Subchemical Accuracy in Combustion Thermochemistry via Ab Initio Methods: From Basis Set Extrapolations to Explicitly-Correlated Wavefunctions

by

Jeremiah J. Wilke

(Under the direction of Henry F. Schaefer III)

Abstract

Wavefunction-based methods in quantum chemistry provide simultaneously the most accurate, but also most expensive method for computing molecular properties. In particular, the realm of "subchemical" accuracy better than 1.0 kcal mol^{-1} has now become feasible even for difficult combustion intermediates. However, accurate methods must treat electron correlation, leading to problems which explode in cost as N^6 or worse in the number of atoms. More compact representations of the wavefunction are necessary to further increase the scope and accuracy. Explicitly-correlated methods have recently emerged which mathematically represent the wavefunction directly in the interelectronic distance r_{12}, providing a much more natural representation of the electron correlation. Herein we develop a series of explicitly-correlated (R12) methods suitable for open-shell radicals in combustion problems. These R12 methods are compared to existing methodologies such as the focal point approach (FPA), demonstrating their utility in combustion thermochemistry.

Index words: coupled cluster theory, Møller-Plesset perturbation theory, basis set extrapolation, R12 methods, explicitly-correlated methods, computational thermochemistry, focal point analysis

Subchemical Accuracy in Combustion Thermochemistry via Ab Initio Methods: From Basis Set Extrapolations to Explicitly-Correlated Wavefunctions

by

Jeremiah J. Wilke

B.S., University of Wisconsin-Madison, 2006

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2011

Subchemical Accuracy in Combustion Thermochemistry via Ab Initio Methods: From Basis Set Extrapolations to Explicitly-Correlated Wavefunctions

by

Jeremiah J. Wilke

Major Professor:	Henry F. Schaefer III
Committee:	Michael A. Duncan Nigel Adams

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2011

DEDICATION

"You don't understand the humiliation of it. To be tricked out of the single assumption that makes our existence bearable, that somebody is watching - we are actors! We are the opposite of people."
-Tom Stoppard, Rosencrantz and Guildenstern Are Dead

"You don't understand the humiliation of it. To be tricked out of the single assumption that makes our existence bearable, that somebody will read our dissertation - we are graduate students! We are the opposite of people."

This dissertation is dedicated to my parents and closest friends whose constant support made this possible and also to "Nanners" who helped me remember why I started the whole thing in the first place. Above all, though, this is for my sister. Most of who I am today is because of you. I'm sorry this dissertation isn't the masterpiece of neurochemistry you asked for 12 years ago. I was never very good in the lab, and even in theoretical chemistry the universe is not so quick to give up its secrets. We'll figure it out eventually, though.

ACKNOWLEDGMENTS

To my adviser, Professor Schaefer, your guidance and encouragement will never be forgotten. You have invested an enormous amount of both time and money in developing me as a scientist, and I hope that as I progress forward I can return that investment in some way. You always tolerated my cynicism and negativity, and managed to keep me at least a little grounded and optimistic. Many thanks to my committee members Nigel Adams and Michael Duncan for tolerating both my missed deadlines and small figure captions.

Academically, David Sherrill, Ed Valeev, and Daniel Crawford will always be appreciated for their advice and input, and Curt Janssen should be especially thanked for tolerating my incessant questions on MPQC for an entire summer. Wim Klopper and David Tew were absolutely essential teachers in helping me along with R12 methods, and I hope the Legos are sufficient compensation for David for all of his help. The input and financial help from Jürgen Gauss, Werner Kutzelnigg, and Andreas Köhn in visiting Deutschland will also always be appreciated.

At CCQC, I should probably thank people specifically. Andy and Jet's never-ending help and advice will always be appreciated along with their rides to the grocery store during my unfortunate carless times. Lucas was always a pleasant and quiet office mate, and managed not to rub it in too much when the Colts beat the Bears. Frank was both a great roommate and Tecmo opponent. Jacqueline was the perfect listener during my sports-venting sessions. Qianyi, Katie, Aaron, and Beulah were excellent badminton partners. I'll always feel bad I couldn't help Bryson more with the lanthanides, and appreciate his tolerance of my seemingly endless drama. Though we haven't overlapped much, I have also enjoyed my time with Martin, Partha, Stefan, David, Alex, Shane, Jay, and Jacob. For Steven, Francesco, and Heather, your brilliance was at times both enjoyable and frustrating to measure up to, and I'm glad we have had the chance to interact scientifically.

And regardless of my publication record, I can leave the CCQC happy knowing Maria and I got to the high Text Twist score. In Karlsruhe, I should thank Florian and Angela for tolerating my terrible German. For Nedko, I should simply say "B-V-B!"

For my friend Annie, our scheduled phone calls were always something to look forward to. For Rachel, Laura, Jenny, Maggie, and Nicole, you were always there to encourage me during my most difficult times. For Tim, I don't want to live in a world where I can't kick you. To all my friends not named, thanks for both tolerating and appreciating my personality.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Towards Subchemical Accuracy from Ab Iniito Quantum Chemistry

With recent advances in both theoretical algorithms and computational power, the "dream" of computing thermochemical parameters entirely from first principles to "subchemical accuracy" better than 1.0 kcal mol^{-1} is now being realized.[1–4] Perhaps nowhere is this more significant than in combustion where massive and complicated reaction schemes make the experimental isolation of individual, elementary rate constants an arduous task. Consider for example the combustion of methanol shown schematically in Figure 1.1.[5] Only the initiation steps are shown, already demonstrating significant mechanistic complexity. At least 84 elementary reactions have been identified in the complete reaction scheme, indicating the enormous difficulty of performing a controlled experiment to determine particular rate constants. Barrier estimates have ranged from 13.8 to 17.1 kcal mol^{-1} even for the seemingly simple reaction of HO_2 addition to ethylene![6,7]

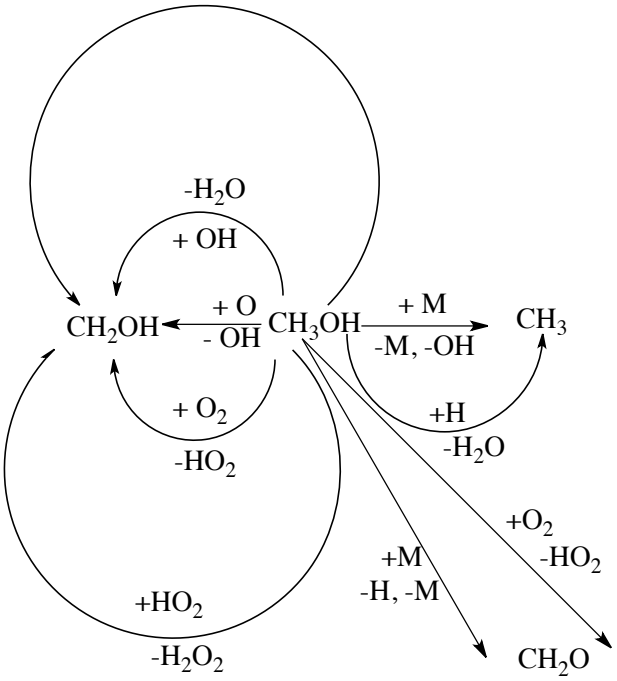

FIGURE 1.1: Reaction scheme for initiation of methanol combustion

Subchemical accuracy is absolutely essential if accurate kinetic models are desired. For combustion in particular, numerous kinetic models have been developed, including master equations with

RRKM,[8,9] QRRK,[10] and more approximate modified strong collider schemes.[11] All of these models are very sensitive to the barrier heights, with errors of even a few kcal mol^{-1} resulting in as much as sevenfold changes in predicted branching fractions.[8] The 3.3 kcal mol^{-1} uncertainty for the HO_2 addition reaction above is therefore clearly not acceptable if combustion models are to be correct within even an order of magnitude. As demonstrated herein, ab initio quantum chemistry has now advanced such that the numerous reaction intermediates and corresponding reaction barriers can now be computed with better than 1.0 kcal mol^{-1} accuracy.

The fundamental challenges of theoretical quantum chemistry can be essentially distilled to the solution of eigenvalue equations involving N-electron wavefunctions. In particular, to every physical observable corresponds a linear, Hermitian operator whose eigenvalues are the observed quantities.[12] In its most famous form, this appears as the time-independent Schrödinger equation

$$\hat{H}\Psi = E\Psi \tag{1.1}$$

where the Hamiltonian, $\hat{H}$, is the Hermitian operator representing the observed energy. The solution of the resulting many-body differential equations is certainly non-trivial. Indeed, the complexity of the exact wavefunction grows exponentially with system size - the problem of the "exponential wall" Walter Kohn emphasized in his Nobel speech.[13] While a computation on molecular oxygen, for example, may take only a few hours, the same computation on combustion intermediates of propane may take longer than the age of the universe! The impracticality of exact solutions to the Schrödinger equation has therefore created the need for approximate wavefunctions which incorporate the most important physical effects without exponential scaling.

In the simplest-possible method, Hartree-Fock,[14,15] the total energy can be computed to better than 99 % accuracy for an individual molecule. However, since even for hydrogen molecule the total electronic energy is more than 600 kcal mol^{-1}, a residual error of one percent is clearly chemically significant. In this regard, ab initio quantum chemistry must rely on cancellation of related errors for computing energy *differences* rather than rigorous computation of total energies. The complexity of the solution required for subchemical accuracy will therefore depend on the "degree of difference" between products and reactants. In the simple case of alkane rotation around a single bond,

for example, the various conformers will have a very similar electronic structure and the ground electronic excited state is well-separated from any excited electronic states. In this regard, even a very primitive wavefunction model may be sufficient. For combustion, however, the systems are open-shell. The transition states between intermediates furthermore require breaking and forming new bonds, leading to an electronic structure that varies significantly from species to species. In this regard, a much more accurate treatment of the wavefunction will be required as subtle new physical effects introduced no longer lead to cancellation of errors.

Again, while brute-force mathematics via the Schrödinger equation leads theoretically to an exact answer, we must develop an approximate wavefunction ansatz which achieves subchemical accuracy while remaining computationally feasible. In this regard, we develop here in the introduction the important concepts related to constructing molecular wavefunctions. Physical insight is used to guide the mathematics, eventually leading to a systematically improvable series of approximations that, even for difficult combustion problems, achieves the gold standard of subchemical accuracy.

1.2 The Many-Body Problem in Wavefunction Mechanics

For the hydrogen atom with only a single electron, an exact, non-relativistic solution can be obtained in which the spectrum of one-electron solutions are a sequence of "orbitals" of gradually increasing diffuseness and angular momentum.[12] Under the clamped nucleus (Born-Oppenheimer) approximation, if the mutual Coulomb repulsion of electrons is omitted from the Hamiltonian, this conceptual picture of one-electron orbitals can be extended to exact, many-body solutions. It is therefore the electron correlation resulting from their Coloumb charge repulsion that presents the fundamental challenge in wavefunction mechanics. One recent review[16] has even dubbed electron correlation "the many-body problem at the heart of chemistry."

The exact mathematical definition of electron correlation in the statistical sense implies[17]

$$\Gamma(\vec{r}_1, \vec{r}_2) \neq \frac{1}{2}\rho(\vec{r}_1)\rho(\vec{r}_2) \tag{1.2}$$

Here $\Gamma(\vec{r}_1, \vec{r}_2)$ is two-particle density distribution given by integrating over all possible positions of electrons $3 \dots N$

$$\Gamma(\vec{r}_1, \vec{r}_2) = \int \psi^*(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)\psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)d\vec{r}_3 \dots d\vec{r}_N \tag{1.3}$$

Equation (1.2) essentially expresses that the exact, spinless two-electron density distribution, Γ, is not expressible as the product of independent one-electron distributions, ρ. Building from the hydrogen atom, the simplest possible representation of an N-electron (fermionic) wavefunction is the Slater determinant, an antisymmetrized product of orbitals.

$$|\phi_1(1)\phi_2(2) \dots \phi_N(N)| = \hat{A}_N\phi_1(1)\phi_2(2) \dots \phi_N(N) \tag{1.4}$$

$$= \phi_1(1)\phi_2(2) \dots \phi_N(N) - \phi_1(2)\phi_2(1) \dots \phi_N(N) + \dots \tag{1.5}$$

where $\hat{A}_N$ is an antisymmetrizer which introduces all possible (antisymmetric) permutations of electrons ensuring that the Slater determinant is a valid fermionic wavefunction. For a single-determinant, the electron-electron interaction is only treated in an average sense with individual electrons correlating only with a mean electric field rather than explicitly correlating with each individual electron. Indeed, for the lowest singlet state of helium in which the $1s$ orbital is doubly-occupied, denoting $\vec{s}$ as the composite spin-space coordinate

$$\psi(\vec{s}_1, \vec{s}_2) = \frac{1}{\sqrt{2}}\chi_{1s}(\vec{r}_1)\chi_{1s}(\vec{r}_2)[\alpha(1)\beta(2) - \beta(1)\alpha(2)] \tag{1.6}$$

which leads after spin-integration to

$$\Gamma(\vec{r}_1, \vec{r}_2) = \frac{1}{2}|\chi_{1s}(\vec{r}_1)|^2|\chi_{1s}(\vec{r}_2)|^2 \tag{1.7}$$

$$= \frac{1}{2}\rho(\vec{r}_1)\rho(\vec{r}_2) \tag{1.8}$$

demonstrating that the spatial distribution of the two electrons is in fact uncorrelated. In contrast, for lowest triplet state of helium, the spatial part of the wavefunction is now antisymmetric with

respect to particle exchange

$$\psi(\vec{s}_1, \vec{s}_2) = \frac{1}{2}\left[\chi_{1s}(\vec{r}_1)\chi_{2s}(\vec{r}_2) - \chi_{2s}(\vec{r}_1)\chi_{1s}(\vec{r}_2)\right]\left[\alpha(1)\beta(2) + \beta(1)\alpha(2)\right] \tag{1.9}$$

leading to a correlated 2-particle density containing interference terms

$$\Gamma(\vec{r}_1, \vec{r}_2) = \frac{1}{2}|\chi_{1s}(\vec{r}_1)|^2|\chi_{2s}(\vec{r}_2)|^2 + \frac{1}{2}|\chi_{2s}(\vec{r}_1)|^2|\chi_{1s}(\vec{r}_2)|^2 - \chi^*_{1s}(\vec{r}_1)\chi^*_{2s}(\vec{r}_2)\chi_{2s}(\vec{r}_1)\chi_{1s}(\vec{r}_2) \tag{1.10}$$

Even the one-particle picture expressed by the Slater determinant therefore incorporates some measure of electron correlation. However, the physical effect encapsulated by this correlation is simply the forced antisymmetry of the electrons as fermions rather than any instantaneous spatial correlation based on their mutual Coulomb repulsion. In this regard, the interference terms in (1.10) are incorporating a "Fermi correlation" rather than a "Coulomb correlation." By choosing to represent the wavefunction as an antisymmetrized product of orbitals, we are able to incorporate the static exchange interaction between electrons, but are still neglecting the important spatial correlation of electrons.

In constructing more exact representations of the wavefunction, the conceptually simple orbital picture can still be maintained. In particular, given a complete a basis set of one-electron orbitals, a complete determinantal basis for the Hilbert space of N-electron wavefunctions can be constructed.[18] In this way, an arbitrary N-electron wavefunction can be constructed as a multideterminantal expansion. As a first approximation to the electron correlation, one can use second-order perturbation theory.[19] The Hamiltonian is partitioned

$$\hat{H} = \hat{H}_0 + \lambda\hat{V} \tag{1.11}$$

into a zeroth order part, $\hat{H}_0$ and a perturbation, $\hat{V}$ which is modulated by a strength parameter λ. A zeroth order reference determinant, Φ_0 is computed, and successive perturbative corrections to the first order wavefunction are gradually included to higher order in the strength parameter λ. In the most popular formulation, second-order Møller Plesset perturbation theory (MP2), the

zeroth order operator is chosen as the Fock operator, $\hat{F}$ which includes the electron kinetic energy, nuclear attraction, and mean-field electron interaction. In this way, the the perturbation is meant to incorporate the missing Coulomb correlation.

Including terms in the multideterminantal expansion up to second order in λ avoids the exponential complexity of the exact wavefunction. However, if the orbital representation of the wavefunction is used, the MP2 ansatz still leads to a system of equations scaling as $\mathcal{O}(N^5)$ in the system size. Even given that dramatic computational cost, the physical correlation captured by the MP2 ansatz is far from exact. Most notably, the MP2 determinantal expansion truncates with only those determinants differing from the reference function Φ_0 by at most two orbitals. In this way, MP2 essentially correlates only two electrons at a time. From physical intuition, three-body and higher effects are expected to be much less significant than pairwise correlation. However, rigorous benchmarking has now shown that even three-particle and four-particle correlations can contribute on the chemically significant scale of 1 to 10 kcal mol^{-1}. However, the already dramatic cost of the two-particle approximations in MP2 becomes an N^7 cost in the simplest formulation, CCSD(T), incorporating three-body effects. As the particle rank (level of many-body effect) increases past three electrons, the computational cost continues to accelerate, rapidly leading to wavefunction ansätze which are only tractable for systems involving a few electrons.

Beyond the intrinsic cost of these approximate wavefunctions, the demands of the underlying one-particle orbital basis set itself become problematic. The wavefunction cannot be expressed exactly as a finite, closed sum over an infinite basis. An approximate, finite computational orbital basis must therefore be introduced. Even for a given particle-rank, the approximate two-body or three-body effects will then not be exactly evaluated. Again, building from the known exact solutions for the hydrogen atom, an arbitrary one-electron function can be written in spherical coordinates

$$F(r,\theta,\phi) = \sum_{n,l,m_l} C_{n,l,m_l} R_{nl}(r) Y_{l,m_l}(\theta,\phi) \tag{1.12}$$

where the expansion coefficients C_{n,l,m_l} are unique for the given function. In practice, the exact

hydrogen radial functions $R_{nl}(r)$ are replaced by Gaussian basis functions[20]

$$R_{nl} = \exp(\alpha_{nl} r^2) \tag{1.13}$$

but any set of complete functions can be chosen for computational convenience. Equation (1.12) expands an arbitrary function as partial waves, separating the function into radial and angular components. The computational cost of the orbital basis rapidly grows with increasing angular momentum. The partial wave expansion must then be rapidly convergent for a tractable, high-accuracy treatment of the wavefunction. For hydrogen-like ions, a closed-form solution exists for a specific l, m_l value. The partial wave expansion is therefore finite (it is a single term), and in that sense rapidly converges. Even though an infinite partial wave expansion is necessary if spherical symmetry is broken, the series converges with an exponential decay for one-electron systems,[21] and high-accuracy can therefore be obtained with only a few terms.

This partial wave expansion can be extended to functions of many-electrons. For helium, exploiting spherical symmetry, we can write

$$\psi(\vec{r}_1, \vec{r}_2) = \psi_l(r_1, r_2) P_l(\theta_{12}) \tag{1.14}$$

where θ_{12} is the angle between electrons and P_l is a Legendre polynomial incorporating the angular dependence. Here, ψ_l depends on the radial position of the two electrons. In this sense, the electron correlation has been partitioned into an angular correlation - the electrons will tend to occupy opposite sides of the helium nucleus - and a radial correlation - the electrons will tend to avoid simultaneously being near the helium nucleus. In contrast to the one-electron, case, however the series is now very slow to converge with increasing values of l. Unlike the exponential decay exhibited for one-electron systems, the individual terms only decay as $(l + \frac{1}{2})^{-1/4}$. Truncating at a particular value of l therefore leads to a cumulative residual error that decays only as l^{-3}. Ensuring a sufficient number of Gaussian functions to saturate the radial component of the wavefunction, the total computational time has been demonstrated to grow empirically as l^{12}. The residual error will therefore decay at the painfully slow rate of $t^{-1/4}$ in the total computational time!

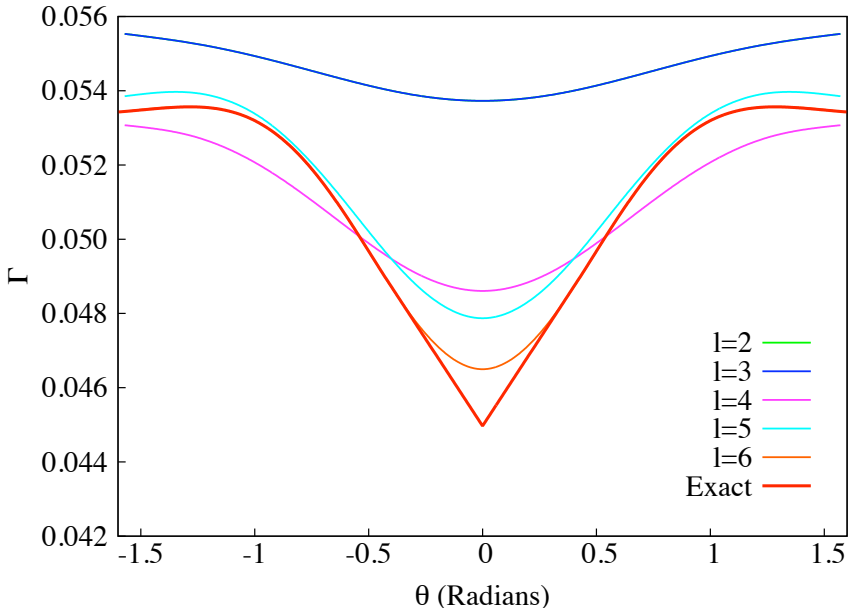

FIGURE 1.2: Coalescence cusp and Coulomb hole shape for gradually increasing levels of angular momentum (l)

The physical reasons for this slow convergence are now understood as the mathematical difficulty of describing the cusp in the wavefunction at the coalescence point between two electrons.[22–24] At coalescence, the singularity in the Coulomb repulsion energy must be offset by a corresponding singularity in the kinetic energy, leading to a cusp and surrounding depletion of electron density or Coulomb hole. Kato would later show that for singlet pairs the correct spherically-averaged shape of the Coulomb hole is given by the cusp condition[24]

$$\lim_{r_{12}\to 0}\left(\frac{\partial\Psi(r_{12})}{\partial r_{12}}\right)_{av} = \frac{1}{2}\Psi(r_{12}=0) \tag{1.15}$$

which demonstrates that the wavefunction shape should depend linearly on the interelectronic distance, r_{12}. The slow convergence can be easily demonstrated by plotting two-electron density near the coalescence point (Figure 1.2). In particular, for Neon, if we place the two electrons on a ring at 1.0 a_0 as shown in Figure 1.3, we can examine the shape of the electron density as the two electrons are brought to coalescence. The shape of the Coulomb hole is demonstrated for a sequence of basis sets with gradually increasing angular momentum in Figure 1.2.

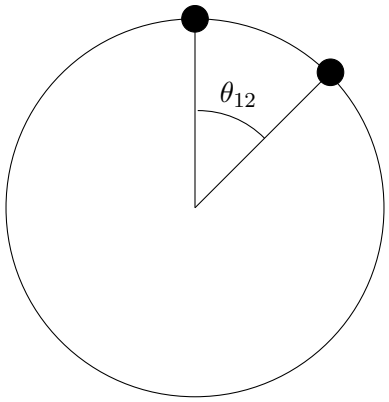

FIGURE 1.3: Schematic of coordinate system for coalescence cusp diagram

The cusp region, while still significant for small l values, rapidly becomes the only significant energetic contribution for higher angular momentum terms in the partial wave expansion. In fact, solving for the first-order wavefunction via second-order perturbation theory demonstrates the limiting result[25]

$$\lim_{l\to\infty} \psi_l = \frac{1}{2}(r_{12})_l \Phi \tag{1.16}$$

$$(r_{12})_l = \frac{1}{2l+3}\frac{r_<^{l+2}}{r_>^{l+1}} - \frac{1}{2l-1}\frac{r_<^{l}}{r_>^{l-1}} \tag{1.17}$$

where Φ is the reference function and $r_<$ is the smaller of r_1 and r_2. In this regard, the only physical effect included in higher partial waves is the electron cusp. As emphasized above, the most conceptually and computationally simple representation of the wavefunction is a multideterminantal expansion since it retains the physical picture of single-electron orbitals. However, due to the limitations imposed by an approximate, finite orbital basis, the cusp problem cannot simply be overcome by a brute force approach. Saturating the orbital basis with high-angular momentum functions will quickly lead to an unreasonable computational cost. Achieving an accurate, but compact representation of the wavefunction is therefore simply not possible by employing only orbital products as N-electron basis functions.

A compact representation of the cusp region can be obtained by employing two-particle basis

functions depending explicitly on the interelectronic distance, r_{12}. In this way, the cusp region can be included exactly through only a few explicitly-correlated terms instead of requiring a cumbersome orbital expansion. Considering again second-order perturbation theory, we can employ the wavefunction ansatz for the first-order correction

$$\psi = \chi + \frac{1}{2} r_{12} \Phi \tag{1.18}$$

in which χ is a perturbative contribution expanded in orbitals and Φ is a single term involving the reference function. This leads to a modified differential equation for χ

$$-(\hat{H}_0 - E_0)\chi = (\hat{U} - E_1)\Phi \tag{1.19}$$

involving a modified, non-singular perturbation

$$\hat{U} = -\frac{1}{2} \frac{\vec{r}_1 - \vec{r}_2}{r_{12}} \cdot (\vec{\nabla}_1 - \vec{\nabla}_2) \tag{1.20}$$

The cusp requirement is therefore removed for χ, and the orbital expansion will be rapidly convergent. Conceptually, we can also consider the r_{12}-dependent terms as a change of coordinates

$$\vec{r} = \vec{r}_1 - \vec{r}_2 \tag{1.21}$$

$$\vec{R} = \frac{1}{2}(\vec{r}_1 + \vec{r}_2) \tag{1.22}$$

from independent to explicitly-correlated electronic coordinates. This aligns with physical intuition as the most natural coordinate for describing electron correlation is the interelectronic vector $\vec{r}$. Conflicting mathematical priorities therefore occur in which the mean-field physical effects are most naturally described through independent-electron orbitals, but the instantaneous electron correlation is most naturally described through explicit r_{12} terms.

While conceptually simple as a wavefunction ansatz, the explicitly-correlated terms introduce new complications. If only Slater determinants are employed, the Hamiltonian matrix elements easily factorize into integrals involving no more than two electrons. In contrast, if explicitly-

correlated functions are included in the wavefunction, numerous and complicated many-electron integrals appear. Consider for example a four-electron wavefunction. Computing the Coulomb repulsion energy for an orbital product yields, for example,

$$E_C = \int \chi_i(\vec{r}_1)\chi_j(\vec{r}_2)\chi_k(\vec{r}_3)\chi_l(\vec{r}_4)\frac{1}{r_{12}}\chi_i(\vec{r}_1)\chi_j(\vec{r}_2)\chi_k(\vec{r}_3)\chi_l(\vec{r}_4)\,d\vec{r}_1\,d\vec{r}_2\,d\vec{r}_3\,d\vec{r}_4 \tag{1.23}$$

$$= \int \chi_i(\vec{r}_1)\chi_j(\vec{r}_2)\frac{1}{r_{12}}\chi_i(\vec{r}_1)\chi_j(\vec{r}_2)\,d\vec{r}_1\,d\vec{r}_2 \tag{1.24}$$

in which the dependence on electrons 3 and 4 can be eliminated by a simple overlap integration. In contrast, if we construct an explicitly-correlated four-electron wavefunction, the Coulomb repulsion term

$$E_C = f_{ij}(\vec{r}_1,\vec{r}_2)f_{kl}(\vec{r}_3,\vec{r}_4)\frac{1}{r_{23}}f_{ij}(\vec{r}_1,\vec{r}_2)f_{kl}(\vec{r}_3,\vec{r}_4)\,d\vec{r}_1\,d\vec{r}_2\,d\vec{r}_3\,d\vec{r}_4 \tag{1.25}$$

cannot be factorized. These many-electron integrals, already complicated for small systems, will quickly become overwhelmingly numerous for larger systems. A rigorous explicitly correlated treatment can therefore only be performed on systems containing at most a few electrons.

We are therefore left with a fundamental dilemma in wavefunction mechanics. Orbital-based representations cannot accurately describe the correlation cusp, but provide a simple mechanism for computing Hamiltonian matrix elements. In contract, explicitly-correlated wavefunctions provide a highly-accurate, compact representation of the wavefunction, but lead to intractable matrix elements. In general, two different approaches to this problem are followed. Orbital-only wavefunctions can be used over a sequence of basis sets, and the effect of higher partial waves can be estimated by extrapolation. This approach is followed in the so-called Focal Point Approach (FPA) and forms the basis for the discussion in the first-half of this dissertation. Alternatively, approximations can be introduced for the difficult matrix elements, which, although no longer computed exactly, become computationally feasible for large molecules. This approach is followed in the so-called R12 methods and forms the basis for the discussion in the second-half of this dissertation.

1.3 Focal Point Approach

As outlined above, is it not computationally feasible to simply follow a brute-force approach in constructing molecular wavefunctions. The computational cost grows far too rapidly with both the particle rank (degree of electron correlation) and size of the underlying orbital basis. The so-called correlation and complete basis set limits are therefore never realized in practice. However, by constructing a hierarchy of gradually improving approximations, the possibility still exists of incrementally approaching the correlation and basis set limits. In this regard, the focal point approach[26–29] employs a systematically-improvable series of approximations meant to extrapolate the exact result, using physical insight where simple computational power fails.

The focal point method is therefore a two-dimensional approximation, estimating the correlation limit (exponential complexity, all electrons simultaneously correlated) and the basis set limit (infinite orbital basis for exactly representing the wavefunction). At its core, the focal point method assumes that as the particle rank is increased to incorporate three-body, four-body, and higher effects, the energetic contributions from each successive rank will monotonically decrease. This is guided by physical intuition. In a system of ten electrons, for example, the probability of even six electrons occupying the same region of space is expected to be virtually zero. In fact, the Pauli exclusion principle demands that the wavefunction have zero amplitude at the coalescence of three electrons since at least two will have the same spin. While for two-particle effects, we have demonstrated above that the partial-wave expansion is very slow to converge, one may ask whether the incremental effect of three-body and higher effects is as demanding. In this regard, if only the three-body correction to the two-body correlation is desired, the partial wave series might be truncated at a much lower level, leading to computations which are computationally feasible.

The first dimension of the focal point, the correlation limit, is treated through the coupled-cluster expansion of the wavefunction[30–33]

$$|\Psi\rangle = \exp(\hat{T})\,|\Phi_0\rangle \tag{1.26}$$

where $\hat{T}$ is an excitation operator that generates a multi-determinantal expansion of the wavefunc-

tion from a given reference function, usually taken to be the Hartree-Fock determinant. The cluster operator can be written schematically as

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \ldots \tag{1.27}$$

where the subscript indicates the particle-rank of the operator, i.e. the number of orbital substitutions from the reference determinant. Conceptually, each successive cluster operator can be considered to incorporate progressively higher correlation effects in the wavefunction. An exact treatment of an N-electron wavefunction would require all cluster operators up to $\hat{T}_N$ with rank N. This is often referred to as the full configuration interaction (FCI) limit. If, for example, an approximate wavefunction which only incorporates two-body effects were desired, the cluster operator could be truncated at $\hat{T} = \hat{T}_1 + \hat{T}_2$. By incrementally increasing the rank of the cluster operator, one converges on the correlation limit in which all electrons are simultaneously correlated. Each level of theory is denoted by an acronym, for example, CCSD(T) which indicates rigorous treatment of single and double substitutions with perturbative inclusion of triple excitations. The focal point method therefore follows the hierarchy HF $\rightarrow$ MP2 $\rightarrow$ CCSD $\rightarrow$ CCSD(T) $\rightarrow$ CCSDT $\rightarrow$ CCSDT(Q). Again, rather than considering total energy changes, ΔE, for a given reaction, the focal point method gauges only incremental contributions, for example

$$\delta[\text{CCSD}] = \Delta E_{\text{CCSD}} - \Delta E_{\text{MP2}} \tag{1.28}$$

the CCSD incremental correction to the MP2 reaction energy.

The second dimension of the focal point method, the basis set limit, is treated through extrapolation techniques. As already noted above, for the Hartree-Fock method (HF), the partial-wave expansion converges exponentially while for correlated methods it converges as l^{-3} in the maximum angular momentum in the basis. The Dunning correlation consistent basis sets[34–36] were carefully calibrated for exactly this purpose, leading to a hierarchy of basis sets denoted cc-pVXZ where X is the "cardinal number" (X=D, T, Q, 5, ...). While for historical reasons the XZ stands for double-ζ, triple-ζ, etc., indicating the number of independent Gaussian functions, the cardinal number

also indicates the maximum angular momentum $l = 2$, $l = 3$, etc. While numerous extrapolation schemes have been proposed in the literature, the standard choice for correlated methods, e.g. MP2 is the formula[37,38]

$$E_{\mathrm{MP2}}^{corr} = E_{\mathrm{MP2}}^{\infty} + BX^{-3}\,. \tag{1.29}$$

where $E_{\mathrm{MP2}}^{\infty}$ and B are parameters to be fit and the ∞ denotes the estimated energy in a complete basis set (CBS). Here only the correlation energy is extrapolated, i.e. the MP2 energy correction to the Hartree-Fock energy

$$E_{\mathrm{MP2}}^{corr} = E_{\mathrm{MP2}} - E_{\mathrm{HF}} \tag{1.30}$$

In a standard treatment, a somewhat brute force approach is applied to the economical Hartree-Fock and MP2 methods employing up through, e.g. a cc-pV6Z basis set. The Hartree-Fock and MP2 correlation energies are then extrapolated, estimating the basis set limit of the MP2 two-body correlation.

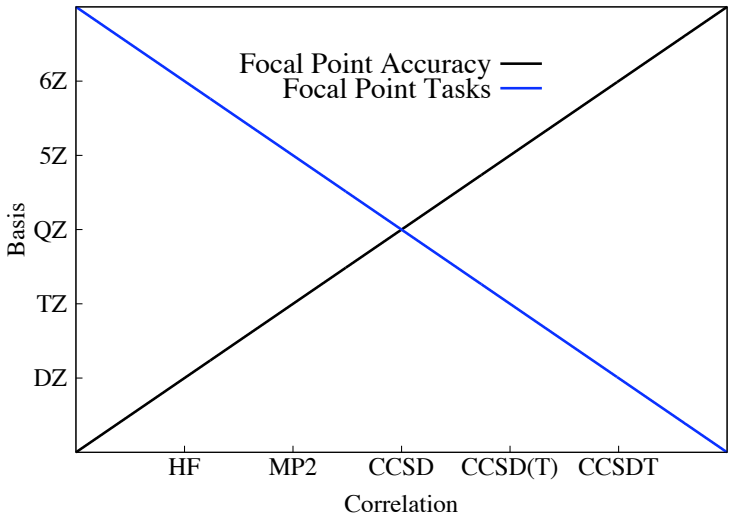

FIGURE 1.4: Schematic of the focal point scheme showing extrapolation to the correlation and basis set limits.

Further corrections along the coupled-cluster hierarchy are then expected to converge much faster. In particular, the Coulomb cusp has already been accounted for by the MP2 extrapola-

tion. In this regard, the increment δ[CCSD] is not rigorously incorporating all two-body effects, but providing only a modest correction to the MP2 perturbative treatment. In this regard, the onerous cc-pV6Z basis is not necessary, and only a cc-pVQZ basis set might be necessary for an accurate δ[CCSD]. Extending this reasoning, the increment δ[CCSD(T)] is not rigorously incorporating all three-body effects, but providing only a modest three-body correction to the CCSD reaction energy. In this regard, only a cc-pVTZ basis set might be necessary. Carrying on, the increment δ[CCSDT(Q)] should further decrease in magnitude, and only a cc-pVDZ basis set might be necessary for chemical accuracy.

The two-dimensional nature of the focal point extrapolation is shown in Figure 1.4. Conceptually, the method is extrapolating down along the black line to larger basis sets and more extensive correlation treatments. Computationally, however, the method is advancing along the blue line, shrinking the basis set as the correlation treatment is expanded. The focal point method, however, reaches a fundamental limit at the cc-pVDZ basis set at which the basis set cannot be further minimized. In this regard, for very expensive methods such as CCSDTQ, even a cc-pVDZ computation might be impossible, and the incremental scheme cannot continue to higher levels of correlation. In this way, the orbital-based methods are no longer sufficient and a more compact representation of the wavefunction would be absolutely required to estimate higher-body effects.

1.4 R12 Methods

As outlined above, despite the seeming conceptual simplicity of using explicitly-correlated terms in the wavefunction, methods employing r_{12}-dependent functions rapidly become intractable due to the numerous many-electron integrals.[39] Directly expanding the wavefunction in explicitly-correlated, two-particle functions (known as geminals) essentially forms the basis of the Gaussian geminals approach.[40–42] This method, while unrivaled in accuracy, can hardly be applied to systems containing more than a few electrons.

Rather than modifying the wavefunction, an alternative explicitly-correlated approach is the transcorrelated Hamiltonian.[43–46] A non-unitary transformation of the Hamiltonian is performed involving r_{12}-dependent terms, leading to a modified differential equation which has no Coulomb

singularity. In this way, the modified Schrödinger equation expressed by the transcorrelated Hamiltonian does not have a cusp requirement, and the partial wave expansion of the wavefunction will be much more rapidly convergent. While not requiring four-electron integrals, the transcorrelated approach still leads to three-electron integrals, again preventing application to large molecular systems.

A possible choice for explicitly-correlated terms would be the simple product of a correlation factor, $F(r_{12})$, and Hartree-Fock orbitals, e.g.

$$f_{ij}(\vec{r}_1, \vec{r}_2) = F(r_{12}) \left| ij \right\rangle \tag{1.31}$$

where Dirac bra-ket notation is now employed. Usually the correlation factor is of Slater form

$$F(r_{12}) = \exp(-\gamma r_{12}) \tag{1.32}$$

which incorporates both the correct cusp and Coulomb hole decay. The many-electron integrals required are therefore of the form

$$\left\langle ijkl \left| \frac{f(r_{12}) f(r_{34})}{r_{23}} \right| ijkl \right\rangle \tag{1.33}$$

By introducing a resolution-of-the-identity (RI), the many-electron integrals can be factorized into products of two-electron integrals, e.g.

$$\left\langle ijkl \left| \frac{f(r_{12}) f(r_{34})}{r_{23}} \right| ijkl \right\rangle = \left\langle ij \left| f(r_{12}) \right| i\alpha \right\rangle \left\langle k\alpha \left| 1/r_{12} \right| \beta j \right\rangle \left\langle \beta l \left| f(r_{12}) \right| kl \right\rangle \tag{1.34}$$

$$\approx \left\langle ij \left| f(r_{12}) \right| ip' \right\rangle \left\langle kp' \left| 1/r_{12} \right| q'j \right\rangle \left\langle q'l \left| f(r_{12}) \right| kl \right\rangle \tag{1.35}$$

where α, β represents the insertion of a formally complete one-electron basis and p', q' is some finite, approximate computational RI basis. If a massive RI basis is required for accurately factorizing the integrals, clearly no computational efficiency is gained. However, if a reasonably-sized RI insertion can be obtained by prudent calibration of the basis set, the many-electron integrals do not need

to be rigorously computed or stored, and only a modest number of two-electron integrals must be computed. This RI insertion forms the basis for most modern explicitly-correlated methods.[39,47–51]

In the focal point approach built on orbital-expansions of the wavefunction, the matrix elements are readily computed and the computational challenge is the determination of a huge number of variational parameters associated with multideterminantal basis functions. Explicitly-correlated methods provide a foil to the conventional methods in which a much smaller number of variational parameters are required, but the Hamiltonian matrix elements provide the computational challenge. A logical extension of these principles is therefore combining the conventional and explicit methods. The explicitly-correlated terms might reduce the number of variational parameters, while the presence of the conventional orbital expansion might simplify the required r_{12}-dependent terms. This ansatz makes physical sense since the explicitly-correlated terms are well-suited for describing the wavefunction for small r_{12} while the orbital-based determinants are better-suited for describing longer-range correlation. The combination of simple Slater determinants and r_{12}-dependent geminals in constructing the wavefunction provides the starting point for the R12 methods. The difficult many-electron integrals are avoided through the RI approximation introduced above. Furthermore, the basis set requirements on the r_{12} terms are significantly relaxed since the explicit geminals are only describing correlation within the Coulomb hole rather than being forced to describe all electron correlation as is done in the Gaussian geminals method.

1.5 Research Prospectus

While the above discussion provides a conceptual overview of the concerns in high accuracy, ab initio quantum chemistry, the physical intuition must always be verified through numerical benchmarks. Additionally, application of these methods in practice requires numerous theoretical subtleties to be accounted for in the implementation. In the first part of the dissertation, the focal point approach is therefore applied to gas-phase conformers of the amino acid cysteine, providing rigorous energetic and structural benchmarks. The physical effects required here are, for example, hydrogen-bonding, long-range electrostatic, dispersion, or hyperconjugation. As emphasized above, all of these treatments depend on cancelation of errors. The accurate calculation of absolute energies

is therefore not necessary as long as relative energies are accurately computed. For cysteine, the overall electronic structure should not dramatically change with conformation. In this regard, the differential physical effects such as hyperconjugation and hydrogen-bonding should converge quickly to the focal point limit down the hierarchy. The conformers of cysteine therefore provide an ideal test case for testing the convergence of the focal point methods.

In the second part of the dissertation, the focal point approach is applied to the reaction of ethyl radical with molecular oxygen. The open-shell radicals present significant new challenges relative to closed-shell species especially for transition states which involve breaking and forming of bonds. This provides an even more stringent test of the focal point method as correlation treatments much more extensive than in cysteine are expected to be necessary for chemical accuracy better than 1.0 kcal mol^{-1}. This difficult case pushes the focal point approach to its limit, demonstrating a need for alternative routes to high accuracy wavefunctions.

In the third part of the dissertation, building on the needs demonstrated by the open-shell ethyl radical combustion, an initial formulation of R12 methods appropriate for radicals is presented at the level of second-order perturbation theory. Significant new challenges are encountered extending the R12 methods to an open-shell coupled-cluster ansatz. As a final component of the dissertation, experience gained from the MP2 derivation is therefore used to formulate an efficient set of R12 approximations suitable for higher-level correlation treatments. The highly accurate, but still efficient open-shell R12 ansatz presented here is therefore the most important result in this dissertation.

BIBLIOGRAPHY

[1] A. Tajti, P. G. Szalay, A. G. Csàszàr, M. Kallay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vazquez, J. F. Stanton, J. Chem. Phys. **121**, 11599 (2004).

[2] D. Feller, D. A. Dixon, J. S. Francisco, J. Phys. Chem. A **107**, 1604 (2003).

[3] A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kallay, J. Gauss, J. Chem. Phys. **120**, 4129 (2004).

[4] A. Karton, E. Rabinovich, J. M. L. Martin, B. Ruscic, J. Chem. Phys. **125**, 144108 (2006).

[5] F. L. Dryer, 'The Phenomenology of Modeling Combustion Chemistry' in *Fossil Fuel Combustion: A Source Book* edited by W. Bartok, A. F. Sarofin. (John Wiley and Sons, New York, NY, 1991).

[6] A. A. Mantashyan, L. A. Khachatryan, O. M. Niazyan, S. D. Arsentyev, Combustion Flame **43**, 221 (1981).

[7] R. R. Baldwin, C. E. Dean, R. W. Walker, J. Chem. Soc. Faraday Transactions Ii **82**, 1445 (1986).

[8] J. A. Miller, S. J. Klippenstein, Int. J. Chem. Kinetics **33**, 654 (2001).

[9] A. F. Wagner, I. R. Slagle, D. Sarzynski, D. Gutman, J. Phys. Chem. **94**, 1853 (1990).

[10] C. Y. Sheng, J. W. Bozzelli, A. M. Dean, A. Y. Chang, J. Phys. Chem. A **106**, 7276 (2002).

[11] H. H. Carstensen, C. V. Naik, A. M. Dean, J. Phys. Chem. A **109**, 2264 (2005).

[12] J. J. Sakurai, *Modern Quantum Mechanics*. (Addison-Wesley, Reading, MA, USA, 1994).

[13] W. Kohn, *Nobel Lectures, Chemistry 1996-2000*. edited by I. Grenthe. (World Scientific Publishing Co., Singapore, 2003).

[14] C. C. J. Roothaan, Rev. Mod Phys. **23**, 69 (1951).

[15] P. Pulay, Mol. Phys. **17**, 197 (1969).

[16] D. P. Tew, W. Klopper, T. Helgaker, J. Comput. Chem. **28**, 1307 (2007).

[17] W. Kutzelnigg, 'Theory of Electron Correlation' in *Explicitly Correlated Wave Functions in Chemistry and Physics: Theory and Applications* edited by J. Rychlewski. (Springer, New York, NY, 2003).

[18] B. MacCluer, *Elementary Functional Analysis*. (Springer, New York, NY, 2010).

[19] A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry*. (Macmillan, New York, NY, 1982).

[20] E. R. Davidson, D. Feller, Chem. Rev. **86**, 681 (1986).

[21] D. Feller, J. Chem. Phys. **96**, 6104 (1992).

[22] R. T. Pack, W. B. Brown, J. Chem. Phys. **45**, 556 (1966).

[23] D. P. Tew, J. Chem. Phys. **129**, 014104 (2008).

[24] T. Kato, Commun. Pure Appl Math. **10**, 151 (1957).

[25] W. Kutzelnigg, Theor. Chim. Acta **68**, 445 (1985).

[26] M. S. Schuurman, S. R. Muir, W. D. Allen, H. F. Schaefer, J. Chem. Phys. **120**, 11586 (2004).

[27] G. Czakò, E. Màtyus, A. C. Simmonett, A. G. Csàszàr, H. F. Schaefer, W. D. Allen, J. Chem. Theor. Comput. **4**, 1220 (2008).

[28] J. M. Gonzales, C. Pak, R. S. Cox, W. D. Allen, H. F. Schaefer, A. G. Csàszàr, G. Tarczay, Chem. Eur. J. **9**, 2173 (2003).

[29] A. L. L. East, W. D. Allen, J. Chem. Phys. **99**, 4638 (1993).

[30] R. J. Bartlett, Annu Rev. Phys. Chem. **32**, 359 (1981).

[31] S. A. Kucharski, R. J. Bartlett, Adv. Quantum Chem. **18**, 281 (1986).

[32] R. J. Bartlett, M. Musiał, Rev. Mod Phys. **79**, 291 (2007).

[33] T. D. Crawford, H. F. Schaefer, Rev. Comp Chem. **14**, 33 (2000).

[34] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

[35] D. E. Woon, T. H. Dunning, J. Chem. Phys. **98**, 1366 (1993).

[36] R. A. Kendall, T. H. Dunning, R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

[37] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).

[38] T. Helgaker, W. Klopper, H. Koch, J. Noga, J. Chem. Phys. **106**, 9639 (1997).

[39] W. Klopper, F. R. Manby, S. Ten-No, E. F. Valeev, Int. Rev. Phys. Chem. **25**, 427 (2006).

[40] K. C. Pan, H. F. King, J. Chem. Phys. **53**, 4397 (1970).

[41] R. Bukowski, B. Jeziorski, K. Szalewicz, J. Chem. Phys. **100**, 1366 (1994).

[42] R. Bukowski, B. Jeziorski, K. Szalewicz, J. Chem. Phys. **110**, 4165 (1999).

[43] H. F. Schaefer, Mol. Phys. **100**, 75 (2002).

[44] O. Hino, Y. Tanimura, S. Ten-No, Chem. Phys. Lett. **353**, 317 (2002).

[45] O. Hino, Y. Tanimura, S. Ten-No, J. Chem. Phys. **115**, 7865 (2001).

[46] S. Ten-No, Chem. Phys. Lett. **330**, 169 (2000).

[47] W. Klopper, W. Kutzelnigg, J. Chem. Phys. **94**, 2020 (1991).

[48] V. Termath, W. Klopper, W. Kutzelnigg, J. Chem. Phys. **94**, 2002 (1991).

[49] W. Kutzelnigg, W. Klopper, J. Chem. Phys. **94**, 1985 (1991).

[50] W. Klopper, C. C. M. Samson, J. Chem. Phys. **116**, 6397 (2002).

[51] W. Klopper, J. Chem. Phys. **120**, 10890 (2004).

CHAPTER 2

CONFORMERS OF GASEOUS CYSTEINE [†]

[†]J. J. Wilke, M. C. Lind, H. F. Schaefer, A. G. Császár, and W. D. Allen, J. Chem. Theory Comput. **5**, 1511 (2009). Reprinted here with permission of the American Chemical Society.

2.1 Abstract

Structures, accurate relative energies, equilibrium and vibrationally averaged rotational constants, quartic and sextic centrifugal distortion constants, dipole moments, ^{14}N nuclear quadrupole coupling constants, anharmonic vibrational frequencies, and double-harmonic infrared intensities have been determined from ab initio electronic structure computations for conformers of the neutral form of the natural amino acid L-cysteine (Cys). A systematic scan located 71 unique conformers of Cys using the MP2(FC)/cc-pVTZ method. The large number of structurally diverse low-energy conformers of Cys necessitates the highest possible levels of electronic structure theory to determine their relative energies with some certainty. For this reason, we determined the relative energies of the lowest-energy eleven conformers, accurate within a standard error (1σ) of about 0.3 kJ mol^{-1}, through first-principles composite focal point analyses (FPA), which employed extrapolations using basis sets as large as aug-cc-pV(5+d)Z and correlation treatments as extensive as CCSD(T). Three and eleven conformers of L-cysteine fall within a relative energy of 6 and 10 kJ mol^{-1}, respectively. The vibrationally averaged rotational constants computed in this study agree well with Fourier-transform microwave spectroscopy results. The effects determining the relative energies of the low-energy conformers of cysteine are analyzed in detail on the basis of hydrogen bond additivity schemes and natural bond orbital analysis.

2.2 Introduction

Because amino acids (AA) are the building blocks of proteins and peptides, the structural investigation of them, extending from solids to the gas phase, has received considerable experimental and theoretical attention.[1] Cysteine (Cys) is the only amino acid with a reactive sulfur moeity. In this regard, cysteine contributes to diverse structures, including disulfide bonds, zinc fingers, and Fe-S coordination in iron-sulfur proteins.[2] Functionally, disulfide bonds formed from cysteine serve a central role in glutathione, a mediator of oxidative stress, and strong nucleophilicity also makes cysteine a key component of the active site in many other enzymes.[3,4]

In the gas phase, amino acids are intrinsically flexible systems, occurring in their neutral form

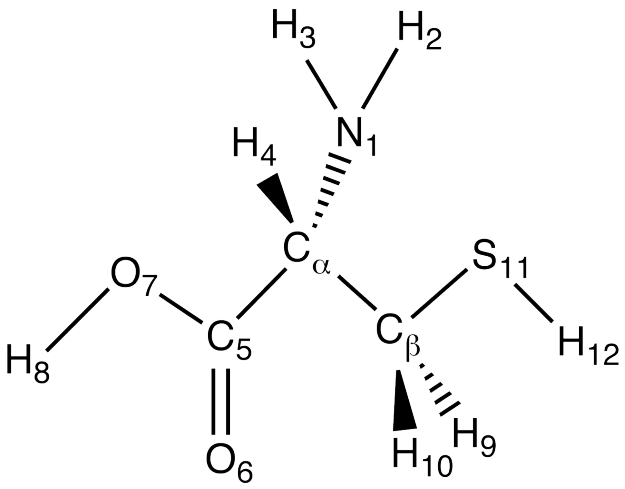

FIGURE 2.1: Labeling scheme for cysteine.

and exhibiting a large number of low-energy conformers. Even glycine, with only three rotatable single bonds, has eight conformers,[1,5,6] five of which have relative energies less than 12 kJ mol^{-1}. The number of natural amino acids is limited, and the relatively small size of these molecules allows the application of highly sophisticated quantum chemical methods to study their equilibrium and dynamical structures and rotational-vibrational spectra. The number of local minima on the respective potential energy surfaces (PES) and the structural properties of the related conformers, including accurate relative energies, are available for a number of amino acids. A review summarizing results before 1999 is provided in Ref 1. Given the accuracy of modern electronic structure techniques, characteriz(ation of the complex PESs of amino acids should precede and supplement related experimental structural and spectroscopic studies. The most accurate equilibrium structures (in cases both Born-Oppenheimer and semiexperimental ones) and relative energies (obtained within the focal point analysis (FPA) approach)[7–11] are available for the amino acids glycine (Gly)[5,12,13] alanine (Ala),[12,14] threonine (Thr),[15] and proline (Pro).[16,17] The present high-level computational study expands the list of structurally well characterized amino acids by careful investigation of all of the low-energy conformers of L-cysteine (Cys). Cysteine is a good representative of those amino acids with the added complexity of a side chain capable of substantial hydrogen-bonding interactions. Substitution of an SH group for one of the H atoms of the methyl

side chain of Ala introduces two new rotators of significance, those around the C_α-C_β and C_β-S bonds (Figure 2.1). A dramatic increase in the number of possible conformers results. The presence of three H-bond donors and four H-bond acceptors in Cys allows for the existence of twelve distinct types of hydrogen bonds (Figure 2.2), including: (a) bifurcated H-bonds between -NH_2 and -COOH, similar to those found in the most stable conformers of Gly[5] and Ala;[12] (b) simple N-H donor bonds, like the adducts to sidechain S-H and to carboxylic acid O-H and C=O; (c) sidechain S-H interactions with nitrogen and the carbonyl or hydroxyl oxygen; and (d) O-H donations to C=O and to sidechain S-H and NH_2. While H-bonds are certainly the main secondary interactions determining the occurence and relative energies of the conformers of Cys, other structural factors are also important. These include electronic effects (exchange, electrostatic and hyperconjugative interactions), as well as steric and dispersive interactions. Detailed investigation of these structural factors is one of the main goals of this study. In principle, Cys could contain the same conformers as serine (Ser), its OH analogue. However, since the interactions in Cys are weaker than in Ser (the thiol group of the side chain has comparatively poorer H-bonding characteristics), the barriers separating the conformers are expected to be smaller and in some instances may even disappear. In such a case, Cys would exhibit fewer unique conformations than Ser.

Previous ab initio electronic structure computations performed on Cys include studies on its conformational behavior[19] and on its various physical properties, including proton affinities and ionization potentials.[20] Schäfer et al.[21] investigated 10 conformers of Cys at the RHF/4-21G level and established conformational trends. Gronert and O'Hair[18] located 42 conformers at several levels of ab initio electronic structure theory, including RHF/6-31G* and MP2/6-31+G*. Recently, Dobrowolski and co-workers[22] located 51 conformers using the B3LYP and MP2 methods in conjunction with the aug-cc-pVDZ (and in some cases aug-cc-pVTZ) basis set. The computed B3LYP/aug-cc-pVDZ frequencies were compared to IR matrix isolation spectra, suggesting the presence of between three and six L-cysteine conformers.

The conformers of cysteine investigated previously range in relative energy by at most 50 kJ mol^{-1}. The six most stable conformers of Cys lie within 7 kJ mol^{-1}, while altogether 33 conformers have been identified within a 17 kJ mol^{-1} range. Alonso et al.[23] recently identified five conformers

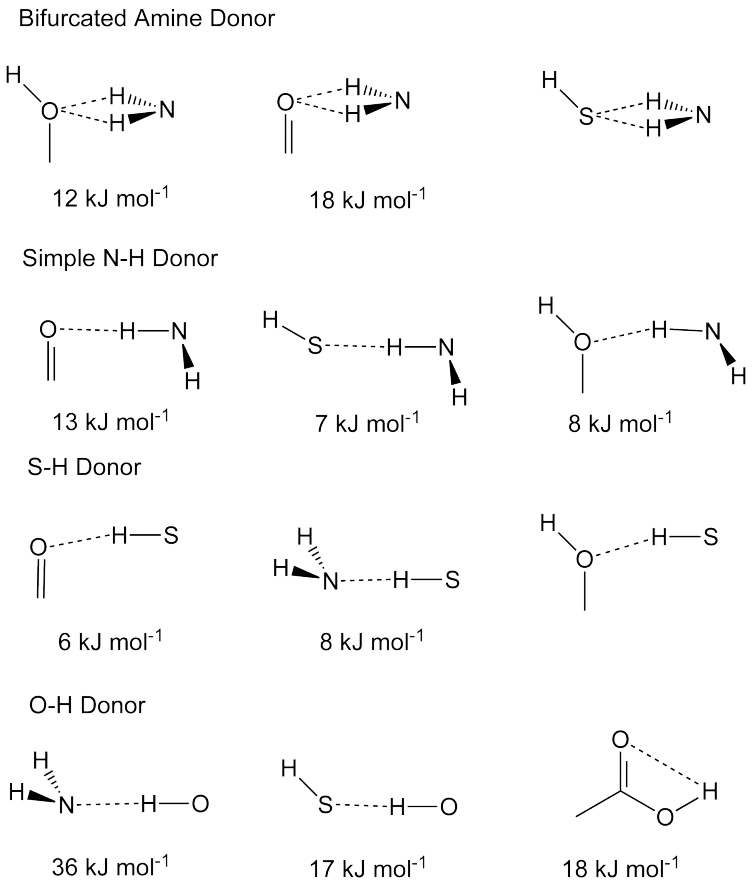

FIGURE 2.2: Common hydrogen bond motifs and approximate interaction strengths, where known.[14,18] Approximate relative energies of conformers can be treated additively as the difference between the sum of near-atom interactions.

within 10 kJ mol^{-1} using laser ablation and Fourier-transform microwave spectroscopy (FTMW). It is clear that the highest possible levels of electronic structure theory must be employed to obtain definitive energetics for these structurally diverse but energetically similar conformers.

The present study yielded, as primary information, accurate equilibrium structures and relative energies, as well as spectroscopic molecular parameters related to the vibrational and rotational spectra of the most important conformers of Cys. In turn, the large number of computed molecular properties allowed the investigation of a number of interesting computational issues. These include systematic errors in the geometries, bracketing the errors in relative energies for different theories, anharmonicity and zero-point vibrational corrections, and the electron correlation effects in prop-

erties such as quadrupole coupling constants. In particular, the focal point convergence is assessed, demonstrating the degree of correlation and basis set size necessary for achieving subchemical accuracy on the order of 1.0 kJ mol^{-1}.

2.3 Computational Details

Most of the atom-centered Gaussian basis sets selected for the electronic structure computations of this study contain both polarization and diffuse functions, both of which are needed for the determination of accurate structures and relative energies of H-bonded systems.[24] The subcompact 3-21G[25,26] basis lacks these functions, and thus it was used only for pre-screening the conformers at the Hartree-Fock[27] level of theory. The correlation-consistent, polarized-valence (aug)-cc-p(C)V(n+d)Z, n = 2(D), 3(T), 4(Q) basis sets of Dunning and co-workers[28–32] were employed extensively for geometry optimizations and single-point energy computations within the FPA approach.[7–11] The augmented (aug) basis sets contain diffuse functions, while tight functions necessary for treating core correlation are contained in the core-polarized (C) basis sets. In addition, the +d notation indicates a set of tight d-functions for second-row atoms to rectify problems with the originally designed correlation-consistent sets and thus smooth basis set extrapolations for sulfur-containing molecules. For Cys, the aug-cc-pV(D+d)Z, aug-cc-pV(T+d)Z, aug-cc-pV(Q+d)Z, and aug-cc-pV(5+d)Z basis sets contain 233, 492, 891, and 1458 CGFs, respectively. Only pure spherical harmonics were employed in all basis sets used in this study.

Electronic wave functions were determined in this study by the single-configuration, self-consistent-field, restricted Hartree-Fock (RHF) method,[27,33] by second-order Møller-Plesset perturbation theory (MP2),[34] and by coupled cluster (CC) methods[35,36] including all single and double excitations (CCSD)[37] as well a perturbative correction for connected triple excitations [CCSD(T)].[38] Additionally, energies and geometries were determined using the hybrid density functional B3LYP.[39–41] Both the T_1 diagnostics of coupled cluster theory[42,43] (0.014) and qualitative bonding principles indicate that the conformers of Cys are well described by single-reference correlation methods. The seven lowest 1s-like orbitals along with the sulfur 2s and 2p orbitals were kept as frozen core (FC) in all post-Hartree-Fock treatments.

The electronic structure program packages MAB-ACESII,[44] MPQC,[45–47] MOLPRO,[48] and Gaussian03[49] were used extensively in this study.

2.3.1 Geometry Optimizations

Initial structures for the geometry optimizations of the conformers of Cys were found by systematically varying the six most important dihedral angles (see Figure 2.1). The thiol carbon, τ($S_1$1-C_β-C_α-C_5), and amine groups, (H_3-N_1-C_α-C_β), were rotated in 30° increments, while the carboxylic acid plane, τ(O_6-O_7-C_5-C_α), thiol hydrogen, τ(H_{12}-S_{11}-C_β-C_α), carboxyl hydrogen, τ(H_8-O_7-C_5O_6), and C_α-C_β bond, τ(S_{11}-C_β-C_α-N_1), were rotated in 120° increments, resulting in a preliminary set of 11664 starting structures. The initial geometries were optimized at the HF/3-21G level until the Cartesian displacements between optimization steps were less than 10^{-4} bohr. Redundant conformers were identified by checking that energies and geometries were equivalent within a given threshold. Energies were considered to be the same if they were within 10^{-7} Eh, while bond lengths and angles were required to be within 0.001 Å and 1.0°, respectively. In total, a set of 90 unique HF/3-21G conformers were found.

The HF/3-21G structures were reoptimized at the frozen-core MP2/cc-pVTZ level. When MP2(FC)/cc-pVTZ geometry optimizations were performed, some of the higher-energy HF/3-21G conformers disappeared, yielding a final set of 71 conformers for Cys, according to the same uniqueness criteria given above. The eleven MP2/cc-pVTZ conformers within 10.0 kJ mol^{-1} of the lowest-energy structure were chosen for a more detailed analysis. These structures were reoptimized at the frozen-core MP2/aug-cc-pV(T+d)Z level. Following a scheme first employed for glycine,[5] the conformers are numbered by Roman numerals (see Figure 2.3), reflecting the energy ordering determined at the MP2(FC)/aug-cc-pV(T+d)Z level. Similarly to Dobrowolski,[22] we choose a series of dihedral angles to uniquely identify conformers, assigning each angle as C (cis, $30 < \tau < +30$), T (trans, $150 < \tau < 210$), G^+ (gauche, $+30 < \tau < +90$), G^- (gauche, $-90 < \tau < -30$), A^+ (anticlinal, $+90 < \tau < +150$), or A^- (anticlinal, $-150 < \tau < -90$). This notation is equivalent to the Klyne-Prelog specification,[50] but we use the terms cis, gauche, and trans instead of syn-periplanar, synclinal, and antiperiplanar, respectively. In particular, we identify the conformers

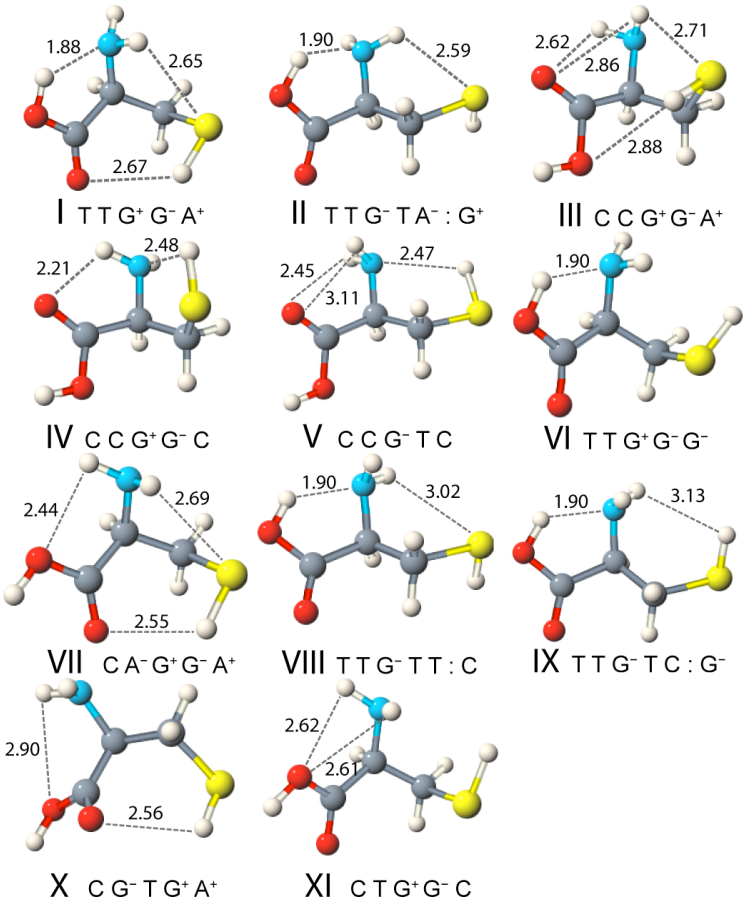

FIGURE 2.3: Pictorial representation of the eleven lowest-energy conformers of L-cysteine (Cys). See Figure 2.1 for numbering. The labeling scheme identifies conformers based on the series of dihedral angles τ(H_8-O_7-C_5-O_6), τ(N_1-C_α-C_5-O_6), τ(N_1-C_α-C_β-S_{11}), τ(C_5-C_α-C_β-S_{11}), and τ(N_1-C_α-S_{11}-H_6). A sixth identifier can be added for the dihedral angle of the nitrogen lone pair relative to the C_α-C_5 bond in cases of ambiguity.

via the dihedral angles τ(H_8-O_7-C_5O_6), τ(N_1-C_α-C_5-O_6), τ(S_{11}-C_β-C_α-N_1), τ(S_{11}-C_β-C_α-C_5), and τ(N_1-C_α-S_{11}-H_{12}). τ(H_8-O_7-C_5O_6) specifies the carboxyl group as cis or trans, τ(N_1-C_α-C_5-O_6) identifies the type of carboxyl-amine hydrogen bond, τ(S_{11}-C_β-C_α-N_1) and τ(S_{11}-C_β-C_α-C_5) define the backbone of the molecule, and τ(N_1-C_α-S_{11}-H_{12}) gives the orientation of the thiol hydrogen relative to the amine. For some conformers, e.g. Cys-II and Cys-VIII, the orientation of the amine lone pair relative to the C_5-C_α bond must also be specified to uniquely identify the conformer.

The 71 conformers located here exceed those of Dobrowolski et al.,[22] and Gronert and O'Hair,[18] who observed only 51 and 42 distinct Cys conformers, respectively. Their initial search tested 324

starting geometries at the AM1 level, resulting in a preliminary set of only 58 conformers. In contrast, our initial search was performed at the HF/3-21G level on an initial set of over 10,000 starting geometries, yielding a preliminary set of 90 conformers. Because it is highly likely that the 71 distinct MP2(FC)/cc-pVTZ conformers located here exist in reality, it seems that the 20 conformers were missed in previous work due to the AM1 preoptimization step.

2.3.2 Focal Point Analysis (FPA)

In order to obtain relative energies as accurately as possible, the focal point analysis (FPA) approach as outlined in the methods section was utilized. The eleven lowest-energy structures of Cys (Figure 2.3), obtained at the MP2(FC)/aug-cc-pV(T+d)Z level, were included in the FPA investigation. Extrapolation of the energies to the complete basis set (CBS) limit at the RHF and MP2 levels was performed, as part of the FPA approach. For HF, the total energy was extrapolated using the formula $E_n = E_{CBS} + A\exp(-Bn)$ with n=$\{3, 4, 5\}$ where A and B are adjustable parameters, E_n is the RHF total energy for a correlation-consistent basis set aug-cc-pV(n+d)Z, and E_{CBS} is the Hartree-Fock limit. For MP2, the correlation energies (E_n^{MP2}) were extrapolated according to $E_n^{MP2} = E_{CBS}^{MP2} + Bn^{-3}$.[51] Coupled-cluster energy corrections were treated additively, since the electron correlation contributions to the conformational energies at higher levels of theory do not change significantly with the size of the basis set.

Since the Born-Oppenheimer, non-relativistic Hamiltonian was employed, additional auxiliary corrections can be included where higher accuracy is desired. Relativistic corrections estimated via the scalar mass-velocity and Darwin terms[52] and diagonal Born-Oppenheimer corrections (DBOC)[53] were deemed negligible. In similar FPA studies performed on the conformers of proline[16] and threonine,[15] the relativistic corrections to the conformational energies turned out to be minuscule. Additionally, only valence electrons are correlated in the focal point treatment. Given the prohibitive cost of a rigorous treatment of core correlation at the CCSD(T) level, an estimate of core correlation effects was obtained at the MP2/aug-cc-pCVTZ level.

2.3.3 Spectroscopic Parameters

For the eleven lowest-energy conformers of Cys, quadratic force constants were computed at the B3LYP/aug-cc-pVTZ level using geometries optimized at the same level. Additionally, quartic force fields in the normal coordinate space were determined at the B3LYP/6-31G* level. No scaling of the force fields or of the resulting vibrational frequencies was attempted. The use of fully optimized reference geometries during the force field determinations helps to avoid the nonzero force dilemma.[54] Anharmonic, fundamental vibrational frequencies of the Cys conformers were computed by applying the VPT2 formalism[55] to the quartic force fields. Whenever a Fermi resonance appeared, the corresponding contribution to the fundamental frequency was evaluated by eliminating the associated terms in the expression for the anharmonic constants and then explicitly diagonalizing the 2 x 2 Hamiltonian matrix for the resonating states. The optimized structures obtained at the MP2(FC)/aug-cc-pVTZ level were used to determine the equilibrium rotational constants, while the quadratic and cubic force fields from B3LYP/6-31G* yield the quartic and sextic centrifugal distortion constants in the A-reduced representation, respectively. Addititionally, quadrupole coupling constants are reported at the MP2 level using a "locally dense" basis set composed of the standard cc-pVTZ functions on carbon, sulfur, oxygen, and hydrogen and the cc-pCV5Z functions on nitrogen. The quadrupole coupling constant is determined by the electric field gradient at nitrogen so that the combination of a locally dense basis and MP2 generally produces good results. This approach is justified with numerical results for ammonia later in the discussion.

2.3.4 Zero-Point Vibrational Corrections

Zero-point vibrational energies (ZPVE) were obtained as harmonic and anharmonic values at the B3LYP/aug-cc-pVTZ and B3LYP/6-31G* levels, respectively. Some ambiguity exists in the computation of zero-point energies based on anharmonic force fields. For molecules with many normal modes, resonances occur in the energy denominators of the related second-order vibrational perturbation theory (VPT2)[55] expressions. Here, we employ the approach of Allen et al.,[56] in which the oft-neglected G_0 term is included to obtain an expression for the zero-point vibrational energy

completely devoid of resonance denominators. The working equations are

$$ZPVE = \frac{1}{2}\sum_i \omega_i - \frac{1}{32}\sum_{ijk}\frac{\phi_{iik}\phi_{jjk}}{\omega_k} - \frac{1}{48}\sum_{ijk}\frac{\phi_{ijk}^2}{\omega_i+\omega_j+\omega_k} + \frac{1}{32}\sum_{iijj}\phi_{iijj} + Z_{kinetic} \tag{2.1}$$

and

$$Z_{kinetic} = -\frac{1}{4}\sum_\alpha B_e^\alpha \left[1 + \sum_{i>j}(\xi_{ij}^\alpha)^2 \frac{B_e^\alpha(\omega_i+\omega_j) - (\omega_i-\omega_j)^2}{\omega_i\omega_j}\right] \tag{2.2}$$

where the ϕ_{ijk} and ϕ_{ijkl} are cubic and quartic force constants in reduced normal coordinates, B_e^α denotes the rotational constant for axis α, and the ξ_{ij}^α are Coriolis coupling constants. The total zero-point vibrational energies obtained directly from Gaussian03 were quite similar to those from (2.1), but contained discrepancies as large as 30 cm^{-1} for some conformers. Both our method and Gaussian03 include the oft-neglected G_0 term,[57] which is a constant term involving a sum over higher-order vibrational force constants and vibration-rotation interaction terms. However, in contrast to our approach, Gaussian03 includes anharmonic effects by summing over the average of the harmonic and fundamental frequencies[58]

$$ZPVE = \frac{1}{4}\sum_i(\omega_i+\nu_i) + G_0 - \frac{1}{4}\sum_i x_{ii} \tag{2.3}$$

where x_{ii} is the diagonal anharmonic constant for mode i. The anharmonic corrections to the ZPVE in Gaussian 03 therefore avoid resonance denominators by including the explicit 2 x 2 matrix diagonalization for resonating states, in contrast to our approach which avoids resonance denominators by computing ZPVE directly in terms of cubic and quartic force constants.

Anomalously large VPT2 anharmonic shifts are observed for rotation of the S-H bond in some conformers, Cys-IV, Cys-VI, and Cys-XI. In general, for conformers IV, VI, and XI, the thiol hydrogen is near the amino group, and the thiol rotation is strongly coupled to the NH_2 wag. Because of this coupling, VPT2 breaks down for these modes, resulting in unphysically large anharmonic shifts. In (2.1) we exclude those force constants involving at least one mode for which VPT2 gives an anomalous fundamental frequency.

2.3.5 Natural Bond Orbital Analysis

The natural bond orbital (NBO) method transforms the molecular orbital picture into a localized description based on the intuitive Lewis structures of molecules.[59–63] The NBO scheme decomposes the 1-particle density matrix into formally occupied 1-center core orbitals and lone pairs (n_X) and 2-center bonding orbitals (σ_{X-Y}, $\pi_{X\text{-}Y}$), which correspond naturally to bonds drawn in a Lewis diagram. The localized, Lewis density is transformed to the exact density through delocalizations into formally unoccupied 2-center antibonding orbitals (σ^*_{X-Y}, $\pi^*_{X\text{-}Y}$) and unoccupied 1-center Rydberg orbitals. These delocalizations can be interpreted physically as energy-stabilizing donor-acceptor interactions between localized orbitals, including conjugation (e.g. $\pi \rightarrow \pi^*$) or hyperconjugation ($\sigma \rightarrow \sigma^*$). Applying perturbation theory gives the leading, second-order energy correction due to these donor-acceptor interactions as

$$E_2 = \frac{F_{ia}^2}{\epsilon_i - \epsilon_a} \tag{2.4}$$

where F_{ia} is the Fock matrix element between occupied orbital i and unoccupied orbital a, and ϵ_i and ϵ_a are the corresponding diagonal Fock matrix elements. Here i and a do not represent canonical doubly occupied and virtual molecular orbitals, but rather "almost doubly occupied" and "almost unoccupied" orbitals, respectively. In this regard, the Brillouin condition does not apply so that the matrix elements F_{ia}, while small, are not rigorously zero. Written this way, the E_2 values indicate the most important relaxation from an idealized, local density to the exact density. For cysteine, we can therefore assess the importance of hyperconjugation in conformational transitions through the NBO formalism. Because E_2 is based on perturbation theory, it may overestimate the absolute magnitude of strong interactions in which the energy denominator is small or the Fock matrix element is large. However, the overall trends should be consistent between conformers.

TABLE 2.1: Summary of focal point analysis for the relative energies of the eleven most stable conformers of cysteine[a]

	Basis	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
ΔE(RHF)[a]	DZ	0.00	3.82	-3.91	-0.09	-5.29	7.62	-2.85	3.23	4.57	-2.75	-0.17
	TZ	0.00	3.67	-4.00	-0.39	-5.40	7.22	-2.34	3.15	4.27	-2.52	-0.13
	QZ	0.00	3.59	-4.02	-0.50	-5.54	7.13	-2.33	2.97	4.06	-2.54	-0.17
	5Z	0.00	3.57	-4.04	-0.54	-5.59	7.11	-2.36	2.94	4.02	-2.57	-0.23
	CBS	0.00	3.56	-4.06	-0.55	-5.61	7.09	-2.39	2.93	4.01	-2.60	-0.25
δ[MP2][a]	DZ	+0.00	+2.20	+10.67	+8.37	+12.77	+1.80	+11.14	+7.24	+6.90	+12.64	+11.25
	TZ	+0.00	+2.83	+11.43	+8.98	+14.38	+2.07	+11.64	+7.87	+7.50	+14.46	+12.32
	QZ	+0.00	+2.94	+11.71	+9.00	+14.40	+2.28	+11.72	+7.82	+7.49	+14.60	+12.62
	5Z	+0.00	+2.98	+11.74	+8.98	+14.41	+2.32	+11.69	+7.84	+7.52	+14.65	+12.63
	CBS	+0.00	+3.02	+11.78	+8.95	+14.43	+2.37	+11.66	+7.86	+7.55	+14.71	+12.64
δ[CCSD]	DZ	+0.00	-0.66	-2.90	-1.84	-3.20	-1.03	-2.92	-2.60	-2.56	-3.32	-3.04
	TZ	+0.00	-0.70	-3.05	-2.03	-3.64	-1.00	-3.14	-2.83	-2.76	-3.60	-3.30
δ[CCSD(T)]	DZ	+0.00	+0.42	+1.88	+1.64	+2.04	+0.30	+1.80	+1.08	+1.01	+1.84	+1.85
	TZ	+0.00	+0.49	+2.06	+1.78	+2.33	+0.32	+1.96	+1.15	+1.07	+2.21	+2.10
ΔE CCSD(T)	CBS	0.00	6.37	6.73	8.15	7.51	8.78	8.09	9.11	9.87	10.72	11.19
core correction		+0.00	+0.13	+0.17	+0.04	+0.33	+0.04	+0.17	+0.21	+0.21	+0.33	+0.21
harmonic ZPVE		+0.00	-0.01	-2.19	-2.38	-2.24	-0.71	-1.73	-0.93	-1.00	-1.85	-1.76
anharmonic correction		+0.00	+0.08	+0.08	+0.14	+0.21	-0.55	-0.01	+0.21	+0.05	+0.32	-0.10
ΔE(FPA)		0.00	6.58	4.79	5.95	5.81	7.56	6.52	8.60	9.13	9.52	9.54

[a] All values given in kJ mol^{-1}. ΔE denotes a relative energy between conformers. δ denotes an incremement or correction to ΔE with respect to the preceding level of theory in the hierarachy RHF $\rightarrow$ MP2 $\rightarrow$ CCSD $\rightarrow$ CCSD(T).

2.4 Results and Discussion

2.4.1 Relative Energies of Conformers

As observed repeatedly for amino acids, the introductory Hartree-Fock level of electronic structure theory, independent of the basis set used, is unable to yield the correct relative energies of the conformers of Cys (Table 2.1). RHF/CBS theory places six conformers (Cys-III, -IV, -V, -VII, -X, and -XI) lower in energy than the global minimum, Cys-I. While in Gly and Ala the inclusion of electron correlation tends to decrease the energy differences between the conformers,[5,14] in the case of Cys it increases the energy differences in almost all cases, which might be attributed to the relatively weak interactions present in Cys. Because sulfur is much more polarizable than oxygen or carbon, the dipole-induced dipole and dispersion forces should be generally more important in Cys than in Ala, Ser, or Pro, while S-H hydrogen bonds should be weaker. Accordingly, the MP2

correlation energy destabilizes all the conformers considered relative to Cys-I, as signified by the positive δ[MP2] values in Table 2.1, which can be as large as 15 kJ mol^{-1}. This observation also serves as a warning that the theoretical results obtained with small basis sets and simple electronic structure methods might change considerably once more rigorous techniques are employed. Compared to δ[MP2], the δ[CCSD] and δ[CCSD(T)] energy increments are relatively small, but can affect relative energies as much as 3.6 and 2.2 kJ mol^{-1} for CCSD and CCSD(T), respectively. Such amounts are clearly substantial when so many conformers are within a window of a few kJ mol^{-1}. Interestingly, but again in line with earlier work on the conformers of Thr[15] and Pro,[16] the relative energies are barely affected by core correlation. Even the largest change is smaller than 0.35 kJ mol^{-1}. In contrast, ZPVE corrections can affect relative energies on the order of 2 kJ mol^{-1}. Anharmonic corrections to the relative energies were less than 0.21 kJ mol^{-1} for all conformers except Cys-VI and Cys-X, for which these shifts were 0.55 and +0.32 kJ mol^{-1}, respectively.

TABLE 2.2: Comparison of conformational energies of L-cysteine (kJ mol^{-1}) for different levels of theory without zero-point vibrational correction

	CCSD(T)/CBS[a,b]	MP2/aV(T+d)Z[b]	CCSD/aV(T+d)Z[b]	RHF/3-21G[c]	B3LYP/aVTZ[d]
Cys-I	0.00	0.00	0.00	1.26	0.00
Cys-II	6.37	6.49	5.82	11.84	3.93
Cys-III	6.73	7.45	4.35	4.35	7.15
Cys-IV	8.15	8.58	6.57	2.43	8.79
Cys-V	7.51	9.00	5.36	0.00	5.44
Cys-VI	8.78	9.29	8.28	12.68	7.53
Cys-VII	8.09	9.29	6.15	5.23	6.36
Cys-VIII	9.11	11.00	8.16	9.08	6.36
Cys-IX	9.87	11.76	9.00	11.30	7.15
Cys-X	10.72	11.97	8.37	13.05	8.91
Cys-XI	11.19	12.18	8.87	10.04	11.09

[a] CCSD(T)/CBS denotes the extrapolated value from the focal point analysis. See Table 2.1. [b] Computed at the MP2/aug-cc-pV(T+d)Z reference geometries. [c] Computed at the RHF/3-21G reference geometries. [d] Computed at the B3LYP/aug-cc-pVTZ reference geometries.

The MP2(FC)/aug-cc-pV(T+d)Z single-point energies happen to be quite accurate (Table 2.2) because the CCSD and CCSD(T) increments are usually of opposite sign and partially cancel (Table 2.1). In this regard, many of the MP2 relative energies in Table 2.2 are closer to the FPA results than their CCSD counterparts. The effect of higher-order correlation from the CCSD(T) perturbative triples is not negligible, however, and the definitive FPA scheme alters the MP2 energy ordering of the conformers. In general, B3LYP performs reasonably well for most conformers, but can be

in error by as much as 2.7 kJ mol^{-1}, as seen in Cys-VIII. While density functional theory can be useful for zero-point corrections and geometry optimizations, obtaining the correct energy ordering of so many conformers in such a small energy range (10 kJ mol^{-1}) clearly requires better accuracy than B3LYP provides. A rigorous energy ordering therefore necessitates correlation treatments as extensive as CCSD(T) and also considerations inherent in the FPA scheme. The FPA scheme allows errors to be bracketed based on the observed convergence to the basis set and correlation limits. The RHF relative energies are converged to better than 0.05 kJ mol^{-1} with the aug-cc-pV(5+d)Z basis, and thus there should be virtually no basis set error in our final RHF/CBS results. Similarly, the MP2/aug-cc-pV(5+d)Z correlation increments match the extrapolated values within 0.12 kJ mol^{-1} for all conformers, and the associated basis set errors should again be negligible. For the coupled-cluster increments, the aug-cc-pV(T+d)Z result matches the aug-cc-pV(D+d)Z result within 0.4 kJ mol^{-1}. In previous work, the CCSD and CCSD(T) increments are essentially converged with a TZ basis,[64] so that the basis set error in the coupled-cluster values should not be greater than 0.2 kJ mol^{-1}. Assesssing the error due to higher-order correlation and zero-point vibrational corrections is more difficult. Corrections due to quadruple and higher excitations are generally an order of magnitude less than CCSD(T) corrections.[64,65] Since most CCSD(T) corrections here are on the order of 12 kJ mol^{-1}, the neglect of higher excitations should introduce an error to the cysteine relative energies of at most 0.2 kJ mol^{-1}. Zero-point vibrational corrections are generally insensitive to the level of theory. For example, even MP2 harmonic zero-point corrections with the Huzinaga-Dunning DZP++ basis[66] (double$-\zeta$ plus polarization and diffuse functions) match the B3LYP/aug-cc-pVTZ values within 0.3 kJ mol^{-1}. Since we have accounted for anharmonicity in the present work, the zero-point error should therefore not be greater than 0.2-0.3 kJ mol^{-1}. In summary, we estimate a standard error (1σ) of 0.3 kJ mol^{-1} for our predicted conformational energies, corresponding to a 95% confidence interval (2σ) of $\pm$ 0.6 kJ mol^{-1}. We emphasize that these uncertainties hold only for the relative energies due to cancellation of errors, and the uncertainty in the absolute energies will therefore be much larger.

2.4.2 Geometric Structures

The sophisticated laser ablation FTMW experiments of Alonso et al.[23] yielded rotational constants of several conformers of cysteine, but only for the parent isotopologues. Therefore, the type of refinement on collections of isotopologues which has yielded semiexperimental equilibrium structures for Gly[13] and Pro[17] cannot be executed at present for any of the conformers of Cys. Consequently, one must rely on otherwise rather accurate[14,17] computed structures when analyzing structural trends among the conformers of Cys. Two major factors seem to determine the general type of conformation that Cys can assume. First, the thiol, amine, and hydroxyl groups adopt different orientations about the C_α-C_β bond as either gauche or trans. In the discussion to follow, unless stated otherwise, gauche and trans identify the orientation about the C_α-C_β bond. Second, the carboxyl group may assume a cis or trans conformation. Depending on these orientations, different hydrogen bonding patterns can form of the types O-H$\cdots$N, N-H$\cdots$O=C, and N-H$\cdots$OH, as clearly seen in Figure 2.3. For the trans carboxyl, a strong O-H$\cdots$N interaction can form as found in conformers Cys-I, Cys-II, Cys-VI, Cys-VIII, and Cys-IX. For the cis carboxyl, bifurcated N-H$\cdots$O bonds form to either the carbonyl oxygen in Cys-III, Cys-IV, and Cys-V or to the hydroxyl oxygen in Cys-VII and Cys-XI. The gauche conformers are more sterically crowded than the trans conformers. However, the gauche conformation brings the thiol group closer to the carboxyl group, allowing S-H$\cdots$O interactions. A ring of hydrogen bonds can therefore form, as in Cys-I, Cys-III, and Cys-VII. For trans conformers, only the amine interacts with the thiol, for example in Cys-II. As observed previously for other amino acids,[14,16] bond lengths and bond angles change little among the low-lying conformers. Most bonds have a standard deviation of less than 0.004 Å while most bond angles have a standard deviation of less than 2.0°. There are, however, a few notable exceptions. The C=O distance has a standard deviation of 0.007 Å with the largest deviation from the mean being 0.012 Å, occurring in Cys-III. The C=O bond in Cys-III forms a bifurcated hydrogen bond with the amine group, lengthening the bond and redshifting the carbonyl stretching frequency. The C_α-C_5 has a standard deviation of 0.008 Å, with the largest deviation from the mean (0.012 Å) occurring in Cys-IV. In general, three strong hyperconjugative interactions are possible for the C_α-C_5 bond with the lone pairs from nitrogen, the hydroxyl oxygen, and

the carbonyl oxygen. The hyperconjugation is strongest when the lone pair is trans to the C_α-C_5 bond (see below), and it will lengthen this bond by increasing the antibond occupation. In Cys-IV, both the amine and hydroxyl groups are unfavorably placed for hyperconjugation, and the C_α-C_5 bond distance is only 1.514 Å. In contrast, both the hydroxyl and amine are favorably placed for hyperconjugation in Cys-I, lengthening the C_α-C_5 bond to 1.534 Å. Similarly, in Cys-III the amine lone pair is trans to the C_α-C_5 bond, but the hydroxyl lone pair is unfavorably placed cis. The C_α-C_5 bond therefore has an intermediate value of 1.521 Å.

For bond angles, the largest standard deviations are for the N-C-C angles. The standard deviation for N_1-C_α-C_β is 2.8° with the largest deviation from the mean being 4.0° for Cys-VI. Similarly, the standard deviation for N_1-C_α-C_5 is 2.8° with the largest deviation from the mean value being 5.4° for Cys-XI. The large spread of N-C-C angles is consistent with the trans angle rule.[67] In general, in primary alcohols and amines, if a C-C bond is trans to the X-H bond, the X-C-C angle will be smaller, owing to both reduced bond repulsion relative to the gauche conformer and weaker hyperconjugation from the nitrogen lone pair. The change in angle depending on bond orientation is clearly evident in cysteine. For conformers Cys-I, Cys-VI, Cys-VIII, and Cys-IX, the N-H bonds are gauche to the C_α-C_β bond, and the angles are in the range 115°-117°. In contrast, for conformers Cys-II, -III, -IV, -V, -VII, -X, and -XI, the N-H bond is trans to the C_α-C_β bond, and the N-C-C angles range from 109° to 111°. The same general trends are observed for the N_1-C_α-C_5 angles. The angle variations are consistent with strong hyperconjugation from the nitrogen lone pair to the C-C antibond ($n_N \rightarrow \sigma^*_{C-C}$). Following arguments rationalizing tilting of the methyl group,[68] the C-C axis generally tilts away from the C-N axis to maximize overlap between the nitrogen lone pair and the backside lobe of the C-C antibond, strengthening the hyperconjugative stabilization. In particular, for both C_α-C_5 and C_α-C_β large hyperconjugative interactions ($E_2 > 20.0$ kJ mol^{-1}) are observed with large C-C-N angles, while weaker interactions ($E_2 < 16.0$ kJ mol^{-1}) are observed with smaller C-C-N angles.

In the same way for the carboxyl group, if the carboxyl group assumes a cis conformation (O-H bond trans to the C-H bond), the O-C-C angle is much smaller. This is observed in conformers Cys-III, Cys-IV, Cys-V, and Cys-VII, all of which have O-C-C angles of approximately 111.5°. In

contrast, for conformers Cys-I, Cys-II, Cys-VI, and Cys-VIII with trans carboxyl (O-H bond cis to the C-C bond), the angles are larger, between 113.5° and 114°. In general, the trans effect seems weaker for the O-H bond than for the N-H bond. The weaker dependence may be attributed to the stronger basicity of the amine and therefore stronger n $\rightarrow \sigma^*$ hyperconjugation. The O-H$\cdots$N hydrogen bonding also seems to offset the trans effect, closing the O-C-C angle to maximize the O-H$\cdots$N interaction.

TABLE 2.3: Summary of hyperconjugative interactions in kJ mol^{-1} for varying conformations of L-cysteine about the C_α-C_β bond. See (2.4) in text.

	X-C-C-Y	E$_2$ $\sigma_{X-C} \rightarrow \sigma^*_{C-Y}$	E$_2$ $\sigma_{Y-C} \rightarrow \sigma^*_{C-X}$	<X-C-C	<C-C-Y	τ(X-C-C-Y)
			Gauche Conformers			
	C_5-C_α-C_β-H_{10}	7.8	12.1	109.7	109.9	175.5
I	H_4-C_α-C_β-S_{11}	20.7	9.6	108.4	112.4	178.4
	N_1-C_α-C_β-H_9	3.8	21.8	115.6	110.3	179.4
	C_5-C_α-C_β-H_{10}	6.4	13.7	110.7	108.1	-179.3
III	H_4-C_α-C_β-S_{11}	22.1	8.5	107.0	113.3	-178.0
	N_1-C_α-C_β-H_9	5.5	16.4	110.2	111.3	-172.4
	C_5-C_α-C_β-H_{10}	6.2	13.2	110.8	109.0	-179.2
IV	H_4-C_α-C_β-S_{11}	20.0	9.0	107.5	112.0	-172.3
	N_1-C_α-C_β-H_9	5.4	17.4	110.7	110.5	-177.6
	C_5-C_α-C_β-H_{10}	8.0	11.2	109.6	111.1	169.5
VI	H_4-C_α-C_β-S_{11}	21.1	9.5	108.2	112.3	178.0
	N_1-C_α-C_β-H_9	3.6	23.0	116.2	109.0	174.1
	C_5-C_α-C_β-H_{10}	7.1	13.1	109.7	108.4	176.0
VII	H_4-C_α-C_β-S_{11}	21.9	9.1	107.5	113.1	179.3
	N_1-C_α-C_β-H_9	5.4	17.5	110.3	111.0	-175.1
	C_5-C_α-C_β-H_{10}	7.9	11.6	108.3	109.8	170.4
XI	H_4-C_α-C_β-S_{11}	21.6	9.0	107.1	112.7	179.5
	N_1-C_α-C_β-H_9	5.2	18.1	110.6	109.8	179.3
			Trans Conformers			
	C_5-C_α-C_β-S_{11}	17.6	7.7	106.3	113.7	-179.4
II	H_4-C_α-C_β-H_{10}	11.9	12.1	108.8	109.6	-177.3
	N_1-C_α-C_β-H11	4.8	17.6	110.2	109.6	175.6
	C_5-C_α-C_β-S_{11}	14.4	8.8	107.4	112.1	172.2
V	H_4-C_α-C_β-H_{10}	11.3	12.1	108.1	109.3	177.6
	N_1-C_α-C_β-H11	5.5	16.6	109.9	110.3	177.2
	C_5-C_α-C_β-S_{11}	14.9	9.5	110.0	113.0	174.7
VIII	H_4-C_α-C_β-H_{10}	12.4	13.2	108.6	109.7	177.2
	N_1-C_α-C_β-H11	3.9	21.5	115.6	110.4	174.9
	C_5-C_α-C_β-S_{11}	15.1	9.2	109.7	113.3	170.8
IX	H_4-C_α-C_β-H_{10}	12.6	12.5	108.4	110.6	179.9
	N_1-C_α-C_β-H11	3.8	22.9	115.8	109.2	177.3
	C_5-C_α-C_β-H_{10}	7.1	13.3	110.6	108.5	-179.8
X	H_4-C_α-C_β-H_9	11.8	12.8	109.8	110.4	179.1
	N_1-C_α-C_β-S_{11}	11.6	15.2	109.1	113.2	-176.3

2.4.3 Structural Effects on Relative Energies

As emphasized previously for alanine[14] and serine,[18] approximate values for the strength of certain types of hydrogen bonds can be computed and used to rationalize energy differences among the conformers. Specifically, the energy of each conformer can be approximated as a sum of stabilizations from near-atom interactions. The interaction strengths are then fit through a linear regression to match as closely as possible the conformational energies. Approximate hydrogen bond strengths are given in Figure 2.2 for the common bonding motifs, as reported in Ref 14. In general, the hydrogen bond donors can be ranked in the order O-H > N-H > S-H and the hydrogen bond acceptors can be ranked in the order N > O > S. Additionally, an additive approximation can be applied to the conformation of the carboxyl group. Based on the formic acid prototype,[14] the cis carboxyl is intrinsically more stable than the trans carboxyl by approximately 18.5 kJ mol^{-1} irrespective of hydrogen bonds to other functional groups.

As shown in Figure 2.2, the O-H$\cdots$N arrangement is the strongest hydrogen bond, matching the strongest hydrogen bond donor, OH, with the best acceptor, N. We therefore find that the structure of the global minimum, Cys-I, is stabilized by a strong O-H$\cdots$N hydrogen bond between the amino and carboxyl groups. The Cys global minimum is in contrast to that of serine,[18] for which the lowest energy conformer is analagous to Cys-V, exhibiting a strong O-H$\cdots$N hydrogen bond with the side chain. The S-H bond in Cys is comparatively a much weaker hydrogen bond donor than the O-H sidechain in serine so that the trans to cis isomerization of the carboxyl is not enough to offset the weaker hydrogen bond. Additionally, Cys-I, because of its gauche conformation, can form three hydrogen bonds whose charge polarization will cooperatively reinforce each other. We emphasize that our focal point conformational energies should be accurate within a standard error of 0.3 kJ mol^{-1} (1σ) or a 95% confidence interval of $\pm$ 0.6 kJ mol^{-1} (2σ). In contrast to previous studies,[22,23] we can therefore definitively say that the energy differences are real physical effects rather than errors in the underlying computational methods.

In addition to hydrogen bonding, gauche and trans conformations are also affected by steric repulsion and hyperconjugation. The gauche conformers of cysteine have all three bulky substituents in close vicinity, increasing steric repulsion. For butane, the trans-gauche difference is

2.6 kJ mol^{-1}.[69] Consequently, steric effects are certainly not negligible in Cys since its eleven lowest-energy conformers lie in an energy range of 10 kJ mol^{-1}. Hyperconjugation is stronger in the gauche configuration since the strongly electronegative groups (amine, carboxyl, thiol) are all antiperiplanar to the electropositive hydrogens. In this regard, the better electron donor orbitals ($\sigma_{\mathrm{C-H}}$) are matched to the better electron acceptors ($\sigma^*_{\mathrm{C-N}}$, $\sigma^*_{\mathrm{C-S}}$). The leading hyperconjugative interactions in cysteine conformers are listed in Table 2.3. For example, the $\sigma_{\mathrm{C-H}} \rightarrow \sigma^*_{\mathrm{C-S}}$ interaction is 20.7 kJ mol^{-1} in the gauche Cys-I while the equivalent $\sigma_{\mathrm{C-H}} \rightarrow \sigma^*_{\mathrm{C-H}}$ interaction is only 11.3 kJ mol^{-1} in Cys-V, although the effect is offset somewhat by stronger hyperconjugation $\sigma_{\mathrm{C-C}} \rightarrow \sigma^*_{\mathrm{C-S}}$ in Cys-V relative to the $\sigma_{\mathrm{C-H}} \rightarrow \sigma^*_{\mathrm{C-H}}$ hyperconjugation in Cys-I. This gauche effect has been observed previously for difluorosubstituted hydrocarbons and hydroxyproline.[70,71] Hyperconjugation and steric effects will therefore tend to offset each other. For some conformers, the relative energies will therefore be dictated mainly by the hydrogen bonding interactions, owing to fortuitious cancellation of competing electronic effects.

The importance of hyperconjugation can be seen in the transformation from Cys-IV to Cys-V, wherein the amine switches from a simple N-H$\cdots$O bond to a bifurcated N-H$\cdots$O bond while the orientation about the C_α-C_β bond simultaneously goes from gauche to trans. Assuming simple hydrogen bond additivity, Cys-V should lie 5 kJ mol^{-1} below Cys-IV due to the larger bifurcated hydrogen bond energy (Figure 2.2). In fact, the energy difference is less than 1 kJ mol^{-1}. The S-H$\cdots$N bond distances are basically equivalent in Cys-IV and Cys-V (2.48 Å versus 2.47 Å), so the S-H$\cdots$N interaction should not contribute significantly to the energy difference. The discrepancy seems to lie in hyperconjugative stabilization of Cys-IV. While steric repulsion is greater in Cys-IV, the much stronger $\sigma_{\mathrm{C-H}} \rightarrow \sigma^*_{\mathrm{C-S}}$ and $\sigma_{\mathrm{C-H}} \rightarrow \sigma^*_{\mathrm{C-C}}$ donations in Cys-IV relative to the $\sigma_{\mathrm{C-C}} \rightarrow \sigma^*_{\mathrm{C-S}}$ and $\sigma_{\mathrm{C-S}} \rightarrow \sigma^*_{\mathrm{C-C}}$ delocalizations in Cys-V (Table 2.3) preferentially stabilize Cys-IV. Perhaps the most interesting conformer is Cys-X, which forms no typical hydrogen bonds in the sense of near-linear X-H$\cdots$Y arrangements. The carboxyl plane is perpendicular to the C-N bond. In this way, the amine seems to form both weak N-H$\cdots$O-H and N-H$\cdots$O=C interactions. Despite positioning the C-N bond trans to the C-S bond, Cys-X exhibits strong $\sigma_{\mathrm{C-S}} \rightarrow \sigma^*_{\mathrm{C-N}}$ hyperconjugation, contributing to its unusual stability.

TABLE 2.4: Anharmonic vibrational fundamentals in cm^{-1} and double-harmonic infrared relative intensities (%) of the eleven lowest-energy conformers of L-cysteine for regions characteristic of particular hydrogen bond patterns[a]

	O-H, N-H stretch			S-H Stretch	C=0 Stretch	O-H bend		C-O stretch	
	ν_1	ν_2	ν_3	ν_7	ν_8	ν_{11}	ν_{12}	ν_{17}	ν_{18}
Cys-I	3397 (4)	3278 (18)	3216 (47)	2542 (0)	1793 (75)	1356 (100)	1353 (2)	1129 (2)	1073 (3)
Cys-II	3383 (6)	3300 (15)	3226 (43)	2543 (0)	1807 (87)	1337 (100)	1337 (8)	1144 (3)	1068 (3)
Cys-III	3539 (22)	3387 (4)	3327 (1)	2551 (0)	1776 (100)	1350 (0)	1314 (3)	1121 (2)	1086 (78)
Cys-IV	3554 (30)	3417 (8)	3374 (5)	2557 (1)	1775 (100)	1385 (4)	1314 (9)	1114 (69)	1084 (22)
Cys-V	3541 (21)	3397 (3)	3324 (1)	2552 (1)	1773 (100)	1380 (2)	1278 (15)	1124 (15)	1084 (58)
Cys-VI	3397 (4)	3378 (54)	3238 (12)	2539 (0)	1798 (84)	1349 (100)	1353 (12)	1130 (4)	1065 (2)
Cys-VII	3545 (28)	3407 (5)	3392 (1)	2548 (0)	1762 (100)	1350 (0)	1330 (8)	1114 (69)	1102 (22)
Cys-VIII	3396 (3)	3305 (10)	3244 (62)	2548 (0)	1794 (100)	1336 (59)	1366 (82)	1120 (3)	1084 (2)
Cys-IX	3410 (2)	3395 (18)	3254 (32)	2525 (0)	1798 (71)	1332 (1)	1362 (100)	1112 (5)	1088 (1)
Cys-X	3538 (25)	3398 (2)	3301 (0)	2560 (0)	1766 (100)	1344 (10)	1354 (0)	1111 (90)	1056 (21)
Cys-XI	3538 (27)	3401 (3)	3315 (0)	2545 (0)	1767 (100)	1359 (2)	1309 (10)	1122 (70)	1102 (18)

[a] Harmonic frequencies and intensities were computed at the B3LYP/aug-cc-pVTZ level. Anharmonic corrections were computed using B3LYP/6-31G*. Intensities are reported as a percentage of the most intense peak for a given conformer.

2.4.4 Vibrational fundamentals

Anharmonic vibrational frequencies in characteristic hydrogen bonding regions of the infrared spectra are given in Table 2.4 for the eleven lowest-energy conformers of Cys. The combination of B3LYP/aug-cc-pVTZ harmonic frequencies and B3LYP/6-31G* anharmonic corrections employed here should be accurate on average to within 30 cm^{-1}, although some deviations may be substantially larger.[72] Various fundamentals can be found where either the intensities or the frequencies distinguish between structural features. For example, Cys-I is the only conformer that has all of the following: two medium/strong bands between 3200-3300 cm^{-1}, no strong bands above 3300 cm^{-1}, and only two medium/weak bands below 1300 cm^{-1}. In matrix isolation experiments on Cys, Dobrowolski et al.[73] noted that broadening and clustering of peaks often prevented assignment of absorptions to individual conformers. In general, the matrix isolation vibrational spectra were only able to distinguish between conformers as being with or without certain intramolecular hydrogen bonds. As a natural extension, Dobrowolski et al. suggested that vibrational circular dichroism (VCD) may be more informative. Because the relatively accurate computation of VCD spectra is straightforward, such experiments hold promise for the detection of conformers present in the gas before matrix deposition.

Despite possible difficulties, four different regions should provide important fingerprints for cysteine conformers to distinguish between O-H$\cdots$N or N-H$\cdots$O=C bonding patterns. No conformers with strong N-H$\cdots$O-H hydrogen bonds appeared in the current study. The first region is the O-H stretch region between 3200-3600 cm^{-1}. For conformers such as Cys-I, Cys-II, Cys-VI, and Cys-VIII with O-H$\cdots$N hydrogen bonds, the O-H stretches are red-shifted so that the highest frequency peaks are the N-H stretches near 3400 cm^{-1}. For free hydroxyl groups, the O-H stretch appears near 3540 cm^{-1}, as seen in conformers Cys-III, Cys-IV, Cys-V, Cys-VII, Cys-X, and Cys-XI. This general band structure was observed in the matrix IR study,[73] and our anharmonic fundamentals agree with the experimental absorptions within 10-20 cm^{-1}, consistent with the uncertainty estimate given above. As noted by Dobrowolski,[73] the C-O single-bond stretch also provides an important diagnostic through differing intensitities. In conformers Cys-I, Cys-II, Cys-VI, and Cys-VII with O-H$\cdots$N bonds, the C-O stretching region does not exhibit any high intensity peaks, presumably because the oscillator strength is smeared out among several modes. In contrast, for hydroxyl groups that do not form hydrogen bonds, the C-O stretch in Cys-III, Cys-IV, Cys-V, Cys-VII, Cys-X, and Cys-XI shows strong features between 1080 and 1125 cm^{-1}, again in excellent agreement with the experimentally observed frequencies. As usual, the most telling band is the carbonyl stretch, ν_8, which ranges from 1762-1807 cm^{-1}. For conformers Cys-III, Cys-IV, Cys-V, Cys-VII, Cys-X, and Cys-XI, which contain a N-H$\cdots$O=C hydrogen bond, the stretch is red-shifted, appearing between 1762-1776 cm^{-1}. In contrast, the free carbonyls in Cys-I, Cys-VI, Cys-VII, and Cys-IX all appear in a narrow range at 1795 cm^{-1}. The highest frequency occurs for Cys-II at 1807 cm-1. However, no peak appears above 1800 cm-1 in the experimental spectrum.[73] Cys-II is essentially identical to Cys-VIII except for 60° rotation of the amine group to form an N-H$\cdots$S hydrogen bond, but the carbonyl stretch for Cys-VIII appears at 1794 cm^{-1}. It is therefore very surprising both that Cys-II and Cys-VIII have such different stretching frequencies, and also that Cys-II, one of the lowest energy conformers, is absent from the matrix.

The S-H stretch varies only over a narrow range of 2539-2560 cm^{-1}. This is consistent with the geometries since S-H seems to form only very weak hydrogen bonds. In general, the peaks are also predicted to be quite weak, so the peak is not likely to be useful in distinguishing conformers.

TABLE 2.5: Equilibrium and ground-state rotational constants, $\{A_e, B_e, C_e\}$ and $\{A_0, B_0, C_0\}$, respectively, quartic centrifugal distortion constants in the A-reduced representation, and dipole moments (μ) of the 11 lowest-energy conformers of L-cysteine.[a] Where available, experimental values are given in parentheses.[23] Ground-state rotational constants are determined from MP2(FC)/aug-cc-pVTZ equilibrium structures with anharmonic B3LYP/6-31G* corrections.

Gauche Conformers						
	Cys-I	Cys-III	Cys-IV	Cys-VI	Cys-VII	Cys-XI
A_e	3059.2	2886.6	2855.8	3094.6	3182.5	2984
B_e	1641.1	1654.3	1715.8	1605.7	1607.4	1610.3
C_e	1357.6	1390.6	1430.2	1335.4	1302.7	1370
A_0	3039.4 (3071.1)	2852.1 (2889.4)	2812.4	3055.1	3174.6 (3216.2)	2953.9
B_0	1623.0 (1606.5)	1642.6 (1623.0)	1698.8	1601.9	1587.9 (1572.7)	1605.2
C_0	1344.5 (1331.8)	1383.6 (1367.8)	1421.0	1331.3	1287.8 (1276.8)	1360.5
D_{JK}	434	449	891	1054	2621	1416
D_J	471	628	628	614	562	749
D_K	1757	2049	2014	2809	9180	5164
μ_a	1.55	1.00	1.08	2.52	2.06	1.22
μ_b	4.08	1.34	0.73	4.94	0.47	1.84
μ_c	1.27	1.45	2.51	1.74	0.04	0.18
Trans Conformers						
	Cys-II	Cys-V	Cys-VIII	Cys-IX	Cys-X	
A_e	4376.2	4235.8	4534.7	4505.9	2977.2	
B_e	1189.3	1195	1181.2	1185.4	1558.3	
C_e	1028.6	1018.5	972.8	970.7	1234.9	
A_0	4341.7 (4359.2)	4204.9 (4235.6)	4502.1	4462.2	2977.1 (3004.2)	
B_0	1180.1 (1178.3)	1187.1 (1187.3)	1172.0	1176.4	1538.7 (1527.4)	
C_0	1018.5 (1015.3)	1007.5 (1003.1)	965.2	963.2	1218.3 (1210.7)	
D_{JK}	898	1036	706	763	1145	
D_J	89	99	59	59	438	
D_K	835	1379	1028	1187	5316	
μ_a	2.84	1.69	2.56	2.46	2.30	
μ_b	2.62	0.34	2.83	3.80	0.36	
μ_c	1.16	-0.66	0.29	0.66	0.24	

[a] Rotational constants given in MHz and belong to the structures optimized at the MP2(FC)/aug-cc-pV(T+d)Z level. Quartic centrifugal distortion constants are given in Hz. Dipole moments are given in Debye and computed at the B3LYP/aug-cc-pVTZ level.

2.4.5 Rotational spectra

Alonso et al.[23] recently reported the identification of six low-energy conformers of Cys through Fourier transform microwave spectroscopy (FT-MW), providing a good opportunity here to compare the computed and the experimental results. Computed rotational constants, centrifugal distortion constants, and dipole moments are reported in Table 2.5 along with experimental values, where available. Some of the conformers (e.g., Cys-I, Cys-II, and Cys-VI) have substantial dipole

moments along the principal axes, thus helping (a) the observation of the related rotational transitions, and (b) the assignment of conformers based on information about which rotational constants correspond to intense transitions. Generally, we can divide the cysteine conformers into two groups based on the orientation about the C_α-C_β bond. In gauche conformers, the thiol is gauche to both the amine and carboxyl groups. The overall geometry is therefore more compact about the B axis, which is reflected in the larger B_e rotational constants for Cys-I and Cys-III in comparison to Cys-II and Cys-V (Table 2.5). In contrast, in a trans conformation, the thiol is positioned antiperiplanar to either the amine or carboxyl group. The overall geometry for the trans conformers is therefore extended along the A axis, leading to much larger A_e rotational constants for Cys-II and Cys-V in comparison to Cys-I and Cys-III.

At the MP2(FC)/aug-cc-pV(T+d)Z level, the rotational constants corresponding to the optimized equilibrium structures should be accurate enough to be useful to deduce the presence of conformers when interpreting experimental microwave spectra. As seen in Table 2.5, this is indeed the case. The mean absolute deviations from experiment are 15, 25, and 21 MHz for A_e, B_e, and C_e, respectively. Vibrational corrections for the rotational constants can be evaluated within the realm of second-order vibrational perturbation theory (VPT2)[55] by taking one-half the sum of the vibration-rotation interaction constants, α_i. The mean absolute deviations for B_0 and C_0 become only 11 and 9 MHz.

For the A axis, the theoretical rotational constants consistently underestimate the experimental ones by as much as 40 MHz. For most conformers, the cysteine molecule lies along the A axis, with the C-S bond running roughly perpendicular along the C axis. The moment of inertia about the A axis is greatly affected by the position of the sulfur atom, and A_0 is therefore very sensitive to the C-S bond length. If the C-S bond length is systematically overestimated, then the computed A_0 constants will be too small. For example, shortening the C-S bond length by 0.005 Å in Cys-I increases A_e by 30 MHz, bringing the corresponding A_0 into nearly exact agreement with experiment. Such bond length discrepancies may be attributed to a number of factors, including neglect of core correlation, basis set incompleteness, or higher excitations that would be included in CCSD or CCSD(T) geometry optimizations. The errors in B_0 and C_0 also appear to be systematic,

with both being consistently overestimated. In all cases, the zero-point vibrational corrections lower the rotational constant, consistent with the vibrationally averaged bond lengths being longer than their equilibrium values. Zero-point vibrational corrections to the rotational constants therefore improve agreement for B_0 and C_0, but actually diminish the agreement for A_0. The source of the systematic error for B_0 and C_0 is more difficult to assess than for A_0, especially without the empirical refinement that was performed for conformers of glycine and proline.[12,13,15,17]

TABLE 2.6: Basis set and electron correlation dependence of computed ammonia quadrupole coupling constants. Values are given in MHz.

		All Electron			Frozen Core		
Basis	HF	MP2	CCSD	CCSD(T)	MP2	CCSD	CCSD(T)
cc-pVDZ	5.05	4.58	4.71	4.69	4.59	4.71	4.69
cc-pCVDZ	5.16	4.71	4.83	4.81	4.71	4.83	4.81
cc-pCVDZ-unc	5.30	5.00	5.09	5.06	4.99	5.08	5.05
cc-pVTZ	4.84	4.42	4.55	4.51	4.41	4.54	4.50
cc-pCVTZ	4.79	4.38	4.51	4.47	4.35	4.48	4.44
cc-pCVTZ-unc	4.81	4.39	4.52	4.48	4.36	4.49	4.45
cc-pVQZ	4.74	4.23	4.40	4.34	4.21	4.39	4.33
cc-pCVQZ	4.64	4.17	4.33	4.28	4.14	4.30	4.25
cc-pCVQZ-unc	4.66	4.19	4.35	4.29	4.15	4.31	4.26
cc-pV5Z	4.65	4.11	4.30	4.24	4.08	4.28	4.21
cc-pCV5Z	4.55	4.04	4.23	4.17	4.00	4.19	4.13
cc-pVTZ-LD	4.56	4.06	4.25	4.19	4.02	4.21	4.15
Experiment				4.09			

In the experiments of Alonso et al.,[23] some ambiguity still remained in differentiating conformers with similar rotational constants. For example, the B_0 and C_0 rotational constants of Cys-I and Cys-III match within 40 MHz while A_0 matches within 180 MHz. Furthermore, the computed rotational constants for Cys-I lie in between the observed values. For example, C_0 for Cys-I is computed to be 1344.5 MHz, in between the observed values of 1331.8 and 1367.8 MHz. Quadrupole coupling constants of the nitrogen nucleus are therefore necessary to uniquely identify such conformers. The quadrupole coupling constants ($\chi_{\alpha\alpha}$) are given as

$$\chi_{\alpha\alpha} = eq_{\alpha\alpha}Q \tag{2.5}$$

where $q_{\alpha\alpha}$ is the the electric field gradient along the α-axis at the nitrogen nucleus, e is the fundamental charge, and Q is the nuclear quadrupole moment. For the nitrogen quadrupole moment, we use the literature value of 20.44 mb.[74] As seen for ammonia, the accurate computation of elec-

tric field gradients at the nuclei presents a difficult theoretical problem. Similar difficulties hold for spin-dependent properties that depend on contact terms since the amplitude and shape of the wavefunction near the nuclei must be very accurately described.[32,75] In particular, Gaussian basis functions have the incorrect shape at the nuclei, so that extremely flexible and carefully designed basis sets are required for accurate results.

In general, double-ζ basis sets and Hartree-Fock methods are not flexible enough to yield good results for quadrupole coupling constants. Since the coupling constant depends only on the nitrogen nucleus, it is possible to use a locally dense basis on the nitrogen atom.[76] The combination of a cc-pCV5Z basis on nitrogen and cc-pVTZ basis on all other atoms (denoted cc-pVTZ-LD) very closely matches both the full cc-pCV5Z result and the experimental coupling constant (4.09 MHz)[77] for ammonia (Table 2.6). Furthermore, probably through fortuitous cancellation of errors, MP2 matches the CCSD(T) results well, better than even CCSD. The combination of MP2 and the locally dense basis therefore seems to offer the best combination of accuracy and efficiency for computing the quadrupole coupling constants of cysteine.

The computed quadrupole coupling constants for the conformers of cysteine are presented in Table 2.7. The MP2/cc-pVTZ-LD approximation generally performs quite well, yielding most coupling constants within 0.1 MHz of experiment with the largest deviation being 0.18 MHz. In particular, the ambiguity in assignment based on rotational constants is now removed. For example, Cys-I has a strong χ_{aa} quadrupole coupling while Cys-III exhibits almost no χ_{aa} coupling, in agreement with the results of Alonso et al.[23]

TABLE 2.7: Quadrupole coupling constants for conformers of L-cysteine computed at the MP2/cc-pVTZ-LD Level (See Text)[a]

Conformer	χ_{aa}	χ_{aa}	χ_{aa}
Cys-I	−3.14 (−3.12)	2.44 (2.44)	0.70 (0.68)
Cys-II	−0.18 (−0.41)	2.19 (2.23)	−2.01 (−1.83)
Cys-III	−0.01 (−0.15)	0.34(0.44)	−0.32 (−0.30)
Cys-IV	−3.09	2.74	0.35
Cys-V	−4.39	2.74	1.65
Cys-VI	−3.26	2.36	0.90
Cys-VII	0.06 (0.00)	−0.48 (−0.45)	0.42 (0.45)
Cys-VIII	−3.02	1.51	1.51
Cys-IX	−3.22	1.61	1.61
Cys-X	0.51	−1.99	1.49
Cys-XI	−1.27	0.88	0.39

[a] Experimental values[23] where known are given in parentheses. All values given in MHz.

2.5 Summary and Conclusions

In the present work, we performd a comprehensive study of the important structural features and spectroscopic signatures of the amino acid L-cysteine. Through the focal point approach, we have established definitive relative energies of the eleven lowest conformers to within a standard error 0.3 kJ mol^{-1} (1σ) or 95% confidence interval of $\pm$ 0.6 kJ mol^{-1} (2σ). Due to the added flexibility of the thiol sidechain, cysteine exhibits 71 unique conformers and eleven conformers within 10 kJ mol^{-1} of the lowest minimum. As observed previously, Hartree-Fock energies are inaccurate. Inclusion of electron correlation with B3LYP greatly improves results, but still fails by more than 2.5 kJ mol^{-1} for some conformers, which becomes significant when so many conformers lie within a narrow 10 kJ mol^{-1} range. In general, B3LYP performs well for geometries and zero-point vibrational corrections, but inclusion of correlation through at least MP2 seems neccessary for accurate energies. Definitive energies to 0.5 kJ mol^{-1} accuracy still require corrections through CCSD(T).

While hydrogen bonding and electrostatics are the most important factors determining structures and energetics, the bond length, bond angle, and energy changes between conformers depend strongly on subtle electronic effects, including hyperconjugation, steric repulsion, hydrogen-bond cooperativity, and dispersion forces. These electronic effects, while subtle conceptually, are still treated with subchemical accuracy by only CCSD(T). It should therefore be noted that any residual error by not including CCSDTQ or higher increments are likely to be less important than other auxiliary corrections for relativistic or Born-Oppenheimer effects. In contrast to previous work, we emphasize features such as the trans angle rule[67] and the gauche effect.[70,71] An additive picture of hydrogen bonds may therefore be overly simplistic for cysteine.

Harmonic frequencies were computed at the B3LYP/aug-cc-pVTZ level with anharmonic corrections at the B3LYP/6-31G* level (Tables 2.4). The vibrational perturbation theory generally performs well (within 20 cm^{-1} of experiment), although it breaks down for a few large-amplitude motions with very low frequencies. The computed fundamentals should aid future IR spectroscopy studies. Since we are aiming for accuracy near 0.5 kJ mol^{-1} in the conformational energies, rigorous anharmonic zero-point vibrational corrections are necessary instead of simply scaling harmonic frequencies.

The extensive ab initio results reported here should serve as an important reference both for calibrating more approximate theoretical methods or future experiments, including circular dichroism or infrared and microwave spectroscopy of isotopologues of cysteine. As more empirical data becomes available (e.g. rotational constants of isotopologues), the structures and energies can be further refined by empirical fitting, as done previously for glycine[13] and proline.[17]

2.6 Acknowledgments

This work was supported by NSF grant CHE-0749868, an NSF-MTA-OTKA grant, and the Hungarian Scientific Research Fund (OTKA, K72885, N77954). Most computations were run at the Pittsburgh Supercomputing Center under TeraGrid grant TG-CHE070039N. I also thank Prof. Steven Wheeler for helpful discussions.

BIBLIOGRAPHY

[1] A. Perczel, A. K. Fuzery, A. G. Csàszàr, J. Comput. Chem. **24**, 1157 (2003).

[2] D. L. Nelson, M. M. Cox, *Principles of Biochemistry.* (W. H. Freeman and Company, New York, NY, 2005).

[3] X. F. Lu, A. Galkin, O. Herzberg, D. Dunaway-Mariano, J. Am. Chem. Soc. **126**, 5374 (2004).

[4] H. M. Li, G. J. Thomas, J. Am. Chem. Soc. **113**, 456 (1991).

[5] A. G. Csàszàr, J. Am. Chem. Soc. **114**, 9568 (1992).

[6] C. H. Hu, M. Z. Shen, H. F. Schaefer, J. Am. Chem. Soc. **115**, 2923 (1993).

[7] A. L. L. East, W. D. Allen, J. Chem. Phys. **99**, 4638 (1993).

[8] A. G. Csàszàr, W. D. Allen, H. F. Schaefer, J. Chem. Phys. **108**, 9751 (1998).

[9] J. M. Gonzales, C. Pak, R. S. Cox, W. D. Allen, H. F. Schaefer, A. G. Csàszàr, G. Tarczay, Chem. Eur. J. **9**, 2173 (2003).

[10] J. M. Gonzales, W. D. Allen, H. F. Schaefer, J. Phys. Chem. A **109**, 10613 (2005).

[11] M. S. Schuurman, S. R. Muir, W. D. Allen, H. F. Schaefer, J. Chem. Phys. **120**, 11586 (2004).

[12] A. G. Csàszàr, J. Mol. Struct. **346**, 141 (1995).

[13] V. Kasalova, W. D. Allen, H. F. Schaefer, E. Czinki, A. G. Csàszàr, J. Comput. Chem. **28**, 1373 (2007).

[14] A. G. Csàszàr, J. Phys. Chem. **100**, 3541 (1996).

[15] T. Szidarovszky, G. Czako, A. G. Csàszàr, Mol. Phys. **107**, 761 (2009).

[16] E. Czinki, A. G. Csàszàr, Chem. Eur. J. **9**, 1008 (2003).

[17] W. D. Allen, E. Czinki, A. G. Csàszàr, Chem. Eur. J. **10**, 4512 (2004).

[18] S. Gronert, R. A. J. Ohair, J. Am. Chem. Soc. **117**, 2071 (1995).

[19] P. R. Laurence, C. Thomson, Theor. Chim. Acta **58**, 121 (1981).

[20] L. R. Wright, R. F. Borkman, J. Am. Chem. Soc. **102**, 6207 (1980).

[21] L. Schäfer, S. Q. Kulpnewton, K. Siam, V. J. Klimkowski, C. Vanalsenoy, THEOCHEM **68**, 373 (1990).

[22] J. C. Dobrowolski, J. E. Rode, J. Sadlej, J. Mol. Struct. **810**, 129 (2007).

[23] M. Eugenia Sanz, S. Blanco, J. C. Lopez, J. L. Alonso, Angew. Chem. Int. Ed. **47**, 6216 (2008).

[24] M. J. Frisch, J. A. Pople, J. E. Delbene, J. Phys. Chem. **89**, 3664 (1985).

[25] M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro, W. J. Hehre, J. Am. Chem. Soc. **104**, 2797 (1982).

[26] J. S. Binkley, J. A. Pople, W. J. Hehre, J. Am. Chem. Soc. **102**, 939 (1980).

[27] C. C. J. Roothaan, Rev. Mod Phys. **23**, 69 (1951).

[28] K. A. Peterson, T. H. Dunning, J. Chem. Phys. **117**, 10548 (2002).

[29] R. A. Kendall, T. H. Dunning, R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

[30] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

[31] A. K. Wilson, T. H. Dunning, J. Chem. Phys. **119**, 11712 (2003).

[32] U. Benedikt, A. A. Auer, F. Jensen, J. Chem. Phys. **129**, 064111 (2008).

[33] P. Pulay, Mol. Phys. **17**, 197 (1969).

[34] R. Krishnan, M. J. Frisch, J. A. Pople, J. Chem. Phys. **72**, 4244 (1980).

[35] J. Čížek, J. Chem. Phys. **45**, 4256 (1966).

[36] T. D. Crawford, H. F. Schaefer, Rev. Comp Chem. **14**, 33 (2000).

[37] G. D. Purvis, R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982).

[38] K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[39] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[40] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[41] C. T. Lee, W. Yang, R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[42] D. Jayatilaka, T. J. Lee, J. Chem. Phys. **98**, 9734 (1993).

[43] T. J. Lee, P. R. Taylor, Int. J. Quantum Chem. Symp **23**, 199 (1989).

[44] J. F. Stanton, J. Gauss, J. D. Watts, G. Szalay, R. J. Bartlett, A. A. Auer, D. B. Bernholdt, O. Christiansen, M. E. Harding, M. Heckert, O. Heun. ACES II, Mainz-Austin-Budapest version. 2005. For the current version, see http://www.aces2.de.

[45] C. L. Janssen, I. M. B. Nielsen, M. L. Leininger, E. F. Valeev, J. P. Kenny, E. T. Seidl. The Massively Parallel Quantum Chemistry Program (MPQC), version 3.0. 2008. Sandia National Laboratories, Livermore, CA, USA. http://www.mpqc.org.

[46] I. M. B. Nielsen, E. T. Seidl, J. Comput. Chem. **16**, 1301 (1995).

[47] I. Nielsen, Chem. Phys. Lett. **255**, 210 (1996).

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

[49] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, J. Montgomery, T. Vreven, K. Kuding, J. Burant, J. Millam, S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, None. Kitao, H. Nakai, M. Klene, X. Li, J. Knox, H. Hratchian, J. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. Stratmann, O. Yazyev, A. Austin, R. Cammi, C. Pomelli, J. Octerski, P. Ayala, K. Morokuma, G. Voth, P. Salvador, J. Dannenber, V. Zkrzewski, S. Dappich, A. Daniels, M. Strain, O. Farkas, D. Malick, A. Rabuck, K. Raghavachari, J. Forseman, J. Ortiz, Q. Cui, A. Baboul, S. Clifford, J. Cioslowski, B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Martin, D. Fox, T. Keith, M. Al-Laham, C. Peng, A. Nanykkara, M. Callacombe, P. Gill, B. Johnson, W. Chen, M. Wong, C. Gonzalez, J. Pople. Gaussian 03. 2004. Gaussian, Inc. Wallingord, CT.

[50] E. L. Eliel, S. H. Wilen, L. N. Mander, *Principles of Biochemistry.* (John Wily and Sons, Inc., New York, NY, 1994).

[51] T. Helgaker, W. Klopper, H. Koch, J. Noga, J. Chem. Phys. **106**, 9639 (1997).

[52] G. Tarczay, A. G. Csàszàr, W. Klopper, H. M. Quiney, Mol. Phys. **99**, 1769 (2001).

[53] N. C. Handy, Y. Yamaguchi, H. F. Schaefer, J. Chem. Phys. **84**, 4481 (1986).

[54] W. D. Allen, A. G. Csàszàr, J. Chem. Phys. **98**, 2983 (1993).

[55] D. A. Clabo, W. D. Allen, R. B. Remington, Y. Yamaguchi, H. F. Schaefer, Chem. Phys. **123**, 187 (1988).

[56] M. S. Schuurman, W. D. Allen, P. V. R. Schleyer, H. F. Schaefer, J. Chem. Phys. **122**, 104302 (2005).

[57] M. Wolfsberg, A. A. Massa, J. W. Pyper, J. Chem. Phys. **53**, 3138 (1970).

[58] V. Barone, J. Chem. Phys. **120**, 3059 (2004).

[59] A. E. Reed, L. A. Curtiss, F. Weinhold, Chem. Rev. **88**, 899 (1988).

[60] A. E. Reed, F. Weinhold, L. A. Curtiss, D. J. Pochatko, J. Chem. Phys. **84**, 5687 (1986).

[61] A. E. Reed, F. Weinhold, J. Chem. Phys. **78**, 4066 (1983).

[62] J. P. Foster, F. Weinhold, J. Am. Chem. Soc. **102**, 7211 (1980).

[63] A. E. Reed, R. B. Weinstock, F. Weinhold, J. Chem. Phys. **83**, 735 (1985).

[64] A. C. Simmonett, F. A. Evangelista, W. D. Allen, H. F. Schaefer, J. Chem. Phys. **127**, 014306 (2007).

[65] A. Tajti, P. G. Szalay, A. G. Csàszàr, M. Kallay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vazquez, J. F. Stanton, J. Chem. Phys. **121**, 11599 (2004).

[66] T. H. Dunning, J. Chem. Phys. **53**, 2823 (1970).

[67] M. Rasanen, A. Aspiala, L. Homanen, J. Murto, J. Mol. Struct. **96**, 81 (1982).

[68] A. Pross, L. Radom, N. V. Riggs, J. Am. Chem. Soc. **102**, 2253 (1980).

[69] N. L. Allinger, J. T. Fermann, W. D. Allen, H. F. Schaefer, J. Chem. Phys. **106**, 5143 (1997).

[70] L. Goodman, H. B. Gu, V. Pophristic, J. Phys. Chem. A **109**, 1223 (2005).

[71] A. Lesarri, E. J. Cocinero, J. C. Lopez, J. L. Alonso, J. Am. Chem. Soc. **127**, 2572 (2005).

[72] P. Carbonniere, T. Lucca, C. Pouchan, N. Rega, V. Barone, J. Comput. Chem. **26**, 384 (2005).

[73] J. C. Dobrowolski, M. H. Jamroz, R. Kolos, J. E. Rode, J. Sadlej, ChemPhysChem **8**, 1085 (2007).

[74] R. Polak, J. Fiser, Chem. Phys. **351**, 83 (2008).

[75] T. Helgaker, M. Jaszunski, K. Ruud, A. Gorska, Theor. Chem. Acc. **99**, 175 (1998).

[76] P. F. Provasi, G. A. Aucar, S. P. A. Sauer, J. Chem. Phys. **112**, 6201 (2000).

[77] P. Grigolin, R. Moccia, J. Chem. Phys. **57**, 1369 (1972).

CHAPTER 3

THE REACTION OF ETHYL RADICAL WITH OXYGEN: A COMBUSTION ARCHETYPE [†]

[†]J J. Wilke, W. D. Allen, and and H. F. Schaefer, J. Chem. Phys. **128**, 074308 (2008). Reprinted here with permission of the American Institute of Physics.

3.1 Abstract

The $C_2H_5 + O_2$ reaction is an archetype for hydrocarbon combustion, and the critical step in the process is the concerted elimination of HO_2 from the ethylperoxy intermediate $C_2H_5O_2$. Master equation kinetic models fitted to measured reaction rates place the concerted elimination barrier 3.0 kcal mol^{-1} below the $C_2H_5 + O_2$ reactants, whereas the best previous electronic structure computations yield a barrier more than 2.0 kcal mol^{-1} higher. We resolve this discrepancy here by means of the most rigorous computations to date. Computations were executed with basis sets as large as cc-pV5Z and correlation treatments as extensive as coupled cluster through full triples with a perturbative inclusion of quadruple excitations. The final predicted barrier is -3.0 kcal mol^{-1}, bringing the concerted elimination mechanism into precise agreement with experiment. This work demonstrates that higher correlation treatments such as CCSDT(Q) are not only feasible on systems of chemical interest but are necessary to supply accuracy beyond 0.5 kcal mol^{-1}, which is not obtained with the gold standard CCSD(T) method. Finally, we compute the enthalpy of formation of $C_2H_5O_2$ to be $\Delta_f H^\circ(298\text{K}) = -5.3 \pm 0.5$ kcal mol^{-1} and $\Delta_f H^\circ(0\text{K}) = -1.5 \pm 0.5$ kcal mol^{-1}.

3.2 Introduction

Alkyl and alkylperoxy radicals appear in many applications, including combustion,[1–4] atmospheric chemistry,[5] and biological processes.[6] Understanding alkyl radical kinetics requires very accurate thermochemistry, especially for the reaction of ethyl with O_2.[1,7,8] Pathways leading to different products have barriers that differ by as little as 2-3 kcal mol^{-1}. Even small errors in barrier heights therefore cause large errors in reactions rates[9] and branching ratios. Unfortunately, the desired accuracy is difficult to obtain experimentally for transient gas-phase radicals, with uncertainties of at least 2.0 kcal mol^{-1} being common.[10] The situation is even more problematic for transition states whose energies must be derived indirectly from elementary reaction rates. As emphasized in the introduction, ab initio schemes can now achieve accuracy of a few tenths of a kcal mol^{-1}, helping to overcome these experimental difficulties.[2,9,11–15]

The reaction of ethyl radical with O_2 has served as a prototype for understanding hydrocarbon combustion.[1,4] A conjunction over the past two decades of experimental measurements,[2,16–22] theoretical computations,[4,11,23–25] and kinetic modeling[2,9,12,26–29] has revealed the key features of the reaction mechanism (Figure 3.1). Despite having only two carbon atoms, the reaction mechanism still involves multiple, competing product channels. Many of the reaction paths such as β-hydrogen abstraction in alkylperoxy intermediates are also relevant in larger hydrocarbons, and a detailed understanding of the ethyl + O_2 mechanism has therefore elucidated more complicated systems.[13,30–33]

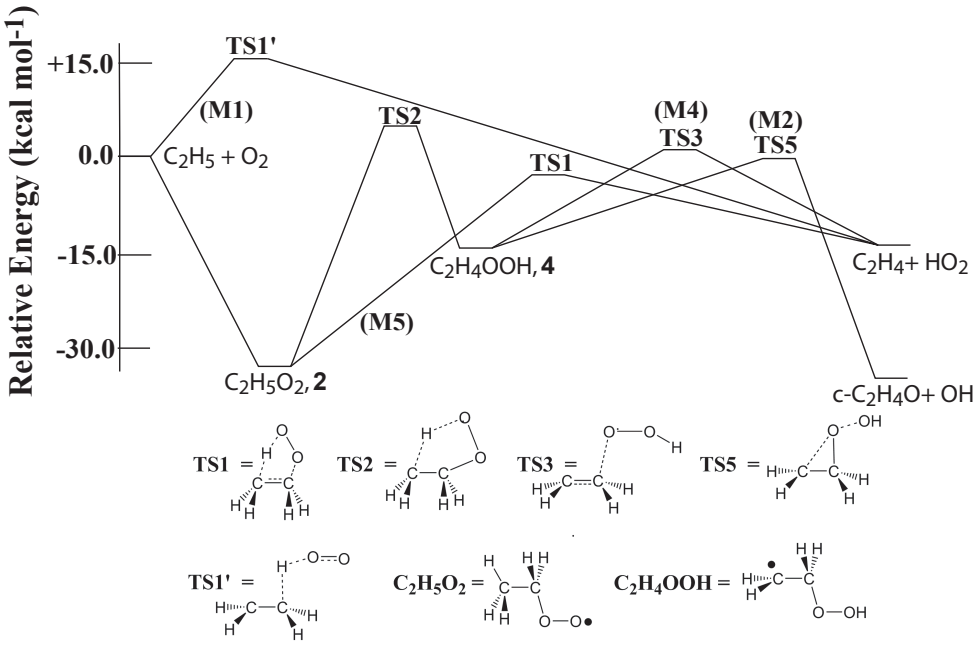

FIGURE 3.1: Mechanisms for reaction of ethyl radical and O_2. In M1, direct hydrogen abstraction from ethyl radical by O_2 through **TS′**) gives ethylene and HO_2. Paths M2, M4, and M5 begin with barrierless addition of O_2 to ethyl radical to form ethylperoxy radical **2**. In M2 and M4, β-hydrogen abstraction in **2** through TS2 forms ethylhydroperoxyl radical **4**, which can proceed through **TS3** to ethylene and HO_2 (M4) or through **TS5** to oxirane and OH (M2). In path M5, concerted elimination of HO_2 from **2** through **TS1** forms ethylene. A previously characterized (Ref. 11) loosely bound complex (**3**) in the C_2H_4 + HO_2 product channel is not shown for simplicity.

Considerable work on the ethyl + O_2 reaction has focused on the production of ethylene and HO_2 radical, the major products in the temperature range of 295-1000 K.[17,18,26,34–37] Two main experimental observations have supported a mechanism involving concerted elimination of HO_2 from ethylperoxy radical to give ethylene. First, the major products, ethylene + HO_2, are accom-

panied by only small amounts of the thermodynamically more favorable oxirane + OH radical.[34,37] Second, the reaction rate actually decreases with increasing temperature.[17] Figure 3.1 shows three possible routes from ethyl radical to ethylene; naming conventions are chosen to match previous work.[11] The simplest mechanism is direct abstraction (M1), in which O_2 removes a hydrogen atom through transition state **TS1′**. However, this process has a significant barrier[11] and could only be relevant at high temperatures. A more feasible path at lower temperatures involves addition of O_2 to ethyl radical (M2, M4, M5) to form an ethylperoxy intermediate ($C_2H_5O_2$, **2**). In earlier work, formation of ethylene and HO_2 was thought to initially proceed from **2** by intramolecular hydrogen abstraction (M2, M4, **TS2**),[26] leading to the hydroperoxy intermediate (C_2H_4OOH, **4**). From **4**, ethylene can form by CO bond scission through **TS3**. However, oxirane formation through elimination of OH radical (**TS5**) should be competitive with ethylene formation in M4, inconsistent with experimental observations. In contrast, concerted elimination of HO_2 through **TS1** (M5) produces only ethylene. Both density functional and coupled cluster computations indicate that the internal H-abstraction barrier (**TS2**) lies 2.3-9.9 kcal mol^{-1} higher in energy than the transition state (**TS1**) for concerted elimination.[9,11,12,38] Furthermore, master equation models have explained the rate coefficient temperature dependence in terms of a $C_2H_5 + O_2 \leftrightarrow C_2H_5O_2$ equilibrium followed by concerted elimination.[9] At higher temperatures, the back reaction to $C_2H_5 + O_2$ competes more effectively with concerted elimination, thereby slowing the overall reaction to form ethylene. Mechanism M5 is then consistent with both the experimentally observed product distribution and negative temperature coefficient, making it the favored mechanism for low-temperate combustion of ethyl radical. Even though the reaction of ethyl radical with O_2 has been extensively studied, disturbing ambiguities still remain for the relative energies of key intermediates in the pathways shown in Figure 3.1.

From the master equation model of Miller and Klippenstein,[9] the best agreement with experiment has been obtained with energies relative to ethyl + O_2 of -33.5 kcal mol^{-1} for ethylperoxy (**2**) and -3.0 kcal mol^{-1} for the transition state **TS1**, similar to the values of -34.0 and -2.8 used by Dean et al.[13] Barriers for **TS1** near -3.0 kcal mol^{-1} have been obtained with a G2-like method[9,39] and the empirical model chemistry CBS-Q.[12] In contrast, the most rigorous ab initio

TABLE 3.1: Some theoretical values (kcal mol^{-1}) for the concerted elimination barrier (**TS1**) of the ethyl + O_2 system.

Level of Theory	Reference	Barrier
CISD/DZP	24	+19.1
CISD + (Q) / DZP	24	+13.3
UBLYP/DZP	23	−10.3
UB3LYP/DZP	23	−2.4
UB3LYP/TZ2Pf	23	−1.9
CCSD/DZP	11	+8.7
CCSD(T)/DZP	11	+3.3
CCSD/TZ2P	11	+6.9
CCSD(T)/TZ2Pf	11	−0.9
G2	9, 39	−3.0
CBS-Q	12	−2.8
Focal point analysis	This work	−3.0

computation in the literature,[11] at the CCSD(T)/TZ2Pf level, gave a barrier of −0.9 kcal mol^{-1}. Published theoretical values for the concerted elimination barrier have varied wildly (Table 3.1), ranging from +19.1 kcal mol^{-1} at the CISD/DZP level to −10.3 kcal mol^{-1} at the UBLYP/DZP level. While the G2 and CBS-Q methods match the optimal values from the kinetic models, Miller and Klippenstein[9] have noted that the nearly exact agreement is "probably fortuitous." Similarly, Dean et al.[13] have stressed that the "good agreement between predictions (from CBS-Q) and experimental data must be seen, at least in part, as a fortunate cancellation of errors." While the G3 method typically obtains accuracy of 1-2 kcal mol^{-1},[40] the accuracy sometimes results not from intrinsic precision but from cancellation of unrelated errors.[41] For several hydrocarbon radical enthalpies of formation, Dixon et al.[42] have even shown that G3 and CBS-Q differ from very precise coupled cluster computations and experimental values by more than 2.0 kcal mol^{-1}. The **TS1** barrier height is further muddled by seeming discrepancies in experimental results. In the reaction of ethyl with O_2, the major products are ethylene and HO_2. Conversely, if the reverse experiment is performed by reacting ethylene with HO_2, the major products are oxirane and OH radical.[4,16,43] Oxirane formation likely proceeds through the hydroperoxy intermediate (**4**). However, the lowest energy path for addition of ground state HO_2 to ethylene is formation of the alkylperoxy radical **2**, which should decompose to reform ethylene and HO_2 rather than oxirane and OH. Stark[4] has rationalized this discrepancy in terms of a conical intersection between A' and A'' surfaces, allowing

different pathways to be followed in the forward and reverse directions. However, Stark suggests that, for this mechanism to be consistent with experimental activation energies, the transition state (**TS1**) should lie higher in energy than the ethyl + O_2 reactants.

Very accurate ab initio computations are needed to resolve the energetic ambiguities of the ethyl + O_2 reaction and, at last, firmly establish the concerted elimination mechanism. The history of scattered predictions for the **TS1** barrier, documented in Table 3.1, highlights the need for rigorous new computations. Obtaining energies with errors of less than 1 kcal mol^{-1} requires correlated methods with large basis sets. Even the ethyl + O_2 system with only four heavy atoms and five hydrogens pushes the limit of what can be feasibly treated. In this regard, the focal point approach of Allen et al.[44–47] outlined in the dissertation introduction minimizes the computational expense by extrapolating the energy to both the basis set and correlation limit. The scheme is entirely free of any possible bias due to empirical corrections, relying only on ab initio values. In the present work, we therefore report focal point computations that yield definitive thermochemistry for the low-temperature pathway of the ethyl + O_2 combustion archetype.

3.3 Theoretical Methods

3.3.1 General Scheme

To obtain energies for each of the reactants and products, the following overall procedure was employed. The geometry was optimized with the CCSD(T) method[48–50] (frozen core) using the Dunning correlation-consistent cc-pVQZ basis set with the standard contractions C,O(*12s6p3d2f1g/5s4p3d2f1g*) and H(*6s3p2d1f/4s3p2d1f*).[51] Energies were then extrapolated with respect to the correlation and basis set limit using the valence focal point approach.[44–46,52] Finally, auxiliary terms were included for core correlation, relativistic effects, diagonal Born-Oppenheimer corrections (DBOC), and zero-point vibrational energies (ZPVEs).

3.3.2 Geometry Optimization

All geometries were optimized using the MOLPRO2002.6 package.[53] In the case of ethylperoxy radical **2** and **TS1**, the geometry was optimized using delocalized internal coordinates via the

OPTKING module in the PSI3 package;[54] a fixed CCSD(T)/cc-pVDZ Hessian computed with the ACESII package[55] was employed in these cases with numerical gradients provided by MOLPRO2002.6.

3.3.3 Valence focal point analysis

The focal point method[44–46,52] invokes a systematic approach towards both the basis set and correlation limit, as explained in the introductory chapter. Here we reiterate the important focal point principles and computational details specific to the ethyl-O_2 system. In the following, E refers to absolute energies, ΔE refers to relative energies between two species or for a reaction, and δ refers to an energy increment, a correction to the relative energy with respect to a preceding level of theory. The scheme proceeds by:

(1) Extrapolation of the Hartree-Fock (HF) energy according to the exponential form

$$E_{\mathrm{HF}}(X) = E_{\mathrm{HF}}^{\infty} + A\exp(-bX) \tag{3.1}$$

where X is the cardinal number of a correlation consistent cc-pVXZ basis.[51] In the current work, the three parameters E_{HF}, A, and b were fit to energy points with $X = 3$, 4 , 5.

(2) Extrapolation to the complete basis set limit of the second-order correlation energy obtained from either closed-shell MP2 or Z-averaged perturbation theory (ZAPT2).[56]

$$\epsilon_{\mathrm{ZAPT2}} = \epsilon_{\mathrm{ZAPT2}}^{\infty} + BX^{-3} \tag{3.2}$$

$$\epsilon_{\mathrm{ZAPT2}} = E_{\mathrm{ZAPT2}}(X) - E_{\mathrm{HF}}(X) \tag{3.3}$$

Because all the relative energies involve one or more open-shell species, we hereafter denote correlation energies and energy increments from second-order perturbation theory as $\epsilon_{\mathrm{ZAPT2}}$ and δ[ZAPT2], respectively, with the implication that for closed-shell species the standard MP2 method was used. In the current work, the two parameters $\epsilon_{\mathrm{ZAPT2}}^{\infty}$ and B were obtained

algebraically from energy points with $X = 4$, 5. The arguments for an X^{-3} dependence have been given elsewhere.[46,57–59]

(3) Computation of the CCSD and CCSD(T) increments with the cc-pVQZ basis,

$$\delta[\text{CCSD}] = \Delta E(\text{CCSD/cc-pVQZ}) - \Delta E(\text{ZAPT2/cc-pVQZ}) \tag{3.4}$$

and

$$\delta[\text{CCSD(T)}] = \Delta E(\text{CCSD(T)/cc-pVQZ}) - \Delta E(\text{CCSD/cc-pVQZ}) \tag{3.5}$$

The focal point scheme adopted here assumes that the CCSD and CCSD(T) increments can be computed as additive corrections.

(4) Inclusion of higher-order electron correlation by performing CCSDT and CCSDT(Q)[60,61] computations with a modest basis, in this case cc-pVDZ, yielding the increments

$$\delta[\text{CCSD(T)}] = \Delta E(\text{CCSDT/cc-pVDZ}) - \Delta E(\text{CCSD(T)/cc-pVDZ}) \tag{3.6}$$

$$\delta[\text{CCSDT(Q)}] = \Delta E(\text{CCSDT(Q)/cc-pVDZ}) - \Delta E(\text{CCSDT/cc-pVDZ}) \tag{3.7}$$

Following the argument above for δ[CCSD(T)], the δ[CCSDT] and δ[CCSDT(Q)] increments converge much more quickly to the basis set limit than the absolute energies. The δ[CCSDT] and δ[CCSDT(Q)] corrections can therefore be well estimated as an additive constant. The accuracy of the cc-pVDZ value is further addressed in the discussion.

(5) Determination of the final valence focal point ΔE_{FP} relative energy given by

$$\Delta E_{\text{FP}} = \Delta E_{\text{HF}}^{\infty} + \delta^{\infty}[\text{ZAPT2}] + \delta[\text{CCSD}] + \delta[\text{CCSD(T)}] + \delta[\text{CCSDT}] + \delta[\text{CCSDT(Q)}] \tag{3.8}$$

With the inclusion of all extrapolations and corrections, (3.8) becomes an estimate of the CCSDT(Q) relative energy at the CBS limit.

Open-shell second-order Z-averaged perturbation theory[56] (ZAPT2) computations were executed with the MPQC package.[62] All CCSD(T) single-point energies were obtained using MOLPRO2006[63] with unrestricted coupled cluster theory using ROHF orbitals.[64] CCSDT and CCSDT(Q) computations were performed using the MRCC program of Kllay.[65] The MRCC program requires a front-end program to compute integrals. For the CCSDT energies, MOLPRO2006 restricted orbitals were used for consistency with the CCSD(T) results. For CCSDT(Q), only unrestricted references are allowed in the MRCC code, and it was therefore necessary to use UHF orbitals from the ACESII program. The different references do not affect the final results since the absolute CCSDT(Q) and CCSDT energies do not enter the extrapolation, but rather the energy differences CCSDT(Q) and CCSDT. The CCSDT(Q) correction is therefore computed as ΔE[UHF-CCSDT(Q)] − ΔE[UHF-CCSDT], while the $\delta[CCSDT]$ increment is computed as ΔE[ROHF-CCSDT] − ΔE[ROHF-CCSD(T)].

3.3.4 Auxiliary Corrections

To account for core correlation, all-electron (AE) ZAPT2 computations were differenced from frozen-core (FC) ones, as evaluated with the cc-pCVQZ basis set,[66] which contains tight functions to recover core correlation.

$$\Delta E_{\text{core}} = \Delta E(\text{AE-ZAPT2/cc-pCVQZ}) - \Delta E(\text{FC-ZAPT2/cc-pCVQZ}) \tag{3.9}$$

A scalar first-order relativistic correction,[67] ΔE_{rel}, from the one-electron mass-velocity and Darwin terms was computed at the CCSD(T)/aug-cc-pVTZ level

$$\hat{H}_{rel} = -\frac{1}{8c^2}\sum_j \nabla_j^4 + \frac{\pi}{2c^2}\sum_{j,A}\delta(r_{jA}) \tag{3.10}$$

where the sums run over all electrons, j, and nuclei, A. ZPVE corrections, ΔE_{ZPVE}, were evaluated from harmonic vibrational frequencies obtained at the frozen-core CCSD(T) level with the cc-pVTZ basis set.[51] Relaxation of the clamped nucleus assumption was considered by computing diagonal

Born-Oppenheimer corrections (DBOC)[68] at the restricted HF/aug-cc-pVTZ level as

$$\langle \Psi_e | - \frac{1}{2} \sum_A \frac{1}{M_A} \nabla_A^2 \Psi_e \rangle_R \tag{3.11}$$

where ∇_A denotes a derivative of the electronic wavefunction Ψ_e with respect to nucleus A evaluated at a particular set of nuclear positions R.

The final energy for each molecule is then given by

$$E_{\text{final}} = \Delta E_{\text{FP}} + \Delta E_{\text{ZPVE}} + \Delta E_{\text{core}} + \Delta E_{\text{rel}} + \Delta E_{\text{DBOC}}. \tag{3.12}$$

Core correlation corrections were computed with the MPQC[62] package, while relativistic and DBOC corrections were computed with ACESII.[55]

3.3.5 Enthalpy of formation at 298 K

Thermal corrections to the molar enthalpies of formation were calculated with the standard formula,

$$H - H_0 = RT^2 \left(\frac{d \ln Q}{dT} \right)_V + RT \tag{3.13}$$

The vibrational and rotational partition functions (Q) were computed using the harmonic-oscillator and rigid-rotor approximations, respectively. The translational partition function was determined by the equipartition theorem. The low-frequency methyl torsions in ethyl radical, ethylperoxy radical, and acetaldehyde were treated as hindered rotors according to the Pitzer-Gwinn procedure[69] using CCSD(T)/cc-pVTZ barriers and moments of inertia.

3.4 Results and Discussion

3.4.1 Geometric Structures

The CCSD(T)/cc-pVQZ reaction intermediate (**2**) and transition state (**TS1**) structures for the concerted elimination pathway are displayed in Figures 3.2 and 3.3. In the case of the ethylperoxy radical, **2**, two different conformers are possible.[11,70] Only conformer **2a** with C_s symmetry is

shown in Figure 3.2. Conformer **2b** involves a rotation about the O_2-C_3 bond, yielding a C_1 gauche structure that is 0.23 kcal mol^{-1} lower in energy.

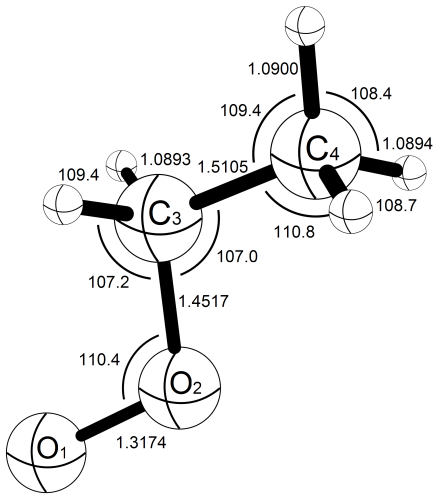

FIGURE 3.2: CCSD(T)/cc-pVQZ optimum geometry of ethylperoxy radical ($^2A''$), conformer **2a** (C_s). Bond lengths are in Å; bond angles in degrees.

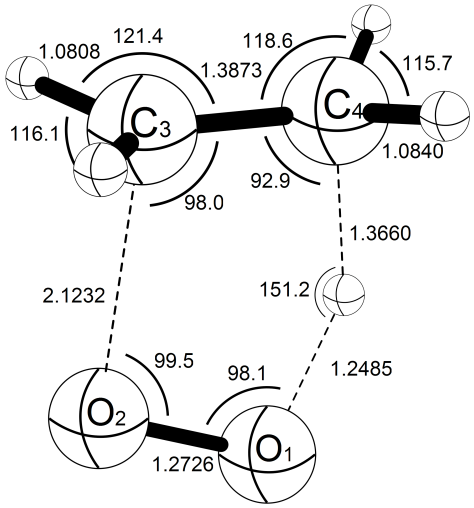

FIGURE 3.3: CCSD(T)/cc-pVQZ optimum geometry of the transition state **TS1** ($^2A''$, C_s), for concerted elimination of ethylene. Bond lengths are in Å; bond angles in degrees.

During the concerted elimination process, three major changes take place. First, a hydrogen atom is transferred from the terminal carbon atom, C_4, to the terminal oxygen, O_1. The O_1-H distance in **TS1** is thus 1.25 Å compared to 0.97 Å in the hydroperoxyl radical, while the C_4-H distance is 1.37 Å compared to 1.09 Å in **2a**. Second, the O_2-C_3 bond is homolytically cleaved. In **TS1**, this bond is therefore significantly extended to 2.12 Å, compared to 1.45 Å in **2a**. Finally, a

double bond forms between C_3-C_4. In **TS1**, the C_3-C_4 bond is accordingly 1.39 Å, compared to 1.51 Å in **2a** and 1.33 Å in ethylene.

The quality of the geometric structures for **2a** and **TS1** may be gauged by comparing our best CCSD(T)/cc-pVQZ parameters with corresponding CCSD(T)/cc-pVTZ results. For **2a**, almost all CCSD(T) bond lengths change by less than 0.003 Å in going from cc-pVTZ to the cc-pVQZ basis set, the largest change being 0.0055 Å for the O_1-O_2 length; all bond angles change by less than 0.3°. Similarly for **TS1**, most bond lengths change by less than 0.004 Å, and all bond angles change by less than 0.5°. The largest bond length change in **TS1** again occurs for the O_1-O_2 distance, which decreases by 0.004 Å from cc-pVTZ to cc-pVQZ CCSD(T). In brief, the CCSD(T)/cc-pVQZ geometries should be well converged with respect to both basis set and correlation limits, making them suitable as reference geometries for higher level computations.

The computation of CCSD(T)/cc-pVQZ harmonic vibrational frequencies for **TS1** is prohibitive in cost; therefore, we rely on CCSD(T)/cc-pVTZ values which confirm that **TS1** is a genuine transition state. Considering the similarity of the cc-pVQZ and cc-pVTZ CCSD(T) structures, significant differences in the corresponding vibrational frequencies are not expected.

TABLE 3.2: Comparative diagnostics for static correlation, CCSD(T)/cc-pVTZ level of theory.

	T_1 diagnostic	Largest T_2 amplitude
HO_2	0.037	0.055
CN	0.053	0.091
TS1	0.034	0.072
2a	0.031	0.022

3.4.2 Static correlation and multireference character

Given the distorted bonds in **TS1**, we must assess the degree of multireference character in the electronic wave function. A common T_1 diagnostic[71] and the largest T_2 amplitude from the coupled-cluster wave function can indicate degrees of static correlation. Single-reference coupled cluster methods give accurate results for the hydroperoxy[72,73] and CN[74] radicals, and we therefore compare **TS1** to these examples. In the present work, the T_1 diagnostic was computed with the MOLPRO2006.1 package, while the largest T_2 amplitude was computed with the ACESII package. Both diagnostics were evaluated at the CCSD(T)/cc-pVTZ level. For the T_1 diagnostic, **TS1** has

a value even lower than the HO_2 and CN benchmarks (Table 3.2). Furthermore, the largest T_2 amplitude in CN radical exceeds the largest amplitude for **TS1**. Similar comparisons hold for the ethylperoxy intermediate **2a**. Therefore, we deem our single-reference coupled cluster methods reliable, especially because we ultimately include electron correlation through the connected quadruple excitation level.

TABLE 3.3: Summary of focal point energies (kcal mol^{-1}) for the reactions of ethyl radical and O_2.

Reaction	E_{HF}	δ^{∞}[ZAPT]	δ^{∞}[CCSD]	δ^{∞}(T)	δ^{∞}[CCSDT]	δ^{∞}(Q)	ΔE_{FP}
$C_2H_5 + O_2 \rightarrow$ TS1	+32.38	−27.59	−1.32	−6.91	−0.13	−0.75	−4.31
$C_2H_5 + O_2 \rightarrow$ 2a	−30.21	+5.32	−12.60	−0.93	−0.22	+0.22	−38.41
$C_2H_5 + O_2 \rightarrow$ c-CH_2CH_2O + OH	−41.68	+12.9	−7.74	+1.28	+0.07	+0.47	−34.70
$C_2H_5 + O_2 \rightarrow CH_3CHO$ + OH	−72.13	+18.54	−8.90	+1.40	−0.05	+0.47	−60.67
$C_2H_5 + O_2 \rightarrow C_2H_4 + HO_2$	−11.82	+11.16	−13.65	−0.67	−0.34	+0.24	−15.08

3.4.3 Focal Point Analysis

The focal point analyses of reactants, intermediates, and products for ethyl + O_2 reactions are summarized in Table 3.3. Auxiliary corrections and final predicted reaction energies appear in Table 3.4. Detailed layouts for the energetics of **TS1** and **2a** are given in Tables 3.5 and 3.6 respectively. For ethylperoxy radical, the focal point analysis was applied only to conformer **2a**, because the CCSD(T)/cc-pVQZ geometry optimizations and subsequent single-point computations were much less demanding in C_s rather than in C_1 symmetry. The relative energy of **2b** was then obtained by a conformer correction of −0.23 kcal mol^{-1} computed with the CCSD(T)/cc-pVTZ method. We found that the **2a**-**2b** energy difference is not sensitive to level of theory, and the conformer correction should be accurate to within 0.05 kcal mol^{-1}.

From the size of the various terms in Table 3.3, we can assess the convergence of the focal point extrapolations. The δ[CCSD] corrections are sizable, larger than 6 kcal mol^{-1} for most cases; however, the subsequent δ[CCSD(T)] increments are much smaller, on the order of 1.0 kcal mol^{-1} except for **TS1**. The higher corrections δ[CCSDT] and δ[CCSDT(Q)] are generally less than 0.5 kcal mol^{-1}, two to threefold smaller than the preceding δ[CCSDT] terms. Assuming continued proportionate reductions, the error in neglecting quintuple and higher excitations should not exceed a few tenths of a kcal mol^{-1}. Nonetheless, some caution is warranted regarding the **TS1** barrier,

TABLE 3.4: Auxiliary corrections and final energetic predictions for the reactions of ethyl radical and O_2. All entries in kcal mol^{-1}. Comparisons to experiment are included where available.

Reaction	E_{FP}	ZPVE[a]	Relativity	Core	DBOC	ΔH° (0 K)	ΔH° (298K)	Exp. ΔH° (298 K)
TS1	−4.31	+1.34	+0.02	−0.03	−0.00	−2.98	−4.56	-
2a	−38.41	+5.45	+0.02	−0.03	−0.03	−32.99	−34.20	-
c-CH2CH2O + OH	−34.70	+1.90	+0.13	−0.04	−0.02	−32.73	−33.10	−32.37
CH3CHO + OH	−60.67	+0.61	+0.10	−0.11	−0.02	−60.10	−59.95	−59.51
C2H4 + HO2	−15.08	+1.37	+0.05	−0.07	−0.01	−13.74	−13.86	−13.07

[a] Computed from harmonic vibrational frequencies at the CCSD(T)/cc-pVTZ level. [b] Based on the following literature values for the standard enthalpies of formation in kcal mol^{-1} at 298 K: ethyl radical, +28.7 ± 0.1 (Ref 76); ethylene, +12.7 ± 0.2 (Ref. 79); oxirane, −12.6 ± 0.2 (Ref. 82); acetaldehyde, −39.7 ± 0.1 (Ref. 81); OH radical, +8.91 ± 0.07 (Ref. 83); and HO_2 radical, +2.94 ± 0.06 (Ref. 77).

which displays a small δ[CCSD] of −1.32 kcal mol^{-1} followed by an unusually large δ[CCSD(T)] of −6.91 kcal mol^{-1} and, finally, a significant δ[CCSDT(Q)] of −0.75 kcal mol^{-1}. The reason that most previous computations with wave function methods (Table 3.1) have overestimated the height of **TS1** is apparent from the occurrence of only negative correlation increments in Tables 3.3 and 3.5.

The full focal point layouts of the concerted elimination barrier (Table 3.5) and the C_2H_5-O_2 bond energy (Table 3.6) indicate that the cc-pV5Z basis set yields Hartree-Fock results within 0.01 kcal mol^{-1} of the complete basis set (CBS) limit. Of course, the limits for the correlation increments are harder to reach, as shown by changes of −0.62 and −0.36 kcal mol^{-1} in the ZAPT2/cc-pV5Z increments obtained by CBS extrapolation in the **TS1** and **2a** cases, respectively. For the δ[CCSD(T)] correction, the cc-pVTZ and cc-pVDZ values differ by 1.21 and 0.33 kcal mol^{-1} for **TS1** and **2a**, but the succeeding cc-pVQZ/cc-pVTZ differences are only 0.26 and 0.14 kcal mol^{-1} in the two cases, respectively. These observations suggest that the difference between cc-pV5Z and cc-pVQZ δ[CCSD(T)] values should be significantly smaller than 0.2 kcal mol^{-1}, supporting the focal point additivity scheme. Finally, for the δ[CCSDT(Q)] and δ[CCSDT] corrections, recent work by Martin et al.[75] has shown that, for atomization energies, the cc-pVDZ values may differ from the CBS limit by as much 0.3 kcal mol^{-1}. However, for intramolecular transition states, the cc-pVDZ and cc-pVTZ values often match within a few hundredths of a kcal mol^{-1} (Ref 80), and our cc-pVDZ

results should be sufficiently accurate for the current system. Combining all sources of error, our final focal point values can be reported as having uncertainties of 0.5 kcal mol^{-1} or smaller.

TABLE 3.5: Focal point layout (kcal mol^{-1}) or the concerted elimination barrier (**TS1**) relative to ethyl + O_2. Computed values are listed without brackets; extrapolated values are shown in brackets.

Basis	ΔE_{HF}	δ[ZAPT2]	δ[CCSD]	$\delta(T)$	δ[CCSDT]	$\delta(Q)$	ΔE[CCSDT(Q)]
cc-pVDZ	+31.93	−19.97	−3.50	−5.44	−0.13	−0.75	+2.15
cc-pVTZ	+32.26	−24.92	−1.74	−6.65	−0.13	[−0.75]	[−1.92]
cc-pVQZ	+32.38	−26.38	−1.32	−6.91	−0.13	[−0.75]	[−3.11]
cc-pV5Z	+32.39	−26.97	[−1.32]	[−6.91]	[−0.13]	[−0.75]	[−3.69]
cc-pV6Z	[+32.39]	[−27.23]	[−1.32]	[−6.91]	[−0.13]	[−0.75]	[−3.95]
CBS limit	[+32.38]	[−27.59]	[−1.32]	[−6.91]	[−0.13]	[−0.75]	[−4.31]
Extrapolation	$a + be^{-cX}$	$a + bX^{-3}$					
Fit Points	3,4,5	3,4					

TABLE 3.6: Focal point layout (kcal mol^{-1}) for the formation of ethylperoxy radical conformer **2a** from ethyl + O_2. Computed values are listed without brackets; extrapolated values are shown in brackets.

Basis	ΔE_{HF}	δ[ZAPT2]	δ[CCSD]	$\delta(T)$	δ[CCSDT]	$\delta(Q)$	ΔE[CCSDT(Q)]
cc-pVDZ	−29.47	+10.80	−12.17	−0.46	−0.22	+0.22	−31.30
cc-pVTZ	−30.37	+7.56	−12.45	−0.79	[−0.22]	[+0.22]	[−36.04]
cc-pVQZ	−30.22	+6.43	−12.60	−0.93	[−0.22]	[+0.22]	[−37.32]
cc-pV5Z	−30.20	+5.89	[−12.60]	[−0.93]	[−0.22]	[+0.22]	[−37.84]
cc-pV6Z	[−30.21]	[+5.65]	[−12.60]	[−0.93]	[−0.22]	[+0.22]	[−38.08]
CBS limit	[−30.21]	[+5.53]	[−12.60]	[−0.93]	[−0.22]	[+0.22]	[−38.41]
Extrapolation	$a + be^{-cX}$	$a + bX^{-3}$					
Fit Points	3,4,5	3,4					

TABLE 3.7: Focal point layout (kcal mol^{-1}) for the formation of oxirane and OH from ethyl + O_2. Computed values are listed without brackets; extrapolated values are shown in brackets.

Basis	ΔE_{HF}	δ[ZAPT2]	δ[CCSD]	$\delta(T)$	δ[CCSDT]	$\delta(Q)$	ΔE[CCSDT(Q)]
cc-pVDZ	−37.36	+19.75	−7.93	+1.20	+0.07	+0.47	−23.80
cc-pVTZ	−40.54	+15.79	−7.86	+1.30	[+0.07]	[+0.47]	−30.77
cc-pVQZ	−41.24	+14.39	−7.74	+1.28	[+0.07]	[+0.47]	−32.77
cc-pV5Z	−41.53	+13.66	[−7.74]	[+1.28]	[+0.07]	[+0.47]	−33.79
cc-pV6Z	[−41.63]	[+13.34]	[−7.74]	[+1.28]	[+0.07]	[+0.47]	−34.21
CBS	[−41.68]	[+12.90]	[−7.74]	[+1.28]	[+0.07]	[+0.47]	−34.70
Extrapolation	$a + be^{-cX}$	$a + bX^{-3}$					
Fit Points	3,4,5	3,4					

As previously discussed, the concerted elimination transition state (**TS1**) best fits the master equation kinetic models with a barrier of −3.0 kcal mol^{-1} relative to ethyl + O_2. By comparison, the Hartree-Fock approximation places **TS1** over 35 kcal mol^{-1} too high (Table 3.5). Even in the ZAPT2/CBS limit, **TS1** lies almost 5 kcal mol^{-1} above the C_2H_5 + O_2 reactants. Simply optimizing geometries at the CCSD(T)/cc-pVQZ level yields a barrier of −0.9 kcal mol^{-1} after

TABLE 3.8: Focal point layout (kcal mol^{-1}) for the formation of ethylene and HO_2 from ethyl + O_2. Computed values are listed without brackets; extrapolated values are shown in brackets.

Basis	ΔE_{HF}	δ[ZAPT2]	δ[CCSD]	$\delta(T)$	δ[CCSDT]	$\delta(Q)$	ΔE[CCSDT(Q)]
cc-pVDZ	−8.92	+15.16	−13.47	−0.30	−0.34	+0.24	−7.63
cc-pVTZ	−11.29	+12.94	−13.66	−0.56	−0.34	[+0.24]	[−12.67]
cc-pVQZ	−11.60	+12.07	−13.65	−0.67	−0.34	[+0.24]	[−13.95]
cc-pV5Z	−11.74	+11.63	−13.65	[−0.67[	[−0.34]	[+0.24]	[−14.53]
cc-pV6Z	[−11.79]	[+11.43]	[−13.65]	[−0.67]	[−0.34]	[+0.24]	[−14.78]
BS	[−11.82]	[+11.16]	[−13.65]	[−0.67]	[−0.34]	[+0.24]	[−15.08]
Extrapolation	$a + be^{-cX}$	$a + bX^{-3}$					
Fit Points	3,4,5	3,4					

ZPVE corrections. Extrapolating to the complete basis set CCSD(T) limit lowers **TS1** to −2.1 kcal mol^{-1}. Inclusion of the CCSDT and CCSDT(Q) corrections finally brings **TS1** to −3.0 kcal mol^{-1}, in excellent agreement with the empirical barrier height. Collectively, the core correlation and relativistic shifts (Table 3.4) alter the **TS1** barrier by only 0.01 kcal mol^{-1}. Our results clearly show that a balanced treatment of the reactants and transition state **TS1** requires basis sets of at least cc-pV5Z quality and correlation treatments beyond CCSDT.

For the ethylperoxy radical **2a**, the C_2H_5-O_2 binding energy is surprisingly good at the Hartree-Fock limit, only about 8 kcal mol^{-1} too low and 5.5 kcal mol^{-1} better than CBS ZAPT2 (Table 3.6). Correlation convergence is achieved once the δ[CCSD] = −12.60 and δ[CCSD(T)] = −0.93 kcal mol^{-1} increments are included; the corresponding δ[CCSDT] and δ[CCSDT(Q)] terms are only 0.22 kcal mol^{-1} in magnitude and precisely cancel each other. Optimizing **2a** at the CCSD(T)/cc-pVQZ level yields a well depth D_0 of 31.87 kcal mol^{-1} after zero-point correction. The focal point analysis yields a CBS CCSD(T) value of D_0 = 32.96 kcal mol^{-1} for **2a**, and with all corrections the final value is only 0.03 kcal mol^{-1} higher. Adding in the **2a**-**2b** energy difference places the lowest energy conformer of the ethylperoxy radical 33.2 kcal mol^{-1} below the C_2H_5 + O_2 reactants. This improved result demonstrates the good accuracy of Miller and Klippenstein's previous value of 33.5 kcal mol^{-1}.[9] Similar focal point trends are observed for the formation of oxirane (Table 3.7) and ethylene (Table 3.8) . The Hartree-Fock energy rapidly converges to the basis set limit, while the residual basis set error for the correlation treatments should not be greater than 0.3 kcal mol^{-1}.

Again, in contrast to **TS1**, correlation increments tend to oscillate, producing a CCSD(T) energy which is fortuitously close to the focal point limit.

3.4.4 Enthalpy of Formation of Ethylperoxy Radical

Our focal point analyses for the reactions of ethyl + O_2 yielding ethylene + HO_2, oxirane + OH, and acetaldehyde + OH yield final reaction energies (Table 3.4) in good accord with the best enthalpies of formation from the literature.[76–78,81–83] Nonetheless, the theoretical exothermicities are systematically greater than the experimental ones by 0.2-0.6 kcal mol^{-1}. If $\Delta_f H(298)$ (C_2H_5) = 28.7 $\pm$ 0.1 kcal mol^{-1} is accepted from Ref. 76, then the disparities are attributable to inexactness of the theoretical treatment of the O_2 molecule.

TABLE 3.9: Enthalpy of formation of ethylperoxy radical **2b** at 298 K computed via four reference reactions. All values in kcal mol^{-1}.

Reaction	$\Delta_f H^\circ$(298 K) reactants[a]	$\Delta_{rxn} H^\circ$(0 K)[b]	$\Delta_{rxn} H^\circ$(298 K)[b]	$\Delta_f H^\circ$(298 K) of **2b**[c]
C_2H_5 + O_2 $\rightarrow$ 2b	+28.70	−33.23	−34.52	−5.82
c-CH_2CH_2O + OH $\rightarrow$ 2b	−3.67	−0.49	−1.42	−5.09
CH_3CHO + OH $\rightarrow$ 2b	−30.81	+26.87	+25.43	−5.38
C_2H_4 + HO_2 $\rightarrow$ 2b	+15.64	−19.49	−20.66	−5.02

[a] See footnote b of Table 3.4
[b] Enthalpies of reaction at 0 K and 298 K computed in this work using the focal point energies and thermal corrections.
[c] Enthalpy of formation of ethylperoxy radical from literature $\Delta_f H^\circ$ (reactants) and computed $\Delta_{rxn} H^\circ$(298 K).

We can employ our thermochemical data to refine the enthalpy of formation of the ethylperoxy radical by averaging results from four different routes that form **2b**, as shown in Table 3.9. Conjoining the selected literature enthalpies of formation of C_2H_5, c-CH_2CH_2O, CH_3CHO, C_2H_4, OH, and HO_2 (Table 3.4) with our computed enthalpies (derived from Table 3.9) for the four reactions that yield **2b**, we obtain $\Delta_f H^\circ$(298K) values for ethylperoxy radical ranging from −5.02 to −5.82 kcal mol^{-1} (Table 3.9). The apparent systematic errors in the theoretical reaction energies in Table 3.4 suggest that the $\Delta_f H^\circ$(298K) value of −5.82 kcal mol^{-1} derived from the C_2H_5 + O_2 reaction may be too negative by roughly 0.5 kcal mol^{-1}. Therefore, we recommend $\Delta_f H^\circ$(298K) = -5.3 ± 0.5 kcal mol^{-1} for ethylperoxy, which corresponds to $\Delta_f H^\circ$(0K)= -1.5 ± 0.5 kcal mol^{-1} using our thermal corrections for **2b** and standard H°(298K) - H°(0K) differences for H_2(g), O_2(g), and C(gr).[84] By comparison, Okumura et al.[10] determined $\Delta_f H^\circ$(298K) = -6.8 ± 2.3 kcal mol^{-1}

for **2** using a combination of experimental bond enthalpies from photoelectron spectroscopy and CBS/APNO computations.

3.5 Summary

The concerted elimination mechanism for the reaction of ethyl + O_2 to yield ethylene + HO_2 has been established by rigorous focal point computations that give a theoretical **TS1** barrier in precise accord with master equation kinetic models fitted to measured reaction rates. The concerted elimination barrier presents an extreme challenge for electronic structure theory, with previous correlated methods placing **TS1** as much as 19 kcal mol^{-1} above and 10 kcal mol^{-1} below the ethyl + O_2 reactants. The most rigorous ab initio result prior to this work was a CCSD(T)/TZ2Pf computation that gave a barrier of -0.9 kcal mol^{-1}.[11] Arduous optimization of the **TS1** geometry and evaluation of the barrier height at the CCSD(T)/cc-pVQZ level does not change the value of -0.9 kcal mol^{-1}. Two improvements past CCSD(T)/cc-pVQZ are essential to converging the **TS1** barrier: first, a 1.2 kcal mol^{-1} lowering by extrapolation to the complete basis set limit of CCSD(T), and second, an additional 0.9 kcal mol^{-1} reduction by incorporation of higher-order correlation via the CCSDT(Q) method. The overall effect of -2.1 kcal mol^{-1} is a critical result, because shifts of this magnitude can engender sevenfold changes in the branching fractions for the ethyl + O_2 reaction.[9] The final concerted elimination barrier from our focal point analysis of the direct $C_2H_5 + O_2 \rightarrow$ **TS1** process is -3.0 kcal mol^{-1}. This result provides a consensus between the highest levels of ab initio theory and optimal values from the kinetic models.[9,39] Nevertheless, minor revisions of the **TS1** barrier may still be necessary in the future. In particular, an alternative, indirect means of determining the concerted elimination barrier for $C_2H_5 + O_2 \rightarrow C_2H_4 + HO_2$ would use our theoretical data to construct a focal point layout for the reverse barrier $C_2H_4 + HO_2 \rightarrow$ TS1 and then append the reaction energy given by the best available enthalpies of formation from the literature. To achieve consistency between the direct, forward computations and the indirect, reverse approach, **TS1** would have to lie -2.6 ± 0.5 kcal mol^{-1} below C_2H_5 + O_2. Therefore, while we must still recommend a concerted elimination barrier of -3.0 kcal mol^{-1}

based on direct computations, this value may migrate upward by a few tenths of a kcal mol^{-1} as existing thermochemical inconsistencies are resolved.

3.6 Acknowledgments

Dr. Justin Turney, Andy Simmonett, and Francesco Evangelista are thanked for helpful discussions. This research was supported by the U.S. Department of Energy DOE Grant No. DE-FG02-97ER14748. The DOE computing resources of the National Energy Research Supercomputing Center at Lawrence Berkeley National Laboratory were essential to this research.

BIBLIOGRAPHY

[1] C. V. Naik, A. M. Dean, Combustion Flame **145**, 16 (2006).

[2] E. W. Kaiser, J. Phys. Chem. A **106**, 1256 (2002).

[3] W. Ludwig, B. Brandt, G. Friedrichs, F. Temps, J. Phys. Chem. A **110**, 3330 (2006).

[4] M. S. Stark, J. Am. Chem. Soc. **122**, 4162 (2000).

[5] U. Platt, G. Lebras, G. Poulet, J. P. Burrows, G. Moortgat, Nature **348**, 147 (1990).

[6] G. Meloni, P. Zou, S. J. Klippenstein, M. Ahmed, S. R. Leone, C. A. Taatjes, D. L. Osborn, J. Am. Chem. Soc. **128**, 13559 (2006).

[7] C. A. Taatjes, J. Phys. Chem. A **110**, 4299 (2006).

[8] T. B. Hunter, T. A. Litzinger, H. Wang, M. Frenklach, Combustion Flame **104**, 505 (1996).

[9] J. A. Miller, S. J. Klippenstein, Int. J. Chem. Kinetics **33**, 654 (2001).

[10] S. Blanksby, T. Ramond, G. Davico, M. Nimlos, S. Kato, V. Bierbaum, W. Lineberger, G. Ellison, M. Okumura, J. Am. Chem. Soc. **123**, 9585 (2001).

[11] R. Rienstra-Kiracofe, W. D. Allen, H. F. Schaefer, J. Phys. Chem. A **104**, 9823 (2000).

[12] H. H. Carstensen, C. V. Naik, A. M. Dean, J. Phys. Chem. A **109**, 2264 (2005).

[13] J. DeSain, S. Klippenstein, J. Miller, C. Taatjes, J. Phys. Chem. A **107**, 4415 (2003).

[14] A. Andersen, E. A. Carter, J. Phys. Chem. A **107**, 9463 (2003).

[15] P. Marshall, J. Phys. Chem. A **103**, 4560 (1999).

[16] R. R. Baldwin, C. E. Dean, R. W. Walker, J. Chem. Soc. Faraday Transactions Ii **82**, 1445 (1986).

[17] I. R. Slagle, Q. Feng, D. Gutman, J. Phys. Chem. **88**, 3648 (1984).

[18] E. P. Clifford, J. T. Farrell, J. D. Desain, C. A. Taatjes, J. Phys. Chem. A **104**, 11549 (2000).

[19] I. C. Plumb, K. R. Ryan, Int. J. Chem. Kinetics **13**, 1011 (1981).

[20] K. G. McAdam, R. W. Walker, J. Chem. Soc. Faraday Transactions Ii **83**, 1509 (1987).

[21] O. Dobis, S. W. Benson, J. Am. Chem. Soc. **115**, 8798 (1993).

[22] D. B. Atkinson, J. W. Hudgens, J. Phys. Chem. A **101**, 3901 (1997).

[23] I. S. Ignatyev, Y. M. Xie, W. D. Allen, H. F. Schaefer, J. Chem. Phys. **107**, 141 (1997).

[24] G. E. Quelch, M. M. Gallo, M. Z. Shen, Y. M. Xie, H. F. Schaefer, D. Moncrieff, J. Am. Chem. Soc. **116**, 4953 (1994).

[25] G. E. Quelch, M. M. Gallo, H. F. Schaefer, J. Am. Chem. Soc. **114**, 8239 (1992).

[26] A. F. Wagner, I. R. Slagle, D. Sarzynski, D. Gutman, J. Phys. Chem. **94**, 1853 (1990).

[27] P. K. Venkatesh, A. M. Dean, M. H. Cohen, R. W. Carr, J. Chem. Phys. **111**, 8313 (1999).

[28] J. W. Bozzelli, A. M. Dean, J. Phys. Chem. **97**, 4427 (1993).

[29] G. Barbieri, F. P. Dimaio, P. G. Lignola, Combustion Sci. Technology **98**, 95 (1994).

[30] J. DeSain, C. Taatjes, J. Miller, S. Klippenstein, D. Hahn, Faraday Discussions **119**, 101 (2001).

[31] E. G. Estupinan, S. J. Klippenstein, C. A. Taatjes, J. Phys. Chem. B **109**, 8374 (2005).

[32] T. S. Dibble, J. Am. Chem. Soc. **123**, 4228 (2001).

[33] E. W. Kaiser, J. Phys. Chem. A **102**, 5903 (1998).

[34] E. W. Kaiser, T. J. Wallington, J. M. Andino, Chem. Phys. Lett. **168**, 309 (1990).

[35] E. W. Kaiser, J. Phys. Chem. **99**, 707 (1995).

[36] E. W. Kaiser, L. Rimai, T. J. Wallington, J. Phys. Chem. **93**, 4094 (1989).

[37] R. R. Baldwin, I. A. Pickering, R. W. Walker, J. Chem. Soc. Faraday Transactions I **76**, 2374 (1980).

[38] C. Y. Sheng, J. W. Bozzelli, A. M. Dean, A. Y. Chang, J. Phys. Chem. A **106**, 7276 (2002).

[39] J. A. Miller, S. J. Klippenstein, S. H. Robertson, Proc. Combustion Institute **28**, 1479 (2000).

[40] J. W. Ochterski, G. A. Petersson, K. B. Wiberg, J. Am. Chem. Soc. **117**, 11299 (1995).

[41] S. E. Wheeler, W. D. Allen, H. F. Schaefer, J. Chem. Phys. **121**, 8800 (2004).

[42] D. Feller, D. A. Dixon, J. S. Francisco, J. Phys. Chem. A **107**, 1604 (2003).

[43] M. S. Stark, J. Phys. Chem. A **101**, 8296 (1997).

[44] A. G. Csàszàr, W. D. Allen, H. F. Schaefer, J. Chem. Phys. **108**, 9751 (1998).

[45] A. L. L. East, W. D. Allen, J. Chem. Phys. **99**, 4638 (1993).

[46] J. M. Gonzales, C. Pak, R. S. Cox, W. D. Allen, H. F. Schaefer, A. G. Csàszàr, G. Tarczay, Chem. Eur. J. **9**, 2173 (2003).

[47] M. S. Schuurman, S. R. Muir, W. D. Allen, H. F. Schaefer, J. Chem. Phys. **120**, 11586 (2004).

[48] K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[49] R. J. Bartlett, J. D. Watts, S. A. Kucharski, J. Noga, Chem. Phys. Lett. **165**, 513 (1990).

[50] R. J. Bartlett, Chem. Phys. Lett. **167**, 609 (1990).

[51] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

[52] R. A. King, W. D. Allen, H. F. Schaefer, J. Chem. Phys. **112**, 5585 (2000).

[53] H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, P. Celani, T. Korona, G. Rauhut, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, A. W. Lloyd, S. J. McNicholas, W. Meyer, M. E. Mura, P. Palmieri, A. Nicklass, R. Pitzer, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson. MOLPRO, version 2002.1, a package of ab initio programs. 2002. http://www.molpro.net.

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

[55] J. F. Stanton, J. Gauss, J. D. Watts, G. Szalay, R. J. Bartlett, A. A. Auer, D. B. Bernholdt, O. Christiansen, M. E. Harding, M. Heckert, O. Heun. ACES II, Mainz-Austin-Budapest version. 2005. For the current version, see http://www.aces2.de.

[56] T. J. Lee, D. Jayatilaka, Chem. Phys. Lett. **201**, 1 (1993).

[57] T. Helgaker, W. Klopper, H. Koch, J. Noga, J. Chem. Phys. **106**, 9639 (1997).

[58] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).

[59] J. M. L. Martin, O. De, J. Chem. Phys. **111**, 1843 (1999).

[60] Y. J. Bomble, J. F. Stanton, M. Kàllay, J. Gauss, J. Chem. Phys. **123**, 054101 (2005).

[61] M. Kàllay, J. Gauss, J. Chem. Phys. **123**, 214105 (2005).

[62] C. L. Janssen, I. M. B. Nielsen, M. L. Leininger, E. F. Valeev, J. P. Kenny, E. T. Seidl. The Massively Parallel Quantum Chemistry Program (MPQC), version 3.0. 2008. Sandia National Laboratories, Livermore, CA, USA. http://www.mpqc.org.

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

[64] P. J. Knowles, C. Hampel, H.-J. Werner, J. Chem. Phys. **99**, 5219 (1993).

[65] M. Kàllay, P. R. Surjàn, J. Chem. Phys. **115**, 2945 (2001).

[66] D. E. Woon, T. H. Dunning, J. Chem. Phys. **103**, 4572 (1995).

[67] S. A. Perera, R. J. Bartlett, Chem. Phys. Lett. **216**, 606 (1993).

[68] N. C. Handy, Y. Yamaguchi, H. F. Schaefer, J. Chem. Phys. **84**, 4481 (1986).

[69] K. Pitzer, W. D. Gwinn, J. Chem. Phys. **10**, 428 (1942).

[70] P. Rupper, E. N. Sharp, G. Tarczay, T. A. Miller, J. Phys. Chem. A **111**, 832 (2007).

[71] D. Jayatilaka, T. J. Lee, J. Chem. Phys. **98**, 9734 (1993).

[72] J. D. Watts, J. Gauss, R. J. Bartlett, Chem. Phys. Lett. **200**, 1 (1992).

[73] W. J. Lauderdale, V. G. Cheng, S. G. Wierschke, J. Phys. Chem. **98**, 4502 (1994).

[74] M. Urban, J. D. Watts, R. J. Bartlett, Int. J. Quantum Chem. **52**, 211 (1994).

[75] A. Karton, P. R. Taylor, J. M. L. Martin, J. Chem. Phys. **127**, 064104 (2007).

[76] M. A. Hanninglee, N. J. B. Green, M. J. Pilling, S. H. Robertson, J. Phys. Chem. **97**, 860 (1993).

[77] B. Ruscic, R. E. Pinzon, M. L. Morton, N. K. Srinivasan, M. C. Su, J. W. Sutherland, J. V. Michael, J. Phys. Chem. A **110**, 6592 (2006).

[78] K. B. Wiberg, L. S. Crocker, K. M. Morgan, J. Am. Chem. Soc. **113**, 3447 (1991).

[79] A. Karton, E. Rabinovich, J. M. L. Martin, B. Ruscic, J. Chem. Phys. **125**, 144108 (2006).

[80] A. C. Simmonett, F. A. Evangelista, W. D. Allen, H. F. Schaefer, J. Chem. Phys. **127**, 014306 (2007).

[81] G. da Silva, J. W. Bozzelli, J. Phys. Chem. A **110**, 13058 (2006).

[82] A. S. Pell, G. Pilcher, Transactions Faraday Soc. **61**, 71 (1965).

[83] B. Ruscic, A. F. Wagner, L. B. Harding, R. L. Asher, D. Feller, D. A. Dixon, K. A. Peterson, Y. Song, X. M. Qian, C. Y. Ng, J. B. Liu, W. W. Chen, J. Phys. Chem. A **106**, 2727 (2002).

[84] M. W. Chase, J. Phys. Chem. Ref Data **14**, 1 (1985).

CHAPTER 4

EXPLICITLY-CORRELATED 2ND-ORDER PERTURBATION THEORY FOR OPEN-SHELLS [†]

[†]J. J. Wilke and H. F. Schaefer, J. Chem. Phys. **131**, 244116 (2009). Reprinted here with permission of the American Institute of Physics.

4.1 Abstract

Explicitly-correlated MP2-R12 and coupled-cluster R12 methods have proven to be effective in achieving the basis set limit of correlated wavefunction methods. However, correlated methods for high-spin open-shell states are typically based on semi-canonical orbitals, requiring a different set of amplitudes to be computed for each spin case. In contrast, Z-averaged perturbation theory redefines the Hamiltonian with a symmetric exchange operator, thereby allowing a spin-restricted formulation with equivalent α and β subspaces. In the current work, we present a preliminary study of explicitly correlated ZAPT for second-order perturbation theory. The superior basis set convergence of R12 methods is demonstrated for a set of atomization energies, showing the R12 results to be competitive with common basis set extrapolation techniques, albeit at a fraction of the cost. Given the efficiency gains associated with the symmetric exchange operator, we suggest ZAPT as a candidate for reducing the cost of current open-shell MP2-R12 and CCSD(T)-R12 computations.

4.2 Introduction

Coupled-cluster (CCSD) and second-order Møller Plesset (MP2) perturbation theories are ubiquitous in quantum chemistry, providing efficient approximations for treating electron correlation. For chemical accuracy better than 1 kcal mol^{-1}, connected triple excitations must be included in the cluster operator,[1–3] which is efficiently accomplished through the CCSD(T) perturbative estimate.[4,5] For larger molecules, MP2 provides a more approximate, but very economical treatment of electron correlation, particularly for hydrogen bonding.[6,7] However, MP2 and CCSD(T), like all wavefunction-based correlated methods, converge very slowly to the basis set limit,[8–10] which is now understood as the inability of orbital-based expansions to properly describe the cusp and Coulomb hole at electron-electron coalescence points. Inclusion of terms in the wavefunction that depend explicitly on the interelectronic coordinate r_{12} can include the cusp exactly, greatly accelerating the basis set convergence. In this direction, explicitly correlated MP2-R12 and CCSD(T)-R12 approaches have been developed, combining the intrinsic accuracy of the correlation treatments with

the superior basis set convergence of the R12 methods. Using CCSD(T)-R12 with modest basis sets, thermochemistry often has an accuracy better than 3.0 kJ mol^{-1} while vibrational frequencies usually agree with experiment within 4 cm^{-1}![11,12]

Given their perturbative nature, MP2 and CCSD(T) require the full Hamiltonian to be partitioned into a zeroth-order Fock operator, $\hat{H}_0$, and first-order fluctuation potential, $\hat{V}$. For RHF and UHF reference functions, the most natural choice of $\hat{H}_0$ is the standard spin-orbital Fock operator (see below). For ROHF reference functions, however, the standard Brillouin variational condition no longer holds, requiring $\hat{H}_0$ to be redefined. As a consequence, properties such as orbital invariance, same spatial orbitals for different spins, and spin-restriction, easily maintained in closed-shell MP2, may no longer be satisfied for ROHF references. Choosing a partition for the Hamiltonian therefore requires numerous subtleties to be considered which are often taken for granted in closed-shell theories. Several methods for partitioning the Hamiltonian were developed in the first half of the 1990s[13] including open-shell perturbation theory (OPT1 and OPT2),[14–16] invariant open-shell perturbation theory (IOPT),[17] restricted open-shell Møller-Plesset (ROMP),[18] restricted Møller-Plesset (RMP),[19,20] and Z-averaged perturbation theory (ZAPT).[21–24] The most straightforward perturbation theory is RMP. However, despite rigorous spin-restriction of the Hartree-Fock wavefunction, the different number of α and β electrons in RMP leads to different exchange operators for the α and β spin-orbital subspaces. For correlated wavefunctions, this asymmetric exchange leads to an essentially unrestricted framework, drastically increasing the cost relative to closed-shell methods. In contrast, ZAPT employs a symmetric exchange operator in $\hat{H}_0$ that restores the equivalence of α and β subspaces, allowing a spin-restricted theory to be formulated which is essentially equivalent in cost to a closed-shell computation.

Given the success of R12 approaches for MP2 and coupled-cluster methods, we present an initial formulation of the symmetric exchange approach for explicitly correlated second-order perturbation theory to establish a ZAPT-R12 framework. We first give a brief summary of the numerous ROHF-based perturbation theories, comparing the advantages and disadvantages of the symmetric exchange approach. We then introduce the MP2-R12 method, considering the additional subtleties caused by the choice of a symmetric exchange operator.

4.3 ROHF Perturbation Theories

In the discussion to follow, the following notation is used to distinguish orbital subspaces.

$i, j, k, l \ldots$	DOCC	Doubly occupied orbitals
$a, b, c, d \ldots$	VIR	Virtual orbitals
$s, t \ldots$	SOCC	Singly occupied orbitals
$p, q \ldots$	ORB	Entire orbital basis set
$p'' q'' \ldots$	CABS	External orbitals

The orbital basis set is the finite computational basis, and is the union of the occupied and virtual orbitals. Lower case letters indicate spin orbitals while upper case letters will indicate spatial orbitals.

The complications in ROHF theories derive from the off-diagonal elements of the spin-orbital Fock operator

$$f_{ia} = \left\langle \Phi_i^a \left| \hat{H} \right| \Phi_0 \right\rangle \tag{4.1}$$

whose matrix elements are given by

$$f_{pq} = h_{pq} + \sum_i \langle pi || qi \rangle \tag{4.2}$$

Here h_{pq} contains the kinetic energy and nuclear attraction terms while the sum over anti-symmetrized two-electron integrals runs over all occupied orbitals. Because of the non-zero f_{ia} block, Φ_0 is no longer an eigenfunction of the Fock operator.

The most straightforward method for redefining $\hat{H}_0$ is RMP[13,19,20] which simply projects out the f_{ia} terms

$$\hat{H}_0 = \hat{o}\hat{f}\hat{o} + \hat{v}\hat{f}\hat{v} \tag{4.3}$$

through the occupied space, $\hat{o}$, and virtual space, $\hat{v}$, projectors. In solving the MP2 equations, the orbitals are usually "semi-canonicalized" by diagonalizing the OCC-OCC and VIR-VIR blocks of the Fock matrix. However, since the α and β Fock matrices differ based on asymmetric exchange with the singly-occupied orbitals, diagonalization within these subspaces breaks the α-β orbital

symmetry, leading to a different orbitals for different spins (DODS) formalism. In particular, the Fock operators for α and β subspaces can be distinguished by

$$\hat{f}_\alpha = \hat{h} + \hat{J} - \hat{K}_c - \hat{K}_o \tag{4.4}$$

$$\hat{f}_\beta = \hat{h} + \hat{J} - \hat{K}_c \tag{4.5}$$

where $\hat{h}$ is the core Hamiltonian, $\hat{J}$ is the Coulomb operator, and $\hat{K}_c$ and $\hat{K}_o$ are the operators for exchange with closed-shell and open-shell orbitals, respectively. If canonical rather than semi-canonical orbitals are used, only one integral transformation is required, but the MP2 amplitudes must be solved iteratively.

The partitioning is spin-dependent, and the RMP wavefunction is therefore spin-contaminated. While the energy is not rigorously spin-restricted, it can be classified as spin-projected.[24] The MP2 energy

$$\left\langle \Phi_0 \left| \hat{H} \right| \Psi^{(1)} \right\rangle \tag{4.6}$$

is therefore equivalent to

$$\left\langle \Phi_0 \left| \hat{H}\hat{P} \right| \Psi^{(1)} \right\rangle \tag{4.7}$$

where $\hat{P}$ is an operator that projects out spin-contaminants. RMP therefore usually provides more meaningful correlation energies than UMP2.[25,26]

In an effort to provide a fully spin-restricted ROHF perturbation theory, Davidson and co-workers introduced the OPT1 method.[16,17] Rather than partition the Hamiltonian through spin-dependent Fock operators, the Hamiltonian is partitioned based on a spin-averaged operator

$$F_{av} = \frac{1}{2}\left(\hat{f}_\alpha + \hat{f}_\beta\right) \tag{4.8}$$

so that the Fock operator becomes

$$\hat{f} = \hat{h} + \hat{J} - \hat{K}_c - \frac{1}{2}\hat{K}_o \tag{4.9}$$

where the open-shell exchange has been symmetrized. $\hat{H}_0$ is therefore spin-independent and rigorously spin-restricted. The spin-independent orbital canonicalization produces equivalent orbitals for different spins, so that OPT1 is equivalent to closed-shell MP2 in computational cost. Unfortunately, OPT1 correlation energies are spuriously large. Both occupied and unoccupied SOCC orbitals have the same eigenvalues, leading to small energy denominators. Furthermore, to avoid mixing occupied and virtual orbitals during canonicalization, $\hat{H}_0$ must take the form

$$\hat{H}_0 = \hat{d}\hat{f}\hat{d} + \hat{s}\hat{f}\hat{s} + \hat{v}\hat{f}\hat{v} \tag{4.10}$$

where $\hat{d}$ and $\hat{s}$ are projectors for the doubly-occupied and singly-occupied spaces. The OPT1 operator therefore does not include the block $\hat{f}_{is}$ between SOCC and DOCC spaces so that, unlike RMP, the theory is not invariant to SOCC-DOCC mixing. To correct the energy denominator problem, a second theory, OPT2, was introduced. Two-electron operators are included in $\hat{H}_0$, producing different eigenvalues for occupied and unoccupied SOCC orbitals. As with OPT1, only a singlet set of orbitals is required, but the correlation energies obtained are much smaller. Unfortunately, rotations among degenerate orbitals within the SOCC space are no longer allowed, producing orbital invariance problems as large as 9 kcal mol^{-1} for some systems. Especially for explicitly correlated methods, a two-electron $\hat{H}_0$ is also undesirable.

Z-averaged perturbation theory (ZAPT) corrects the energy denominator problem of OPT1 by defining separate Fock operators for doubly-occupied and singly-occupied orbitals. For DOCC orbitals, the symmetric OPT1 operator is used, maintaining equivalence of α and β orbitals within the DOCC space. Asymmetric exchange is used for the SOCC orbitals, giving

$$\hat{f} = \hat{h} + \hat{J} - \hat{K}_c - \hat{K}_o \tag{4.11}$$

for occupied SOCC orbitals and

$$\hat{f} = \hat{h} + \hat{J} - \hat{K}_c \tag{4.12}$$

for unoccupied SOCC orbitals. For multiple SOCC orbitals within the same irreducible representation, the different exchange operators for unoccupied and occupied SOCC orbitals will produce different ZAPT canonical orbitals. The majority of open-shell applications will be doublets or multiplet states occupying orbitals of different symmetries.[22] Even if SOCC spatial symmetry is broken, the DOCC orbitals dominate the computation and the cost will still be essentially equivalent to closed-shell MP2.

In contrast to OPT1, the ZAPT F_{av} operator is only spin-averaged, but not spin-independent. As in RMP, the correlation energies will not be spin-restricted, but rather spin-projected. In contrast to RMP, though, ZAPT can be solved non-iteratively with only a single integral transformation. Furthemore, Lee has argued that the ZAPT wavefunction contains "less spin contamination" than RMP, and should therefore give slightly improved correlation energies.[24,26] For conventional perturbation theory, ZAPT therefore presents an excellent compromise between efficiency and accuracy. For a full review of of ROHF-based perturbation theories, we refer the reader to references 13 and 24.

Despite the simplifications associated with a single set of spatial orbitals, closed-shell MP2 and coupled-cluster theories are dramatically simplified by the amplitude relation

$$T_{i_\alpha j_\alpha}^{a_\alpha b_\alpha} = T_{i_\alpha j_\beta}^{a_\alpha b_\beta} - T_{j_\alpha i_\beta}^{a_\alpha b_\beta} \tag{4.13}$$

that allows the theory to be formulated through a single set of intermediates. For high-spin open-shell states, α-β amplitude symmetry is broken in coupled-cluster theories regardless of the orbital canonicalization due to the unequal number of α and β electrons. Although ZAPT is intrinsically a redefinition of $\hat{H}_0$ based on open-shell exchange, it can be equivalently derived by using symmetric spin functions in the SOCC space

$$\begin{aligned} \sigma^+ &= \frac{1}{\sqrt{2}}(\alpha + \beta) \\ \sigma^- &= \frac{1}{\sqrt{2}}(\alpha - \beta) \end{aligned} \tag{4.14}$$

so that the SOCC orbitals are "Z-averaged" with respect to the DOCC orbitals. In this regard, symmetric exchange for α and β subspaces results not from a particular choice of $\hat{H}_0$, but directly from the symmetric spin basis. The symmetric spin basis produces a Z-averaged generalization[27] of (4.13)

$$T_{i_\alpha j_\alpha}^{a_\alpha b_\alpha} + T_{i_\alpha j_\alpha}^{a_\beta b_\beta} = T_{i_\alpha j_\beta}^{a_\alpha b_\beta} - T_{j_\alpha i_\beta}^{a_\alpha b_\beta} \tag{4.15}$$

where, despite the appealing symmetry of the amplitudes, the theory is now complicated by α,β exchange amplitudes $T_{i_\alpha j_\alpha}^{a_\beta b_\beta}$. In particular, exchange matrix terms of the form

$$K_{i_\alpha}^{j_\beta} = \sum_s \langle s_{\sigma+} s_{\sigma+} | i_\alpha j_\beta \rangle \tag{4.16}$$

are non-zero due to the symmetric spin basis. At first order, as in the MP2 wavefunction, such α,β exchange amplitudes are zero, first appearing only at second-order. If such amplitudes are neglected, the Z-averaged approach returns to a closed-shell formalism. Even if the α,β exchange amplitudes are rigorously included, spin-adapted equations can still be derived which represent a drastic reduction in computational cost relative to other high-spin open-shell approaches.[27]

To ensure well-defined orbital eigenvalues such that

$$\hat{F}_{av} |i\rangle = \epsilon_i |i\rangle \tag{4.17}$$

α, β exchange terms are not included in $\hat{H}_0$. In this way, it is the perturbation, $\hat{V}$, and not $\hat{H}_0$ that introduces α, β exchange terms in the wavefunction.

4.4 MP2-R12

Until recently, most ab initio methods have relied upon an n-electron basis comprised of antisymmetrized products of orbitals. Despite the conceptual appeal of theories formulated in one-electron orbitals and potential for high accuracy through partial wave extrapolations,[28–32] one-particle approaches converge very slowly to the basis set limit.[8–10] As noted in the introductory sections, the correlation energy generally converges as $(L+1)^{-3}$ where L denotes the highest angular momentum

included in the orbital basis.[32] For correlated wavefunction methods with N^4 basis set scaling, the residual error decays at the painfully slow rate of $t^{-\frac{1}{4}}$ with respect to total computational time![33] While the exact wavefunction requires a cusp and depletion of electron density (Coulomb hole) in the coalescence region between two electrons,[8,34,35] orbital-based expansions are smooth at the coalescence point and require very high angular momentum functions to provide an accurate shape. This troubling convergence originally led Hylleraas to formulate an explicitly correlated expansion of the Helium wavefunction[36] that included functions depending on the interelectronic distance, r_{12}. By expanding the wavefunction in two-particle functions that depend explicitly on r_{12}, the correct Coulomb hole shape can be enforced independent of any underlying one-particle expansion.

As alluded to in the introductory sections, explicitly correlated theories can be grouped into roughly two different approaches. In transcorrelated approaches[37–43] the Hamiltonian is transformed by an r_{12}-depedent operator, leading to a modified differential equation equivalent to the Schrödinger equation but involving a non-singuar Hamiltonian with no cusp requirement. In contrast, R12[9,10,44–49] and gaussian geminal methods[50–54] employ the standard Hamiltonian, but expand the wavefunction in an explicitly-correlated two particle basis. The fundamental drawback of both approaches is the appearance of numerous, difficult 3e and 4e integrals that would otherwise limit application of the techniques to small systems.[33] R12 theories have therefore achieved the greatest success by allowing the difficult integrals to be factored into products of much simpler 2e integrals through an ingenious resolution of the identity approximation[44,55] originally introduced by Klopper and Kutzelnigg. R12 methods are therefore unique amongst the explicitly correlated methods in escaping the restriction to small molecules, and, even without local[56–58] or density-fitting approximations,[59,60] can be routinely be applied to molecules as large as benzene dimer.[61] Explicitly correlated methods have now been implemented for coupled cluster,[12,62–75] configuration interaction,[76–78] and multireference perturbation theory.[79] Generally, R12 computations with the aug-cc-pVTZ[80] basis set can now achieve the accuracy of conventional aug-cc-pV5Z computations at a fraction of the cost.[81]

In addition to the previous notation, we must introduce the following notation for subspaces in R12 methods

$x, y, \ldots$		Geminal space
$p''q'' \ldots$	CABS	External orbitals
$\alpha, \beta, \ldots$	CBS	Complete basis set
p', $q', \ldots$	RI	Resolution of the identity

The geminal space generates the R12 pair correlation functions as the product of an orbital pair and the correlation factor

$$|ij\rangle \rightarrow \hat{Q}_{12}\hat{F}_{12}\,|xy\rangle \tag{4.18}$$

where $\hat{F}_{12}$ is a correlation factor depending explicitly on the interelectronic coordinate r_{12}, usually of Slater, $\exp(-r_{12})$, or Gaussian geminal, $\exp(-r_{12}^2)$, form.[82,83] $\hat{Q}_{12}$ is a projection operator necessary to ensure that $\Psi^{(1)}$ is orthogonal to Φ_0. Typically, only occupied orbitals are necessary in the geminal space, although for excited state and response properties virtual orbitals must also be included.[63,64,72] The RI space approximates the complete basis obtained within a finite computational basis, usually obtained as the union of the orbital basis set and an auxiliary basis set. The CABS (complementary auxiliary basis set) space represents those orbitals in the RI that are complementary or external to the orbital basis set. The indices α, β denote orbitals in a formally complete basis which may be used in derivations, but never enter the programmable equations.

The MP2-R12 energy can be solved variationally starting from the Hylleraas functional

$$E_2[\Psi^{(1)}] = \left\langle \Psi^{(1)} \left| \hat{H}_0 - E_0 \right| \Psi^{(1)} \right\rangle + 2\,\mathrm{Re}\left\langle \Psi^{(1)} \left| \hat{H} \right| \Phi_0 \right\rangle \tag{4.19}$$

where $\Psi^{(1)}$ is the first order perturbative wavefunction correction and E_2 is the second order correlation energy, written here as a functional of $\Psi^{(1)}$. The wavefunction $\Psi^{(1)}$ is chosen by the ansatz

$$\left|\Psi^{(1)}\right\rangle = \frac{1}{4}\left|\Phi_{ij}^{ab}\right\rangle T_{ab}^{ij} + \frac{1}{4}\left|\Phi_{ij}^{\alpha\beta}\right\rangle F_{\alpha\beta}^{kl} T_{kl}^{ij} + |\Phi_i^a\rangle\, T_a^i \tag{4.20}$$

Here repeated indices imply summation over spin orbitals. The above ansatz differs slightly from the closed-shell case by the inclusion of singly-substituted determinants. Here the matrix element

$F^{kl}_{\alpha\beta}$ is defined as

$$F^{kl}_{\alpha\beta} = \left\langle kl \left| \hat{F}_{12}\hat{Q}_{12} \right| \alpha\beta \right\rangle \tag{4.21}$$

From the form of (4.20), we see that $\Psi^{(1)}$ involves a conventional contribution, $\left|\Phi^{ab}_{ij}\right\rangle T^{ij}_{ab}$, supplemented by an R12 correction, $\left|\Phi^{\alpha\beta}_{ij}\right\rangle F^{kl}_{\alpha\beta}T^{ij}_{kl}$ that helps complete the determinantal basis.

In spin-orbital form with i, j denoting all occupied orbitals DOCC and SOCC and a, b denoting all virtual orbitals, the amplitude residuals become

$$R^{ab}_{ij} = \langle ij||ab\rangle + T^{ab}_{ik}f^{k}_{j} + T^{ab}_{kj}f^{k}_{i} + T^{ac}_{ij}f^{b}_{c} + T^{cb}_{ij}f^{a}_{c} + T^{xy}_{ij}C^{ab}_{xy} \tag{4.22}$$

$$R^{xy}_{ij} = V^{xy}_{ij} + B^{xy}_{wz}T^{wz}_{ij} - X^{xy}_{wz}(f^{k}_{i}T^{wz}_{kj} + f^{k}_{j}T^{wz}_{ik}) + C^{xy}_{ab}T^{ab}_{ij} \tag{4.23}$$

The matrix elements in (4.22) and (4.23) are defined by

$$V^{xy}_{ij} = \left\langle ij \left| \frac{1}{r_{12}}\hat{Q}_{12}\hat{F}_{12} \right| xy \right\rangle \tag{4.24}$$

$$B_{wz,xy} = \left\langle wz \left| \hat{F}_{12}\hat{Q}_{12}\left(\hat{f}_1 + \hat{f}_2\right)\hat{Q}_{12}\hat{F}_{12} \right| xy \right\rangle \tag{4.25}$$

$$X_{wz,xy} = \left\langle wz \left| \hat{F}_{12}\hat{Q}_{12}\hat{F}_{12} \right| xy \right\rangle \tag{4.26}$$

$$C^{ab}_{xy} = \left\langle xy \left| \hat{F}_{12}\hat{Q}_{12}\left(\hat{f}_1 + \hat{f}_2\right) \right| ab \right\rangle \tag{4.27}$$

V is the R12 generalization of the electron repulsion integrals, where instead of correlating an occupied i, j pair with a virtual a, b pair the occupied orbitals are correlated with an R12 geminal.

$$\left\langle ij \left| \tfrac{1}{r_{12}} \right| ab \right\rangle \Rightarrow \left\langle ij \left| \tfrac{1}{r_{12}}\hat{Q}_{12}\hat{F}_{12} \right| xy \right\rangle \tag{4.28}$$

B is a generalization of the MP2 energy denominator. X is the overlap matrix between R12 geminals, and C is the matrix coupling conventional and explicit terms. After computing the matrix elements, the equations may be solved iteratively for the amplitudes. Alternatively, for canonical orbitals, the equations can be inverted and solved non-iteratively. Once amplitudes have

been obtained, the pair energies are

$$E_{ij} = \sum_{ab} T_{ij}^{ab}(\langle ij||ab\rangle + R_{ij}^{ab}) + \sum_{xy} T_{ij}^{xy}(V_{ij}^{xy} + R_{ij}^{xy}) \tag{4.29}$$

By inclusion of the residuals in the energy expression, the error in the correlation energy will become quadratic in the amplitude error, which is important for methods which do not fully optimize the amplitudes. From (4.29), the R12 pair energies are seen to be the conventional MP2 energy plus an R12 correction. However, we emphasize that if the coupling matrix C is non-zero, the amplitudes T_{ij}^{ab} in MP2-R12 are not equal to the conventional MP2 value. By effectively treating the short range correlation, the explicitly correlated terms relax the conventional MP2 amplitudes.

4.5 Variations and Approximations within R12 Methods

4.5.1 Projection operators

The projection operator is required to make the correlating R12 geminals orthogonal to the occupied space. For the projection operator $\hat{Q}_{12}$ in (4.21), the most common choice is[45,81]

$$\hat{Q}_{12} = (1 - \hat{o}_1)(1 - \hat{o}_2)(1 - \hat{v}_1\hat{v}_2) \tag{4.30}$$

The operators $\hat{o}_1$, $\hat{o}_2$ ensure orthogonality of the R12 geminals on the reference function, while the virtual space projector additionally enforces orthogonality against the conventional MP2 terms in (4.20). In ZAPT, the complete basis wavefunction would require singly-substituted determinants of the form $\Phi_i^{p''}$, which are annihilated by the above projectors. However, it can be shown that the singles contribution to the correlation energy converges rapidly to the basis set limit. In this regard, the explicitly correlated terms are not necessary for spanning the space of functions $\Phi_i^{p''}$. In addition, single excitations into external orbitals of the form $\Phi_i^{p''}$ contribute not only to the ZAPT2 correlation energy, but also perturbatively correct the inexact Hartree-Fock reference due to the finite computational basis.[65] For use in the context of focal point[84] or other additive thermochemistry schemes,[1,3] we wish to separately obtain basis set limit Hartree-Fock and MP2

correlation energies. Separating Hartree-Fock and MP2 contributions from single-excitations can be difficult.[65] Further, if the basis set limit of the correlation energy is desired rather than the total energy, Valeev has shown that inexact Hartree-Fock orbitals can yield very accurate correlation energies.[85] We therefore suggest that the cleanest approach for obtaining correlation energies is to neglect single excitations in the projection operator.

4.5.2 The general and extended Brillouin conditions

Two approximations can be made on the exactness of the Fock operator computed in the finite basis.[9,45,86] First, the generalized Brillouin condition (GBC) assumes that the occupied space is closed under the action of the Fock operator

$$\hat{f}\,|\phi_i\rangle = \sum_j f_{ij}\,|\phi_j\rangle \tag{4.31}$$

The errors associated with the GBC approximation are generally negligible compared with the other approximations.[86] Second, the extended Brillouin condition (EBC) assumes that the virtual space is also closed under the action of the Fock operator

$$\hat{f}\,|\phi_a\rangle = \sum_b f_{ab}\,|\phi_b\rangle \tag{4.32}$$

The EBC approximation is somewhat poorer than the GBC approximation. However, for reference functions computed with a sufficiently large orbital basis set, the associated errors are again very small.[86]

The basis of the GBC approximation for closed shells is the Brillouin condition, since all off-diagonal terms f_{ia} are rigorously zero. Because of the approximate nature of the Hartree-Fock wavefunction in a finite basis, Fock matrix terms of the form $f_{ip''}$ are non-zero, but would otherwise vanish in the limit of an exact reference. For an ROHF reference, such terms do not vanish even in the limit of a complete basis. However, because of our choice of partition, the terms $f_{ip''}$ are chosen to be zero in $\hat{H}_0$, instead being included in the perturbation. Therefore, while in closed-shell

MP2-R12 the GBC terms are completely neglected (neither in $\hat{H}_0$ or $\hat{V}$), in ROHF the GBC is no longer an approximation, but rather a choice of partition.

4.5.3 Amplitude ansätze

The original R12 formulation[9] only included diagonal excitations of the form

$$|ij\rangle^{\pm} \rightarrow \hat{Q}_{12}\hat{F}\,|ij\rangle^{\pm} \tag{4.33}$$

where $\pm$ indicates singlet and triplet spin-adapted pairs. The energy is not invariant to orbital rotations, which led Klopper to introduce an orbitally invariant formulation (IJKL ansatz) where off-diagonal excitations

$$|ij\rangle \rightarrow \hat{Q}_{12}\hat{F}_{12}\,|kl\rangle \tag{4.34}$$

are also included. More recently, a fixed-amplitude ansatz has been favored in which only diagonal excitations are again included, but the amplitudes T_{ij}^{ij} are not variationally optimized. The amplitudes are instead fixed at the singlet and triplet complete basis asymptotic values of 1/2 and 1/4, restoring orbital invariance. Although the diagonal and fixed amplitude methods are "less variational" than the full orbital-invariant approach, the diagonal approximations often gives superior results.[87] Since the number of R12 excitations per ij pair scales as N_{occ}^2 in the orbital-invariant approach, larger molecules will have a disproportionate number of R12 pair correlation functions, leading to geminal basis set superposition errors

Spin-adapted triplet and singlet pairs cannot be defined for DOCC-SOCC pairs, and the fixed-amplitude approach must therefore be recast in a spin-orbital formulation. The rational generator of Ten-no can be applied, which properly generates the asymptotic cusp conditions in a spin-orbital formulation.[68,73,88] The rational generator, G, is applied as

$$G\,|ij\rangle = \frac{3}{8}\,|ij\rangle + \frac{1}{8}p(12)\,|ij\rangle \tag{4.35}$$

where $p(12)$ permutes the spatial orbitals of electrons 1 and 2, but not the spins. Applied to spin-adapted triplet and singlet pairs, the rational generator can easily be verified to yield the

asymptotic amplitudes of $\frac{1}{4}$ and $\frac{1}{2}$. For DOCC-SOCC pairs, the rational generator yields geminal excitations of the form

$$|i_\alpha s_{\sigma+}\rangle \rightarrow G\,|i_\alpha s_{\sigma+}\rangle \tag{4.36}$$

$$G\,|i_\alpha s_{\sigma+}\rangle = \frac{3}{8}\,|i_\alpha s_{\sigma+}\rangle + \frac{1}{8}\,|s_\alpha i_{\sigma+}\rangle \tag{4.37}$$

In contrast to closed-shell, open-shell R12 methods must then include unoccupied orbitals (s_α) in the geminal generating space. The rational generator is indepedent of the spin functions involved, and the symmetric spin basis in ZAPT therefore introduces no additional complications with respect to previous open-shell approaches.[68]

Because of the symmetric spin basis and the presence of the occupied-space projector in $\hat{Q}_{12}$, α,β exchange terms such as $V_{i_\alpha j_\alpha}^{x_\alpha y_\beta}$ are non-zero, leading to non-zero α,β exchange amplitudes $T_{i_\alpha j_\alpha}^{x_\alpha y_\beta}$. This differs from the conventional MP2 case in which all α,β exchange excitations are zero by spin-integration

$$T_{i_\alpha j_\alpha}^{a_\beta b_\beta} = \frac{\langle i_\alpha j_\alpha || a_\alpha b_\beta\rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} = 0 \tag{4.38}$$

We must therefore clearly identify what is meant by a double excitation in the context of ZAPT. By first order interacting space arguments, Lee and Jayatilaka have argued that an α,β exchange should correspond to an excitation.[22] Although determinants of the form

$$\Phi_{i_\alpha}^{a_\beta} \tag{4.39}$$

are single *substitutions*, they should therefore be considered double *excitations*. Matrix elements $f_{i_\alpha}^{a_\alpha}$ corresponding to true single excitations are zero in ZAPT, and, unlike other ROHF perturbation theories, the ZAPT first-order wavefunction can be said to contain only doubly excited determinants. A very subtle point arises for matrix elements of the form $V_{i_\alpha j_\alpha}^{x_\alpha y_\beta}$ for α,β exchanges. The term initially seems to correspond to an artificially included α,β exchange excitation which is not included in conventional ZAPT2 and would be classified as a triple excitation. However, all

α,β exchange terms are projected out by $\hat{Q}_{12}$. The only surviving correlations in the R12 geminal are of the form

$$|i_\alpha j_\alpha\rangle \rightarrow \left|s_{\sigma_-} a_\alpha\right\rangle \tag{4.40}$$

Despite the appearance of only α and β spin-orbitals, the correlation factor is actually generating σ_- virtual pairs so that no inconsistency exists with the first order interacting space definition of double excitations, and the ZAPT2-R12 ansatz does not artificially include any α,β exchange excitations absent in conventional ZAPT2.

4.5.4 R12 Matrix Elements

In general, three different approximations have been introduced for computing integrals, denoted A′,[44] B[9,44] and C.[89] We here present the procedure for approximation C, since it does not require the commutator integrals

$$\left\langle wz \left| [\hat{T}_1, \hat{F}_{12}] \right| xy \right\rangle \tag{4.41}$$

and therefore allows integrals to be performed with minimal modification of an existing Obara-Saika code.[90,91] For all approximations, the V matrix is straightforwardly computed by expanding the projection operator as

$$\hat{Q}_{12} = 1 - \hat{o}_1\hat{p}_2^\perp - \hat{p}_1^\perp\hat{o}_2 - \hat{p}_1\hat{p}_2 \tag{4.42}$$

where $\hat{o}$, $\hat{p}^\perp$, $\hat{p}$ are the projectors for the occupied, CABS, and orbital subspaces. By directly substituting $\hat{Q}_{12}$ into the definition (4.24), we then compute

$$V_{ij}^{xy} = (Fg)_{xy}^{ij} - F_{ij}^{p''j} g_{p''j}^{xy} - F_{ij}^{iq''} g_{iq''}^{xy} - F_{ij}^{pq} g_{pq}^{xy} \tag{4.43}$$

where $(Fg)_{ij}^{xy}$ is the product of the correlation factor and electron repulsion operator

$$\left\langle ij \left| \hat{F}_{12} \frac{1}{r_{12}} \right| xy \right\rangle \tag{4.44}$$

The matrix elements X_{wz}^{xy} are computed in the same fashion as

$$X_{wz}^{xy} = \tilde{F}_{xy}^{wz} - F_{wz}^{p''j} F_{p''j}^{xy} - F_{wz}^{iq''} F_{iq''}^{xy} - F_{wz}^{pq} F_{pq}^{xy} \tag{4.45}$$

where have denoted

$$\tilde{F}_{xy}^{wz} = \left\langle wz \left| F_{12}^2 \right| xy \right\rangle \tag{4.46}$$

The computation of the matrix elements B_{kl}^{mn} is much more complicated. For approximation C, a commutator approach is used in which we begin by writing the projection operator as $\hat{Q}_{12} = 1 - \hat{P}_{12}$, so that the B matrix expression can be expanded to

$$\begin{aligned} B_{wz}^{xy} = & \left\langle wz \left| \hat{F}_{12}(\hat{f}_1 + \hat{f}_2)\hat{F}_{12} \right| xy \right\rangle \\ & - \left\langle wz \left| \hat{F}_{12}(\hat{f}_1 + \hat{f}_2)\hat{P}_{12}\hat{F}_{12} \right| xy \right\rangle \\ & - \left\langle wz \left| \hat{F}_{12}\hat{P}_{12}(\hat{f}_1 + \hat{f}_2)\hat{F}_{12} \right| xy \right\rangle \\ & + \left\langle wz \left| \hat{F}_{12}\hat{P}_{12}(\hat{f}_1 + \hat{f}_2)\hat{P}_{12}\hat{F}_{12} \right| xy \right\rangle \end{aligned} \tag{4.47}$$

For the terms involving $\hat{P}_{12}$, the projection operator is directly written out, and programmable equations are obtained by replacing complete basis set indices with the approximate, computational RI basis. A more detailed derivation can be found in Ref 89, and we here only report the final expression

$$\begin{aligned} & \left\langle wz \left| \hat{F}_{12}(\hat{P}_{12}\hat{f}_1\hat{P}_{12} - \hat{P}_{12}\hat{f}_1 - \hat{f}_1\hat{P}_{12})\hat{F}_{12} \right| xy \right\rangle \\ = & - F_{wz}^{p'm} f_{p'}^{r'} F_{r'm}^{xy} + F_{wz}^{p''m} f_m^n F_{p''n}^{xy} \\ & - F_{wz}^{pb} f_p^r F_{rb}^{xy} \\ & - F_{wz}^{p''m} f_m^{q'} F_{p''q'}^{xy} - F_{xy}^{p''m} f_m^{q'} F_{p''q'}^{wz} \\ & - F_{wz}^{p''a} f_{p''}^q F_{qa}^{xy} - F_{xy}^{p''a} f_{p''}^q F_{qa}^{wz} \end{aligned} \tag{4.48}$$

where the analogous expression for $\hat{f}_2$ is obtained by permutation of indices. We note here that all matrix elements f_{pq} in ZAPT are not simply matrix elements of the standard Fock operator, but rather the spin-averaged $\hat{H}_0$ given above.

For the initial term in (4.47), a commutator approach is applied in which we rewrite

$$\hat{F}_{12}\hat{f}_1\hat{F}_{12} = \hat{F}_{12}(\hat{f}_1 + \hat{K}_1)\hat{F}_{12} - \hat{F}_{12}\hat{K}_1\hat{F}_{12} \tag{4.49}$$

$$\hat{F}_{12}(\hat{f}_1 + \hat{K}_1)\hat{F}_{12} = \tfrac{1}{2}[[\hat{F}_{12}, \hat{f}_1 + \hat{K}_1], \hat{F}_{12}] + \tfrac{1}{2}[\hat{f}_1 + \hat{K}_1, \hat{F}_{12}^2]_+ \tag{4.50}$$

Because the correlation factor is a multiplicative operator it commutes with all local operators, and, for the closed-shell Fock operator, the commutator is therefore nonzero only for the exchange and kinetic energy terms. In this way, the double commutator term becomes

$$[[\hat{F}_{12}, \hat{f}_1 + \hat{K}_1], \hat{F}_{12}] = [[\hat{F}_{12}, \hat{T}_1], \hat{F}_{12}] = \frac{1}{2}\left(\vec{\nabla}_1 F_{12}\right)^2 \tag{4.51}$$

for which the integrals can be computed analytically for Gaussian correlation factors. For the anticommutator term, the operator $\hat{f} + \hat{K}$ is expanded in the approximate RI basis, yielding

$$\left\langle wz \left| \hat{F}_{12}^2(\hat{f}_1 + \hat{K}_1) \right| xy \right\rangle = \tilde{F}_{wz}^{p'y}(f + K)_{p'}^{x} \tag{4.52}$$

where again $\tilde{F}$ denotes matrix elements of $\hat{F}_{12}^2$.

As noted by Knizia,[65] the approach must be reassessed in the context of ROHF methods, since matrix elements of the form f_{ia} are ignored in $\hat{H}_0$. In this regard, the equality in (4.51) is no longer valid since the terms T_i^a are not included in the kinetic energy operator. Ideally, the kinetic energy contribution to the B matrix should still be computed analytically rather than by some form of direct RI insertion. In this regard, we write the ZAPT zeroth order operator as

$$\hat{H}_0 = \hat{f} + \delta\hat{f} \tag{4.53}$$

and consider the ZAPT Fock operator to be the standard Fock operator plus a difference term. We

then rewrite, inserting $\hat{H}_0$ to distinguish the ZAPT Fock operator

$$\hat{F}_{12}\hat{H}_0\hat{F}_{12} = \hat{F}_{12}(\hat{f} + \hat{K})\hat{F}_{12} - \hat{F}_{12}(\hat{K} - \delta\hat{f})\hat{F}_{12} \tag{4.54}$$

The approach allows the double commutator term to be computed exactly as before. The second term on the right-hand side is computed by direct RI insertion

$$\left\langle wz \left| \hat{F}_{12}(\hat{K}_1 - \delta\hat{f}_1)\hat{F}_{12} \right| xy \right\rangle = F^{p'q'}_{wz}(\hat{K}_1 - \delta f_1)^{r'}_{p'}F^{xy}_{r'q'} \tag{4.55}$$

The contribution from $\hat{K}_2 - \delta\hat{f}_2$ is obtained by permutation of indices. Constructing the B matrix for ZAPT then proceeds in exactly the same way as for the closed-shell case, complicated only by replacement of the exchange matrix with the modified operator $\hat{K} + \delta\hat{f}$. For ZAPT, however, we must again note the inclusion of α,β exchange excitations. Because of the symmetric spin basis, matrix elements of the form

$$(K + \delta f)^{q_\beta}_{p_\alpha} \tag{4.56}$$

are non-zero. If the diagonal ansatz is chosen, these α,β exchange excitations are excluded from the geminal space.

The most computationally demanding step in approximation C is the double RI index integral transformation necessary in computing the B matrix. We therefore employ a hybrid approximation which avoids the double RI transformation, greatly reducing computational requirements. In particular, two terms in the B matrix require a double RI transform.

$$F^{p'q'}_{ij}K^{r'}_{p'}F^{kl}_{r'q'} \tag{4.57}$$

$$F^{mq''}_{ij}f^{p'}_{m}F^{kl}_{p'q''} \tag{4.58}$$

We therefore propose approximating each term by a single RI analog

$$F_{ij}^{p'q} K_{p'}^{r'} F_{r'q}^{kl} \tag{4.59}$$

$$F_{ij}^{mq''} f_m^p F_{pq''}^{kl} \tag{4.60}$$

For the exchange term, we are essentially expanding the identity operator in the orbital basis instead of the full RI basis. For the second term, we are neglecting Fock matrix terms of the form

$$f_i^{p''} \tag{4.61}$$

which is equivalent to the GBC approximation for closed shell molecules. For ROHF, as explained above, these terms are chosen as rigorously zero by the Hamiltonian partition. The RI approximation is very similar to the hybrid approaches proposed by Klopper[92] for Approximation B and Werner for Approximation C.[45,65]

4.6 Computational Methods

The ZAPT2-R12 method described herein was added to the MPQC package[61,93] for a Slater-type correlation factor with exponent 1.4 approximated as a sum of six Gaussian geminals. In the present work, all results are reported for approximation C so that the kinetic energy commutator integrals do not need to be computed. Unless otherwise stated, all R12 results are computed with fixed amplitudes, assuming the EBC condition and hybrid RI approximation. Integrals were implemented in the CINTS/LIBINT package[94] within MPQC according to the Obara-Saika scheme with the fundamentals modified accordingly.[90,91,95,96] All conventional ZAPT2 computations were performed with the Dunning aug-cc-pVnZ[80,97] family of basis sets (hereafter denoted aVnZ). The ZAPT2 basis set limit is estimated from an aV5Z-aV6Z extrapolation using an n^{-3} extrapolation.[30,32] All ZAPT2-R12 computations use the Peterson F12 and F12-RI basis sets,[98] which have been shown to be well-calibrated for R12 computations.[99] For all R12 computations, we shall abbreviate the basis sets as DZ, TZ, and QZ, being understood as the Peterson R12 basis sets. R12 complete basis limits are estimated according to the n^{-4} extrapolation suggested by Peterson. All molecules

in the current work were optimized at the B3LYP/aVTZ level, since the goal of the current work is not to establish rigorous basis set limit correlation energies, but only to assess the R12 basis set convergence properties.

4.7 Results and Discussion

In the following we do not report total energy differences, but rather only the contributions from the correlation energy.

$$\delta[\mathrm{ZAPT2}]_{A\to B} = E^{corr}_{\mathrm{ZAPT2}}(B) - E^{corr}_{\mathrm{ZAPT2}}(A) \tag{4.62}$$

With the R12 corrections, the basis set error can often be larger for the Hartree-Fock energy than the correlation energy. Schemes for perturbative corrections to the Hartree-Fock energy have therefore been proposed using the auxiliary basis.[65] In the present work, however, we focus only on the error inherent in the ZAPT2 correlation increment.

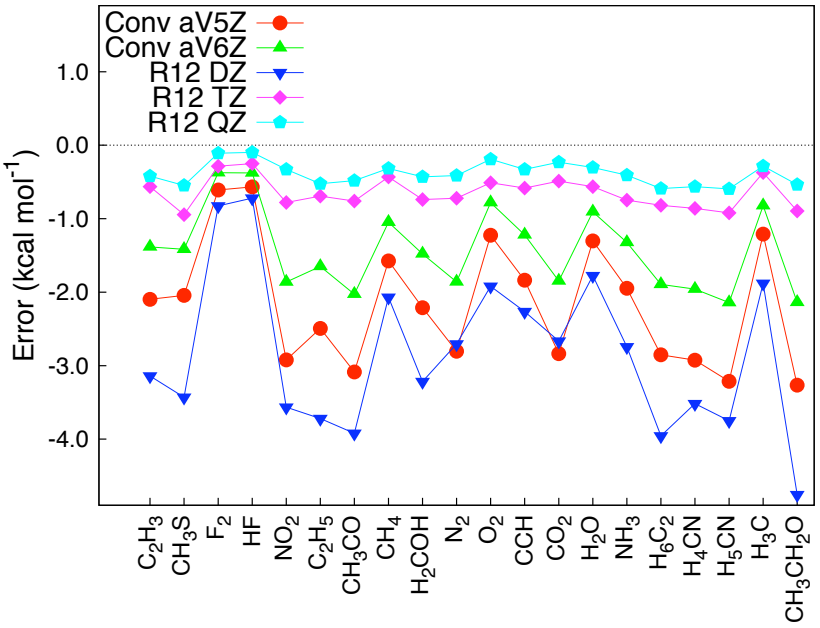

FIGURE 4.1: Error of atomization energy δ[ZAPT2] increments with respect to ZAPT2 aV5Z-aV6Z extrapolation for conventional ZAPT2 (Conv) and ZAPT2-R12 (R12) computations.

The error in the δ[ZAPT2] correlation energy increments for both conventional ZAPT2 and R12 computations are shown in Figure 4.1 for a set of 20 atomization energies. Even the R12 TZ

computations consistently outperform the conventional aV6Z result. The R12 DZ results, while somewhat worse than the ZAPT2 aV5Z results, are still remarkably accurate, especially considering the approximations involved: EBC, hybrid RI, and fixed amplitudes. In extending the basis set from TZ to QZ, the R12 results generally change by only 0.1 to 0.2 kcal mol^{-1}, suggesting the R12 correlation energies are essentially converged at the QZ level. In contrast, for conventional ZAPT2, the aV5Z to aV6Z δ[ZAPT2] increment usually changes by as much as 0.5 to 1.0 kcal mol^{-1}. The difference between R12 TZ and R12 QZ results is generally smaller than the difference between the R12 QZ and 5-6 extrapolation, suggesting the R12 QZ results may actually be closer to the basis set limit than the aV5Z-aV6Z extrapolation.

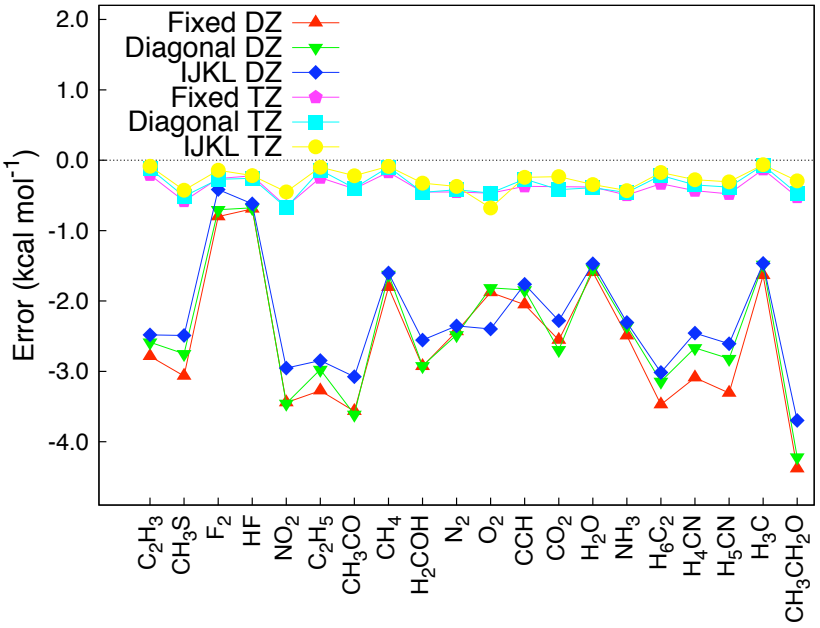

FIGURE 4.2: Error of atomization energy δ[ZAPT2] increments with respect to R12 T-Q extrapolation for R12 ansätze.

To validate our choice of R12 ansatz, the errors for the different amplitude ansätze are displayed in Figure 4.2, with the basis set limit now estimated by an R12 T-Q extrapolation. While the IJKL ansatz generally performs slightly better with the DZ basis, all results are essentially identical for the TZ basis. For the DZ basis, the IJKL results actually benefit from compensating errors in which geminal basis set superposition errors tends to offset the basis set incompleteness error. For

situations in which only a DZ computation is feasible, we suggest that errors associated with the use of fixed amplitudes are negligible compared with the intrinsic basis set error.

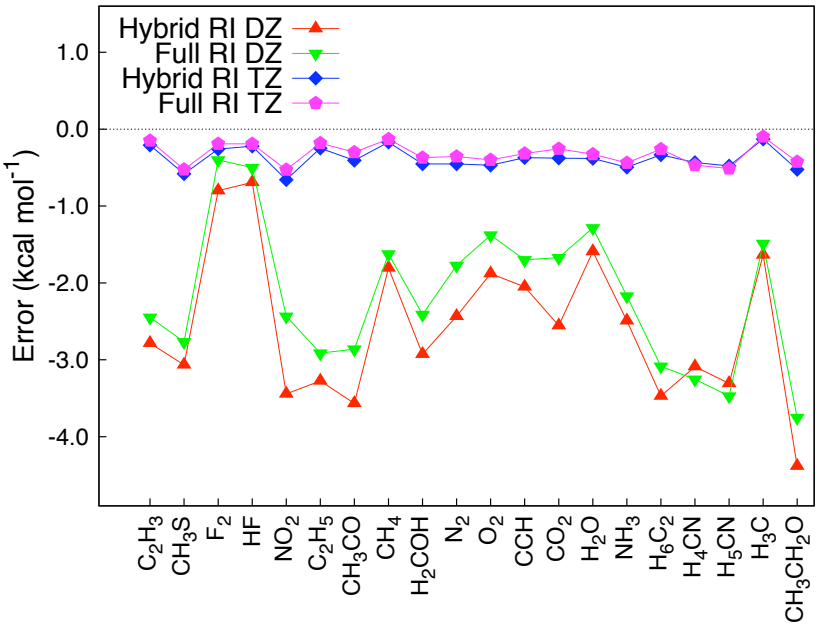

FIGURE 4.3: Error of atomization energy δ[ZAPT2] increments with respect to R12 T-Q extrapolation for RI approximations

As mentioned previously, the most expensive part of an R12 computation is the double RI index transformation necessary for the exchange matrix term of the B intermediate. The double RI can be replaced by a single RI approximation, drastically reducing memory and CPU requirements. The effect of the approximation is shown in Figure 4.3. For the TZ basis, the results are equivalent within 0.1 kcal mol^{-1}. For the DZ basis, the hybrid RI is clearly a more severe approximation. However, the difference from the full RI approach is usually less than 0.5 kcal mol^{-1}, and, again, for computations in which only a DZ computation is feasible, the hybrid RI error is probably negligible compared to the total error in the computation. To further assess the accuracy of the R12 T-Q limit versus the conventional aV5Z-aV6Z limit, we can examine the basis set convergence of the conventional and R12 increments. Figure 4.4 shows the conventional δ[ZAPT2] increments for the aVQZ, aV5Z, aV6Z progression. Even the conventional aV5Z δ[ZAPT2] increments are usually 2.0 to 3.0 kcal mol^{-1} from the CBS estimate while aV6Z increments are 1.0 to 2.0 kcal mol^{-1} from the CBS estimate. As a consequence, the aVQZ-aV5Z extrapolation usually differs from aV5Z-aV6Z

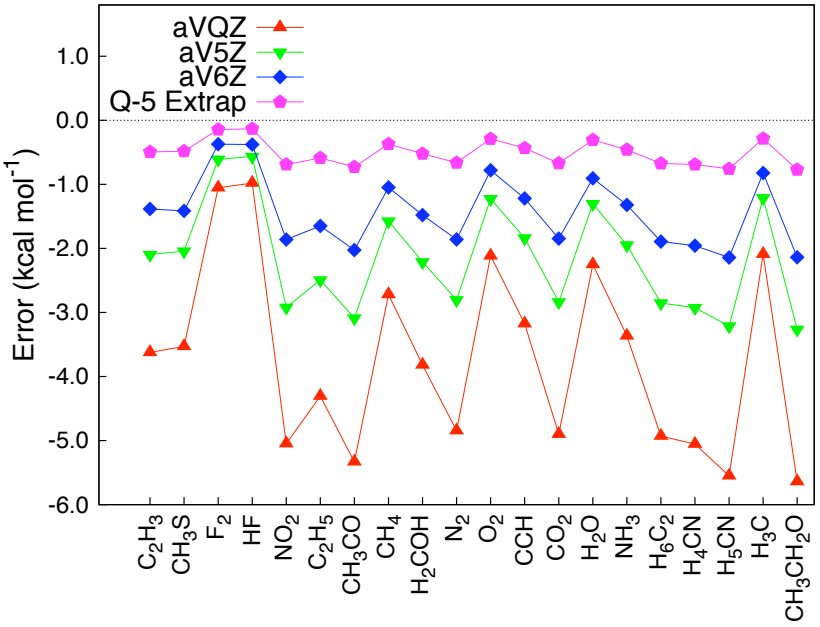

FIGURE 4.4: Error of atomization energy δ[ZAPT2] increments for conventional ZAPT2 with respect to aV5Z-aV6Z extrapolation.

result by as much as 0.5 kcal mol^{-1}, demonstrating a significant indeterminacy in the exact CBS limit.

In contrast, the R12 QZ results differ from the R12 T-Q limit by less than 0.2 kcal mol^{-1} (Figure 4.5), suggesting the R12 T-Q CBS estimate is essentially converged. The D-T extrapolation improves little on the accuracy of a single R12 TZ computation given the relatively large error in the R12 DZ result, and there seems to be little benefit in performing a D-T extrapolation.

4.8 Conclusions

An explicitly-correlated second-order perturbation theory for high-spin open-shell states has been presented based on a symmetric (Z-averaged) exchange operator. Excellent basis set convergence for atomization energies is observed with the recently developed R12 basis sets of Peterson. This convergence is competitive with even an aV5Z-aV6Z extrapolation. Through the use of symmetric

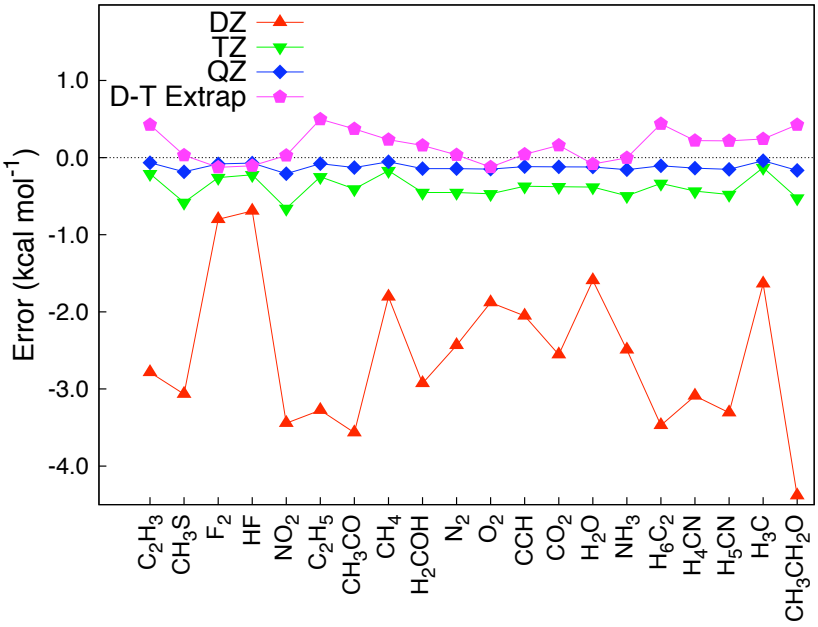

FIGURE 4.5: Error of atomization energy δ[ZAPT2] increments with respect to R12 T-Q extrapolation for ZAPT2-R12 computations.

spin-orbitals in the singly-occupied space, equivalence of α and β subspaces is maintained, greatly reducing the computational cost. Since computational cost at the level of second-order perturbation theory is determined almost exclusively by the atomic orbital (AO) to molecular orbital (MO) integral transformation, ZAPT represents a significant computational savings over approaches based on semi-canonical orbitals. For restricted Møller Plesset (RMP) perturbation theories, even if canonical ROHF orbitals are used so that only one set of integrals is required, different amplitudes are still required for each spin case. In this regard, ZAPT reintroduces amplitude symmetry relations between the different spin cases, allowing a spin-restricted formalism which will be similar in cost to closed-shell computations. ZAPT should therefore be useful in greatly reducing the cost of explicitly-correlated, high-spin open-shell computations.

BIBLIOGRAPHY

[1] A. Tajti, P. G. Szalay, A. G. Csàszàr, M. Kallay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vazquez, J. F. Stanton, J. Chem. Phys. **121**, 11599 (2004).

[2] A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kallay, J. Gauss, J. Chem. Phys. **120**, 4129 (2004).

[3] A. Karton, E. Rabinovich, J. M. L. Martin, B. Ruscic, J. Chem. Phys. **125**, 144108 (2006).

[4] K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[5] R. Krishnan, M. J. Frisch, J. A. Pople, J. Chem. Phys. **72**, 4244 (1980).

[6] J. Šponer, P. Jurečka, P. Hobza, J. Am. Chem. Soc. **126**, 10142 (2004).

[7] P. Jurečka, P. Hobza, Chem. Phys. Lett. **365**, 89 (2002).

[8] W. Kutzelnigg, Theor. Chim. Acta **68**, 445 (1985).

[9] W. Kutzelnigg, W. Klopper, J. Chem. Phys. **94**, 1985 (1991).

[10] W. Klopper, W. Kutzelnigg, Chem. Phys. Lett. **134**, 17 (1987).

[11] G. Rauhut, G. Knizia, H.-J. Werner, J. Chem. Phys. **130**, 054105 (2009).

[12] G. Knizia, T. B. Adler, H.-J. Werner, J. Chem. Phys. **130**, 054104 (2009).

[13] T. D. Crawford, H. F. Schaefer, T. J. Lee, J. Chem. Phys. **105**, 1060 (1996).

[14] C. W. Murray, E. R. Davidson, Int. J. Quantum Chem. **43**, 755 (1992).

[15] C. W. Murray, N. C. Handy, J. Chem. Phys. **97**, 6509 (1992).

[16] C. W. Murray, E. R. Davidson, Chem. Phys. Lett. **187**, 451 (1991).

[17] P. M. Kozlowski, E. R. Davidson, Chem. Phys. Lett. **226**, 440 (1994).

[18] R. D. Amos, J. S. Andrews, N. C. Handy, P. J. Knowles, Chem. Phys. Lett. **185**, 256 (1991).

[19] W. J. Lauderdale, J. F. Stanton, J. Gauss, J. D. Watts, R. J. Bartlett, Chem. Phys. Lett. **187**, 21 (1991).

[20] P. J. Knowles, J. S. Andrews, R. D. Amos, N. C. Handy, J. A. Pople, Chem. Phys. Lett. **186**, 130 (1991).

[21] D. Jayatilaka, T. J. Lee, Chem. Phys. Lett. **199**, 211 (1992).

[22] T. J. Lee, D. Jayatilaka, Chem. Phys. Lett. **201**, 1 (1993).

[23] D. Jayatilaka, T. J. Lee, J. Chem. Phys. **98**, 9734 (1993).

[24] T. J. Lee, A. P. L. Rendell, K. G. Dyall, D. Jayatilaka, J. Chem. Phys. **100**, 7400 (1994).

[25] P. M. W. Gill, J. A. Pople, L. Radom, R. H. Nobes, J. Chem. Phys. **89**, 7307 (1988).

[26] S. E. Wheeler, W. D. Allen, H. F. Schaefer, J. Chem. Phys. **128**, 074107 (2008).

[27] T. D. Crawford, T. J. Lee, H. F. Schaefer, J. Chem. Phys. **107**, 7943 (1997).

[28] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).

[29] D. Feller, J. Chem. Phys. **96**, 6104 (1992).

[30] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).

[31] T. H. Dunning, J. Phys. Chem. A **104**, 9062 (2000).

[32] T. Helgaker, W. Klopper, H. Koch, J. Noga, J. Chem. Phys. **106**, 9639 (1997).

[33] W. Klopper, F. R. Manby, S. Ten-No, E. F. Valeev, Int. Rev. Phys. Chem. **25**, 427 (2006).

[34] T. Kato, Commun. Pure Appl Math. **10**, 151 (1957).

[35] D. P. Tew, J. Chem. Phys. **129**, 014104 (2008).

[36] E. A. Hylleraas, Z Phys. **48**, 469 (1929).

[37] S. F. Boys, N. C. Handy, Proc. R. Soc. Lon Ser A **311**, 309 (1969).

[38] J. O. Hirschfelder, J. Chem. Phys. **39**, 3145 (1963).

[39] S. Ten-No, Chem. Phys. Lett. **330**, 169 (2000).

[40] N. Umezawa, S. Tsuneyuki, J. Chem. Phys. **119**, 10015 (2003).

[41] O. Hino, Y. Tanimura, S. Ten-No, Chem. Phys. Lett. **353**, 317 (2002).

[42] O. Hino, Y. Tanimura, S. Ten-No, J. Chem. Phys. **115**, 7865 (2001).

[43] S. Ten-No, Chem. Phys. Lett. **330**, 175 (2000).

[44] W. Klopper, C. C. M. Samson, J. Chem. Phys. **116**, 6397 (2002).

[45] H.-J. Werner, T. B. Adler, F. R. Manby, J. Chem. Phys. **126**, 164102 (2007).

[46] W. Klopper, Chem. Phys. Lett. **186**, 583 (1991).

[47] V. Termath, W. Klopper, W. Kutzelnigg, J. Chem. Phys. **94**, 2002 (1991).

[48] W. Klopper, W. Kutzelnigg, J. Chem. Phys. **94**, 2020 (1991).

[49] W. Klopper, W. Kutzelnigg, J. Phys. Chem. A **94**, 5625 (1990).

[50] E. F. Valeev, J. Chem. Phys. **125**, 244106 (2006).

[51] R. Polly, H.-J. Werner, P. Dahle, P. R. Taylor, J. Chem. Phys. **124**, 234107 (2006).

[52] R. Bukowski, B. Jeziorski, K. Szalewicz, J. Chem. Phys. **110**, 4165 (1999).

[53] B. J. Persson, P. R. Taylor, J. Chem. Phys. **105**, 5915 (1996).

[54] R. Bukowski, B. Jeziorski, K. Szalewicz, J. Chem. Phys. **100**, 1366 (1994).

[55] E. F. Valeev, Chem. Phys. Lett. **395**, 190 (2004).

[56] T. B. Adler, H.-J. Werner, F. R. Manby, J. Chem. Phys. **130**, 054106 (2009).

[57] H.-J. Werner, J. Chem. Phys. **129**, 101103 (2008).

[58] F. R. Manby, H.-J. Werner, T. B. Adler, A. J. May, J. Chem. Phys. **124**, 094103 (2006).

[59] F. R. Manby, J. Chem. Phys. **119**, 4607 (2003).

[60] S. Ten-No, F. R. Manby, J. Chem. Phys. **119**, 5358 (2003).

[61] E. F. Valeev, C. L. Janssen, J. Chem. Phys. **121**, 1214 (2004).

[62] H. Fliegl, W. Klopper, C. Hättig, J. Chem. Phys. **122**, 084107 (2005).

[63] H. Fliegl, C. Hättig, W. Klopper, J. Chem. Phys. **124**, 044112 (2006).

[64] C. Neiss, C. Hättig, W. Klopper, J. Chem. Phys. **125**, 064111 (2006).

[65] G. Knizia, H.-J. Werner, J. Chem. Phys. **128**, 154103 (2008).

[66] T. Shiozaki, M. Kamiya, S. Hirata, E. F. Valeev, Phys. Chem. Chem. Phys. **10**, 3358 (2008).

[67] T. Shiozaki, M. Kamiya, S. Hirata, E. F. Valeev, J. Chem. Phys. **129**, 071101 (2008).

[68] D. Bokhan, S. Ten-No, J. Noga, Phys. Chem. Chem. Phys. **10**, 3320 (2008).

[69] M. Torheyden, E. F. Valeev, Phys. Chem. Chem. Phys. **10**, 3410 (2008).

[70] D. P. Tew, W. Klopper, C. Hättig, Chem. Phys. Lett. **452**, 326 (2008).

[71] E. F. Valeev, Phys. Chem. Chem. Phys. **10**, 106 (2008).

[72] A. Köhn, J. Chem. Phys. **130**, 201103 (2009).

[73] D. Bokhan, S. Bernadotte, S. Ten-No, Chem. Phys. Lett. **469**, 214 (2009).

[74] A. Köhn, J. Chem. Phys. **130**, 131101 (2009).

[75] T. Shiozaki, M. Kamiya, S. Hirata, E. F. Valeev, J. Chem. Phys. **130**, 054101 (2009).

[76] R. J. Gdanitz, Chem. Phys. Lett. **210**, 253 (1993).

[77] R. J. Gdanitz, R. Röhse, Int. J. Quantum Chem. **55**, 147 (1995).

[78] W. Klopper, R. Röhse, W. Kutzelnigg, Chem. Phys. Lett. **178**, 455 (1991).

[79] S. Ten-No, Chem. Phys. Lett. **447**, 175 (2007).

[80] R. A. Kendall, T. H. Dunning, R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

[81] D. P. Tew, W. Klopper, C. Neiss, C. Hättig, Phys. Chem. Chem. Phys. **9**, 1921 (2007).

[82] D. P. Tew, W. Klopper, J. Chem. Phys. **123**, 074101 (2005).

[83] S. Ten-No, Chem. Phys. Lett. **398**, 56 (2004).

[84] J. M. Gonzales, C. Pak, R. S. Cox, W. D. Allen, H. F. Schaefer, A. G. Csàszàr, G. Tarczay, Chem. Eur. J. **9**, 2173 (2003).

[85] E. F. Valeev, Chem. Phys. Lett. **418**, 333 (2006).

[86] A. J. May, E. F. Valeev, R. Polly, F. R. Manby, Phys. Chem. Chem. Phys. **7**, 2710 (2005).

[87] D. P. Tew, W. Klopper, J. Chem. Phys. **125**, 094302 (2006).

[88] S. Ten-No, J. Chem. Phys. **121**, 117 (2004).

[89] S. Kedžuch, M. Milko, J. Noga, Int. J. Quantum Chem. **105**, 929 (2005).

[90] S. Obara, A. Saika, J. Chem. Phys. **89**, 1540 (1988).

[91] S. Obara, A. Saika, J. Chem. Phys. **84**, 3963 (1986).

[92] W. Klopper, J. Chem. Phys. **120**, 10890 (2004).

[93] C. L. Janssen, I. M. B. Nielsen, M. L. Leininger, E. F. Valeev, J. P. Kenny, E. T. Seidl. The Massively Parallel Quantum Chemistry Program (MPQC), version 3.0. 2008. Sandia National Laboratories, Livermore, CA, USA. http://www.mpqc.org.

[94] E. F. Valeev, J. T. Fermann. Libint 1.0. 2007. Virginia Tech Department of Chemisry, Blacksburg, VA, 24061. http://www.files.chem.vt.edu/chem-dept/valeev/software/libint/libint.html.

[95] R. Ahlrichs, Phys. Chem. Chem. Phys. **8**, 3072 (2006).

[96] S. Höfener, F. A. Bischoff, A. Gloess, W. Klopper, Phys. Chem. Chem. Phys. **10**, 3390 (2008).

[97] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

[98] K. A. Peterson, T. B. Adler, H.-J. Werner, J. Chem. Phys. **128**, 084102 (2008).

[99] F. A. Bischoff, S. Wolfsegger, D. P. Tew, W. Klopper, Mol. Phys. **107**, 1 (2009).

CHAPTER 5

SUMMARY AND CONCLUSIONS

5.1 Concluding Remarks

Even the Schrödinger equation for the non-relativistic, Born-Oppenheimer Hamiltonian presents daunting mathematical challenges if highly-accurate solutions are desired. In particular, it is the exponential complexity of the electron correlation which rapidly becomes intractable for large systems. In this regard, we have discussed a series of approximations and extrapolations which can be feasibly applied to combustion intermediates containing more than ten atoms. In particular, the coupled-cluster ansatz provides a truncated model of the electron correlation which can be systematically improved. This model is exploited in the focal point approach (FPA), extrapolating to the so-called complete basis set and correlation limits at which the wavefunction becomes mathematically an exact eigenfunction of the Hamiltonian. The focal point method has been applied here to two separate systems: conformers of gaseous cysteine and combustion intermediates of ethyl radical. For cysteine, the realm of subchemical accuracy better than 1.0 kcal mol^{-1} is already achieved with triple excitations and orbital basis sets including g-functions. In contrast, for ethyl radical, a much more extensive treatment is required through quadruple excitations and orbital basis sets with at least h-functions.

Much more compact representations of the electron correlation can be obtained with wavefunctions directly incorporating the interelectronic distance r_{12}. Although new challenges related to the computation of Hamiltonian matrix elements arise, we have demonstrated that the R12 methods provide a set of efficient but still highly accurate approximations. In particular, we have derived an R12 formalism appropriate for open-shell radicals in combustion via the so-called Z-averaged or symmetric exchange approach. The Z-averaged R12 methods have been tested at the level of both second-order perturbation theory and coupled-cluster with singles and doubles, demonstrating the validity of the approximations. In this regard, the R12 terms essentially only treat two-body effects, almost exclusively in modeling the Coulomb hole near the coalescence point of two electrons. In this regard, a foundation has been built for extending these open-shell R12 methods to even more accurate wavefunction models, potentially providing compact mathematical representations of both the connected three-body and four-body electron correlation effects.