triplet, and quintet states to less than 0.10 Å LMRSD. The comparison of the Ru(II)-salen singlet geometries incurs the greatest error of 0.33 Å LMRSD. Although some of the model 1 geometries differ by a few tenths of Å LMRSDs compared to the model 2 geometries, they are close enough to capture the correct electronic structure of the complexes considered here. Moreover, preliminary investigations suggest that the model 2 ligand does a good job reproducing the geometry of the full salen ligand.

3.2.4 Conclusions

The systematic study of the reliability of DFT methods for organometallic complexes has been extended to the d^6 metal-salens. DFT methods rarely predict the correct ordering of states as compared to high-level CASPT3 results. Moreover, in general, the splitting between states seem to be underestimated by the B3LYP and BP86 functionals.

For the highly charged Ni(IV)- and Pd(IV)-salens, however, either the quintet states are understabilized or the singlet and triplet spin states are overstabilized, which leads to overestimated energy gaps. A curious non-convergence of the perturbative correction series for the Co(III)- and Rh(III)-salens was observed. Further investigation of these systems is underway.

Whereas the reliability of DFT for energetic ordering of states is still unpredictable, DFT geometry optimizations are in good agreement with CASSCF geometry optimizations. This result is consistent with previous findings for d^0 and d^2 metal-salen complexes. The results for the energetic ordering of spin states for the larger model 2 salen, furthermore, show no appreciable difference from the smaller model 1 salen results. Therefore, the model 1 system and DFT geometry optimizations can be considered appropriate for most computations concerning d^6 metal-salens. It has been recognized that new DFT functionals may improve the accuracy and reliability when considering metal-ligand complexes and as such, further research is necessary and also currently underway.

3.3 Assessing the Performance of Density Functional Theory for the Electronic Structure of Metal-Salens: The M06 Suite of Functionals and the d^4-Metals

3.3.1 Introduction

This section is based on submitted material by Takatani, Sears, and Sherrill.[7]

Previous DFT studies on the metal-salens have employed functionals that depend

on the local spin density and its gradient (*i.e.*, generalized gradient approximation, GGA, functionals) sometimes mixed with Hartree-Fock exchange (hybrid functionals). More recently developed functionals also include the kinetic-energy density and are referred to as the meta-GGA functionals. Some meta-GGA functionals have been parameterized against vast databases of highly accurate energies, which include databases of transition metal systems.[29–31] Zhao and Truhlar report that the M0-suite outperforms GGA, hybrid-GGA, and other meta-GGA functionals, such as the BMK functional,[28] for thermochemistry, kinetics, noncovalent interactions, and electronic excitations.[31] Moreover, the M06 and M06-L functionals performed the best for metal atom excitations, transition metal bonding, and organometallic and inorganometallic thermochemistry. The M06-L functional is even more attractive due to the exclusion of the Hartree-Fock exchange which decreases its computational cost compared to the other functionals within the M0-suite. This M0-family of meta-GGA functionals will be tested here for a series of metal-salens.

This work serves to extend the previous systematic study on the accuracy of DFT methods for the metal-salen complexes, where the d^4 metal-salens are now under investigation. These metal-centers include the Cr(II)-, Mn(III)-, Fe(IV)-, Mo(II)-, Tc(III)-, and Ru(IV)-metals. The same procedure to obtain the optimized geometries and relative energies with DFT and with high-level multireference *ab initio* methods is followed as in previous studies.[6,184,185] As work on the d^6 metal-salens suggests, there is little difference in energetics when considering the small model salen opposed to the larger model salen.[6] Therefore, here only the optimized geometries and relative energies are benchmarked with the small model 1 salen. This is the first test of the M0-suite for the very theoretically challenging salen catalysts. As before, the intent is not to study the chemical properties of the metal-salen systems *per se*, but to test the reliability of DFT methods for use in such studies. The d^4 metal-salen series does, however, contain the important Mn(III)-salen catalyst that is extensively utilized for

various oxidation[218–221] and epoxidation reactions.[222–229]

3.3.2 Specialized Theoretical Methodology

The metal centers, M of the model 1 complex, studied were Cr(II), Mn(III), Fe(IV), Mo(II), Tc(III), and Ru(IV). All DFT computations were performed with the Q-Chem 3.2 suite of programs.[110] The standard SG-1 grid was used with the Los Alamos basis sets and corresponding effective core potentials of Hay and Wadt (LANL2DZ) for all transition metal atoms[215] and the 6-31G* basis sets for all other atoms[216] (LANL2DZ*) to obtain optimized singlet, triplet, and quintet geometries and energies. The BP86 functional combined Becke's 1988[213] exchange functional with Perdew's 1986[214] correlation functional and the B3LYP functional combined Becke's three parameter hybrid functional[27] with the correlation functional of Lee, Yang, and Parr.[99] This study also included the analysis of the M0-suite (M05, M05-2X, M06, M06-2X, M06-HF, and M06-L) of meta-GGA DFT functionals.[29–31,113] An unrestricted reference was used in DFT computations of all triplet and quintet states. Subsequent frequency computations were performed to ensure the structures corresponded to potential energy minima.

3.3.3 Results and Discussion

Figure 32 represents the active space orbitals used in the wavefunction-based computations. The computed active space orbitals match chemical intuition in that the highest lying occupied and partially occupied molecular orbitals are four out of the five non/anti-bonding d-orbitals of the metal center. Previous work, furthermore, outlines the importance of the two C-C-C (three-center-two-electron) π bonds referred to as the Rπ orbitals hereafter. All wavefunction-based computations, except for the Ru(IV)-salen complex, used 4 active electrons in 4 d-orbitals. The highly positive Ru(IV)-salen complex included four d-orbitals as well as the in-phase and out-of-phase combinations of the Rπ orbitals and therefore required an active space
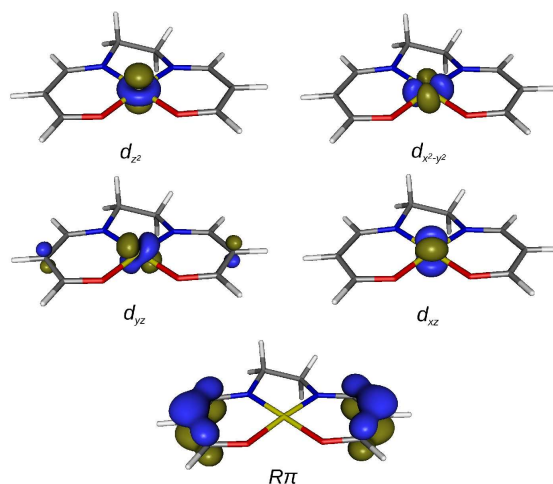
101

Figure 32: Representation of the active space orbitals for the d^4 metal-salen catalysts.

of 8 electrons in 6 orbitals. For the other complexes, the $R\pi$ orbitals remain doubly occupied. As with the d^6 metal-salen series, the bonding nature of the metal center with the salen ligand mainly relies on the interactions between the d_{xy} metal orbital with the p orbitals of the surrounding nitrogen and oxygen atoms. The anti-bonding d_{xy} combinations, however, are not occupied for any of the d^4 metal-salen series.

Table 19 provides the wavefunction-based relative energies for the lowest singlet, triplet, and quintet states of the d^4 metal-salen series and the corresponding coefficients of the leading determinant for each state. The high-spin quintet state is the ground state for all metal-salen complexes tested here except for the Ru(IV)-salen complex. This signifies that the energy spacings between the d-orbitals are small compared to the Coulomb repulsion of the d-electrons. Furthermore, these quintet states are dominated by a single determinant reference, thereby suggesting that ground state properties for these systems can be properly characterized by single reference methods. In most cases, the singlet states are somewhat higher-lying (with CASPT3 relative energies of 44-84 kcal mol^{-1}) and may not be very relevant to the electronic structure of these systems. However, the lowest singlet state of the Fe(IV)-salen complex lies lower at 28 kcal mol^{-1}, and the singlet state is the ground state for the

Table 19: Wavefunction Based Relative Energies (kcal mol^{-1}) for the Low-Lying Electronic States of the d^4-Metal Salens.

	State	CASPT3[a]	CASPT2[a]	CASSCF	Det[b]
Cr(II)	1^{1}A	76.53	77.87	83.26	0.561
	1^{3}A	51.41	78.02	58.71	0.619
	1^{5}A	0.00	0.00	0.00	1.000
Mn(III)	1^{1}A	83.71	90.11	93.00	0.822
	1^{3}A	60.13	63.94	66.73	0.619
	1^{5}A	0.00	0.00	0.00	1.000
Fe(IV)	1^{1}A	28.17	58.52	71.84	0.969
	1^{3}A	20.65	45.57	54.70	0.963
	1^{5}A	0.00	0.00	0.00	1.000
Mo(II)	1^{1}A	50.76	54.77	60.91	0.545
	1^{3}A	34.25	35.64	43.71	0.888
	1^{5}A	0.00	0.00	0.00	1.000
Tc(III)	1^{1}A	43.88	54.41	60.34	0.946
	1^{3}A	27.56	35.66	41.83	0.948
	1^{5}A	0.00	0.00	0.00	1.000
Ru(IV)	1^{1}A	0.00	0.00	0.00	0.469
	1^{3}A	1.83	1.88	0.95	0.549
	1^{5}A	5.97	6.64	3.01	0.894

[a]Relative energies computed at the CASSCF optimized geometries. [b]Coefficient for the leading determinant in the natural orbital basis from SA-CASSCF computations.

Ru(IV)-salen complex. The triplet states are lower-lying with CASPT3 relative energies of 21-60 kcal mol^{-1} above the quintet states except for the Ru(IV)-salen, where the triplet state is only 2 kcal mol^{-1} above the ground state. The highly positive metal-center of the Ru(IV)-salen complex decreases the d-orbital energies compared to the Rπ orbitals, creating the situation where the singlet and triplet states are nearly isoenergetic and the quintet state remains slightly higher in energy. The leading determinant for the Ru(IV)-salen quintet state is only 0.894 and it is, therefore, unlikely that single reference methods will be able to accurately characterize the ground state of this system.

Figure 33 summarizes the DFT performance for the singlet-triplet (S-T), triplet-quintet (T-Q), and singlet-quintet (S-Q) energy gaps compared to the CASPT3 results. Most of the spin state energy gaps are tens of kcal mol^{-1} for these model complexes, suggesting that one should expect significant errors in DFT approximations
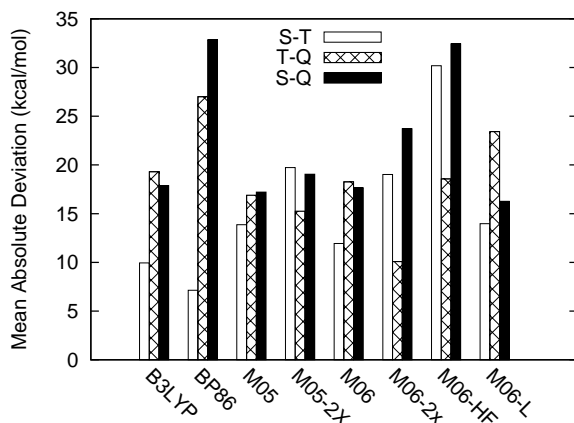
Figure 33: Mean absolute deviations (kcal/mol) for various DFT functionals compared against CASPT3 singlet-triplet (S-T), triplet-quintet (T-Q), and singlet-quintet (S-Q) energy gaps.

to these gaps. Nevertheless, the large mean absolute deviations (MADs) in Figure 33 (generally in the range of 10-20 kcal mol^{-1}) are surprisingly large and comparable, in some cases, to the size of the energy gaps themselves. In particular, the BP86 and the M06-HF functionals perform rather poorly with MADs greater than 30 kcal mol^{-1} for the S-Q energy gaps; the average CASPT3 S-Q energy gap is only around 48 kcal mol^{-1}. Furthermore, the T-Q energy gap MAD for the BP86 functional is over 25 kcal mol^{-1} and the S-T energy gap MAD for the M06-HF functional is around 30 kcal mol^{-1}. The BP86 functional only predicts the correct ground state for 2 out of the 6 systems presented here (the Cr(II)- and Mn(III)-salen complexes). Clearly, these two functionals cannot be considered reliable to properly characterize the low-lying spin states of the d^4 metal-salen systems. Considering the other functionals, none consistently outperform all the rest. Therefore, the hybrid B3LYP functional appears to perform just as well as the meta-GGA functionals for these systems even though the M06 and M06-L functionals are the suggested methods for organometallic and transition metal chemistry in other contexts. It should be noted that no functional tested here obtained the correct ground state for the Ru(IV)-salen complex, and only

the M06-2X functional obtained the correct ground state for the Fe(IV)-salen complex. The more positively charged metal-centers seem to present greater difficulties for these DFT methods.

The performance of the DFT geometry optimizations compared to CASSCF optimized geometries is presented in Figure 34 as the average LRMSD for the singlet, triplet, and quintet geometries in Å. Although the average LRMSDs for the singlet
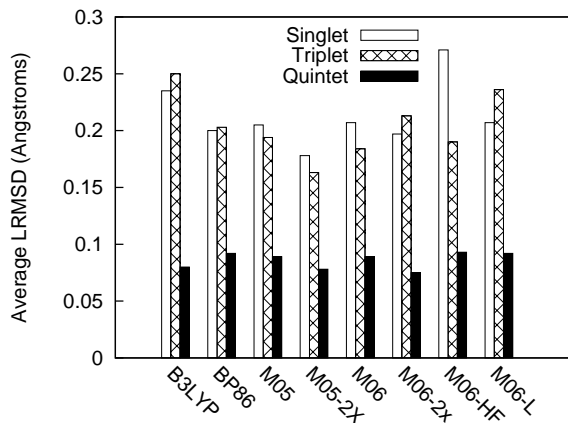

Figure 34: Average LRMSD (Å) in the 1^1A, 1^3A, and 1^5A state molecular geometries for various DFT functionals compared to CASSCF optimized geometries.

and triplet state geometries are only around 0.2-0.3 Å, the average LRMSD for the quintet state geometries are within 0.1 Å. This better general performance for the quintet state geometries is most likely due to the single reference nature of the wavefunction for the quintet states (except for the Ru(IV)-salen complex). The highly charged Fe(IV)- and Ru(IV)-salen complexes, however, present the exceptions to this generalization. The quintet state of the Fe(IV)-salen complex relies on a single determinant, but DFT optimizations incur 0.2-0.3 Å LRMSD. In contrast, the quintet state of the Ru(IV)-salen complex has a coefficient of only 0.894 for the leading determinant, yet DFT optimizations perform extremely well incurring less than 0.1 Å LRMSD. Overall, the M05-2X functional slightly outperforms the other functionals for geometry predictions, with singlet and triplet average LRMSDs well under 0.2 Å.

3.3.4 Conclusions

DFT methods exhibit alarmingly large MADs when comparing DFT and CASPT3 spin-state energy gaps; in general, CASPT3 energy gaps are on the order of tens of kcal mol^{-1}, while the error in the DFT energy gaps are also on the order of tens of kcal mol^{-1}. Surprisingly, none of the M0-suite of meta-GGA functionals appear statistically better than the popular B3LYP functional for the computation of energy gaps for the systems studied. Furthermore, no functional tested here obtains the correct ground state for either the Fe(IV)- or Ru(IV)-salen complexes (except the M06-2X functional on the Fe(IV)-salen complex). It should also be stressed that the BP86 functional performs extremely poorly and is definitely not recommended for use on these systems. Fortunately, as concluded in the previous work on the d^0, d^2, and d^6 metal-salen complexes,[6,184,185] DFT optimizations are reasonably reliable compared to CASSCF geometries. In this case, the DFT and CASSCF optimized structures match to within 0.3 Å LRMSD for the singlet and triplet states and 0.1 Å LRMSD for the quintet states.

Combined with previous systematic study results, it can be concluded that no DFT functional (tested here) can be considered reliable over all metal-salen complexes. This is extremely unfortunate due to the system sizes one might encounter while attempting to model the metal-salen mechanisms, *e.g.*, the tethered Co(bisalen) catalyst.[230] It is highly recommended that the accuracy of any given DFT functional should be assessed on a case-by-case basis. For the Mn(III)-salen complex, for example, CASPT3 results suggest a very low-lying quintet ground state followed by a much higher-lying triplet state around 60 kcal mol^{-1} higher in energy. Of the DFT functionals tested here, the M06-2X functional provides the smallest error in the T-Q energy gap, with a quintet ground state followed by a triplet state around 45 kcal mol^{-1} higher in energy. The other functionals provide T-Q energy gaps of around 6-33 kcal mol^{-1}. Moreover, the M06-2X LRMSDs for the quintet and triplet geometries

are 0.02 and 0.05 Å, respectively. Therefore, if attempting to model the Mn(III)-salen complex (with no counter-ions), then the M06-2X functional should provide reliable results. However, since many metal-salen catalysts incorporate counter-ions (*e.g.* Cl$^-$, OH$^-$, etc.), further research is necessary to explore how the multireference character may be affected by the inclusion of counter-ions and, of course, to test how these changes affect the performance of DFT functionals for these systems.

3.4 Mechanistic Investigation of the Conjugate Addition of Cyanide to α,β-Unsaturated Imides Catalyzed by the Al(Cl)-Salen Complex

3.4.1 Introduction

Among the most studied catalytic reactions involving the metal-salen complex are the oxidation and epoxidation reactions catalyzed by the Mn(III)-salen complex,[218–223,225–229] the hydrolytic kinetic resolution (HKR) reaction catalyzed by the Co(III)-salen complex,[164,165,206–208,210–212,230] and the addition of cyanide to aldehydes, ketones, and imides catalyzed by the Al(III)-salen complex.[231–238] Furthermore, the catalytic activity of both the Co(III)- and Al(III)-salen catalysts have been shown to vary with different immobilization schemes and molecular supports.[164,165,208,210,211,233] Understanding why this happens and how to tune the reactivity through these variations would be much easier if detailed mechanisms were available through electronic structure computations.

Theoretical research on the metal-salen complexes, however, has proven to be quite difficult. The application of two common density functional theory (DFT)[24,25] methods to the olefin epoxidation mechanism catalyzed by the Mn-salen complex yields large qualitative discrepancies.[170] Additionally, research specifically concerning the Mn-salen complex[181] and subsequent systematic studies[6,7,184,185] on various transition metal-salen complexes suggest that, overall, DFT methods do not predict quantitatively accurate spin state energy gaps and do not always obtain the

correct spin ground states. As a result, mechanistic studies utilizing DFT methods involving transition metal-salen catalysts[170,239–241] must always be scrutinized carefully to extract meaningful insights. This is unfortunate because DFT is the most advanced computationally feasible for these studies due to the size of the metal-salen complexes and the common occurrence of cooperative catalysis involving two of these complexes.[208,230,234,236–238] On the other hand DFT methods, as applied to the Al(III)-salen complex, should not suffer from the same difficulties as for transition metal-salens because the problems are primarily due to partially occupied d-orbitals, which are not an issue for aluminum.

Figure 35 depicts the (R,R)-(salen)AlCl (further referred to as the Al(Cl)-salen) catalyst utilized in the enantioselective conjugate addition of cyanide to imide reaction in Scheme 36. Experimental research shows that the reaction rate is second-order
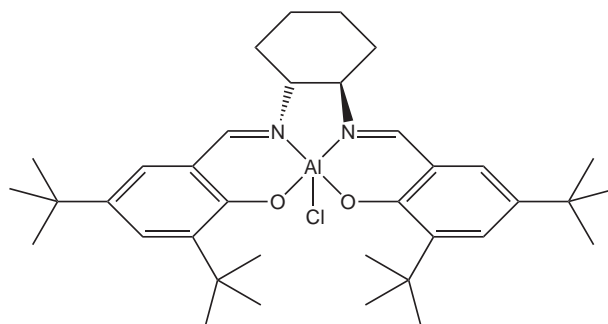

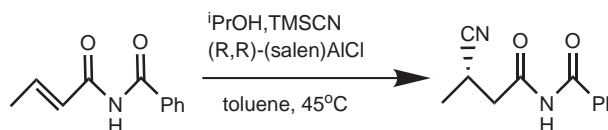**Figure 35:** The (R,R)-(salen)AlCl catalyst.

Figure 36: The addition of cyanide to the α,β-unsaturated imide catalyzed by the Al(Cl)-salen catalyst.

with respect to the catalyst, and thus at least one step of the reaction mechanism involves the interaction between two catalysts.[232,237] The salen ligand, however, exhibits significant π-aromaticity. Theoretical research on the benzene dimer and other

π-aromatic noncovalent interactions[3,43–52] suggest that robust and expensive electron correlation methods must be utilized to obtain quantitatively reliable results noncovalent π-π interactions. Standard DFT methods do not model noncovalent London dispersion interactions, and hence it is highly unlikely that DFT methods will predict the correct interaction of two stacked Al(Cl)-salen complexes.

Fortunately, DFT methods can obtain semi-quantitatively correct results for noncovalent interactions with a simple addition of a R^{-6} component as implemented by the DFT-D methods.[116,117] Moreover, the spin-component scaled Møller-Plesset second-order perturbation theory (SCS-MP2) method[54] often yields noncovalent interaction energy errors of only around 10-20%.[1,4,5,83,88] The cost of the MP2 method can be further reduced by the resolution-of-the-identity[65,68] and dual-basis[118] approximations without any significant loss of accuracy.[3] Furthermore, recent mechanistic studies show an improvement of the SCS-MP2 method over the canonical MP2 method.[242,243] Using these techniques, this study attempts to elucidate on the enantioselective conjugate addition of cyanide to α,β-unsaturated imides with the Al(Cl)-salen catalyst.

3.4.2 Specialized Theoretical Methodology

All geometry and transition state optimizations, solvent corrections, and single point SCS-MP2[54] computations were performed with the Q-Chem 3.2 suite of programs.[110] DFT optimizations employed the BP86 functional[213,214] with the DFT-D[116,117] term (BP86-D) and the standard SG-1 grid. The 6-31G* basis set[216] was used for all optimizations except for complexes greater than 70 atoms, where the 3-21G* basis set was used instead to make the computations practical. Subsequent frequency computations ensured that optimized structures and transition states corresponded to potential energy minima and first-order saddle points (one imaginary frequency), respectively.

Utilizing the optimized BP86-D structures, single point SCS-MP2 energies in the gas phase were obtained with the aug-cc-pVTZ (aTZ) basis set[66,67] in conjunction with the dual-basis approach[118] with the truncated aTZ basis set and the resolution-of-the-identity approximation[65,68] for the MP2 correlation energy with the corresponding aTZ auxiliary MP2-fitting[69] basis set. The frozen core approximation was used for all SCS-MP2 computations. The toluene solvent (dielectric constant of 2.38) was approximated by the self-consistent reaction field (SCRF)[244,245] method as implemented in conjunction with the Hartree-Fock (HF) method. Due to the size of certain complexes, only the cc-pVDZ (DZ) basis set could be used consistently for all SCRF computations and thus final SCS-MP2 interaction energies with the SCRF corrections were calculated via Equation 1,

$$E^{SCRF}_{SCS-MP2}(aTZ) \approx E^{gas}_{SCS-MP2}(aTZ) + E^{SCRF}_{HF}(DZ) - E^{gas}_{HF}(DZ). \tag{64}$$

Figure 37 shows the Al(III)-salen models utilized to reduce computational costs. The model 1 salen was used throughout the mechanistic study, whereas the model
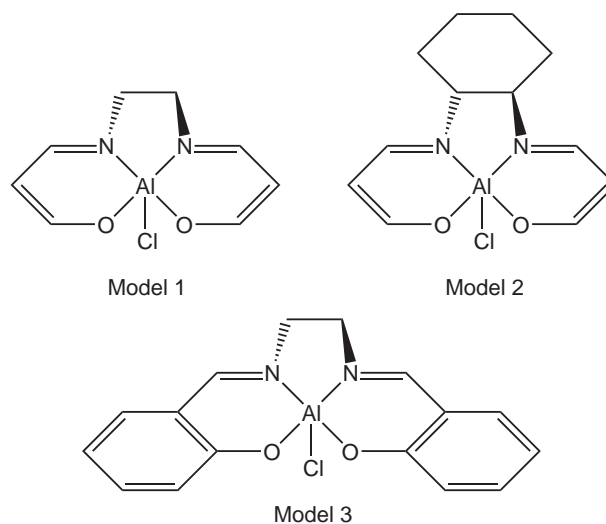

Figure 37: Model systems for the Al(Cl)-salen catalyst.

2 and model 3 salens were used only to enforce chirality and increase π-aromaticity, respectively, in specific cases (see below). To rationalize certain experimental results,

atomic charges were computed by fitting to BP86/6-31G* electrostatic potentials with the Jaguar suite of programs.[135]

3.4.3   Results and Discussion

Through spectroscopic studies and titrations, Sammis and Jacobsen concluded that the Al(Cl)-salen catalyst is converted into two distinct species: the Al(CN)-salen upon mixing the Al(Cl)-salen catalyst with TMSCN and the Al(imidate)-salen upon mixing the Al(CN)-salen with the imide reactant.[237] Therefore in the following discussion, it will be assumed that the Al(Cl)-salen catalyst is first converted into the Al(CN)-salen catalyst and the subsequent catalytic cycle will only involve the Al(CN)-salen catalyst as shown in figure 38.

The major steps of the catalytic cycle for the addition of cyanide to the unsaturated imide are shown in figure 38. Steps A and B are energetically favorable (-20.2 and -42.9 kcal mol^{-1}, respectively), where in step A the Al(CN)-salen catalyst is converted into the Al(imidate)-salen. HCN is produced as a byproduct of step A, and this *in situ* production of HCN was also hypothesized by Sammis and Jacobsen, although through the reaction of TMSCN and 2-propanol.[237] Step B involves a dual catalyst activation that adds cyanide to imidate, while simultaneously adding hydrogen also to imidate. Experimentally, this cooperativity between catalysts has been suggested due to the second-order dependence of the reaction with respect to the Al(Cl)-salen catalyst.[232,237] Step C is the only unfavorable reaction (19.9 kcal mol^{-1}), where the S-adduct of the cyanide added to the imide (referred to as imide-CN, hereafter) is recovered through the reaction with 2-propanol. The final step recovers the Al(CN)-salen catalyst through the reaction of the Al(^{i}PrO)-salen with TMSCN and does not require or yield any additional energy.
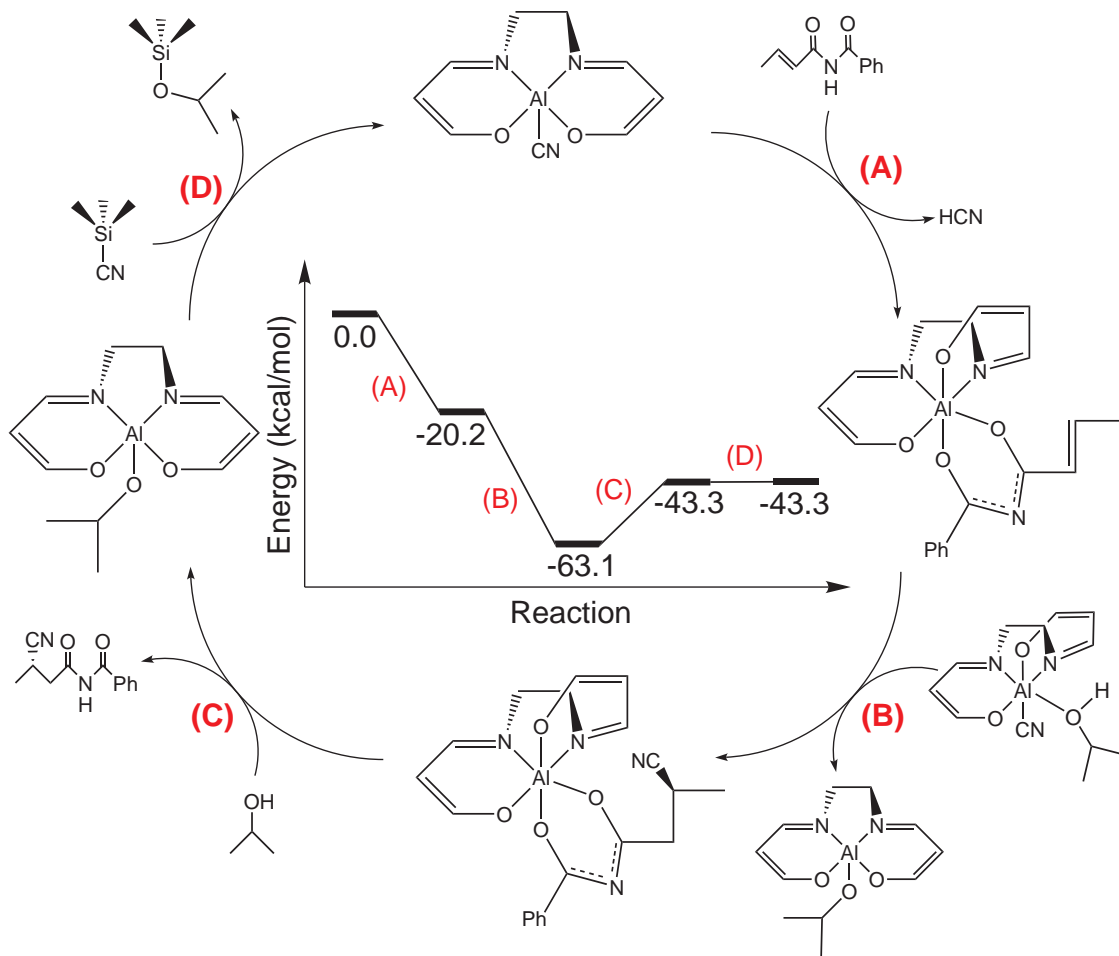
Figure 38: Catalytic cycle for the conjugate addition of cyanide to α,β-unsaturated imide (the S-adduct) with the Al(CN)-salen catalyst.

3.4.3.1 Al(Cl)-Salen to Al(CN)-Salen

Figure 39 shows two possible coordination sites of TMSCN to the Al(Cl)-salen catalyst. In this case, the model 2 salen ligand is used to investigate the chiral effects of the (R,R)salen-ligand. The coordination to the 'right' side of the Al(Cl)-salen catalyst, where the 'right' oxygen is displaced into the axial position, is preferred with a transition state barrier nearly half that of the 'left' side coordination. Moreover, the coordination energy of the 'right' side is 7.3 kcal mol^{-1} less than the 'left' side. For these reasons, only the coordination to the 'right' side of the salen ligand are further considered in the catalytic cycle with the (achiral) model 1 salen ligand.
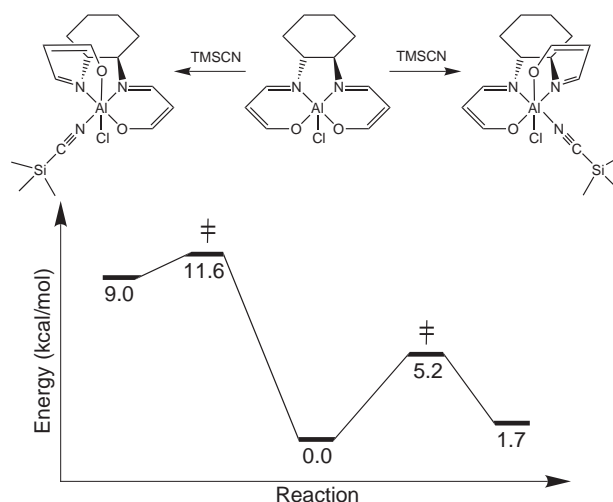
Figure 39: Reaction energy profile of the TMSCN coordination to the Al(Cl)-Salen catalyst with the model 2 salen ligand where the left- and right-most structures are the products of the TMSCN coordination to the Al(Cl)-salen complex.

Figure 40 depicts the production of the Al(CN)-salen catalyst through the reaction of the Al(Cl)-salen catalyst and TMSCN. As shown by the transition state of figure 40, the cyanide counter-ion must rotate to coordinate to the Al(III)-salen. Therefore, the transfer of the cyanide counter-ion is facilitated through the sufficient activation of TMSCN. A single catalyst (figure 40a) can undergo this reaction, however with a 38.6 kcal mol^{-1} activation barrier. Favorable London dispersion forces can decrease the transition state energy by 2 kcal mol^{-1} as suggested by utilizing the model 3 salen ligand in figure 40b. The chloride of a second catalyst can aid the activation of TMSCN by reducing the barrier by around 4 kcal mol^{-1} as depicted by figure 40c with the model 1 salen ligand. Furthermore, if we assume that all Al(Cl)-salen catalysts are primarily coordinated to TMSCN, similar to the right side structure of figure 39, then the dual catalyst transition state can be decreased to 28.8 kcal mol^{-1}. We also predict that the activation barrier of the cooperative catalyst transition state is an upperbound to the energy due to the neglect of favorable London dispersion interactions between the two catalyst by utilizing the model 1 salen ligand. The reaction, resulting with the Al(CN)-salen and TMSCl products, is only slightly endothermic
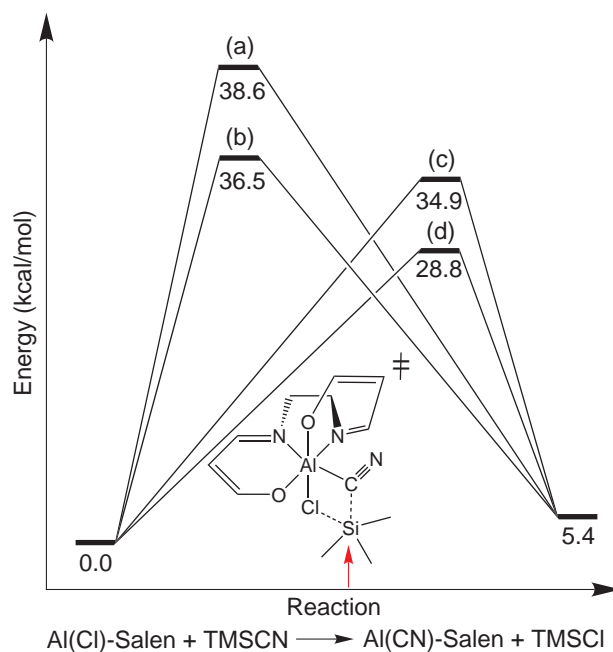
Figure 40: Reaction energy profile for the production of the Al(CN)-Salen catalyst where (a) is a single catalyst transition state with the model 1 ligand, (b) is a single catalyst transition state with the model 3 ligand, (c) is a dual catalyst transition state with the model 1 ligand and one TMSCN molecule, and (d) is a dual catalyst transition state in combination with two TMSCN molecules (the red arrow indicates the position of the Cl counter-ion of the second catalyst to activate TMSCN).

with a reaction energy of 5.4 kcal mol^{-1}.

3.4.3.2 Step A

Before elucidating the production of the Al(imidate)-salen, two conformations of imide will be examined, as shown in figure 41. Figure 41a is the normal depic-
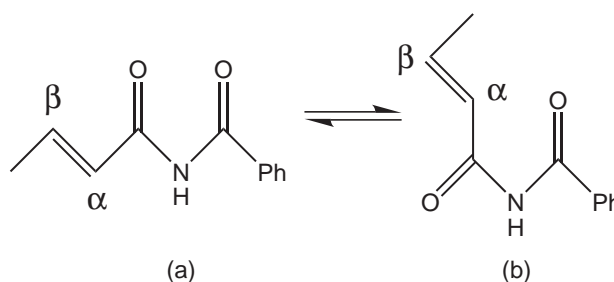

Figure 41: Two conformations for the imide molecule.

tion of imide where both oxygens are spatially close. However, figure 41b is favored

114

over figure 41a by 3.9 kcal mol^{-1}. While both conformations correspond to potential energy minima (no imaginary frequencies), figure 41b represents the global minimum.

Figure 42 shows the reaction profile for the production of Al(imidate)-salen and the corresponding transition state structures. The transition state, depicted by figure
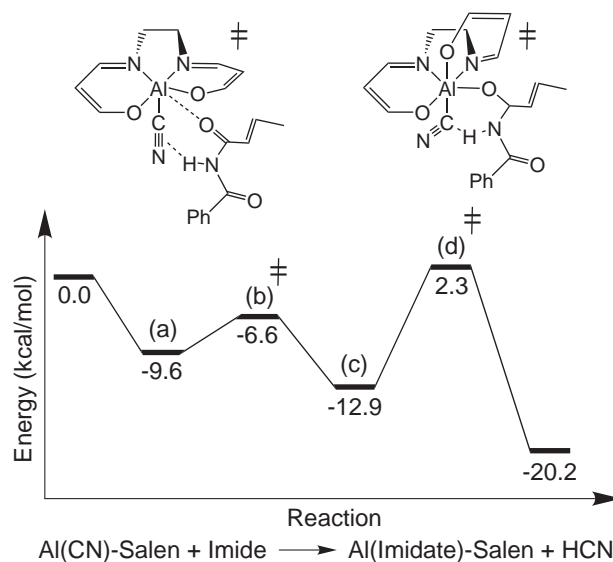

Figure 42: Reaction energy profile for the production of Al(imidate)-salen where (b) and (d) are the left and right transition state structures, respectively.

42b, illustrates the preferred positioning of the imide relative to the Al(CN)-salen catalyst. The chirality of the (R,R)salen-ligand favors the displacement of the 'right' side oxygen into the axial position, while conformation 41b of imide allows a simultaneous electrostatic attraction between the hydrogen (connected to the nitrogen of imide) to the cyanide counter-ion and the oxygen of imide to the aluminum center. In this case, imide is more than sufficient to activate the Al(CN)-salen; the energy of the transition state is only 2.3 kcal mol^{-1} above that of the reactants, as depicted by figure 42d (note the rotation of the cyanide counter-ion in the transition state). As stated, the production of Al(imidate)-salen with the *in situ* production of HCN is favorable by -20.2 kcal mol^{-1} (see figure 38 for the depiction of the products and the reactants).

3.4.3.3 Step B

Figure 43 shows the reaction profile for the addition of cyanide to the β position of imidate. The first three steps (structures a-c) produce the 2-propanol-Al(CN)-salen
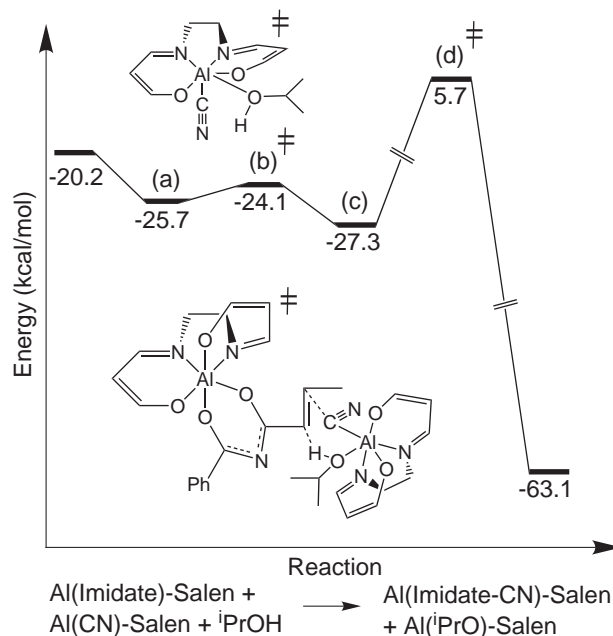

Figure 43: Reaction energy profile for the addition of cyanide to imidate where (b) and (d) are the top and bottom transition state structures, respectively.

coordinated complex (referred to as the second catalyst hereafter) with a small activation barrier of 1.6 kcal mol^{-1} depicted by figure 43b. The dual catalyst transition state structure is illustrated by figure 43d, where the hydrogen of the 2-propanol and the cyanide counter-ion are both activated in a single step thus leading to the addition of hydrogen to the α position and cyanide to the β position of imidate. The resulting reaction produces the S-adduct of imide-CN. In the course of this study, the transition state structure for the production of the R-adduct was also obtained. While this structure yields a similar activation energy to figure 43d, it was only possible by utilizing the model 1 salen ligand. Comparing both dual catalyst transition state structures, the R-adduct transition state structure suffers from large amounts

116

of steric strain caused by the cyclohexane backbone of the second catalyst and therefore, is much less favored than the more spatially unconstrained S-adduct transition state structure. The large 25.9 kcal mol^{-1} activation barrier is high relative to the intermediate structures of figure 43c; however, relative to the energy of the initial reactant, the energy of this transition state is only 5.7 kcal mol^{-1}. Hence, as long as none of the previous intermediates is so long-lived that the excess energy of the reaction is dissipated, this barrier should not pose a significant problem. Again, the energy of the cooperative catalysis transition state may be an upperbound due to the the truncation of the salen ligand. Overall, step B is the most favored step thermodynamically, with a reaction energy of -42.9 kcal mol^{-1} and the products of this step lying -63.1 kcal mol^{-1} below the initial reactants.

Experimentally substituting the β substituents of imide, Sammis and Jacobsen concluded that, whereas aliphatic substituents have little effect on the reaction, conjugated substituents render the system unreactive.[237] In attempts to rationalize this observation, figure 44 shows the atomic charges for two β substituted imides (in the Al(imidate)-salen form). The methyl β substituent generates atomic charges of -0.6
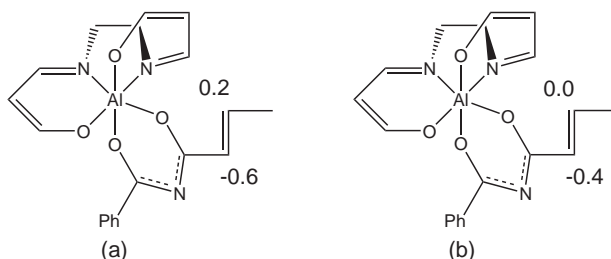
Figure 44: Atomic charges for the Al(imidate)-salen at the α and β positions where (a) is a methyl β substituent and (b) is a vinyl β substituent.

and 0.2 for the α and β position carbons, respectively. These charges correspond well with the attack of a positive hydrogen to the α position, and the attack of a negative cyanide counter-ion to the β position (as described above). The vinyl β substituent, however, generates a less negative α position carbon (-0.4) and a neutral β position

carbon. This change in atomic charges could sufficiently alter the electrostatic appeal of the imidate to the second catalyst thereby creating an activation barrier which is too large.

3.4.3.4   Step C

Figure 45 represents the reaction energy profile for the recovery of the imide-CN product through the reaction of Al(imidate-CN)-salen with 2-propanol. The first step
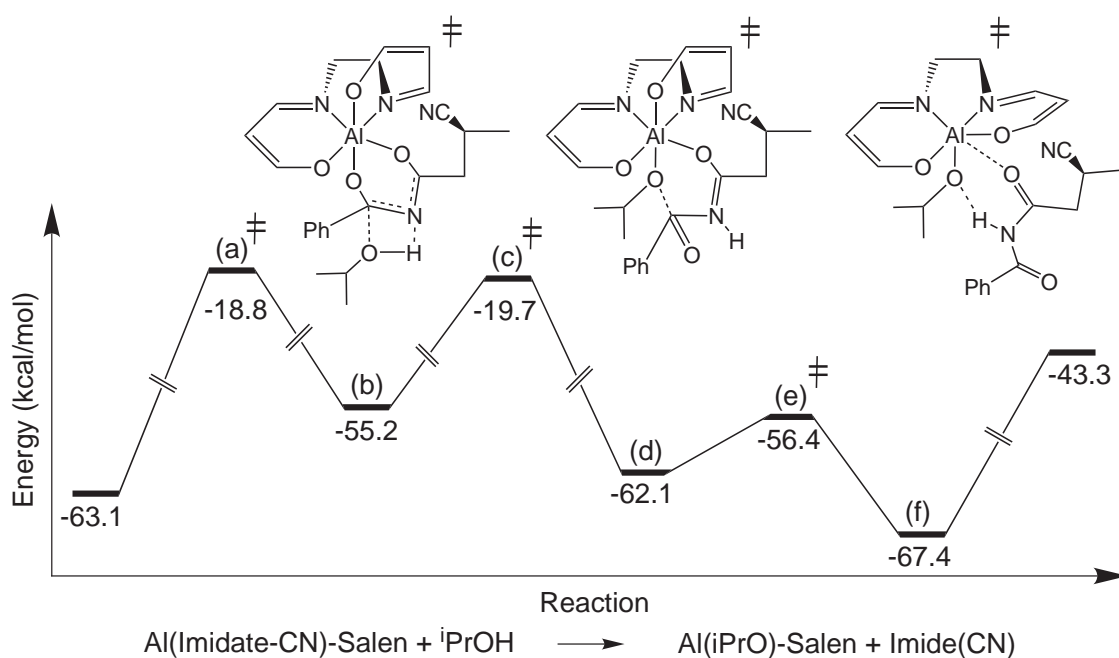
Figure 45: Reaction energy profile for the recovery of imide-CN where (a), (c), and (e) are the left, middle, and right transition state structures, respectively.

in this reaction is the simultaneous deposition of the hydrogen (from 2-propanol) to the nitrogen of imidate and the oxygen of 2-propanol to the carbon of imidate (see figure 45a). Figure 45c corresponds to the transition state where the oxygen of 2-propanol coordinates to the aluminum center and the uncoordinated oxygen of imide twists away from the aluminum center. The outcome of this twisting motion produces an intermediate structure that is highly favored due to an intramolecular hydrogen bond. Figure 45e illustrates the intramolecular hydrogen bond between the

118

nitrogen of imide and the oxygen of 2-propanol while the axial oxygen of the salen ligand reverts to the equatorial position. The final intermediate, figure 45e, is highly favorable due London dispersion forces, hydrogen bonding, and other electrostatic effects. Therefore, the recovery of the imide-CN product requires 24.2 kcal mol^{-1} of energy to overcome the noncovalent interactions for an overall endothermic process of 19.9 kcal mol^{-1}. The resulting reactants, however, are still favored with a net catalytic reaction energy of -43.3 kcal mol^{-1}.

3.4.3.5 Step D

Figure 46 represents the reaction energy profile for the recovery of the Al(CN)-salen catalyst from the Al(^{i}PrO)-salen intermediate and TMSCN. The activation barrier,
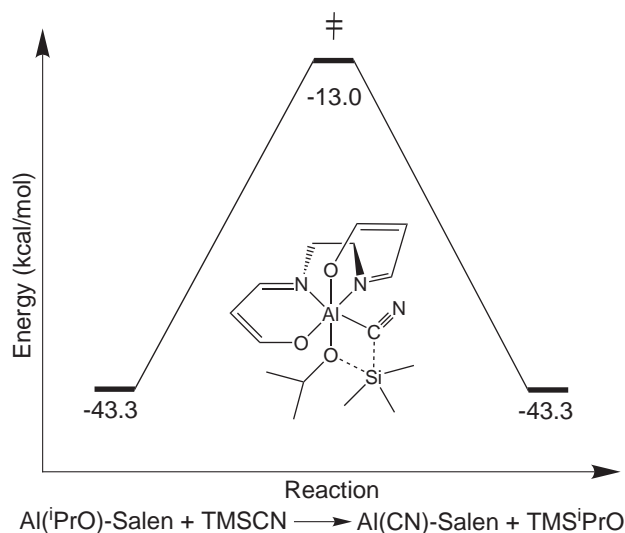
Figure 46: Reaction energy profile for the recovery of the Al(CN)-salen catalyst with the corresponding transition state structure.

like the production of Al(CN)-salen from Al(Cl)-salen, is quite large at 30.3 kcal mol^{-1}. However in this case, a single catalyst is sufficient to activate TMSCN due to the energy released from the previous catalytic reactions, and the barrier is -13.0 kcal mol^{-1} relative to the initial reactants. Again like the reaction of Al(Cl)-salen and TMSCN, step 4 may also be facilitated by cooperative catalysis. The resulting

reaction neither yields nor requires any net energy, and thus the final energy produced by the catalytic reaction is -43.3 kcal mol^{-1}.

3.4.4 Conclusions

Through the course of this study, certain experimental observations on enantioselectivity, the second-order rate dependence, and the unreactivity of conjugated β substituents were rationalized. The S-adduct is preferred over the R-adduct due to the chirality of the (R,R)salen-ligand, specifically the tendency of the 'right' oxygen of the salen ligand to be displaced into the axial position, and the steric effects during the dual activation transition state when cyanide is added to imide. The identification of the dual activation transition state, with an energy of only 5.7 kcal mol^{-1} with respect to the initial products, suggests that cooperative catalysis is indeed plausible and is consistent with the experimental findings that the reaction rate is second-order with respect to the catalyst. Finally, adding a conjugated β substituent reduces the atomic charge of the β carbon from 0.2 to neutral and, therefore, does not facilitate an attack of the cyanide at the β position of imide. It is expected that the insights gained from this detailed mechanistic study of the Al(Cl)-salen complex will aid future attempts to improve bimetallic metal-salen catalysis.

CHAPTER IV

CONCLUSION

As necessity dictates, current quantum chemistry methods are built upon layers of approximations for solving the molecular Schrödinger equation. Indeed, it is tempting to utilize only the most tractable quantum computational methods to extract results. However, in many situations, the most tractable methods do not generate even qualitatively correct results. In the computation of noncovalent interactions, both HF and standard DFT methods fail to predict favorable London dispersion forces. Moreover, utilizing post-HF methods only up two-body electron correlation, as in the MP2 and CCSD methods, dramatically overbinds and underbinds π-π interactions, respectively. For transition metal-ligand complexes, standard DFT methods do not consistently predict correct spin-state energy gaps and can even predict the incorrect ground state, as seen in the systematic studies on the metal-salen complexes. Even more dramatic are the artificially large correlation energies from MP2 and CCSD computations that can arise from the nondynamical electron correlation inherent in the transition metal-salen complexes.

Empirically scaling both standard wavefunction and DFT methods provides at least semi-quantitative approximations to benchmark quality CCSD(T) noncovalent interaction energies. Across potential energy curves for prototype systems (benzene and methane dimers, and methane-benzene and H_2S-benzene complexes), the SCS-MP2 and SCS-CCSD methods dramatically improve upon their canonical counterparts. In general, both methods incur at most only a few tenths of one kcal mol^{-1} error. The SCS-CCSD method consistently outperforms the SCS-MP2 method over all prototype PECs, single point energies for the intricate indole-benzene complex,

121

and the mean absolute deviation for the S22 test set. The SCS-MP2 method, however, is much less computationally demanding, especially when density fitting and dual-basis approaches are utilized. DFT methods can be improved by adding an empirical dispersion correction (DFT-D) and tends to perform just as well or better than new meta-GGA DFT methods for the prototype noncovalent PECs. DFT-D methods are advantageous because they are no more computationally demanding than standard DFT methods and still yield reasonable results.

In the case of nondynamical correlation effects, simply scaling standard methods will not yield improved results, as indicated by the seemingly random results obtained with the B3LYP and BP86 methods for the spin-state energies of transition metal-salen complexes. For the spin-state energy gaps of the d^4 and d^6 series of metal-salen complexes, both B3LYP and BP86 methods incur up to 40 kcal mol^{-1} error compared to the robust CASPT3 method. In some cases, DFT methods do not even obtain the correct ground state. This problem is not ameliorated by the new M0-family of meta-GGA functionals despite much advocacy for their application on metal-ligand systems. On the other hand, the geometries obtained by all DFT methods match CASSCF optimized geometries to, normally, no more than a couple tenths of an Å root mean squared deviation. This indicates that DFT methods may be reliable for geometry optimizations of metal-salen complexes, but that it must be used with extreme caution for energetics.

Although the CASPT3 method was very useful in providing reliable benchmark energies for evaluating the quality of various DFT methods for metal-salen complexes, it is too computationally demanding for use in detailed mechanistic studies. The problematic nondynamical correlation, however, originates from partially occupied d-orbitals of the transition-metal centers in the metal-salen complexes. Fortunately, the Al(III)-salen complex can also catalyze a variety of chemical reactions, but because an aluminum center does not contain partially occupied d-orbitals, single reference

methods can properly describe its electronic structure. Due to the significant degree of π-aromaticity of the salen-ligand, DFT-D optimizations with subsequent SCS-MP2 single points were obtained for the mechanistic study on the conjugate addition of cyanide to imide catalyzed by the Al(Cl)-salen complex. The proposed mechanism supports the production of the S-adduct over the R-adduct, and a process of cooperative catalysis. Experimentally, the Al(Cl)-salen catalyst is highly enantioselective favoring the S-adduct, and a second-order rate dependence on the Al(Cl)-salen is observed. Only with these new empirically scaled quantum chemistry methods and a good understanding of the Al(Cl)-salen electronic structure was a reliable reaction energy profile generated.

For future mechanistic investigations of the nucleophilic addition to unsaturated imides catalyzed by the Al(Cl)-salen, it would be interesting to vary the nucleophile to predict whether other additions to the β position of imide are possible. Since the dual activation transition state seems to be the rate determining step, this investigation may only require computations of this transition state with different nucleophiles. It would also be interesting to utilize a larger model salen (or the full salen ligand) to understand the effect of various molecular supports. For mechanistic investigations concerning transition metal-salen catalysts, an appropriate DFT method should be chosen for each specific system, if one exists. However, since the metal-salen catalysts often contain counter-ions, it is necessary to understand the effects of the counter-ion to the nature of the wavefunction, *i.e.* the amount of nondynamical correlation. An important future study would be to obtain benchmark CASPT3 spin-state energy gaps for the Co(III)-salen complex with various counter-ions (Cl$^-$, OH$^-$, MeO$^-$, acetate, etc.) to test DFT methods. If a particular DFT method is found to perform reliably, then a subsequent study would be to apply that DFT functional to the epoxide ring opening reaction catalyzed by the Co(III)-salen complex (which includes a variety of counter-ions), while taking measures to appropriately capture London

dispersion forces, most likely in the form of a DFT-D method.

Bibliography

[1] Takatani, T.; Sherrill, C. D. *Phys. Chem. Chem. Phys.* **2007**, *9*, 6106–6114.

[2] Takatani, T.; Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2008**, *128*, 124111.

[3] Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. *J. Phys. Chem. A* **2009**, *113*, 10146–10159.

[4] Geng, Y.; Takatani, T.; Hohenstein, E. G.; Sherrill, C. D. *J. Phys. Chem. A* **2010**, *114*, 3576–3582.

[5] Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *132*, 144104.

[6] Takatani, T.; Sears, J. S.; Sherrill, C. D. *J. Phys. Chem. A* **2009**, *113*, 9231–9236.

[7] Takatani, T.; Sears, J. S.; Sherrill, C. D. *J. Phys. Chem. A* **2010**, *Submitted.*

[8] Takatani, T.; Sherrill, C. D. **2010**, *In Preparation.*

[9] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory;* McGraw-Hill: New York, 1989.

[10] Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910–1918.

[11] Rittby, M.; Bartlett, R. J. *J. Phys. Chem.* **1988**, *92*, 3033.

[12] Scuseria, G. E.; Scheiner, A. C.; Lee, T. J.; Rice, J. E.; Schaefer, H. F. *J. Chem. Phys.* **1987**, *86*, 2881.

[13] Scuseria, G. E.; Janssen, C. L.; Schaefer, H. F. *J. Chem. Phys.* **1988**, *89*, 7382.

[14] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

[15] Lee, T. J.; Scuseria, G. E. In *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy;* Langhoff, S. R., Ed.; Kluwer Academic Publishers: Dordrecht, 1995; pages 47–108.

[16] Schmidt, M. W.; Gordon, M. S. *Annu. Rev. Phys. Chem.* **1998**, *98*, 233–266.

[17] Abrams, M. L.; Sherrill, C. D. *J. Chem. Phys.* **2003**, *118*, 1604–1609.

[18] Abrams, M. L.; Sherrill, C. D. *J. Chem. Phys.* **2004**, *121*(19), 9211–9219.

[19] Hinze, J. *J. Chem. Phys.* **1973**, *59*, 6424–6432.

[20] Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. *Chem. Phys.* **1980**, *48*, 157.

[21] Celani, P.; Werner, H.-J. *J. Chem. Phys.* **2000**, *112*(13), 5546–5557.

[22] Andersson, K.; Malmqvist, P.-Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

[23] Werner, H. J. *Mol. Phys.* **1996**, *89*, 645–661.

[24] Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, *136*, B864.

[25] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, 1133.

[26] Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974–12980.

[27] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

[28] Boese, A. D.; Martin, J. *J. Chem. Phys.* **2004**, *121*, 3405–3416.

[29] Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103.

[30] Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

[31] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

[32] Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23–28.

[33] Hunter, C. A.; Singh, J.; Thornton, J. M. *J. Mol. Bio.* **1991**, *218*, 837–846.

[34] Zimm, B. H. *J. Chem. Phys.* **1960**, *33*, 1349–1356.

[35] Muraki, M. *Protein Peptide Lett.* **2002**, *9*, 195–209.

[36] Morgan, R. S.; Tatsch, C. E.; Gushard, R. H.; Mcadon, J. M. *Int. J. Pept. Protein Res.* **1978**, *11*, 209–217.

[37] Askew, B.; Ballester, P.; Buhr, C.; Jeong, K. S.; Jones, S.; Parris, K.; Williams, K.; Rebek, J. *J. Am. Chem. Soc.* **1989**, *111*, 1082–1090.

[38] Claessens, C. G.; Stoddart, J. F. *J. Phys. Org. Chem.* **1997**, *10*, 254–272.

[39] Dahl, T. *Acta Chem. Scand.* **1994**, *48*, 95–106.

[40] Lerman, L. S. *J. Mol. Biol.* **1961**, *3*, 18.

[41] Smithrud, D. B.; Diederich, F. *J. Am. Chem. Soc.* **1990**, *112*, 339–343.

[42] Watanabe, T.; Suzuki, T.; Umezawa, Y.; Takeuchi, T.; Otsuka, M.; Umezawa, K. *Tetrahedron* **2000**, *56*, 741–752.

[43] DiStasio, R. A.; vonHelden, G.; Steele, R. P.; Head-Gordon, M. *Chem. Phys. Lett.* **2007**, *437*, 277–283.

[44] Gräfenstein, J.; Cremer, D. *J. Chem. Phys.* **2009**, *130*, 124105.

[45] Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Am. Chem. Soc.* **1994**, *116*, 3500–3506.

[46] Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790–18794.

[47] Lee, E. C.; Kim, D.; Jurečka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. *J. Phys. Chem. A* **2007**, *111*, 3446–3457.

[48] Pitoňák, M.; Neogrády, P.; Řezáč, J.; Jurečka, P.; Urban, M.; Hobza, P. *J. Chem. Theory Comput.* **2008**, *4*, 1829–1834.

[49] Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A* **2006**, *110*, 10345–10354.

[50] Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887–10893.

[51] Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*(46), 10200–10207.

[52] Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*(1), 104–112.

[53] Hopkins, B. W.; Tschumper, G. S. *J. Phys. Chem. A* **2004**, *108*(15), 2941–2948.

[54] Grimme, S. *J. Chem. Phys.* **2003**, *118*(20), 9095–9102.

[55] Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

[56] Molpro, version 2006.1, a package of ab initio programs. Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. **2006**.

[57] Gauss, J.; Stanton, J. F. *J. Phys. Chem. A* **2000**, *104*, 2865–2868.

[58] Ringer, A. L.; Figgs, M. S.; Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10822–10828.

[59] Edwards, T. H.; Moncur, N. K.; Snyder, L. E. *J. Chem. Phys* **1967**, *46*, 2139–2142.

[60] Weigend, F. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

[61] Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Ringnalda, M. N.; Pollard, W. T.; Dunietz, B. D.; Cao, Y. *J. Phys. Chem. A* **1999**, *103*, 1913–1928.

[62] Martinez, T. J.; Carter, E. A. In *Modern Electronic Structure Theory*; Yarkony, D. R., Ed., Vol. 2 of *Advanced Series in Physical Chemistry;* World Scientific: Singapore, 1995; pages 1132–1165.

[63] Vahtras, O.; Almlöf, J.; Feyereisen.; M, W. *Chem. Phys. Lett.* **1993**, *213*, 514–518.

[64] Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359–363.

[65] Werner, H.-J.; Manby, F. R.; Knowles, P. J. *J. Chem. Phys.* **2003**, *118*(18), 8149–8160.

[66] Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

[67] Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

[68] Vahtras, O.; Almlöf, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514–518.

[69] Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175–3183.

[70] Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*(4), 553–566.

[71] Halkier, A.; Klopper, W.; Helgaker, T.; Jørgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 9157–9167.

[72] Hill, J. G.; Platts, J. A. *J. Chem. Theory Comput.* **2007**, *3*, 80–85.

[73] Distasio, R. A.; Head-Gordon, M. *Mol. Phys.* **2007**, *105*, 1073–1083.

[74] Pulay, P.; Saebø, S. *Theor. Chim. Acta* **1986**, *69*, 357–368.

[75] Saebø, S.; Pulay, P. *Chem. Phys. Lett.* **1985**, *113*, 13–18.

[76] Saebø, S.; Pulay, P. *J. Chem. Phys.* **1988**, *88*, 1884–1890.

[77] Hetzer, G.; Pulay, P.; Werner, H.-J. *Chem. Phys. Lett.* **1998**, *290*, 143–149.

[78] Lee, M. S.; Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **2000**, *112*(8), 3592–3601.

[79] Maslen, P. E.; Head-Gordon, M. *Chem. Phys. Lett.* **1998**, *283*, 102–108.

[80] Schutz, M. *Phys. Chem. Chem. Phys.* **2002**, *4*(16), 3941–3947.

[81] Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330–8343.

[82] Venkatnathan, A.; Szilva, A. B.; Walter, D.; Gdanitz, R. J.; Carter, E. A. *J. Chem. Phys.* **2004**, *120*(4), 1693–1704.

[83] Hill, J. G.; Platts, J. A.; Werner, H. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4072–4078.

[84] Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*(9), 4916–4926.

[85] Boughton, J. W.; Pulay, P. *J. Comput. Chem.* **1993**, *14*, 736–740.

[86] Saebø, S.; Tong, W.; Pulay, P. *J. Chem. Phys.* **1993**, *98*(3), 2170–2176.

[87] Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*(20), 9793–9802.

[88] Antony, J.; Grimme, S. *J. Phys. Chem. A* **2007**, *111*, 4862–4868.

[89] Saebø, S.; Pulay, P. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213–236.

[90] Schütz, M.; Rauhut, G.; Werner, H. *J. Phys. Chem. A* **1998**, *102*, 5997–6003.

[91] Schütz, M.; Hetzer, G.; Werner, H. *J. Chem. Phys.* **1999**, *111*, 5691–5705.

[92] Tauer, T. P.; Derrick, M. E.; Sherrill, C. D. *J. Phys. Chem. A* **2005**, *109*(1), 191–196.

[93] Williams, H. L.; Szalewicz, K.; Jeziorski, B.; Moszynski, R.; Rybak, S. *J. Chem. Phys.* **1993**, *98*, 1279–1292.

[94] SAPT2006: An ab initio program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies. see: http://www.physics.udel.edu/~szalewic/SAPT. Bukowski, R.; Cencek, W.; Jankowski, P.; Jeziorski, B.; Jeziorska, M.; Kucharski, S. A.; Lotrich, V. F.; Misquitta, A. J.; Moszynski, R.; Patkowski, K.; Rybak, S.; Szalewicz, K.; Williams, H. L.; Wheatley, R. J.; Wormer, P. E. S.; Zuchowski, P. S.

[95] Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887–1930.

[96] Szabados, A. *J. Chem. Phys.* **2006**, *125*, 214105.

[97] Šponer, J.; Jurečka, P.; Marchan, I.; Juque, F. J.; Orozco, M.; Hobza, P. *Chem. Eur. J.* **2006**, *12*, 2854–2865.

[98] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[99] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

[100] Crawford, T. D.; Sherrill, C. D.; Valeev, E. F.; Fermann, J. T.; King, R. A.; Leininger, M. L.; Brown, S. T.; Janssen, C. L.; Seidl, E. T.; Kenny, J. P.; Allen, W. D. *J. Comput. Chem.* **2007**, *28*, 1610–1616.

[101] Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656–10668.

[102] Rendell, A. P.; Lee, T. J. *J. Chem. Phys.* **1994**, *101*, 400–408.

[103] Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1996**, *104*, 6286–6297.

[104] Schütz, M.; Werner, H.-J. *J. Chem. Phys.* **2001**, *114*, 661–681.

[105] Schütz, M. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3914–3947.

[106] Schütz, M.; Manby, F. R. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3349–3358.

[107] Ringer, A. L.; Senenko, A.; Sherrill, C. D. *Protein Sci.* **2007**, *16*, 2216–2223.

[108] Elsohly, A. M.; Tschumper, G. S. *Int. J. Quantum Chem.* **2009**, *109*, 91.

[109] Janowski, T.; Pulay, P. *Chem. Phys. Lett.* **2007**, *447*, 27–32.

[110] Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; Jr., R. A. D.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Voorhis, T. V.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

[111] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

[112] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[113] Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364–382.

[114] Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett.* **2004**, *394*, 334–338.

[115] Johnson, E. R.; Becke, A. D.; Sherrill, C. D.; DiLabio, G. A. *J. Chem. Phys.* **2009**, *131*, 034111.

[116] Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

[117] Grimme, S. *J. Comput. Chem.* **2006**, *27*(15), 1787–1799.

[118] Steele, R. P.; DiStasio, R. A.; Shao, Y.; Kong, J.; Head-Gordon, M. *J. Chem. Phys.* **2006**, *125*, 074108.

[119] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *123*, 154101.

[120] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2006**, *124*, 014104.

[121] Heßelmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 014103.

[122] Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*, 024101.

[123] Podeszwa, R.; Szalewicz, K. *Chem. Phys. Lett.* **2005**, *412*, 488–493.

[124] Puzder, A.; Dion, M.; Langreth, D. C. *J. Chem. Phys.* **2006**, *124*, 164105.

[125] von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*(15), 153004.

[126] von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. B* **2005**, *71*, 195119.

[127] Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289–300.

[128] Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525–5534.

[129] Hunter, C. A.; Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 837–846.

[130] Razmutė-Razmė, I.; Kuodis, Z.; Eicher-Lorka, O.; Niaura, G. *Chemija* **2007**, *18*, 16–20.

[131] Volbeda, A.; Garcin, E.; Piras, C.; de Lacey, A. L.; Fernandez, V. M.; Hatchikian, E. C.; Frey, M.; Fontecilla-Camps, J. C. *J. Am. Chem. Soc.* **1996**, *118*, 12989–12996.

[132] Braun, J.; Neusser, H. J.; Hobza, P. *J. Phys. Chem. A* **2003**, *107*, 3918–3924.

[133] Pavone, M.; Rega, N.; Barone, V. *Chem. Phys. Lett.* **2008**, *452*, 333–339.

[134] Spartan '08. Wavefunction, Inc.. **2008**.

[135] Jaguar, version 5.5. Schrödinger, L. L. C. **1991-2003**.

[136] Hättig, C. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59–66.

[137] manuscript in preparation. Hohenstein, E. G.; Sherrill, C. D.

[138] Hohenstein, E. G.; Sherrill, C. D. *J. Phys. Chem. A* **2009**, *113*, 878–886.

[139] Börnsen, K. O.; Selzle, H. L.; Schlag, E. W. *J. Chem. Phys.* **1986**, *85*, 1726–1732.

[140] Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802–11808.

[141] Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28*, 555–569.

[142] Benighaus, T.; DiStasio, R. A.; Lochan, R. C.; Chai, J.; Head-Gordon, M. *J. Phys. Chem. A* **2008**, *112*, 2702–2712.

[143] Goll, E.; Leininger, T.; Manby, F. R.; Mitrushchenkov, A.; Werner, H.; Stoll, H. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3353–3357.

[144] Marchetti, O.; Werner, H. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3400–3409.

[145] Marchetti, O.; Werner, H. *J. Phys. Chem. A* **2009**, *113*, 11580–11585.

[146] McNamara, J. P.; Hillier, I. H. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2362.

[147] Morgado, C.; Vincent, M. A.; Hillier, I. H.; Shan, X. *Phys. Chem. Chem. Phys.* **2007**, *9*, 448–451.

[148] Riley, K. E.; Hobza, P. *J. Phys. Chem. A* **2007**, *111*, 8257–8263.

[149] Sato, T.; Tsuneda, T.; Hirao, K. *J. Chem. Phys.* **2007**, *126*, 234114.

[150] Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. *Chem. Phys. Chem.* **2009**, *10*, 282–289.

[151] Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397–3406.

[152] Molnar, L. F.; He, X.; Wang, B.; Merz, K. M. *J. Chem. Phys.* **2009**, *131*, 065102.

[153] Pitoňák, M.; Janowski, T.; Neogrády, P.; Pulay, P.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1761–1766.

[154] Pitoňák, M.; Riley, K.; Neogrády, P.; Hobza, P. *Chem. Phys. Chem.* **2008**, *9*, 1636–1644.

[155] NWChem, A Computational Chemistry Package for Parallel Computers, version 5.1 (2007), Pacific Northwest National Laboratory, Richland, Washington, 99352-0999, USA. Bylaska, E. J.; de Jong, W. A.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; Wang, D.; Apra, E.; Windus, T. L.; Hammond, J.; Nichols, P.; Hirata, S.; Hackler, M. T.; Zhao, Y.; Fan, P.-D.; Harrison, R. J.; Dupuis, M.; Smith, D. M. A.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Wu, Q.; Voorhis, T. V.; Auer, A. A.; Nooijen, M.; Brown, E.; Cisneros, G.; Fann, G. I.; Fruchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, J. A.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Pollack, L.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z.

[156] Kendall, R. A.; Apra, E.; Bernholdt, D. E.; Bylaska, E. J.; Dupuis, M.; Fann, G. I.; Harrison, R. J.; Ju, J. L.; Nichols, J. A.; Nieplocha, J.; Straatsma, T. P.; Windus, T. L.; Wong, A. T. *Comput. Phys. Commun.* **2000**, *128*, 260–283.

[157] Blaser, H. U.; Spindler, F.; Studer, A. *Appl. Catal., A* **2001**, *221*, 119–143.

[158] Katsuki, T. *Adv. Synth. Catal.* **2002**, *344*, 131–147.

[159] Noyori, R. *Adv. Synth. Catal.* **2003**, *345*, 15–32.

[160] Yoon, T. P.; Jacobsen, E. N. *Science* **2003**, *299*, 1691–1693.

[161] Canali, L.; Sherrington, D. C. *Chem. Soc. Rev.* **1999**, *28*, 85–93.

[162] McGarrigle, E. M.; Gilheany, D. G. *Chem. Rev.* **2005**, *105*, 1563–1602.

[163] Venkataramanan, N. S.; Kuppuraj, G.; Rajagopal, S. *Coord. Chem. Rev.* **2005**, *249*, 12491268.

[164] Zheng, X. L.; Jones, C. W.; Weck, M. *Chem. Eur. J.* **2005**, *12*, 576–583.

[165] Zheng, X. L.; Jones, C. W.; Weck, M. *J. Am. Chem. Soc.* **2007**, *129*, 1105–1112.

[166] Abashkin, Y. G.; Burt, S. K. *Org. Lett.* **2004**, *6*, 59–62.

[167] Abashkin, Y. G.; Collins, J. R.; Burt, S. K. *Inorg. Chem.* **2001**, *40*, 4040–4048.

[168] Cavallo, L.; Jacobsen, H. *Angew. Chem. Int. Ed.* **2000**, *39*, 589–592.

[169] Cavallo, L.; Jacobsen, H. *J. Org. Chem.* **2003**, *68*, 6202–6207.

[170] Cavallo, L.; Jacobsen, H. *J. Phys. Chem. A* **2003**, *107*, 5466.

[171] Cavallo, L.; Jacobsen, H. *Inorg. Chem.* **2004**, *43*, 2175–2182.

[172] Ivanic, J. *J. Chem. Phys.* **2003**, *119*, 9377–9385.

[173] Jacobsen, H.; Cavallo, L. *Chem. Eur. J.* **2001**, *7*, 800–807.

[174] Jacobsen, H.; Cavallo, L. *Phys. Chem. Chem. Phys.* **2004**, *6*, 3747–3753.

[175] Khavrutskii, I. V.; Musaev, D. G.; Morokuma, K. *J. Am. Chem. Soc.* **2003**, *125*, 13879–13889.

[176] Khavrutskii, I. V.; Rahim, R. R.; Musaev, D. G.; Morokuma, K. *J. Phys. Chem. B* **2004**, *108*, 3845–3854.

[177] Khavrutskii, I. V.; Musaev, D. G.; Morokuma, K. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5743–5748.

[178] Khavrutskii, I. V.; Musaev, D. G.; Morokuma, K. *Inorg. Chem.* **2005**, *44*, 306–315.

[179] Linde, C.; Arnold, M.; Norrby, P. O.; Akermark, B. *Angew. Chem. Int. Ed.* **1997**, *36*, 1723–1725.

[180] Linde, C.; Arnold, M.; Norrby, P. O.; Akermark, B.; Svensson, M. *J. Am. Chem. Soc.* **1999**, *121*, 5083–5084.

[181] Sears, J. S.; Sherrill, C. D. *J. Chem. Phys.* **2006**, *124*, 144314.

[182] Strassner, T.; Houk, K. N. *Org. Lett.* **1999**, *1*, 419–421.

[183] Ivanic, J.; Collins, J. R.; Burt, S. K. *J. Phys. Chem. A* **2004**, *108*, 2314–2323.

[184] Sears, J. S.; Sherrill, C. D. *J. Phys. Chem. A* **2008**, *112*(15), 3466–3477.

[185] Sears, J. S.; Sherrill, C. D. *J. Phys. Chem. A* **2008**, *112*, 6741–6752.

[186] Anthon, C.; Bendix, J.; Schäffer, C. E. *Inorg. Chem.* **2004**, *43*, 7882–7886.

[187] Chermette, H. *Coord. Chem. Rev.* **1998**, *178*, 699–721.

[188] Petrie, S.; Stranger, R. *Inorg. Chem.* **2003**, *43*, 2597–2610.

[189] Schäffer, C. E.; Anthon, C.; Bendix, J. *Coord. Chem. Rev.* **2009**, *253*, 575–593.

[190] Vlček, A.; Zàliš, S. *Coord. Chem. Rev.* **2007**, *251*, 258–187.

[191] Barden, C. J.; Rienstra-Kiracofe, J. C.; Schaefer, H. F. *J. Chem. Phys.* **2000**, *113*, 690–700.

[192] Ghosh, A. *J. Biol. Inorg. Chem.* **2006**, *11*, 712–724.

[193] Harvey, J. N. *Struct. and Bonding* **2004**, *121*, 151–183.

[194] Nakao, Y.; Hirao, K.; Taketsugu, T. *J. Chem. Phys.* **2001**, *114*, 7935–7940.

[195] Yanagisawa, S.; Tsuneda, T.; Hirao, K. *J. Chem. Phys.* **2000**, *112*, 545–553.

[196] Quintal, M. M.; Karton, A.; Iron, M. A.; Boese, A. D.; Martin, J. M. L. *J. Phys. Chem. A* **2006**, *110*, 709–716.

[197] Wang, S. G.; Schwarz, W. H. E. *J. Chem. Phys.* **1998**, *109*, 7252–7262.

[198] Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *124*, 224105.

[199] Schmidt, M. W.; Gordon, M. S. *Ann. Rev. Phys. Chem.* **1998**, *49*, 233–266.

[200] Sherrill, C. D.; Schaefer, H. F. *Adv. Quantum Chem.* **1999**, *34*, 143–269.

[201] Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657–2664.

[202] Hehre, W. J.; Ditchfie, R.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1970**, *52*, 2769.

[203] Pietro, W. J.; Hehre, W. J. *J. Comp. Chem.* **1983**, *4*, 241–251.

[204] Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics Modell.* **1996**, *14*, 33.

[205] Portmann, S.; Luthi, H. P. *Chimia* **2000**, *54*, 766–770.

[206] Cozzi, P. G. *Chem. Soc. Rev.* **2004**, *33*, 410–421.

[207] Larrow, J. F.; Hemberger, K. E.; Jasmin, S.; Kabir, H.; Morel, P. *Tetrahedron Asymm* **2003**, *14*, 3589–3592.

[208] Madhavan, N.; Jones, C. W.; Weck, M. *Accounts of Chemical Research* **2008**, *41*, 1153–1165.

[209] Schaus, S. E.; Brandes, B. D.; Larrow, J. F.; Tokunaga, M.; Hansen, K. B.; Gould, A. E.; Furrow, M. E.; Jacobsen, E. N. *J. Am. Chem. Soc.* **2002**, *124*, 1307–1315.

[210] Gill, C. S.; Venkatasubbaiah, K.; Phan, N. T. S.; Weck, M.; Jones, C. W. *Chem. Eur. J.* **2008**, *14*, 7306–7313.

[211] Jain, S.; Zheng, X. L.; Jones, C. W.; Weck, M.; Davis, R. J. *Inorg. Chem.* **2007**, *21*, 8887–8896.

[212] Zheng, X. L.; Jones, C. W.; Weck, M. *Advanced Synthesis and Catalysis* **2008**, *350*, 255–261.

[213] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

[214] Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824.

[215] Hay, P. J.; Wadt, W. R. *J. Chem. Phys* **1985**, *82*, 270–283.

[216] Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654–3665.

[217] Chasman, D.; Beachy, M. D.; Wang, L. M.; Friesner, R. A. *J. Comp. Chem.* **1998**, *19*, 1017–1029.

[218] Cicchi, S.; Cardona, F.; Brandi, A.; Corsi, M.; Goti, A. *Tetrahedron Letters* **1999**, *40*, 1989–1992.

[219] Kokubo, C.; Katsuki, T. *Tetrahedron* **1996**, *52*, 13895–13900.

[220] Lee, N. H.; Lee, C. S.; Jung, D. S. *Tetrahedron Letters* **1998**, *39*, 1385–1388.

[221] Palucki, M.; Hanson, P.; Jacobsen, E. N. *Tetrahedron Letters* **1992**, *33*, 7111–7114.

[222] Brandes, B. D.; Jacobsen, E. N. *J. Org. Chem.* **1994**, *59*, 4378–4380.

[223] Jacobsen, E. N.; Deng, L.; Furukawa, Y.; Martinez, L. E. *Tetrahedron* **1994**, *50*, 4323–4334.

[224] Katsuki, T. *J. Molecular Catalysis A-Chemical* **1996**, *1-2*, 87–107.

[225] Lee, N. H.; Jacobsen, E. N. *Tetrahedron Letters* **1991**, *32*, 6533–6536.

[226] Palucki, M.; Finney, N. S.; Pospisil, P. J.; Guler, M. L.; Ishida, T.; Jacobsen, E. N. *J. Am. Chem. Soc.* **1998**, *5*, 948–954.

[227] Pietikainen, P. *Tetrahedron Letters* **1994**, *35*, 941–944.

[228] Yamada, T.; Imagawa, K.; Nagata, T.; Mukaiyama, T. *Chem. Lett.* **1992**, *11*, 2231–2234.

[229] Zhang, W.; Jacobsen, E. N. *J. Org. Chem.* **1991**, *56*, 2296–2298.

[230] Venkatasubbaiah, K.; Gill, C. S.; Takatani, T.; Sherrill, C. D.; Jones, C. W. *Chem. Eur. J.* **2009**, *15*, 3951–3955.

[231] Goodman, S. N.; Jacobsen, E. N. *Adv. Synth. Catal.* **2002**, *344*, 953–956.

[232] Madhavan, N.; Takatani, T.; Sherrill, C. D.; Weck, M. *Chem. Eur. J.* **2009**, *15*, 1186–1194.

[233] Madhavan, N.; Weck, M. *Adv. Synth. Catal.* **2008**, *350*, 419–425.

[234] Mazet, C.; Jacobsen, E. N. *Angew. Chem. Int. Ed.* **2008**, *47*, 1762–1765.

[235] Myers, J. K.; Jacobsen, E. N. *J. Am. Chem. Soc.* **1999**, *121*, 8959–8960.

[236] North, M.; Villuendas, P.; Williamson, C. *Tetrahedron* **2010**, *66*, 1915–1924.

[237] Sammis, G. M.; Jacobsen, E. N. *J. Am. Chem. Soc.* **2003**, *125*, 4442–4443.

[238] Sammis, G. M.; Danjo, H.; Jacobsen, E. N. *J. Am. Chem. Soc.* **2004**, *126*, 9928–9929.

[239] Cavallo, L.; Jacobsen, H. *Inorg. Chem.* **2004**, *43*, 2175–2182.

[240] Jacobsen, H.; Cavallo, L. *Phys. Chem. Chem. Phys.* **2004**, *6*, 3747–3753.

[241] Sun, K.; Li, W.; Feng, Z.; Li, C. *Chem. Phys. Lett.* **2009**, *470*, 259–263.

[242] Piacenza, M.; Hyla-Kryspin, I.; Grimme, S. *J. Comput. Chem.* **2007**, *28*, 2275–2285.

[243] Schwabe, T.; Grimme, S. *J. Phys. Chem. A* **2009**, *113*, 3005–3008.

[244] Kirkwood, J. G. *J. Chem. Phys.* **1939**, *7*, 908–911.

[245] Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.