Theoretical Studies on the Periodic Trends of Small Systems

by

Boyi Zhang

(Under the Direction of Henry F. Schaefer III)

Abstract

Several projects involving the periodic trends of various small systems were studied using rigorous computational methods. The first project involves comparing the spin-orbit splittings of p-block elements B–I as computed by various four-component methods. Recommendations are given with regards to the best basis set and method to use for computing spin-orbit splittings. In the second project, ground state singlet and triplet states of cyanomethylene derivatives $HBCN^-$, HCCN, $HNCN^+$, $HAlCN^-$, HSiCN, $HPCN^+$, $HGaCN^-$, HGeCN, and $HAsCN^+$ were studied. Singlet-triplet gaps for these species were computed using a focal point analysis approach and insight is given for the observed trends. In the third project, the isomerization of HXYH to H_2XY (X, Y = O, S, Se, Te) was explored. Various means of modelling the relativistic effects of the system are compared. Periodic trends in geometries and isomerization barriers were established and analyzed. Finally, the applicability of quantum mechanical methods as applied to larger system sizes is investigated through the implementation of an open-source adaptive QM/MM software.

Index words: Electronic Structure Theory, Coupled Cluster Theory, Focal Point Analysis, Spin-Orbit Splitting, Cyanomethylene, Oxywater, Adaptive QM/MM

Theoretical Studies on the Periodic Trends of Small Systems

by

Boyi Zhang

B.S., University of Richmond, 2016

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

Doctor of of Philosophy

Athens, Georgia

2021

Theoretical Studies on the Periodic Trends of Small Systems

by

Boyi Zhang

Major Professor: Henry F. Schaefer

Committee: Gary E. Douberly
Steven E. Wheeler

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
May 2021

Acknowledgments

There are many to whom I am grateful:

First and foremost, I would like to thank my husband Adam Abbott for his constant support. It has been a joy to go through graduate school with him by my side.

My advisor, Dr. Henry Schaefer, has taught me invaluable lessons during my time at UGA. I thank him for his guidance and wisdom.

My committee members, Dr. Steven Wheeler and Dr. Gary Douberly, for all the feedback and advice I have received on my research. They have taught me a lot about what it means to be a good scientist, and I am so grateful for their time.

My undergraduate advisor Dr. Carol Parish for introducing me to computational chemistry and believing in me. Her encouragement through the years has meant the world.

All members of the CCQC for their friendship and stimulating scientific discussions.

My family for all the love you have given me.

"The wonder of the world is not its complexity but its simplicity. Given enough color and canvas, anybody can make a mess; that, we do ourselves. More to admire is the artist who makes do with little, the artist whose art is to conceal an economy of form and design. That, nature does unsuspected - in a world hidden from the senses... There, I think, is the surprise and irony of the universe and also its incomparable beauty."

- Michael Munowitz, Principles of Chemistry

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Schrödinger Equation

The experiments of the early twentieth century drastically changed our perception of light and matter. Light was no longer just a wave. Matter was no longer just made of particles. And electrons, whose interaction is so fundamental to the microscopic world, fell into the realm of something not completely defined, not completely understood. It was both a particle and a wave. It was neither a particle nor a wave. As a consequence of this duality, we could no longer talk about the path of the electron with certainity. Instead, we assign a probablity amplitude with varies with position and time. We call this probability amplitude a "wavefunction". This wavefunction, which we denote Ψ, contains all knowable information about the system, and when acted upon by an operator, will return the observable properties of the system.

Our primary interest in quantum mechanics is to solve for this wavefunction. We do it through the Schrödinger equation. We start with the time-depedent Schrödinger equation:

$$\hat{H}\Psi(\mathbf{x},t) = i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{x},t)$$

With some assumptions and mathematical manipulations, we can derive a time-independent Schrödinger equation:

$$\hat{H}\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

For a molecular system $\hat{H}$ is:

$$\hat{H} = -\sum_{i}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i}^{N}\sum_{A}^{M}\frac{Z_A}{|\mathbf{r_i}-\mathbf{R_A}|} + \sum_{i}^{N}\sum_{j>i}^{N}\frac{1}{|\mathbf{r_i}-\mathbf{r_j}|} + \sum_{A}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{|\mathbf{R_A}-\mathbf{R_B}|}$$

The first two terms are the kinetic energy terms for the electrons and nuclei, respectively, while the last three terms are the electron–nucleus, electron–electron, and nucleus–nucleus pairwise potential energy interactions.

The attraction and repulsion terms in the Hamiltonian means that the motions of electrons and the nuclei are not independent, but correlated. This prevents the operator from being separable, and makes solving the Schrödinger equation considerably more difficult. To simplify matters, we use the idea that the nucleus moves much slower than the electrons, and invoke the Born-Oppenheimer approximation. We "fix" the nucleus and solve for the movement of the electrons only. The kinetic energy term for the nucleus becomes zero, and our electron-nuclear potential energy terms now become a constant which is added onto the final energy.

In this way, we obtain our clamped-nuclei electronic Hamiltonian:

$$\hat{H}_e = -\sum_i^N \frac{1}{2}\nabla_i^2 - \sum_i^N \sum_A^M \frac{Z_A}{|\mathbf{r_i} - \mathbf{R_A}|} + \sum_i^N \sum_{j>i}^N \frac{1}{|\mathbf{r_i} - \mathbf{r_j}|}$$

We can also express the Hamiltonian as a sum of the one-electron $[\hat{h}(i)]$ and two-electron operators $[\hat{g}(i,j)]$, where

$$\hat{h}(i) = -\frac{1}{2}\nabla_i^2 - \sum_A^M \frac{Z_A}{|\mathbf{r_i} - \mathbf{R_A}|}$$

$$\hat{g}(i,j) = \frac{1}{|\mathbf{r_i} - \mathbf{r_j}|}$$

and

$$\hat{H}_e = \sum_i^N \hat{h}(i) + \sum_{i<j}^N \hat{g}(i,j)$$

We can use the electronic Schrödinger equation to solve for the electronic wavefunction:

$$\hat{H}_e \Psi_e = E_e \Psi_e$$

Ψ_{elec} describes the motion of electrons, explicitly depend on electronic coordinates but parametrically on nuclear coordinates. The Schrödinger equation can be alternatively represented using a state vector notation:

$$\langle \Psi | H | \Psi \rangle = E$$

Once we have solved for the electronic wavefunction, we can then solve for the positions of the nuclei in the average field of the electrons. The electronic Schrödinger equation is the primary equation used in electronic structure theory. The electron-electron repulsion term in the electronic Hamiltonian prevents the electronic Schrödinger equation from being solvable for systems with more than one electron. The primary goal of the most common methods in electronic structure theory is thus, for a given system, to define a wavefunction which is an accurate approximation to the unknownable, "true" wavefunction Ψ.

1.2 Slater Determinants

The task, then, is to find an appropriate form of the wavefunction. A one-electron wavefunction, ψ_i, also known as a molecular orbital, obeys the one-electron Hamiltonian:

$$\hat{h}(i)\psi_i = \epsilon_i \psi_i$$

This wavefunction is made up of a spatial function ϕ_i and a spin function ω_i:

$$\psi_i = \phi_i \omega_i$$

An N-electron wavefunction Φ is some product of N one-electron wavefunctions. Specifically, due to the Pauli exclusion principle, it is the antisymmetric product of N one-electron wavefunctions. We can express the antisymmetric product as a Slater determinant:

$$\Phi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1^1 & \psi_2^1 & \psi_3^1 & \dots & \psi_N^1 \\ \psi_1^2 & \psi_2^2 & \psi_3^2 & \dots & \psi_N^2 \\ \psi_1^3 & \psi_2^3 & \psi_3^3 & \dots & \psi_N^N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_1^N & \psi_2^N & \psi_3^N & \dots & \psi_N^N \end{vmatrix}$$

Using the state vector notation, the Schrödinger equation using the wavefunction Φ is

$$\langle \Phi | H | \Phi \rangle = E$$

Given the form of Φ as an antisymmetric product, and using the separability rules for integrals, the energy expression can be written as:

$$E = \sum_i^N \langle \psi_i^i | \hat{h}(i) | \psi_i^i \rangle + \sum_{i<j}^N \langle \psi_i^i \psi_j^j | \hat{g}(i,j) | \psi_i^i \psi_j^j \rangle - \langle \psi_i^i \psi_j^j | \hat{g}(i,j) | \psi_j^i \psi_i^j \rangle$$

This is known as the first Slater-Condon rule.

We now have an energy expression given a set of orbitals functions. If we choose some random functions, we will get an energy. But how will we know if it is the correct energy? Given an arbitrary wavefunction Φ, the Variational Principle necessitates that the energy given by that wavefunction will always be greater or equal to the energy of the true ground state E_0:

$$\frac{\int \Phi H \Phi d\mathbf{r}}{\int \Phi^2 d\mathbf{r}} \geq E_0$$

Although we do not know the true form of the wavefunction, we do know that any approximate wavefunction will never give us an energy that is lower than the true energy. This gives us a criteria by which to optimize a guess wavefunction: the lower the energy given by that wavefunction, the better it is. Thus, we can minimize the energy given by the first Slater-Condon rule through changing the parameters that define our orbitals.

1.3 Hartree-Fock Theory

The Hartree-Fock (HF) equations give us a way to optimize the orbitals such that we have the best Slater determinant wavefunction to describe a chosen system. The key constraint that gives rise to the Hartree-Fock equations is that the orbitals are kept orthonormal. If we assume a system with $N = 2n$ electrons in n doubly-occupied molecular orbitals, we can re-express our energy expression for this special, closed-shell case.

$$E = 2\sum_{i}^{N/2} \langle \phi_i | \hat{h}_i | \phi_i \rangle + \sum_{i}^{N/2}\sum_{j}^{N/2} 2\,\langle \phi_i \phi_j | \hat{g}(i,j) | \phi_i \phi_j \rangle - \langle \phi_i \phi_j | \hat{g}(i,j) | \phi_j \phi_i \rangle$$

Here, we have integrated out our spin-orbitals to just leave spatial orbitals. This is the restricted HF formalism. An unrestricted HF formalism uses the first Slater-Condon rule as is. We can define coulomb and exhange operators $\hat{J}$ and $\hat{K}$ in terms of how they act on a spatial orbital:

$$\hat{J}_i \, |\phi^\mu\rangle = \langle \phi_i^\nu | \hat{g}(\mu,\nu) | \phi_i^\nu \phi^\mu \rangle$$

$$\hat{K}_i \, |\phi^\mu\rangle = \langle \phi_i^\nu | \hat{g}(\mu,\nu) | \phi_\nu \phi_i^\mu \rangle$$

and substitute them into our energy expression to obtain:

$$E = 2\sum_{i}^{N/2} \langle \phi_i | \hat{h}_i | \phi_i \rangle + \sum_{i}^{N/2}\sum_{j}^{N/2} 2\,\langle \phi_i | \hat{J}_j | \phi_i \rangle - \langle \phi_i | \hat{K}_j | \phi_i \rangle$$

Notice that now, we have gotten rid of the two electron operators in the equation, and every operator is a one-electron operator. This the the basis for the Hartree-Fock (HF) equations, which gives that the best set of molecular orbitals for a given system is given by:

$$\left[\hat{h} + \sum_j (2\hat{J}_j - \hat{K}_j) \right] \phi_i = \varepsilon_i \phi_i$$

The operator on the lefthand side is also known as the Fock operator $\hat{f}$. The Fock operator is an effective one-electron operator. It describes an effective one-body interaction that comes from a two-body operator. In other words, it does not describe the interaction of any two individual electrons, but rather the "mean-field" effect of all the electrons on a particular electron. This is the key assumption and shortcoming

of Hartree-Fock theory. To solve this equation, one goes through an iterative process of picking some set of ϕ_i, building the Fock operator, and solving for ε_i until the orbitals converge. The process is referred to as a self-consistent field (SCF) procedure.

For molecular systems, solving the Hartree-Fock equations is impossibly expensive and we have to apply an approximation to the "true" molecular orbitals. Namely, we represent the molecular orbitals as a linear combination of atomic orbitals (LCAO). The atomic orbital functions (χ) form a basis for representing the approximated molecular orbitals. Our approximated MO can be mathematically represented as a linear combination of the atomic orbital functions multiplied by a coefficient C:

$$\phi_i = \sum_q \chi_q C_{qi}$$

If we substitute this definition of ϕ_i into our Hartree-Fock equation and reformulate the expression in matrix notation, we get the expression that is used in any practical implementation of Hartree-Fock theory:

$$\mathbf{F}_{pq}\mathbf{C}_{qi} = \mathbf{S}_{pq}\mathbf{C}_{qi}\varepsilon_i$$

where $\mathbf{F}$ is the Fock matrix, $\mathbf{S}$ is the overlap matrix, and $\mathbf{C}$ is the coefficient matrix. This is known as the Roothaan-Hall equation.

1.4 Correlated Methods

The Hartree-Fock approximation is a good approximation. By definition, it leaves out electron correlation and instead treats electron-electron interactions in a mean-field manner. The true energy is thus

$$E_{true} = E_{HF} + E_{correlation}$$

On average, the Hartree-Fock energy makes up 95% of the true electronic energy. Various post-HF theories attempt to recover the correlation energy. There are two types of electron correlation: dynamical and static. We will first discuss dynamical correlation, which refers to the correlated motion arising from electron-electron interactions.

Configuration Interaction

For a given set of spin-orbitals, the exact N-electron wavefunction Ψ can be expressed as a linear combination of all possible Slater determinants that can be formed from that set of spin-orbitals:

$$|\Psi\rangle = \sum_{p_1<\cdots<p_N} c_p \, |\Phi_{p_1\cdots p_N}\rangle$$

We can choose a "reference" determinant, where the occupied spin-orbitals are denoted by indices $i, j, k \ldots$ and the unoccupied spin-orbitals are denoted by indices $a, b, c \ldots$. The set of all possible Slater determinants in this case would be given by every possible excitation of electrons from the occupied orbitals of the reference determinant to the unoccupied orbitals:

$$|\Psi\rangle = c_0 |\Phi\rangle + \sum_{ia} c_a^i |\Phi_i^a\rangle + \sum_{\substack{i<j \\ a<b}} c_{ab}^{ij} |\Phi_{ij}^{ab}\rangle + \sum_{\substack{i<j<k \\ a<b<c}} c_{abc}^{ijk} |\Phi_{ijk}^{abc}\rangle + \cdots$$

In the above equation, Φ_i^a would indicate a Slater determinant in which the electron from orbital i was excited to orbital a, and c_a^i would be the contribution of this Slater determinant to the exact wavefunction Ψ. We can define a set of excitation operators, which contains the coefficient c as well as some means of showing the excitation of electrons. This is best done using the second quantization formalism, although we will not go through a detailed derivation or discussion of second quantization presently. We use $\tilde{a}_i^a$ to represent the excitation from orbital i to a and so on, and also represent the $(n!)^2$ redundancy of summing over unrestricted orbitals:

$$|\Psi\rangle = c_0 |\Phi\rangle + \left(\frac{1}{1!}\right)^2 \sum_{ia} c_i^a \tilde{a}_i^a |\Phi\rangle + \left(\frac{1}{2!}\right)^2 \sum_{\substack{ij \\ ab}} c_{ij}^{ab} \tilde{a}_{ij}^{ab} |\Phi\rangle + \left(\frac{1}{3!}\right)^2 \sum_{\substack{ijk \\ abc}} c_{ijk}^{abc} |\Phi_{ijk}^{abc}\rangle + \cdots$$

We can define excitation operators C_N:

$$C_1 = \left(\frac{1}{1!}\right)^2 c_a^i \tilde{a}_i^a \qquad C_2 = \left(\frac{1}{2!}\right)^2 c_{ab}^{ij} \tilde{a}_{ij}^{ab} \qquad C_3 = \left(\frac{1}{3!}\right)^2 c_{abc}^{ijk} \tilde{a}_{ijk}^{abc} \qquad \cdots$$

where the subscript represents the type of excitation. For example, C_1 is a single excitation operator and C_2 is a double excitation operator. We now rewrite our equation as:

$$|\Psi\rangle = (c_0 + C_1 + C_2 + C_3 + \cdots + C_N) |\Phi\rangle$$

This is the final form of the wavefunction for the full configuration interaction (FCI) method. The reference determinant Φ used is the determinant given by Hartree-Fock. For a system with more than a few electrons, FCI is very computationally expensive. In practice, a truncated version is used, where we cut off the excitations that we consider. For example, CID only considers double excitations. CISD only considers singles and doubles excitaions.

One consequence of truncating terms in the full CI expansion is that we lose the property of size consistency: the energy of non-interacting systems A and B calculated together will not equal adding the energies of A and B calculated separately. Various corrections are added to CI to account for the loss of size consistency. The most popular is the Davidson correction.

Coupled Cluster

In coupled cluster (CC) theory, we employ a similar excitation operator to the one seen in CI. We will use T_N instead of C_N. Instead of a linear expansion of excitation operators, coupled cluster takes on an exponential form:

$$|\Psi_{CC}\rangle = exp\left(T_1 + T_2 + T_3 + \cdots + T_N\right)|\Phi\rangle$$

or more simply,

$$|\Psi_{CC}\rangle = exp\left(T\right)|\Phi\rangle$$

Unlike CI, coupled cluster is size consistent when truncated due to its exponential form. In addition to straightforward trucations such as CCSD, which include singles and doubles excitations, commonly used truncated coupled cluster methods also partially include higher excitation terms. For example, CCSD(T) adds selected terms from the triples excitation. We point out here that untruncated CI and CC will give us the exact answer for a set of chosen spin-orbitals. That is, it recovers 100% of the correlation energy. However, is not feasible for most systems of practical interest.

Basis Sets

As a side note, we refer to the set of spin-orbitals as a basis set. A spin-orbital is made of a linear combination of atomic orbitals. Atomic orbitals behave most like Slater-type functions, which has the form e^{-r}. However, this functional form is computationally difficult. Instead, we approximate the form of a Slater-type orbital using Gaussian functions of the form e^{-r^2}. Thus, basis sets vary in both the number of gaussian functions used to approximate atomic orbitals, as well as the number of atomic orbitals it includes in the formation of molecular spin-orbitals

Multireference Methods

Truncated configuration interaction and coupled cluster methods remedy dynamical correlation, and rely on the assumption that the Slater determinant given by the HF solution is a good first approximation to the system. The contribution of the reference determinant in the wavefunction expansion is very large compared to the contribution of the excited determinants. Static correlation is essentially the idea that a single reference determinant is not enough to describe the electronic structure of a system. For example, in the case of degenerate orbitals, electrons may equally choose to occupy one or the other. The two Slater determinants that represent the electrons in the two degenerate orbitals have an equally large contribution to the true wavefunction. In this case, the Hartree-Fock approach fails.

Multiconfigurational self-consistent field (MCSCF) is similar to Hartree-Fock in its purpose, which is to give a set of optimized orbitals for the system. However, multiple configurations are taken into account instead of just one single determinant. In addition to optimizing orbitals, the weights of each configuration is also optimized. In MCSCF, we limit the number of initial configurations considered. For example,

one approach is to choose a complete active space (CAS), and only consider the configurations that arise from the electrons in the CAS. Post-MCSCF methods like MRCI or MRCC use the Slater determinants as optimized in MCSCF and consider excitations out of those determinants

1.5 Focal Point Analysis

Computational accuracy for non-relativistic methods have two axes on which to improve. One is the type of theory used. We have already discussed that for a given system, FCI will give the exact wavefunction for the set of spin-orbitals included. The more excitations that are included in truncated methods like CI or CC, the closer the energy is to the true energy. We make the assumption that for most systems, the contribution of high-order excitations become increasingly small. For example, CCSDTQP will have a much smaller energy contribution than CCSDTQ, which in turn has a much smaller energy contribution than CCSDT. Due to this, we can include an approximation to these higher-order terms by adding a correction, which is computed at a much smaller basis set.

The second axis is the actual number of spin-orbitals included, in other words, the quality of the basis set used. Some basis sets, such as the Dunning cc-pVXZ (X = D, T, Q, 5, etc.) basis sets are designed to converge as the size of the basis set increases. Thus, at a given level of theory, the energy change from a cc-pVQZ basis set to a cc-pV5Z will be smaller than the energy change from a cc-pVTZ basis set to a cc-pVQZ basis set. Extrapolation schemes have been derived based on the convergence behavior of these basis sets. For a given level of theory, we can compute the energy using several basis sets, and then uses these energies to extrapolate the energy if infinite basis sets were used. This is known as the complete basis set (CBS) limit.

The focal point analysis takes advantage of the convergence behaviors for both the level of theory and the basis set. First, a high-level geometry computation is performed. Using this geometry, single-point energy computations are computed to, on one axis, extrapolate to the complete basis set limit for a given level of theory, and on the other axis, to add higher-order energy corrections to approximate a FCI solution. Thus, a FCI/CBS answer can be reached.

1.6 Adaptive QM/MM

The methods discussed above entail a highly mathematically rigorous approach that gives accurate approximations to various properties. The trade-off is the massive computational cost that makes it impossible to apply these methods to systems of a few hundred atoms or more. One way that we can apply the accuracy of quantum mechanical methods to bigger system sizes is to use a hybrid approach known as quantum mechanics/molecular mechanics (QM/MM). Just as the name implies, QM/MM treats a small part of the system with QM while modelling the rest with a classical molecular mechanics approach. In traditional quantum mechanics/molecular mechanics (QM/MM), a small, chemically relevant portion (the primary subsystem, PS) of the entire system of interest (S), is described with quantum mechanics

while the rest is treated with molecular mechanics. In a protein, for example, the active site would be modelled with QM since it is there that a proper treatment of electronic structure is needed. This allows us to model many-body effects on a large-scale while accurately treating the electronic structure of the system.

The potential energy $V(S)$ for the QM/MM method is given by

$$V(S) = V_{MM}(S) + V_{QM}(PS) - V_{MM}(PS)$$

where $V_{MM}(S)$ is the energy of the system S computed with molecular mechanics, $V_{QM}(PS)$ is the energy of the PS computed with quantum mechanics, and $V_{MM}(PS)$ is the energy of the PS computed with molecular mechanics.

The potential energy using QM is given with a chosen method as discussed above. The potential energy in molecular mechanics is computed as a sum of bonded and non-bonded terms:

$$V_{MM} = V_{bond} + V_{angle} + V_{dihedral} + V_{electrostatic} + V_{van\,der\,Waals}$$

The potential energy is given as a function of position and is known as a force field.

One shortcoming of traditional QM/MM is that the atoms making up the PS are fixed. This is not a problem if we are computing one snapshot of the system. However, QM/MM is often used in conjunction with molecular dynamics (MD) or Monte Carlo (MC) simulations to study the time evolution or bulk properties for a system. In cases such as solvent diffusion and ion transport where the active center is non-localized, traditional QM/MM is inappropriate. For example, there is no guarantee that an initial set of solvent molecules around a target system described with QM will stay near the target system in a time simulation.

Adaptive quantum mechanics/molecular mechanics (QM/MM) is an extension of traditional QM/MM that addresses this problem. In adaptive QM/MM, the PS region is determined on-the-fly: its definition is reevaluated and redefined at each step of a simulation. However, switching the description of an atom from QM to MM, or vice versa, between two simulation frames can cause abrupt changes in the forces and energies of the system. In most adaptive QM/MM methods, the discontinuity between the frames is addressed through the introduction of a buffer zone between the PS and the rest of the system. The atoms in the buffer zone can be assigned QM or MM descriptions, and the permutations associated with different assignment combinations give rise to different QM/MM configurations. These configurations are then interpolated through a smoothing function to effectively give dual QM and MM character to buffer group atoms. For example, an energy-based interpolation is shown below as a linear combination of the potential energy from the n^{th} QM/MM partition ($V^n(S)$) weighted by a smoothing function (λ):

$$V_{adaptive}(S) = \sum_n \lambda^n V^n(S)$$

This in theory gives way to a smoother transition between the QM and MM regions. Various adaptive QM/MM methods differ in how they define the buffer zone, how they determine QM/MM configurations, the form of the smoothing function, and whether the interpolation of QM/MM configurations is energy- or force-based.

Chapter 2

Spin-Orbit Coupling via Four-Component Multireference Methods: Benchmarking on p-Block Elements and Tentative Recommendations[1]

[1]Reproduced with permission from Zhang, B; Vandezande, J. E.; Reynolds, R. D.; Schaefer, H. F. *J. Chem. Theory Comput.* 2018, 14, 1235–1246, Copyright 2018 American Chemical Society. Download available via ACS Articles on Request at https://pubs.acs.org/articlesonrequest/AOR-jjjzy3BGvbQhKKTbxGYv.

2.1 Abstract

Within current electronic structure theory methods, fully relativistic four-component (4c) approaches based on the Dirac Hamiltonian treat spin-orbit coupling with the most rigor. The spin treatment arises naturally from the formulation and does not need to be included *ad hoc*. Spin-orbit splittings can provide insightful benchmark criteria for the assessment of 4c methods; however, there have not been extensive studies in this respect. Spin-orbit splittings of the p-block elements B–I are computed using the 4c-CASSCF, 4c-CASPT2, and 4c-MR-CISD+Q methods, recently implemented in BAGEL, with uncontracted Dunning basis sets. Comparison with experiment reveals that the four-component methods yield good results, with most computed splittings falling within 15% of experiment. A large basis set is needed to obtain accurate splittings of the light elements B–F, while splittings of heavier elements show little basis dependence. The 4c-MR-CISD+Q method gave the best splittings for light elements while 4c-CASSCF showed the best splittings for elements beyond fluorine. The 4c-CASPT2 method gave the best splittings for group 13 atoms.

2.2 Introduction

The inclusion of relativistic effects in quantum chemistry methods has led to large improvements in the accuracy of quantum chemical computations, explaining the color of gold and the shrinkage of s- and p-orbitals in heavy atoms.[1] Spin-orbit coupling (SOC) can only arise with the inclusion of relativity in the Hamiltonian, as it couples different spin states that are otherwise independent in the Schrödinger equation. This is most notable for open-shell species, where previously degenerate states are split due to the coupling of spin and orbital angular momentum.[2] While this effect is small for light atoms (404 cm^{-1} for F),[3] it becomes increasingly significant for heavier atoms (7603 cm^{-1} for I).[4] This is important in both astrochemistry and atmospheric chemistry when dealing with the binding of free atoms, particularly the binding of halogens.[5,6] SOC can significantly affect the dissociation energies of systems as simple as water-halogen complexes, and plays a large role in their reaction pathways.[7–9]

A common way to treat SOC is through a spin-orbit (SO) operator applied perturbatively on top of a non-relativistic computation. Fedorov, Koseki, Schmidt, and Gordon[2] give an excellent overview of SO operators that have been used to treat SOC, and we will not belabor this paper with further discussion. Such *ad hoc* approaches are not entirely satisfactory if one seeks to arrive at an answer from a theoretically rigorous ansatz where spin arises naturally, which can only happen in a relativistic framework. Numerous reviews have been written that detail relativistic effects on chemical systems and the current progress on the development of relativistic methods to treat these problems.[1,10–13] Four-component (4c) methods (and the derived exact two-component methods) are fully relativistic and explicitly couple spin and orbital angular momentum; thus they are an ideal way to treat relativistic effects like SOC.[2,14] However, the computationally demanding nature of 4c methods has lead to many approximate approaches that seek to capture relativistic effects at reduced computational costs.

Although SOC frequently serves as a benchmark for the assessment of relativistic methods, these benchmarks have often been limited to a few select systems at a time.[15–68] We highlight some studies here, and include a more detailed list of previous studies in the supporting information. Ermler, Christiansen, and coworkers looked at atoms Li through element 118 with SO operators based on relativistic effective potentials in a series of papers.[37–44] Eliav and Kaldor studied row 1 atoms with their relativistic Fock-space coupled-cluster (FSCC) method.[30] They have also used this approach to treat many other atoms and give a comprehensive summary of their work elsewhere.[33] In a later study, Kaldor and coworkers adapted the four-component intermediate-Hamiltonian FSCC method to a two-component treatment with relativistic pseudopotentials to study the SO splittings of select atoms from groups 11 and 12.[32]

Werner and coworkers[46] studied how the SO splittings of F, Cl, and Br computed with the Breit–Pauli operator changed with various basis sets. Gordon and coworkers[48] studied the SO splittings of third-row transition elements with scalar and two-component methods. Roos and coworkers[50] computed SO splittings of a few atoms based on a restricted active space SCF method with a spin-orbit Hamiltonian. In a later study, Roos and coworkers[51] used the same approach to perform a more comprehensive survey of p-block elements B–At (excluding groups 15 and 18) with a relativistically recontracted ANO basis set. Fleig[55] reported splittings for group 13 atoms using an MRCC method with the Dirac–Coulomb (DC) Hamiltonian. Jones and Albers used tight binding methods to study the splittings of select actinide elements.[58]

More recently, approaches to treat spin-orbit coupling with equation-of-motion coupled-cluster (EOM-CC) methods have been developed. Wang and coworkers study the SO splitting of select group 13 and 17 atoms using EOM-CC with the inclusion of a SO operator at the post-Hartree-Fock level.[61,62] Pal and coworkers[67] computed the SO splitting of alkali metals as a benchmark for relativistic EOM-CC methods, starting from either a closed-shell 4c or an exact two-component (X2C) reference determinant. Liu and coworkers[68] looked at group 13 and 17 atoms with an EOM-CC method treating SOC with the first-order Douglas–Kroll–Hess SO operator. They also computed the SO splittings of the same atoms using spin-adapted time-dependent density functional theory, treating SOC as a perturbation with the X2C one-electron spin-orbit operator.[65]

There has been no previous large-scale study on the effectiveness of 4c methods in determining the spin-orbit splittings of atoms or molecules. In this research, the SO splittings of p-block elements from B–I (excluding the closed-shell noble gases) are computed via a 4c multireference approach. A four-component Dirac–Coulomb–Breit Hamiltonian is used to incorporate relativistic effects. Static correlation in these inherently multiconfigurational systems is treated with the complete active space self-consistent field (CASSCF) method, while dynamic correlation effects are assessed with both perturbation theory and the configuration interaction method.

2.3 Method

The four-component (4c) complete active space self-consistent field (4c-CASSCF)[69–71] method used here employed a 4-orbital active space corresponding to the valence s- and p-orbitals of the atoms. The

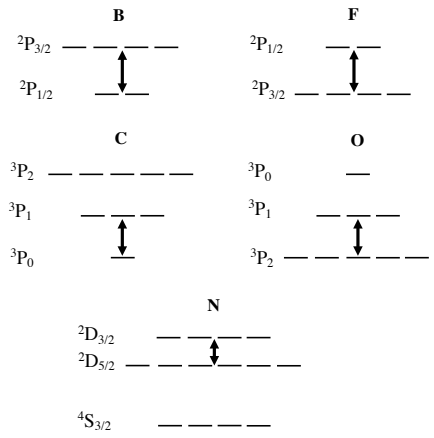

Figure 2.1: States considered for the B–F atoms. The ordering of these states is consistent for the next three rows except for a switch in energy between the two ^{2}D components of P, As, and Sb. The arrows represent splittings that are discussed explicitly in this work. All splittings and absolute energies are provided in the supporting information.

Dirac–Coulomb–Breit Hamiltonian was used. The 4c-CASSCF computations were state-averaged among all targeted and lower states to avoid biasing the orbitals for any specific states and to prevent root flipping problems (see Figure 2.1 for a representation of all the states computed). 4c-Dirac–Hartree–Fock (4c-DHF) closed-shell computations were used to generate starting guess orbitals for 4c-CASSCF computations; if the atom contained an odd number of electrons, the corresponding anion was used.[72]

Breit terms were included in the computation of the two-electron integrals[73]. A point nucleus was employed and both electronic and positronic orbitals were optimized using a second-order optimization algorithm[71]. The all-electron 4c-CASSCF computations used the default convergence parameters (10^{-8} for the root mean square of the orbital gradient).

The fully internally contracted 4c multireference configuration interaction method with singles, doubles, and the Davidson correction (ic-4c-MR-CISD+Q)[74–76], as well as the fully internally contracted 4c complete active space second order perturbation theory (ic-4c-CASPT2) [76–79] method were employed to include dynamic correlation. The virtual space was not truncated. Multi-state ic-4c-CASPT2 was used along with a multi-state, multi-reference (MS-MR) parameterization, and did not include a level shift. The 4c-CASPT2 and 4c-MR-CISD+Q computations were converged according to the default parameters (10^{-6} for the root mean square of the residual vector). Core orbitals, defined as the non-valence s-, p-, and d-orbitals (see Table 2.1) were frozen. Non-frozen-core (all-electron) computations were also

Table 2.1: Frozen orbitals for the frozen-core computations. Analogous all-electron computations were also carried out for select atoms.

Atoms	Frozen Orbitals
B–F	1
Al–Cl	5
Ga–Br	14
In–I	23

performed for select atoms to determine the effects of core correlation. The virtual space was not truncated. We will omit the prefixes in the following sections and simply refer to the four-component methods used as CASSCF, CASPT2, and MR-CISD+Q. All computations were performed with the BAGEL relativistic quantum chemistry program[80, 81]. The program does not take advantage of molecular symmetry, thus all computations were run with C_1 symmetry. The all-electron cc-pVXZ basis sets were used: atoms B–F employed X = D, T, Q, 5[82]; elements Al–Br employed X = D, T, Q[83, 84]; and elements In–I employed X = T, Q[85] (due to the absence of available cc-pVDZ basis sets). For elements Al–Cl, computations using the cc-pV(X+d)Z basis sets[86] were also performed. All basis sets used were uncontracted to allow for flexibility in the core region.[47]

Although relativistically optimized basis sets exist and have been in active development, we chose the correlation consistent(cc) basis sets of Dunning and coworkers due to their extensive use in the computational community and availability for all the considered atoms. To compare the performance of uncontracted nonrelativistic basis sets against relativistic ones, the uncontracted relativistically optimized basis sets of Dyall were also used. CASSCF and CASPT2 computations were performed for Ga–Br with Dyall's cc-pVTZ basis set[87–89] and In–I with Dyall's cc-pVTZ and cc-pVDZ[89–91] basis sets.

Since BAGEL employs density fitting for the generation of two-electron integrals and there are no available fitting basis sets for uncontracted cc-pVXZ basis sets (henceforth referred to as uc-cc-pVXZ), auxiliary basis sets were constructed using AutoAux[92] as implemented in Orca 4.0.0.2[93]. These are included in the supporting information for reference. The energies and splittings for all computations are also shown in the supporting information.

2.4 Results and Discussion

To determine if the generated fitting basis sets using the AutoAux feature in Orca adequately saturated the auxiliary space, the uc-cc-pV5Z-AutoAux fitting basis was used in conjunction with the uc-cc-pVDZ basis set and compared to the results using the uc-cc-pVDZ-AutoAux fitting basis. These computations were performed with the CASSCF, CASPT2, and MR-CISD+Q methods, as shown in Table 2.2. The largest change (Δ) in the absolute splittings was 0.3 cm^{-1}, while the percent change ($\%$) is consistently below 0.5% and significantly smaller in most cases. Thus the error in SO splittings due to the choice of fitting

Table 2.2: The differences (Δ, in cm^{-1}) and relative errors (%) of SO splittings between a uc-cc-pVDZ computation with an uc-cc-pVDZ-AutoAux (DZA) and uc-cc-pV5Z-AutoAux (5ZA) fitting basis.

	CASSCF		CASPT2		MR-CISD+Q	
Atom	Δ	%	Δ	%	Δ	%
B	0.002	0.018	-0.003	-0.026	-0.004	-0.034
C	0.002	0.015	-0.074	-0.522	-0.064	-0.438
N	0.000	0.000	-0.025	-0.242		
O	0.015	0.010	-0.214	-0.151	-0.204	-0.139
F	0.031	0.008	-0.308	-0.083	-0.305	-0.082

Δ is computed by subtracting SO splittings computed with the DZA fitting basis from those with the 5ZA fitting basis. The relative error is compared against splittings with the DZA fitting basis. The largest error for CASSCF is 0.03 cm^{-1}, while for CASPT2 and MR-CISD+Q the largest error is 0.3 cm^{-1}.

basis set is negligible when compared to the other errors that arise in these computations. Accordingly, the AutoAux generated fitting basis sets are able to saturate the auxiliary space, and could perhaps even be pruned to increase the speed of these computations.

In all following figures, experimental splittings form the reference point (zero on the y-axis) against which theoretically computed SO splittings are compared. Thus, a negative error means that the computation underestimates experimentally observed splittings. MR-CISD+Q computations had to be omitted for some atom and basis set combinations due to the large memory requirements.

Boron Through Fluorine

The SO splittings of atoms B–F monotonically approach the experimentally observed results as the basis set size increases, as seen in Figure 2.2. The nitrogen results approach the experimental from above, while the other atoms approach from below. All-electron computations tended to marginally outperform the corresponding frozen-core computations, except for in fluorine where some aberrant cancellation of errors results in the frozen-core computations being more accurate. The trend seen for MR-CISD+Q all-electron computations in fluorine may indicate that the splitting will be overestimated at the complete basis set limit, however, no attempt was made at an extrapolation. The difference between all-electron and frozen-core computations increases with increasing basis set size, which may be due to the fact that a larger basis set will include more core-core and core-valence correlation. The similarity between the all-electron and frozen-core computations indicates that the contribution of core correlation is small for these elements. The inclusion of dynamic correlation in the CASPT2 and MR-CISD+Q methods tended to improve the spin-orbit splittings, and MR-CISD+Q consistently outperformed or matched the CASPT2 results. In

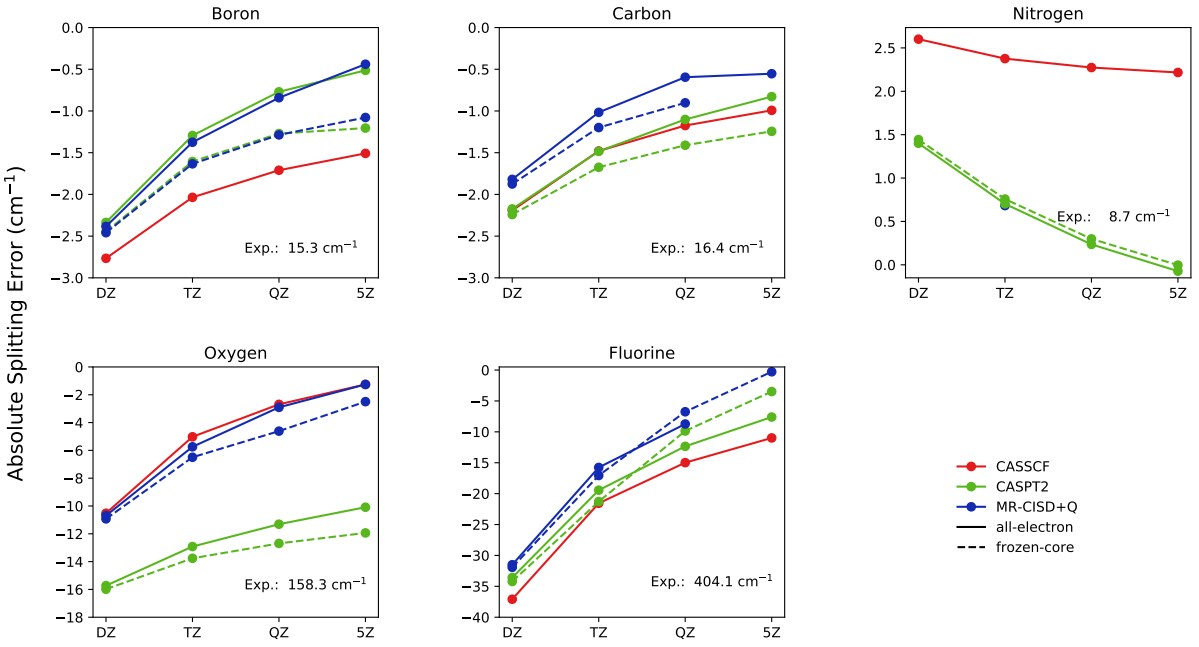

Figure 2.2: Spin-orbit splitting errors for elements B–F in cm^{-1} across the uc-cc-pVXZ (X = D, T, Q, 5) basis sets. Absolute experimental splittings[94] are shown in text on the lower right of each panel. The computations approach the experimental values as the basis set size is increased. For these atoms only, the correlated methods tend to outperform CASSCF, but not consistently. For correlated methods, freezing the core marginally increases the absolute error of most computations.

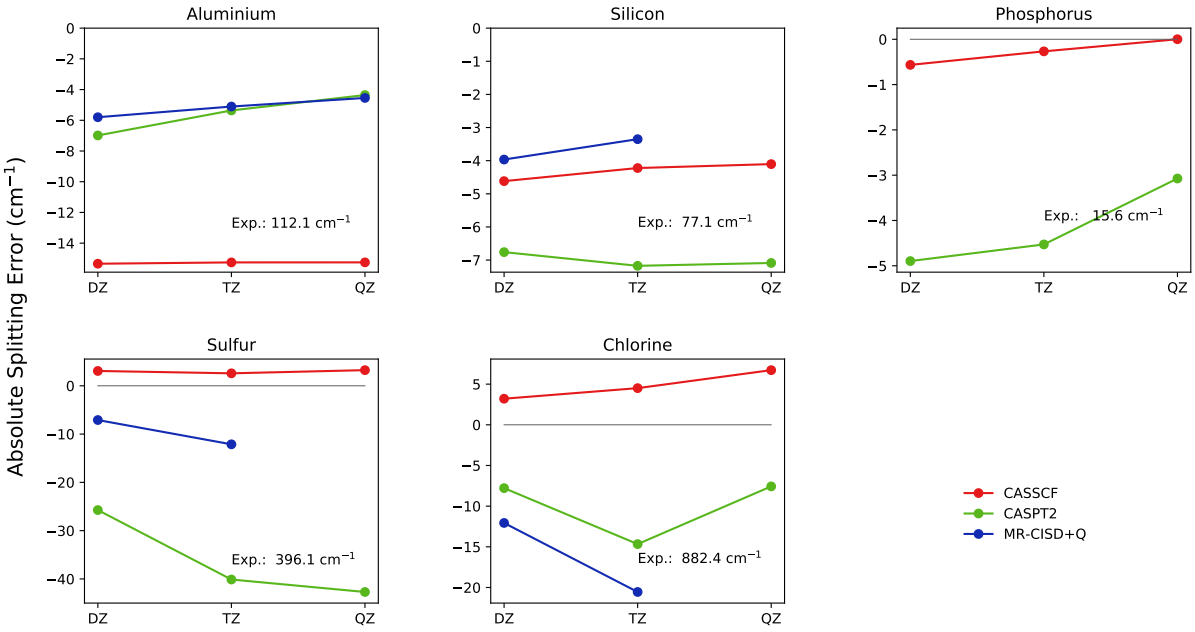

Figure 2.3: Spin-orbit splitting errors for elements Al–Cl in cm^{-1} across the uc-cc-pVXZ (X = D, T, Q) basis sets. Core orbitals were frozen in all computations. Absolute experimental splittings[94] are shown in text on the lower right of each panel. The correlated methods, CASPT2 and MR-CISD+Q, tended to perform worse than CASSCF (except for aluminum). Computations with the uc-cc-pV(X+d)Z basis sets show similar splitting errors, and are reported in the supporting information.

the case of oxygen, CASPT2 gave a larger splitting error compared to CASSCF and MR-CISD+Q gave similar splittings to CASSCF.

Aluminum Through Iodine

Unlike the lighter elements, there is no clear basis-dependent trend to the SO splittings in later rows (see Figures 2.3, 2.4, and 2.5). The CASSCF (red) computations showed minimal change with increasing basis set size, while for the post-CASSCF methods increasing the basis sets size had inconsistent effects on the splitting errors. The CASSCF method outperformed most computations that included dynamic correlation (CASPT2 (green) and MR-CISD+Q (blue)) for all atoms except group 13 elements aluminum, gallium and indium, and yielded especially accurate results for group 17 elements(δ Cl = 6.7 cm^{-1}, δ Br = 1.2 cm^{-1}, δ I = 63.8 cm^{-1} at QZ)

It is interesting to note that CASSCF and CASPT2 computations for tellurium places the $^3P_1$ state below the $^3P_0$ state in energy, while experiment places the $^3P_0$ state below $^3P_1$. Experimentally,[94] the difference in the two states is around 44 cm^{-1}. With an uc-cc-pVQZ basis set, we find the difference to be

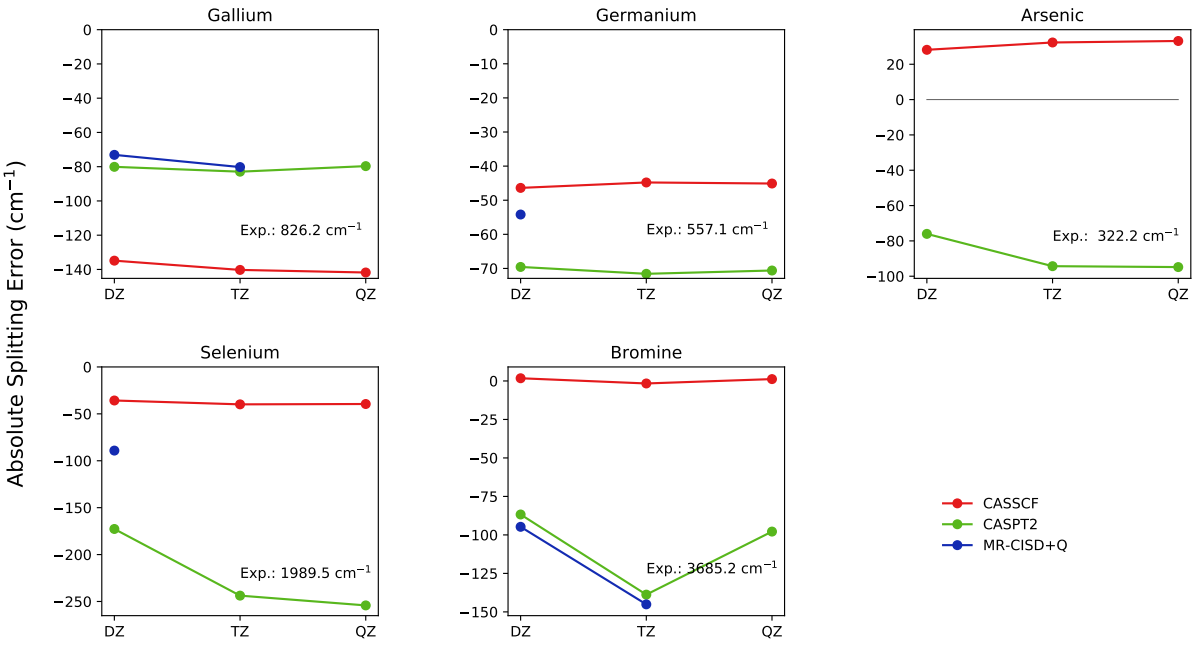

Figure 2.4: Spin-orbit splitting errors for elements Ga–Br in cm^{-1} across the uc-cc-pVXZ (X = D, T, Q) basis sets. Core orbitals were frozen in all computations. Absolute experimental splittings[94] are shown in text on the lower right of each panel. Just as for elements Al–Cl, the correlated methods, CASPT2 and MR-CISD+Q, tended to perform worse than CASSCF (except for gallium).

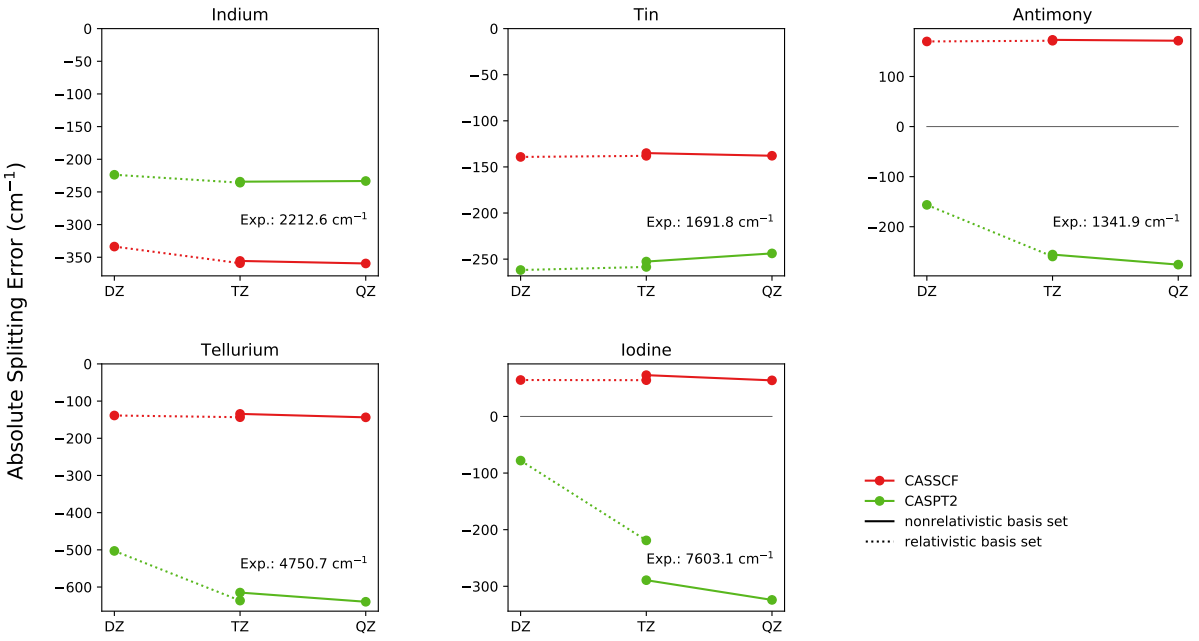

Figure 2.5: Spin-orbit splitting errors for elements In–I in cm^{-1} across the uc-cc-pVXZ (X = T, Q) basis sets. Computations using a uc-cc-pVDZ basis set are absent for atoms In–I due to the lack of available Dunning basis sets. Computations using relativistically optimized uc-cc-pVDZ and uc-cc-pVTZ basis sets are also shown. The MR-CISD+Q method was intractable for elements In–I due to the large memory requirements. Core orbitals were frozen in all computations. Absolute experimental splittings[94] are shown in text on the lower right of each panel. Just as for the previous two rows, the correlated method CASPT2 tended to perform worse than CASSCF (except for indium).

around 103 cm^{-1} with CASSCF and 204 cm^{-1} with CASPT2. Roos et al.[52] also found the $^3P_1$ state to lie below the $^3P_0$ state and computed the difference between the two states to be around 508 cm^{-1} with their perturbative approach to SOC. As mentioned in the methods section, the splittings shown in Figure 5 are from the $^3P_2$ state to the $^3P_1$ state.

Although all-electron Dunning cc-pVDZ basis sets do not exist for these elements, we show the DZ and TZ computations using the relativistically optimized basis sets of Dyall. Table 2.3 shows a comparison of the absolute and relative errors obtained by using both nonrelativistic and relativistic TZ basis sets for Ga–I. The two types of basis sets give marginal differences in SO splittings, and one does not give consistently better splittings than the other. CASPT2 splittings for iodine shows the biggest absolute and relative difference, but the difference in relative error is still less than 1%. Except for In, the difference for CASPT2 computations were larger than for CASSCF computations. Across the atoms listed, CASSCF gives an RMSD value of 0.17% and CASPT2 gives an RMSD value of 0.43% for the relative error as compared against experiment. Given the similarities in the splittings given by the two basis sets at the TZ level (except for perhaps iodine), it would be reasonable to use the splittings given by the relativistically optimized DZ basis set as an approximation for splittings given by a nonrelativistic DZ basis set.

Group Comparison

Due to the large increase in spin-orbit splittings down the periodic table, relative splitting errors (%) must be used to compare the effectiveness of methods in different rows. Figure 2.6 shows the relative splitting errors of all atoms studied at the CASSCF level of theory. As mentioned previously, increasing the basis set size makes a significant difference for B–F but not for Al–I. For all groups except group 15, CASSCF performed reasonably consistently down each group. Early studies of SO splitting[15–19] have shown that SCF methods give good SO splittings, thus the relatively accurate splittings for groups 16 and 17 especially are not too surprising. Figure 2.7 shows the relative splitting errors at the CASPT2 level of theory for all atoms studied.

CASPT2 shows similar trends as CASSCF save for group 15. However, it seems to give larger splitting errors down a column for any group, especially group 14. This trend may be due to the fact that our computations were done with frozen-core and the importance of core correlation increases as one goes down the periodic table. The observation that splittings were better for group 17 may be due to the decrease in the importance of core correlation going across a given p-block; however, there was no systematic improvement in the splittings when comparing the different groups. MR-CISD+Q shows similar results to the CASPT2 method, but the number of computations is limited due to the high memory requirements, and thus comparisons need to be made with caution.

Visscher and Dyall[95] have observed that for row 1 atoms additional tight functions are needed in the uc-cc-pVDZ and cc-pVTZ basis sets for better description of the core region. Accurate representation of the core region is especially important for studying SOC since SO interaction scales as $1/r^3$ and is largest in the core. They have also observed that later rows show less dependence on basis set quality. Thus, the basis dependence seen in elements B–F could be a reflection of inadequate core description given by the

Table 2.3: Absolute (Abs) and relative (%) errors of SO splittings using a nonrelativstic uc-cc-pVTZ basis set (TZ) and a relativistically optimized uc-cc-pVTZ basis set (rTZ) for elements Ga–Br and In–I.

	Ga		Ge		As		Se		Br	
	Abs	%	Abs	%	Abs	%	Abs	%	Abs	%
CASSCF(TZ)	-141.27	-16.98	-44.76	-8.03	32.33	10.03	-39.88	-2.00	-1.63	-0.04
CASSCF(rTZ)	-140.40	-17.11	-44.98	-8.07	33.40	10.37	-38.52	-1.94	2.97	0.08
Difference	-1.12	-0.14	-0.21	-0.04	1.07	0.33	1.35	0.07	4.60	0.12
CASPT2(TZ)	-82.91	-10.04	-721.54	-12.84	-94.31	-29.27	-243.76	-12.25	-138.78	-3.77
CASPT2(rTZ)	-80.01	-9.68	-72.84	-13.07	-92.45	-28.69	-246.38	-12.38	-134.03	-3.64
Difference	2.90	0.35	-1.30	-0.23	1.86	0.58	-2.62	-0.13	4.76	0.13

	In		Sn		Sb		Te		I	
	Abs	%	Abs	%	Abs	%	Abs	%	Abs	%
CASSCF(TZ)	-355.65	-16.07	-134.99	-7.98	173.20	12.91	-134.41	-2.83	72.96	0.96
CASSCF(rTZ)	-359.11	-16.23	-137.98	-8.16	171.43	12.78	-143.30	-3.02	64.10	0.84
Difference	-3.46	-0.16	-2.99	-0.18	-1.77	-0.13	-8.90	-0.19	-8.86	-0.12
CASPT2(TZ)	-234.33	-10.59	-252.64	-14.93	-255.65	-19.05	-614.89	-12.94	-289.21	-3.80
CASPT2(rTZ)	-235.85	-10.66	-258.46	-15.28	-259.48	-19.34	-636.47	-13.40	-218.95	-2.88
Difference	-1.52	-0.07	-5.83	-0.34	-3.83	-0.29	-21.58	-0.45	70.26	0.92

Abs in cm^{-1} is compared against experiment. The difference is calculated by subtracting TZ relative errors against experiment from rTZ ones. Both CASSCF and CASPT2 levels of theory are shown. Numbers in bold indicate instances where rTZ splittings had a smaller relative error.

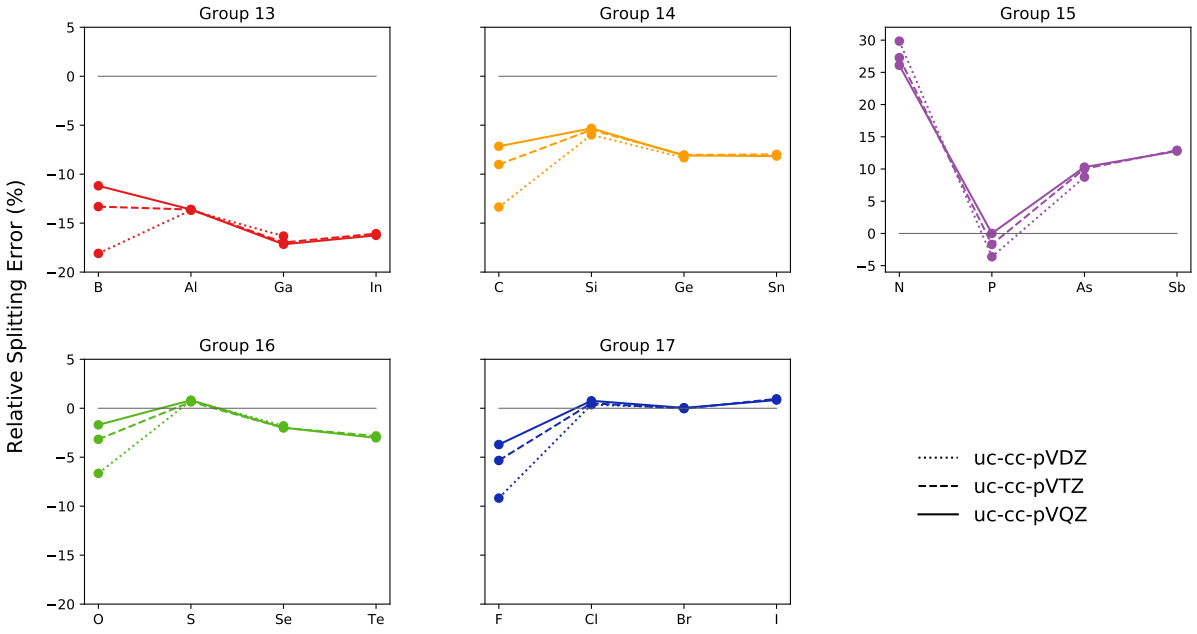

Figure 2.6: Relative spin-orbit splitting errors at the CASSCF level of theory down individual groups of the periodic table. Computations using a uc-cc-pVDZ basis set are absent for atoms In–I due to the lack of available Dunning basis sets. Group 15 had much larger relative errors due to their smaller SO splittings relative to the other members of their period.

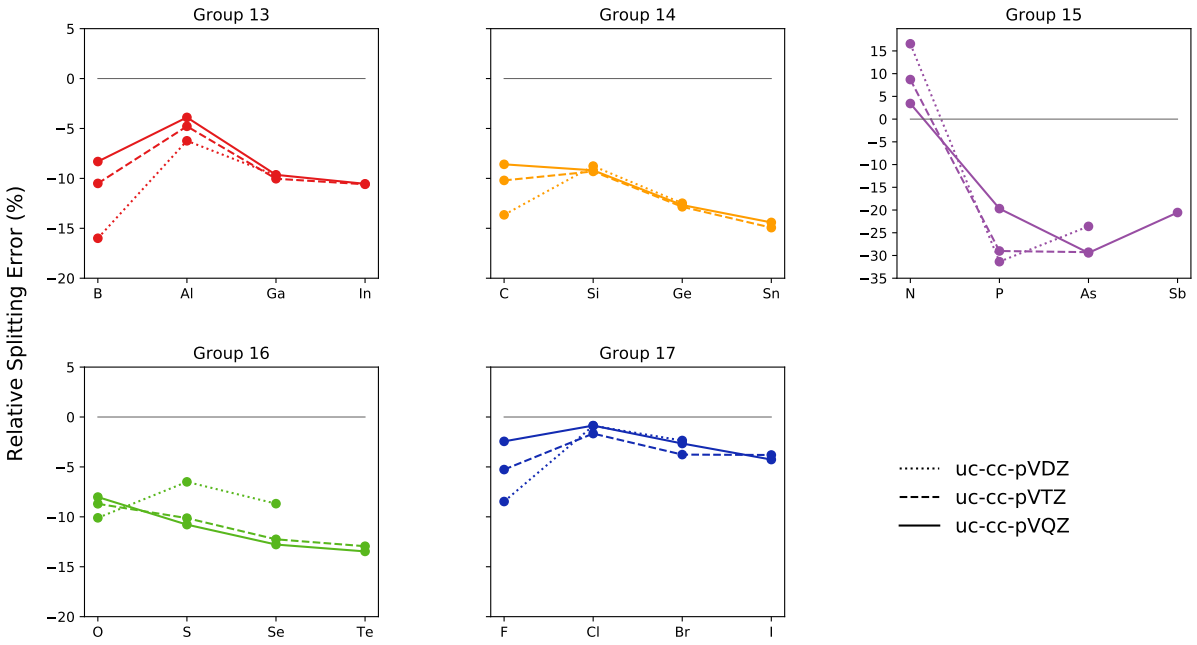

Figure 2.7: Relative spin-orbit splitting errors at the CASPT2 level of theory down individual columns of the periodic table. Computations using a uc-cc-pVDZ basis set are absent for atoms In–I due to the lack of available Dunning basis sets. Group 15 had much larger relative errors due to their smaller SO splittings relative to the other members of their period.

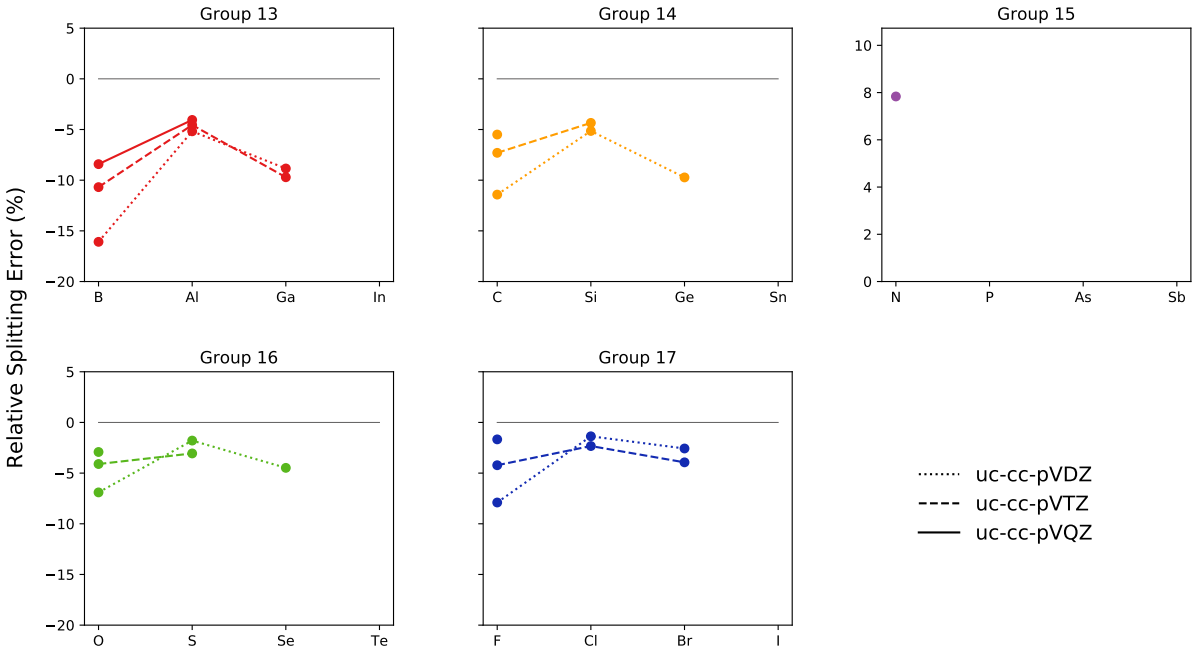

Figure 2.8: Relative spin-orbit splitting errors at the MR-CISD+Q level of theory down individual groups of the periodic table. Computations using a uc-cc-pVDZ basis set are absent for atoms In–I due to the lack of available basis sets. The errors mirror those of CASPT2: larger basis sets reduce the error for B–F and have minimal effect for later rows.

smaller basis sets, while the lack of dependence in later elements, especially in CASSCF, may be due to sufficient saturation of the core region by these basis sets.

In several studies, Fleig[53–55] and Visscher[47] have explored the effect of correlated methods on SO splittings. Inclusion of the Gaunt term has been shown to reduce SO splitting in light elements.[54] For heavier elements, an imbalanced contribution of the p^3 configuration to the different ground-state components could outweigh the improvement gained by introducing correlation and decrease the splittings.[47, 54] Correlated methods with valence-only excitations reduce the weight of the p-orbital, and hence will in general reduce the spin-orbit splitting. These reasons might offer explanations for the underestimation of SO splittings by the CASPT2 and MRCI methods.

To assess the importance of core correlation for computing the spin-orbit splittings of heavier atoms, a number of different computations varying the degree of core correlation included in the CASPT2 computations were performed. These results are shown in Table 2.4. As discussed earlier, there is not a large difference between frozen-core and all-electron CASPT2 computations for B and F. For atoms in the next row (i.e. Al and Cl), there is not a large difference in the splittings if one uses a regular Dunning basis set. However, the addition of a tight d function for these two atoms causes a large change in the relative error when comparing freezing five core orbitals to freezing one orbital. Although the SO splitting did not change much for Al and Cl between TZ and (T+d)Z when five cores were frozen, there is a bigger difference for computations with more core correlation. For example, the absolute splitting error for Cl with five frozen-core orbitals is -14.66 cm^{-1} with TZ and -14.85 cm^{-1} with(T+d)Z, while for Cl with one frozen orbital the error is -10.40 cm^{-1} for TZ and -5.06 cm^{-1} for (T+d)Z. Thus the addition of the tight d function, which provides a better description of the core region, makes more of a difference when more core correlation is included.

Overall, group 13 shows a larger change in computed splittings with increasing core correlation than group 17. For Al, Ga, and In, the difference in relative error between valence-only computations and some inclusion of the core orbitals is around 4-5%. For group 17 atoms, this difference is only around 1%. For In and I, computations were performed up to fourteen frozen-core orbitals due to computational cost. However, for the heavier atoms, the difference between including the $(n-1)$d shell and additional core orbitals is minimal, and we can assume there will not be a significant improvement in SO splittings of In and I with more core correlation. Since core correlation becomes less important as one moves across the p-block, it is probable that including core orbitals in correlated methods for all the atoms studied will at most improve the results by 5% for CASPT2. We expect that including more orbitals in MRCI computations will show similar or slightly better improvement.

2.5 Conclusions

Spin-orbit splittings of p-block elements B–I were computed via the fully relativistic four-component methods CASSCF, CASPT2, and MR-CISD+Q. Comparisons to experimentally measured splittings showed good agreement for elements B–F, with successively larger absolute errors for later rows. Most computations were within 15% of the experimental results and no method consistently outperformed the

Table 2.4: Absolute (Abs) and relative (%) errors of SO splittings at the CASPT2 level with an uc-cc-pVTZ basis set for group 13 and 17 atoms.

Orbitals Frozen	Group 13			Group 17		
	Atom	Abs	%	Atom	Abs	%
1	B	-1.61	-10.50	F	-21.28	-5.27
0		-1.29	-8.47		-19.44	-4.81
5	Al	-5.36	-4.78	Cl	-14.66	-1.66
1		-3.47	-3.10		-10.40	-1.18
0		-3.41	-3.04		-10.14	-1.15
5	Al(+d)	-5.12	-4.57	Cl(+d)	-14.85	-1.68
1		-0.39	-0.35		-5.06	-0.57
0		-0.34	-0.30		-4.82	-0.55
14	Ga	-82.91	-10.04	Br	-138.78	-3.77
9		-51.85	-6.28		-108.04	-2.93
5		-46.94	-5.68		-111.17	-3.02
1		-44.65	-5.40		-105.49	-2.86
0		-44.26	-5.36		-104.14	-2.83
23	In	-234.33	-10.59	I	-289.21	-3.80
18		-151.40	-6.84		-261.95	-3.45
14		-133.13	-6.02		-249.64	-3.28

Absolute errors are in cm^{-1} and are compared against experiment. The number of orbitals frozen indicates the extent of core correlation. Both uc-cc-pVTZ and uc-cc-p(V+d)Z values are shown for Al and Cl. A negative value means an underestimation of SO splitting.

others across the periodic table. In general, increasing the basis set size yielded more accurate results for elements B–F across all methods; however, no such trend was observed for the later rows.

The CASSCF method outperformed the correlated methods (except for group 13) in both time and accuracy for p-block elements beyond fluorine, and gives consistent relative errors going down the column (except for group 15). For groups 16 and 17 the errors were within 5%, and sometimes much less. Furthermore, CASSCF computations show few basis set effects, except for in elements B–F, with uc-cc-pVDZ basis sets yielding similar results to uc-cc-pVQZ. Thus a large basis set does not have to be used for the most accurate results. Given that this is the least computationally demanding method, taking at most one fifth the time of a CASPT2 computation, CASSCF is recommended for all cases beyond fluorine with the exception of group 13 atoms.

For elements beyond fluorine, the CASPT2 method gave better results than CASSCF for group 13 atoms only. For elements beyond fluorine, we show that including core orbitals can improve CASPT2 results by 5% for group 13 atoms and 1% for group 17 atoms with the uc-cc-pVTZ basis set. CASPT2 can thus give SO splittings within 5-6% for group 13 elements with inclusion of core orbitals.

The MR-CISD+Q method yielded the best results for elements B–F (within 1–2% with a 5Z basis set). However, it fails to outperform the other methods for the heavier elements. Although it is the most computationally demanding method, taking at least ten times longer than CASPT2 and has large memory demands, there is no significant increase in reliability to offset this cost. For light elements, MR-CISD+Q can give accurate results with reasonable cost; however, a large basis set (QZ or 5Z) is needed to fully realize these benefits.

Since the correlated methods used in this study are internally contracted, there is a high dependence on the number of correlated orbitals included. Unfreezing core orbitals for group 13 and 17 atoms with the 4c-CASPT2 method did not show significant changes. The use of core-valence basis sets might further improve splittings, however, we did not use these due to convergence issues. The inclusion of more core description, through the addition of tight functions by using core-valence basis sets and including more orbitals in the correlation, in principle should improve the SO splittings for these correlated methods, especially MR-CISD+Q. The offset of this is cost. Our present results show that although MR-CISD+Q is a more accurate theory, a frozen-core treatment with a valence-only basis set will not outperform CASSCF.

A relativistic basis set at the TZ level did not give significantly different results than a nonrelativistic one for CASSCF and CASPT2. Thus, an uncontracted nonrelativistic basis set seem to give a good description of the systems studied.

In general, fully 4c methods give more accurate splittings than HF or SCF methods that treat SOC perturbatively[46, 68], but perform comparably with their correlated methods that incorporate SOC. The 4c results are also generally comparable with the DC Hamiltonian approach of Fleig.[55]

It is clear that although various current four-component methods can predict reasonably accurate SO splittings, the development of new, scalable methods is by no means complete. Until then, the above recommendations have been given with the goal of achieving the best balance between computational cost and reliability.

Chapter 3

Relatives of Cyanomethylene: Replacement of the Divalent Carbon by B^-, N^+, Al^-, Si, P^+, Ga^-, Ge, and As^+ [1]

[1]Reproduced from Abbott, B.Z.; Hoobler, P.R.; Schaefer, H.F.*Phys. Chem. Chem. Phys.*, 2019, 21, 26438-26452, with permission from the PCCP Owner Societies.

3.1 Abstract

The lowest lying singlet and triplet states of $HBCN^-$, HCCN, $HNCN^+$, $HAlCN^-$, HSiCN, $HPCN^+$, $HGaCN^-$, HGeCN, and $HAsCN^+$ were studied using the CCSDT(Q)/CBS//CCSD(T)/ aug-cc-pVQZ level of theory. Periodic trends in geometries, singlet-triplet gaps, and barriers to linearity were established and analyzed. The first row increasingly favors the triplet state, with a singlet-triplet gap (ΔE_{ST} = $E_{\text{singlet}} - E_{\text{triplet}}$) of 3.5 kcal mol^{-1} , 11.9 kcal mol^{-1} , and 22.6 kcal mol^{-1} , respectively, for $HBCN^-$, HCCN, and $HNCN^+$. The second row increasing favors the singlet state, with singlet-triplet gaps of -20.4 kcal mol^{-1} ($HAlCN^-$), -26.6 kcal mol^{-1} (HSiCN), and -26.8 kcal mol^{-1} ($HPCN^+$). The third row also favors the singlet state, with singlet-triplet gaps of -26.8 kcal mol^{-1} ($HGaCN^-$), -33.5 kcal mol^{-1} (HGeCN), and -33.1 kcal mol^{-1} ($HAsCN^+$). The HXCN species have larger absolute singlet-triplet energy gaps compared with their parent species XH_2 except for the case of $X{=}N^+$. The effect of the substitution of hydrogen with a cyano group with was analyzed with isodesmic analysis and NBO.

3.2 Introduction

Background

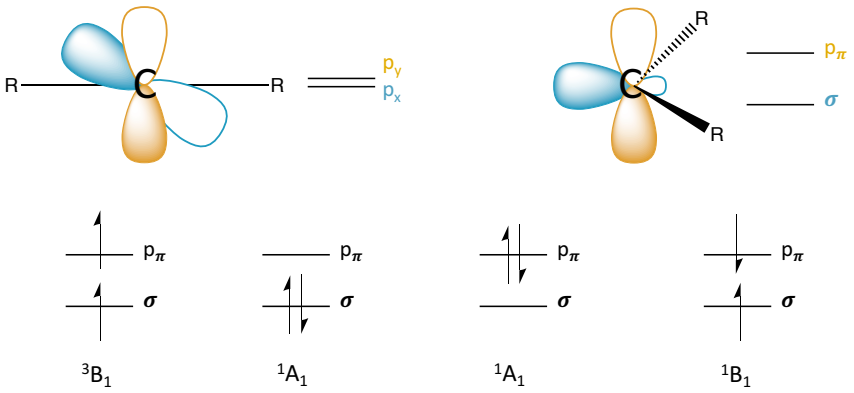

Figure 3.1: Linear and bent geometries of carbenes and their electronic configurations.

It is well known that the spin states of carbenes dictate their reactivities.[96–110] Carbenes with linear geometries possess two degenerate p orbitals (p_x, p_y). As the carbene bends, the p_x orbital becomes increasingly sp^2-like in hybridization while the out-of-plane p_y retains its p-character (Figure 3.1), lifting the previous degeneracy. We will refer to the sp^2-like orbital as σ and the out-of-plane p orbital as p_π. The two carbon-centered valence electrons can either be spin-unpaired, leading to a triplet B_1 state (in the case of C_{2v} symmetry), or spin-paired, leading to singlet A_1 and B_1 states. The lowest-lying singlet

state is typically the $^1A_1$ state with electron occupation in the σ orbital, whose energy is lowered as the geometry moves away from linearity. The ground state multiplicity of carbenes is qualitatively dependent on the energy separation between the σ and p_π orbitals, where a triplet ground state is more favorable if the separation is small. Equilibrium geometries of triplet states thus almost always have larger R–C–R (where C is the divalent carbon) angles than the corresponding singlet states.[111–113]

Singlet carbenes have been utilized to achieve stereospecific one-step addition to carbon-carbon double bonds, as well as insertion into O–H bonds.[100,101] Triplet carbenes, in contrast, undergo a non-stereospecific two-step addition to carbon double-bonds and participate in hydrogen abstraction reactions. Controlling the spin state of carbenes through various substituent and solvent substitutions has thus been a subject of study since the 1950s. Numerous recent reviews have been written summarizing both experimental[103,105,106,108,109] and computational[99,110] studies on the influence of subsituents and solvents on the spin state of carbenes, and we refer the reader to these for a more comprehensive overview of carbene history.

Also of interest is the effect of substitution of the divalent carbon by other atoms. A natural choice of replacement would be with atoms in the same group, such as silicon (silylenes) and germanium (germylenes), and indeed there are many such studies.[112–118] A different substitution involves the replacement of carbon with the isoelectronic nitrogen cation[119,120] as well as others in that group (P^+, As^+). More recently, there has been increasing interest in group 13 anions such as B^- and Al^- as carbene substitutes.[121–123] Few of these studies focus on comparing more than two different divalent atom species and instead focus on substituent effects on singlet-triplet (ST) gaps. The most comprehensive comparison of periodic trends for the substitution of the divalent atom comes from Cramer and coworkers,[124] who used DFT and multireference methods to study the simpler XH_2 species, X= B^-, C, N^+, Al^-, Si, P^+, Ga^-, Ge, As^+.The substituted valence-isoelectronic atoms form a three by three block on the periodic table, and gives an adequate representation of row and column effects.

In the present study, we look at the same nine valence-isoelectronic atoms of Cramer and coworkers, but with the cyano-substituted HXCN species instead of XH_2. Carbenes with cyano (CN) substituents are interesting not only due to their relevance in organic chemistry,[125] but because of their importance in combustion and atmospheric chemistry. Cyanomethylene (HCCN) is a major product of pyridine decomposition and a precursor of molecules found on Titan's atmosphere.[126–129] Due to its relevance to atmospheric chemistry, its reaction with O_2,[129,130] H,[131] CH_3,[131] and NO[130,132] have all been investigated. Additionally, HCCN is thought to be an intermediate in HCN and HNC related reactions which are important in interstellar chemistry,[133–140] and has been identified as a candidate to forming acrylonitrile in the interstellar medium through a reaction with CH_2.[141] Given the important applications of cyanocarbenes, it would be of interest to study how substitution of the divalent carbon with other atoms affects their electronic structures.

Previous work

Group 13

Reactions of B, Al, and Ga atoms with HCN, in which HXCN (X=B, Al, Ga) acts as an intermediate, have been subjects to a variety of combined experimental and theoretical studies. Andrews and coworkers have looked at reactions of B – Ti atoms with HCN in a solid argon matrix with accompanying DFT computations.[142,143] Miller, Douberly and coworkers have analyzed the reactions of Al, Ga, and In with HCN via spectroscopy in helium nanodroplets accompanied with *ab initio* computations.[144,145] However, these studies focused on the radical, not the anion that is isoelectronic with the cycanocarbene species. The only theoretical studies involving HXCN ($X=B^-$, Al^-, Ga^-) is the 1996 work of Liebman and coworkers on the singlet and triplet states of $HBCN^-$ with MBPT2 geometries and coupled cluster single point energies.[146,147]

Group 14

The HCCN radical has long been the subject of experimental[148–167] and computational[128,154,168–185] study. Its interesting history has been characterized by disagreements regarding its triplet ground state geometry. Early experiments from the 1960s–1980s[148–153] supported a linear equilibrium geometry while theory[154,169–173,175,178] led by Kim, Schaefer, Pople, Radom, and coworkers consistently found that a bent equilibrium geometry is lower in energy. Since then, theory has been confirmed based on more recent experimental studies. HCCN is a bent molecule, with an estimated barrier to linearity of between 240 and 300 cm^{-1}(0.67 and 0.86 kcal mol^{-1}).[157,158,160–163,178,182]

Although silicon is also considered to be abundant in space[186–189] and the cyanocarbene relative cyanosilylene (HSiCN) is thought to be detectable,[190,191] there have been fewer studies on this molecule.[159,190–197] In 1998, Maier and coworkers studied the reaction between Si and HCN in an argon matrix and observed HSiCN as an isomerization product.[159] HSiCN was detected by Thaddeus and coworkers in 2002 via microwave spectroscopy.[190] Flores and Carballeria used density functional theory (DFT) and multireference methods to study the lowest-lying singlet and triplet isomers of HSiCN.[192] Ding and coworkers have considered the potential energy surface of HSiCN and its isomers with DFT.[193] Kalcher studied various cyanosilylenes with RHF-ACPF and CAS(2,2)-ACPF methods.[194,195] Thorwirth and Harding have examined various isomers of singlet HSiCN with coupled cluster theory..[196] Zhang and coworkers have looked at the low-lying excited states of HSiCN and HSiNC with CASSCF and CASPT2 methods.[191] Various experimental[198–200] and computational[201–203] studies exist on species containing germanium and cyano groups, but the only theoretical study on HGeCN that we know of is by Kassaee and coworkers (2005), who use various *ab initio* and DFT theories to study the singlet and triplet states of HGeCN (X=H,F,Cl,Br).[204]

Like HCCN, the HNCN radical is believed to be an important molecule in atmospheric chemistry.[205–209] Its reaction with OH,[205] NO,[206,207] and CN,[208] as well as its role as an intermediate in the reaction CH + N_2[209] have all been investigated theoretically. The HNCN cation, however, has significantly fewer studies. Puzzarini and Gambi performed a coupled-cluster study on the triplet state of $HNCN^+$ along with the parent HNCN radical and HNCN anion.[210] Antoniotti and coworkers performed a multireference study on the singlet and triplet states of $HNCN^+$ and its isomers.[211] Betterton performed a DFT study on singlet and triplet $HPCN^+$ and the HPCN radical.[212] There has been no theoretical study on the $HAsCN^+$ molecule.

In this work, we use rigorous *ab initio* methods to study the lowest lying singlet and triplet states of $HBCN^-$, HCCN, $HNCN^+$, $HAlCN^-$, HSiCN, $HPCN^+$ $HGaCN^-$, HGeCN, and $HAsCN^+$ to gauge periodic affects on structure and energetics. Periodic trends for geometry parameters, ST gaps, and barriers to linearity will be discussed, and quantitative chemical analysis will be provided to understand these trends. In addition, we will compare the ST gaps of HXCN with those of XH_2 and use isodesmic substitution to analyze the differences in ST gaps between the two species. Throughout the study we will use "row one" to refer to HXCN where X=B^-, C, or N^+, "row two" for X=Al^-, Si, or P^+, and "row three" for X=Ga^-, Ge, or As^+.

3.3 Methods

Geometries and Vibrational Frequencies

Geometry optimizations and harmonic frequencies were obtained for the lowest lying singlet and triplet states of the nine cyanocarbene-like species using coupled cluster theory with singles, doubles, and perturbative triples excitations [CCSD(T)].[213–215] The $1s$ electrons of boron, carbon, and nitrogen were not correlated, the $1s$, $2s$, and $2p$ electrons of aluminum, sillicon, and phosphorus were not correlated, and the $1s$, $2s$, $2p$, $3s$, and $3p$ electrons of gallium, germanium, and arsenic were not correlated. These frozen core settings were used for all subsequent computations. For species containing the atoms B, C, and N, the Dunning correlation consistent basis sets aug-cc-pVXZ (X=T,Q)[216,217] were used. For species containing the atoms Al, Si, and P, the aug-cc-pV(X+d)Z (X=T,Q)[218,219] basis sets were used. In our preliminary computations, we found that correlating the $4s$ and $4p$ electrons only for row 3 atoms Ge, Ga, and As was not sufficient to accurately capture the singlet-triplet gap and the additional correlation of $3d$ electrons was needed (see Table 3.1). Thus, the weighted core-valence basis sets [aug-cc-pwCVXZ (X=T,Q)][220] were used for Ge, Ga, and As containing species because the aug-cc-PVXZ[221] basis set parameters for Ge and Ga were obtained with the correlation of the $4s$ and $4p$ electrons only. Moving forward, the set of aug-cc-pVXZ (X=T,Q), aug-cc-pV(X+d)Z (X=T,Q) and aug-cc-pwCVXZ (X=T,Q) basis sets will be collectively referred to as XZ.

The Hartree–Fock density matrix elements, coupled cluster amplitudes, and lambda coefficients were converged to 10^{-9} for all molecules except singlet $HGaCN^-$ at CCSD(T)/aug-cc-pwCVQZ, which was

Table 3.1: Comparison of the effects of correlation on the energies and singlet-triplet (ST) gap of $HGaCN^-$. A positive ST gap (ΔE_{ST}) implies a triplet ground state.

Correlation	Singlet (E_h)	Triplet (E_h)	ΔE_{ST} (kcal mol^{-1})
$4s4p$	$-2016.632\,961$	$-2016.649\,159$	10.164
$3d4s4p$	$-2017.120\,277$	$-2017.075\,290$	-28.230
$3s3p3d4s4p$	$-2017.419\,249$	$-2017.373\,822$	-28.506
all	$-2017.593\,425$	$-2017.547\,691$	-28.698

$\Delta E_{ST} = E_{\text{singlet}} - E_{\text{triplet}}$. Correlating only the $4s$ and $4p$ electrons gives a qualitatively wrong singlet-triplet gap and reverses the energy ordering of the states. The addition of $3d$ correlation is needed to describe the correct energy ordering. At higher levels of correlation the singlet state is the ground state. Energies shown are from single point computations of $HGaCN^-$ at the CCSD(T)/aug-cc-pwCVTZ level of theory with CCSD(T)/aug-cc-pVTZ optimized singlet and triplet geometries.

converged to 10^{-8}. The RMS gradient for the geometry optimizations was converged to 10^{-8}. A restricted Hartree–Fock (RHF) reference was used for singlet states, while a restricted open-shell Hartree–Fock (ROHF) reference was used for triplet states. The electronic structure program CFOUR 2.0[222] was used to obtain all reported geometries and harmonic vibrational frequencies except for the vibrational frequencies for triplet HGaCN, which was computed using Molpro.[223,224]

Energetics

Singlet-triplet (ST) energy gaps (ΔE_{ST}) defined as $E_{\text{singlet}} - E_{\text{triplet}}$, where a positive gap indicates a triplet ground state, were obtained via the focal point approach of Allen and coworkers.[225–228] This approach allows us to obtain ST gaps at the CCSDT(Q)/CBS level of theory. Single point energies at the HF, MP2, CCSD, CCSD(T), CCSDT, and CCSDT(Q) levels of theory were performed using the CCSD(T)/QZ optimized structures. The HF through CCSD(T) computations were performed using Molpro[223,224] at the DZ, TZ, QZ, and 5Z levels. The basis set type used for each species was the same as that employed for geometry optimizations. Full CCSDT computations were performed at the TZ level using the NCC[229] module for the singlets and the VCC module for the triplets in CFOUR 2.0. CCSDT(Q) computations were performed at the DZ level using the NCC module in CFOUR 2.0 for the singlet states. CCSDT(Q) is not implemented for an ROHF reference in CFOUR 2.0, and thus for the triplet states MRCC[230] was used (CCSDT(Q)/B[231] energy was taken). Computations were performed to ensure that CFOUR 2.0 and MRCC gave identical CCSDT(Q) results for the singlet states. HF energies were extrapolated using Feller's three-point formula[232] and correlation energies with Helgaker's two-point formula.[233] For $HNCN^+$,

the focal point table did not satisfactorily converge at the CCSDT(Q) level of theory, and CCSDTQ computations at the DZ level was performed using the NCC module in CFOUR 2.0 for the singlets and MRCC[230,234] for the triplets. Thus the ST energy gap reported for HNCN is at the CCSDTQ/CBS level of theory.

Additional corrections were added to the (ΔE_{ST}), including the Diagonal Born–Oppenheimer correction (ΔE_{DBOC} at the Hartree–Fock level) and harmonic zero-point vibrational energy correction (ΔE_{ZPVE}), both with CFOUR 2.0. Frozen-core corrections ($\Delta E_{\mathrm{core}} = E_{\mathrm{AE-CCSD(T)}} - E_{\mathrm{CCSD(T)}}$) were computed with Dunning's aug-cc-pwCVQZ[220,235,236] basis sets in Molpro. Scalar relativistic corrections ($\Delta_{\mathrm{rel}} = E_{\mathrm{AE-CCSD(T)/SF-X2C-1e}} - E_{\mathrm{AE-CCSD(T)}}$) for the singlet and triplet states were determined by correlating all electrons using the SF-X2C-1e[237] Hamiltonian as implemented in CFOUR 2.0 with the uncontracted aug-cc-pVTZ (B^-, C, N^+, Ge^-, As^+), aug-cc-pV(T+d)Z (Al^-, Si, P^+), and aug-cc-pwCVTZ (Ga^-) basis sets. The uncontracted aug-cc-pVTZ basis set was used for Ge^- and As^+ containing species instead of aug-cc-pwCVTZ because of convergence issues arising from linear dependency. To test the validity of this choice, we compared the relativistic corrections for singlet HGeCN using an uncontracted aug-cc-pVTZ and an uncontracted aug-cc-pwCVTZ basis set and found that there was very little difference between the two (0.8 mE_h).

The Cartesian coordinates, harmonic vibrational frequencies, and full focal point tables with absolute energies for all optimized species are included in the Supporting Information.

Assessment of Multi-Reference Character

To assess our use of single-reference methods for geometries and energetics, we performed full-valence CASSCF single point computations with 14 electrons in 13 active orbitals ($14e^-$, 130) using CCSD(T)/QZ optimized geometries for each HXCN species in Molpro. The aug-cc-pVTZ basis was used for X=B^-, C, N^+, aug-cc-pV(T+d)Z for X=Al^-, Si, P^+, and aug-cc-pwCVTZ for X= Ga^-, Ge, As^+. Most of the singlet and triplet HXCN species showed a dominant configuration of 85% or more. The second largest configuration was mostly small (contribution of under 2%); a few had a contribution around 3.5% and one had a contribution of 5.2% (singlet $HAlCN^-$). Since none of the species displayed strong multi-reference character, our highly-correlated single-reference approach is more than sufficient to capture the electronic structure of the system. The species with the most multi-reference character from our computations is singlet $HNCN^+$, with the dominant configuration having a contribution of 84% and the second dominant configuration having a contribution of 3.6%. The lack of convergence for the focal point energies of $HNCN^+$ at CCSDT(Q) mentioned in the Energetics section might be due to $HNCN^+$ having slightly more multi-reference character. However, the additional CCSDTQ correction recovers some of the missing correlation and is sufficient for the treatment of $HNCN^+$. The two largest CASSCF coefficients and the corresponding occupation vectors for all HXCN species are included in the Supporting Information.

H–X–C Angle Scan

A relaxed scan of the H–X–C angle, where X is the central divalent atom, of all HXCN species in the singlet and triplet states was performed to compare the barriers of linearity between species. Constrained geometry optimizations were performed with the H–X–C angle varying between 90° and 180° at 15° intervals. The CCSD(T) level of theory was used for the constrained geometry optimizations. To assess the multi-reference nature of the HXCN species as it changed geometries, MR-CISD+Q was also used to compute single point energies on top of CASSCF ($10e^-$, 8o) constrained geometry optimizations. The aug-cc-pVTZ basis was used for X=B^-, C, N^+, aug-cc-pV(T+d)Z for X=Al^-, Si, P^+, and aug-cc-pwCVTZ for X= Ga^-, Ge, As^+. All computations for the scans were done using Molpro.

NBO

Natural bonding orbital theory[238] was used to analyze the hybridization of relevant orbitals and give the natural bond orders. Important orbital donor-acceptor interactions were also analyzed via second order perturbation theory analysis of the NBO Fock matrix. The interaction energy between a donor orbital i and acceptor orbital j is defined as:

$$E(2) = q_i \left[\frac{F(i,j)^2}{\epsilon_j - \epsilon_i} \right]$$

where q is the charge, ϵ is the orbital energy, and $F(i,j)$ is the matrix element of the NBO Fock matrix. All NBO computations were performed using NBO 5.0 as interfaced in the Q-Chem[239] software package. B3LYP[240] was used for the NBO analysis with the aug-cc-pVDZ (X=B^-, C, N^+), aug-cc-pV(D+d)Z (X=Al^-, Si, P^+), and aug-cc-pwCVDZ (X= Ga^-, Ge, As^+) basis sets.

3.4 Results and Discussion

Geometries

Figure 3.2 shows the geometric parameters of the lowest lying singlet and triplet states in the nine cyano-species studied, while Figure 3.3 gives a pictorial representation of the comparison of the geometry parameters across these species summarized below. A comparison of QZ geometry parameters with TZ parameters, as well as structures obtained from previous studies (when available) can be found in the Supporting Information.

H–X–C Angle. For both the singlet and triplet states, the H–X–C angle generally increases as X traverses across a row. The difference is most dramatic for the first row (X=B^-, C, N^+), which showed a difference of 12.6° and 48.3° from X=B^- to X=N^+, for the singlet and triplet states respectively. For triplet $HNCN^+$ the lowest energy structure is no longer bent but linear. For the second row (X=Al^-, Si, P^+) in the singlet state, the angle stays almost the same (down by 0.01°) as X goes from Al^- to Si, then increases slightly by 1.8° for X=P^+. In the triplet state the angle increases by 3.68° from X=Al^- to X=P^+. For the third row (X=Ga^-, Ge, As^+) in the singlet state, the angle goes down slightly (0.13°) as X goes from

Ga^- to Ge, then increases by 1.24° for X=As^+. In the triplet state the angle increases by only 0.89° from X=Ga^- to X=As^+. As X goes down the column the angles are generally decreasing. However, the change in the H–X–C angle from the first row to the second row (average decrease of 18.82° for the singlets and 36.43° for the triplets) is larger than the change from the second row to the third row (average decrease of 0.94° for the singlets and actually an increase of 0.70° for the triplets). Thus there is some convergence of the H–X–C angle with increasing atom size. For all nine species the H–X–C angle of the singlet state is smaller than that for the triplet state.

X–C–N Angle. The range of the X–C–N angle across the nine species is narrower than that of the H–X–C angle, varying between 171.3° and 174.6° in the singlet states and 175.1° and 180.0° in the triplet states. However, there is no discernible trend for how the X–C–N angle changes between species, other than the fact that the singlet state angles are always smaller than triplet state angles. We note that none of the angles is perfectly linear, except in the case of triplet $HNCN^+$.

X–H and X–C Bonds. The X–H and X–C bonds follow the same periodic trend and will be discussed together. For both the singlet and triplet states, the X–H and X–C bonds decrease as X traverses across a row. The net decrease in all three rows where about the same. In the singlet state, the X–H and X–C bond respectively decreased by 0.188 Å and 0.312 Å as X changed from B^- to N^+, 0.245 Å and 0.366 Å as X changed from Al^- to P^+, and 0.153 Å and 0.295 Å as X changed from Ge^- to As^+. In the triplet state, the X–H and X–C bond respectively decreased by 0.177 Å and 0.329 Å as X changed from B^- to N^+, 0.189 Å and 0.301 Å as X changed from Al^- to P^+, and 0.066 Å and 0.175 Å as X changed from Ge^- to As^+. The bond length increases as X goes down a column but at a decreasing rate, similar to the H–X–C angle. The only exception is the slight decrease in the X–C bond length from 1.992 Å to 1.987 Å going from X=Al^- to Ga^-. Predictably, the X–H bond is always shorter compared to the X–C bond in a given species.

C–N Bond. The C–N bonds show the smallest variation between species, lying between 1.17 Å and 1.18 Å in the singlet states and 1.17 Å and 1.19 Å in the triplet states for all species except $HNCN^+$, which has longer C–N bonds (1.223 Å for the singlet and 1.306 Å for the triplet). The exception is because $HNCN^+$ does not have a carbene-like electronic structure, which will be discussed in detail in later sections. It is interesting to note that the shortest C–N bond is seen in HSiCN and HGeCN, both with bonds of around 1.167 Å for the singlet and triplet states. The C–N bond is shorter than the H–X bond in all cases except for singlet and triplet HCCN and $HNCN^+$, and shorter than the X–C bond in all species except triplet $HNCN^+$. The C–N bond length in triplet $HNCN^+$ especially suggests that the bonding is less sp in nature like the other species and has more sp^2 character brought about by its linear geometry. This observation will be discussed more thoroughly in later sections.

Singlet-Triplet Gaps

Table 3.2 shows the focal pointed singlet-triplet (ST) gap ($\Delta E_{ST} = E_{\text{singlet}} - E_{\text{triplet}}$) for all nine HXCN species. The ST gap shows convergence at the CCSDT(Q) level of theory for all species except $HNCN^+$, which required an additional CCSDTQ computation to reach convergence (see Methods section). The magnitude of ZPVE corrections to the ST gap are below 0.4 kcal mol^{-1} for all of the species except

Table 3.2: Incremental focal point singlet-triplet (ST) gaps for the HXCN species. Energies are in kcal mol^{-1}. A positive ST gap implies a triplet ground state. $\Delta E_{ST} = E_{\text{singlet}} - E_{\text{triplet}}$. Single point energies were performed on the singlet and triplet HXCN geometries optimized at the CCSD(T)/QZ level of theory. Each column shows the incremental change (δ) in ST energy gap from the previous column. Numbers shown in brackets were not directly computed. The final ΔE_{ST} is given with corrections as follows: $\Delta E_{ST} = \Delta E_{ST}$[CCSDT(Q)/CBS] + ΔE_{ZPVE} + ΔE_{core} + ΔE_{rel} + ΔE_{DBOC}. Additional details on these corrections are given in the Methods section. The ST gap was extrapolated to the CCSDTQ/CBS level of theory for $HNCN^+$

HBCN$^-$

	HF	$+\delta$ MP$_2$	$+\delta$ CCSD	$+\delta$ CCSD(T)	$+\delta$ CCSDT	$+\delta$ CCSDT(Q)	NET
aug-cc-pVDZ	+13.33	−5.84	−2.76	−0.93	−0.13	−0.18	[+3.48]
aug-cc-pVTZ	+13.48	−6.56	−2.14	−0.93	−0.11	[−0.18]	[+3.55]
aug-cc-pVQZ	+13.51	−6.93	−1.84	−0.93	[−0.11]	[−0.18]	[+3.51]
aug-cc-pV$_5$Z	+13.52	−7.11	−1.70	−0.94	[−0.11]	[−0.18]	[+3.48]
CBS Limit	[+13.52]	[−7.30]	[−1.55]	[−0.94]	[−0.11]	[−0.18]	[+3.44]

$\Delta E_{ST} = 3.44 - 0.33 + 0.37 - 0.04 + 0.06 = 3.49$

HCCN

	HF	$+\delta$ MP$_2$	$+\delta$ CCSD	$+\delta$ CCSD(T)	$+\delta$ CCSDT	$+\delta$ CCSDT(Q)	NET
aug-cc-pVDZ	+13.96	+1.31	−2.25	−1.13	−0.17	−0.19	[+11.53]
aug-cc-pVTZ	+14.55	+0.38	−1.97	−1.10	−0.19	[−0.19]	[+11.48]
aug-cc-pVQZ	+14.59	−0.03	−1.78	−1.08	[−0.19]	[−0.19]	[+11.32]
aug-cc-pV$_5$Z	+14.60	−0.26	−1.66	−1.08	[−0.19]	[−0.19]	[+11.22]
CBS Limit	[+14.61]	[−0.50]	[−1.54]	[−1.08]	[−0.19]	[−0.19]	[+11.12]

$\Delta E_{ST} = 11.12 + 0.25 + 0.47 - 0.07 + 0.09 = 11.86$

$HNCN^+$

	HF	$+\delta$ MP$_2$	$+\delta$ CCSD	$+\delta$ CCSD(T)	$+\delta$ CCSDT	$+\delta$ CCSDT(Q)	$+\delta$ CCSDTQ	NET
aug-cc-pVDZ	+43.03	−13.70	−0.61	−4.54	−0.20	−1.08	+0.24	[+23.14]
aug-cc-pVTZ	+43.94	−14.97	−0.29	−4.78	−0.15	[−1.08]	[+0.24]	[+22.91]
aug-cc-pVQZ	+43.97	−15.46	−0.07	−4.82	[−0.15]	[−1.08]	[+0.24]	[+22.63]
aug-cc-pV$_5$Z	+43.99	−15.75	+0.06	−4.84	[−0.15]	[−1.08]	[+0.24]	[+22.47]
CBS Limit	[+44.00]	[−16.06]	[+0.21]	[−4.86]	[−0.15]	[−1.08]	[+0.24]	[+22.30]

$\Delta E_{ST} = 22.30 - 0.20 + 0.42 - 0.08 + 0.20 = 22.64$

$HAlCN^-$

	HF	$+\delta$ MP$_2$	$+\delta$ CCSD	$+\delta$ CCSD(T)	$+\delta$ CCSDT	$+\delta$ CCSDT(Q)	NET
aug-cc-pVDZ	−5.56	−9.49	−3.57	−0.41	+0.02	−0.06	[−19.08]
aug-cc-pVTZ	−5.28	−10.87	−2.70	−0.43	+0.05	[−0.06]	[−19.29]
aug-cc-pVQZ	−5.13	−11.34	−2.26	−0.43	[+0.05]	[−0.06]	[−19.17]
aug-cc-pV$_5$Z	−5.12	−11.63	−2.02	−0.43	[+0.05]	[−0.06]	[−19.21]
CBS Limit	[−5.12]	[−11.92]	[−1.78]	[−0.44]	[+0.05]	[−0.06]	[−19.26]

$\Delta E_{ST} = -19.26 - 0.36 - 0.52 - 0.22 + 0.02 = -20.34$

HSiCN

	HF	$+\delta$ MP$_2$	$+\delta$ CCSD	$+\delta$ CCSD(T)	$+\delta$ CCSDT	$+\delta$ CCSDT(Q)	NET
aug-cc-pVDZ	−11.56	−9.43	−3.47	−0.43	+0.06	−0.07	[−24.90]
aug-cc-pVTZ	−11.39	−10.99	−2.70	−0.45	+0.09	[−0.07]	[−25.52]
aug-cc-pVQZ	−11.18	−11.61	−2.25	−0.46	[+0.09]	[−0.07]	[−25.48]
aug-cc-pV$_5$Z	−11.19	−11.91	−2.02	−0.46	[+0.09]	[−0.07]	[−25.55]
CBS Limit	[−11.21]	[−12.22]	[−1.78]	[−0.45]	[+0.09]	[−0.07]	[−25.64]

$\Delta E_{ST} = -25.64 - 0.19 - 0.45 - 0.39 + 0.03 = -26.63$

$HPCN^+$

	HF	$+\delta$ MP$_2$	$+\delta$ CCSD	$+\delta$ CCSD(T)	$+\delta$ CCSDT	$+\delta$ CCSDT(Q)	NET
aug-cc-pVDZ	−15.28	−6.21	−3.06	−0.65	+0.14	−0.09	[−25.14]
aug-cc-pVTZ	−15.07	−7.96	−2.51	−0.61	+0.17	[−0.09]	[−26.07]
aug-cc-pVQZ	−14.95	−8.61	−2.16	−0.59	[+0.17]	[−0.09]	[−26.23]
aug-cc-pV$_5$Z	−14.98	−8.90	−1.96	−0.58	[+0.17]	[−0.09]	[−26.34]
CBS Limit	[−15.01]	[−9.21]	[−1.74]	[−0.57]	[+0.17]	[−0.09]	[−26.44]

$\Delta E_{ST} = -26.44 + 0.34 - 0.24 - 0.44 + 0.06 = -26.73$

$HGaCN^-$

	HF	$+\delta$ MP$_2$	$+\delta$ CCSD	$+\delta$ CCSD(T)	$+\delta$ CCSDT	$+\delta$ CCSDT(Q)	NET
aug-cc-pVDZ	−12.13	−9.60	−1.55	−0.63	+0.06	−0.13	[−23.99]
aug-cc-pVTZ	−11.96	−10.93	−0.53	−0.78	+0.11	[−0.13]	[−24.22]
aug-cc-pVQZ	−11.87	−11.40	−0.11	−0.83	[+0.11]	[−0.13]	[−24.22]
aug-cc-pV$_5$Z	−11.85	−11.61	+0.07	−0.84	[+0.11]	[−0.13]	[−24.25]
CBS Limit	[−11.85]	[−11.83]	[+0.26]	[−0.85]	[+0.11]	[−0.13]	[−24.29]

$\Delta E_{ST} = -24.29 - 0.68 - 0.15 - 1.69 + 0.02 = -26.78$

HGeCN

	HF	$+\delta$ MP$_2$	$+\delta$ CCSD	$+\delta$ CCSD(T)	$+\delta$ CCSDT	$+\delta$ CCSDT(Q)	NET
aug-cc-pVDZ	−18.78	−9.71	−1.18	−0.61	+0.10	−0.10	[−30.27]
aug-cc-pVTZ	−18.49	−11.33	−0.20	−0.75	+0.14	[−0.10]	[−30.73]
aug-cc-pVQZ	−18.45	−11.83	+0.20	−0.79	[+0.14]	[−0.10]	[−30.82]
aug-cc-pV$_5$Z	−18.44	−12.06	+0.38	−0.80	[+0.14]	[−0.10]	[−30.88]
CBS Limit	[−18.44]	[−12.30]	[+0.56]	[−0.81]	[+0.14]	[−0.10]	[−30.95]

$\Delta E_{ST} = -30.95 - 0.17 - 0.29 - 2.09 + 0.04 = -33.46$

$HAsCN^+$

	HF	$+\delta$ MP$_2$	$+\delta$ CCSD	$+\delta$ CCSD(T)	$+\delta$ CCSDT	$+\delta$ CCSDT(Q)	NET
aug-cc-pVDZ	−21.24	−7.58	−0.71	−0.70	+0.20	−0.12	[−30.15]
aug-cc-pVTZ	−20.94	−9.37	+0.07	−0.76	+0.23	[−0.12]	[−30.88]
aug-cc-pVQZ	−20.91	−9.85	+0.39	−0.78	[+0.23]	[−0.12]	[−31.04]
aug-cc-pV$_5$Z	−20.91	−10.09	+0.55	−0.79	[+0.23]	[−0.12]	[−31.12]
CBS Limit	[−20.91]	[−10.33]	[+0.71]	[−0.79]	[+0.23]	[−0.12]	[−31.20]

$\Delta E_{ST} = -31.20 + 0.34 - 0.28 - 1.99 + 0.05 = -33.08$

$HGaCN^-$, which has a ZPVE correction of −0.68 kcal mol^{-1}. The magnitude of frozen core corrections fall below 0.5 kcal mol^{-1} for all species. The relativistic corrections to the ST gap varied the most. In the first row ($X=B^-$, C, N^+), the relativistic corrections were the smallest (less than 0.1 kcal mol^{-1} in magnitude). In the second row ($X=Al^-$, Si, P^+), the relativistic corrections increased significantly, and varied between −0.22 kcal mol^{-1} for $HAlCN^-$ and −0.44 kcal mol^{-1} for $HPCN^+$. In the third row ($X=Ga^-$, Ge, As^+), the relativistic corrections were the largest, and varied between −1.69 kcal mol^{-1} for $HGaCN^-$ and −2.09 kcal mol^{-1} for HGeCN. The singlet state was more affected by the relativistic correction compared to the triplet state in all cases. This increase in relativistic correction is expected as relativistic effects become more important with heavier atoms.[1] The DBOC corrections are the smallest in magnitude, with most having a correction below 0.1 kcal mol^{-1} and $HNCN^+$ having a DBOC correction of 0.2 kcal mol^{-1}.

As X moves from B^- to C to N^- in the first row, the ST gap increases from 3.49 kcal mol^{-1} for $HBCN^-$ to 11.86 kcal mol^{-1} for HCCN to 22.64 kcal mol^{-1} $HNCN^+$, showing an increasing preference for the triplet state. Our computed CCSDT(Q)/CBS ST gap for $HBCN^-$ is slightly lower than the value obtained by Liebman and coworkers [3.94 kcal mol^{-1} with CCSD(T)/DZP(d)//MBPT(2)/DZP(d)].[146] For HCCN, the computed CCSDT(Q)/CBS ST gap is lower than the 13.8 kcal mol^{-1} gap computed at the QCISD(T)/D95** level of theory,[173] but higher than the gaps obtained by two other studies [9.93[181] at QCISD(T)/6-311++G** and 7.61[183] with the G2 composite theory]. For $HNCN^+$, the computed CCSDT(Q)/CBS ST gap is lower than a multireference result by Antoniotti and coworkers (24.6 kcal mol^{-1} at MR-CISD(Q)/6-311G**//CASSCF($8e^-$,8o)/6-31G*) by 2 kcal mol^{-1}.[211]

Moving from the first row to the second row, there is a switch in the ordering of the states. Unlike all the species in the first row, in the second and third rows the singlet state is now the favored state, and the sign of the ST gap is now negative instead of positive. As X moves from Al^- to Si to P^-, the ST gap increases in magnitude from −20.34 kcal mol^{-1} for $HAlCN^-$ to −26.63 kcal mol^{-1} for HSiCN to −26.73 kcal mol^{-1} for $HPCN^+$. Like the first row, the magnitude of the gap in the second row is increased moving from left to right, but this time the singlet state is increasingly favored instead of the triplet. The change in the ST gaps between species are not as dramatic in the second row. The magnitude of change in the ST gap from $HBCN^-$ to HCCN is 8.37 kcal mol^{-1}, and from HCCN to $HNCN^+$ is 10.78 kcal mol^{-1}. In contrast, the magnitude of change in the ST gap from $HAlCN^-$ to HSiCN is 6.29 kcal mol^{-1}, and from HSiCN to $HPCN^+$ is only 0.10 kcal mol^{-1}. Compared to previous results, our computed CCSDT(Q)/CBS ST gap for HSiCN (−26.63 kcal mol^{-1}) is close to that of Flores and coworkers[192] (−26.51 kcal mol^{-1} at CCSD(T)/CBS//B3LYP/6-311G**) and Kalcher[194] (−26.22 kcal mol^{-1} at ACPF/aug-cc-pVTZ). It is larger than both previous DFT [24.55 kcal mol^{-1} at B3LYP/6-31G(d)][193] and multireference [22.83 kcal mol^{-1} at CASPT2 ($14e^-$,13o)/aug-cc-pVTZ][191] results. The CCSDT(Q)/CBS ST gap for $HPCN^+$ (−26.73 kcal mol^{-1}) is much smaller than magnitude than a previously obtained result of −30.44 kcal mol^{-1} using the G3 composite method.[212]

The ST gaps of the third row show a further increase in preference for the singlet state compared to the second row, evidenced by the more negative ST gaps in the HXCN ($X= Ga^-$, Ge, As^+) species. Like the second row, there is a decreasing rate of change in the ST gaps moving from left to right. The ST gap changed by 6.29 kcal mol^{-1} between $HGaCN^-$ and HGeCN, decreasing from −26.78 kcal mol^{-1} to

−33.46 kcal mol^{-1}. Going from HGeCN to HAsCN^{+}, the change is now only 0.38 kcal mol^{-1}, increasing from −33.46 kcal mol^{-1} to −33.08 kcal mol^{-1}. The change from HGeCN to HAsCN^{+} is a break from the trend of increasing ST gap magnitudes seen in the first and second rows. It is interesting to note that the CCSDT(Q)/CBS ST gaps without additional corrections do show an increase in ST gap magnitude from HGeCN to HAsCN^{+} (−30.95 kcal mol^{-1} to −31.20 kcal mol^{-1}). The two main sources for an overall lower ST gap in HGeCN thus come from: 1. a negative contribution from the ZPVE correction in HGeCN (−0.17 kcal mol^{-1}) combined with a positive ZPVE correction in HAsCN^{+} (0.34 kcal mol^{-1}); and 2. a relativistic correction in HGeCN that is more negative (−2.09 kcal mol^{-1}) than the relativistic correction in HAsCN^{+} (−1.99 kcal mol^{-1}). Our computed CCSDT(Q)/CBS ST gap for HGeCN (−33.46 kcal mol^{-1}) is slightly lower than the −32.83 kcal mol^{-1} gap reported by Kassaee and coworkers using the G2 composite method.[204] A discussion of these general trends will be made in the following section.

Origins of the ST Gap

As mentioned in the introduction, the magnitude of the ST gap in carbenes can largely be explained by understanding the nature of the two relevant orbitals for the two nonbonding electrons on the divalent atom. In our case, the HXCN species has C_s geometry, and subsequently the two relevant orbitals are an in-plane a' orbital that has mixed s and p character, and an out-of-plane a'' orbital that has largely p character. The relation between these orbitals and the ST gap can be summarized as follows: the more s-character the a' orbital has (leading to a more bent H–X–C angle), the larger the energy separation between the a' and a'' orbitals, and the more the singlet state is favored.

Schleyer and coworkers in their comparative study on silylenes vs. carbenes used these principles to explain the differences between the two species.[112,113] First, they pointed out that the nonbonding electrons in silicon are often in orbitals that have more s-character compared to carbon. Second, they compare the energy gap between the singly occupied molecular orbitals (SOMO) in the triplet state for CH_2 and SiH_2, citing the larger gap in SiH_2 as the reason for its preference for a singlet ground state.

In Figure 3.4, we quantify these two characteristics, namely, the percentage of s-character in the a' orbital in the singlet state and the energy gap between the two SOMOs in the triplet state and plot them against the ST gap for the nine species we are presently studying. We see that the observations made for carbenes and silylenes carry through to the other carbene derivatives. There is a strong correlation between the ST gap with both the percentage of s-character in the a' orbital (R^2=0.975) and SOMO energy gap (R^2=0.945) in the triplet state. As expected, the species that strongly favor a singlet ground state have both a higher percentage of s-character in their a' orbital, and a larger SOMO energy gap.

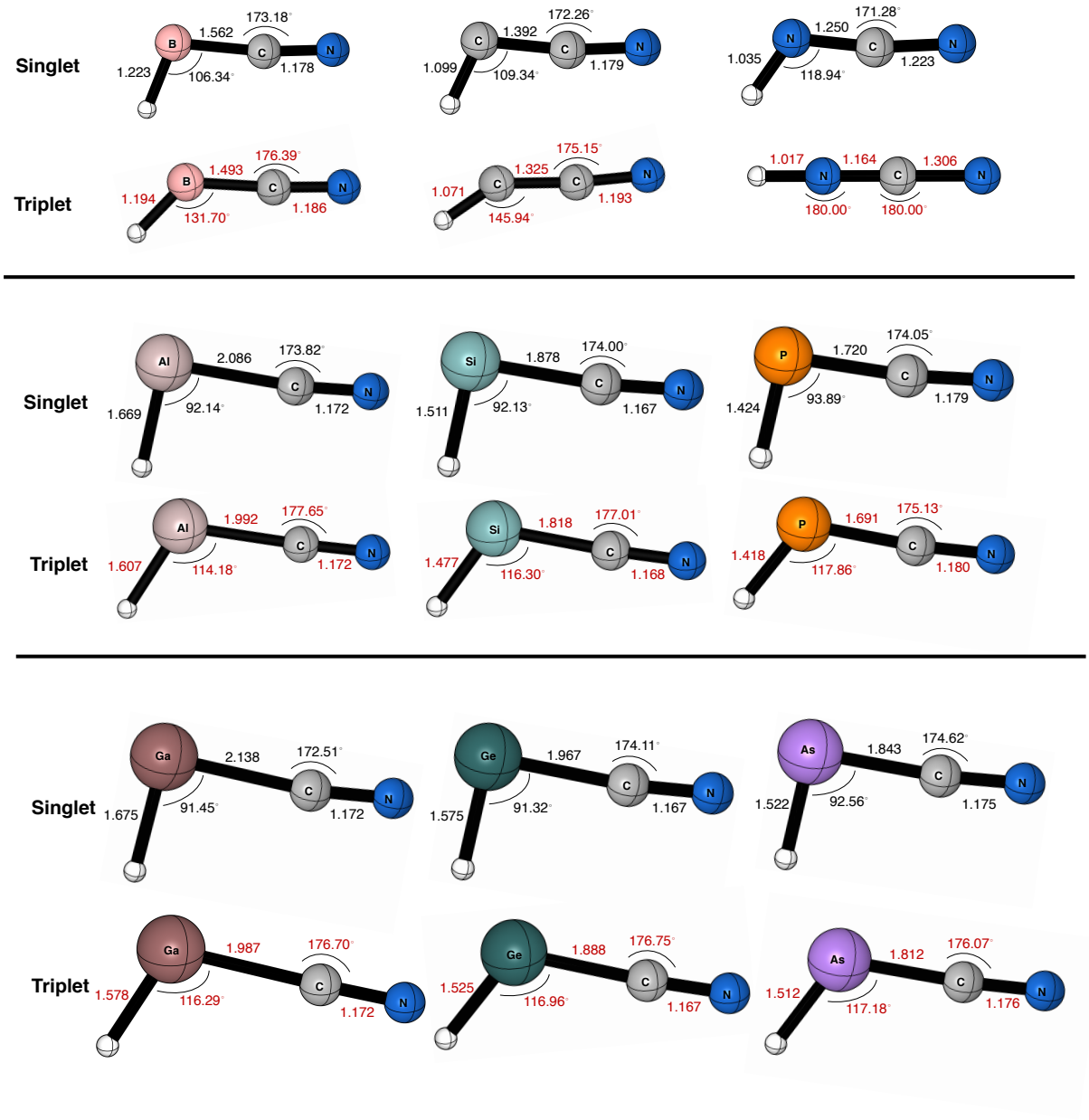

Figure 3.2: Geometries at the CCSD(T) level of theory with the aug-cc-pVQZ basis sets for the first row, aug-cc-pV(Q+d)Z for the second row, and aug-cc-pwCVQZ for the third row. In a given row, the top geometry shows values (black) for the lowest lying $^1A'$ state while the bottom show values (red) for the lowest lying $^3A''$ state. The H–X–C–N dihedral angle is 180° in all cases, i.e., each structure is planar. Graphics were made with CYLview[241]

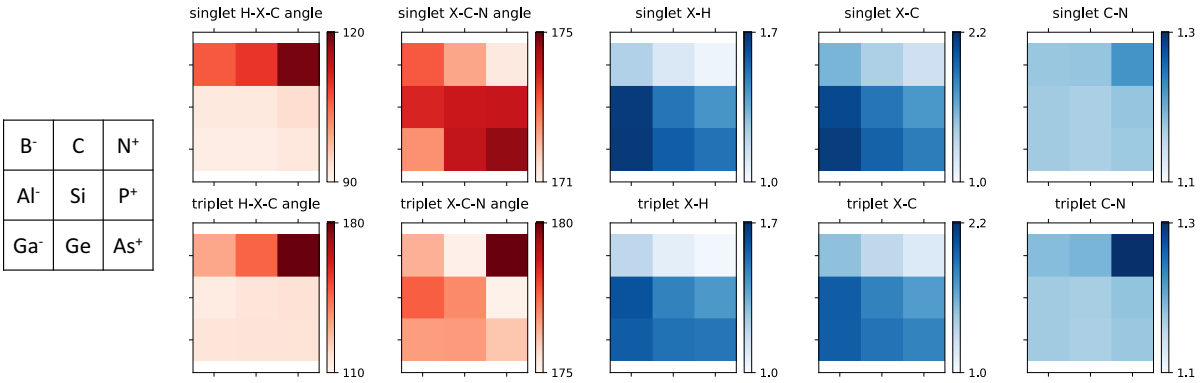

Figure 3.3: Pictorial comparison of HXCN geometric parameters. The three by three grid shown for each geometrical parameter represents the nine species HXCN studied, with X shown on the leftmost grid. The color bar beside each grid shows the range of values depicted. Angles are shown in degrees (°) and bond distances are shown in angstroms (Å). Increasing color intensity represents an increase in the value of the geometry parameter.

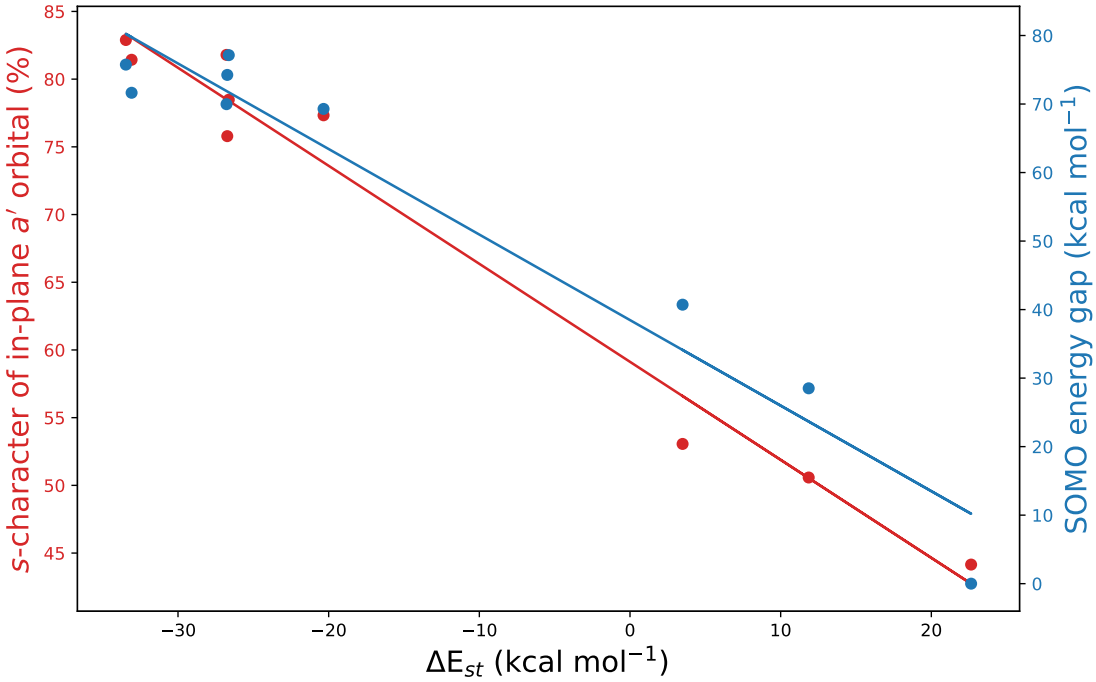

Figure 3.4: The correlation between HXCN ST gap and both the percentage of s-character in the in-plane a' lone pair orbital in the singlet state and the SOMO energy gap computed for the triplet state. s-character is obtained through NBO computations as described in the methods section. Orbital energies for the SOMO energy gap were computed at the RHF/QZ level of theory. R^2=0.975 for the percentage of s-character in the in-plane a' lone pair orbital vs. ST gap, R^2=0.945 for SOMO energy gap vs. ST gap

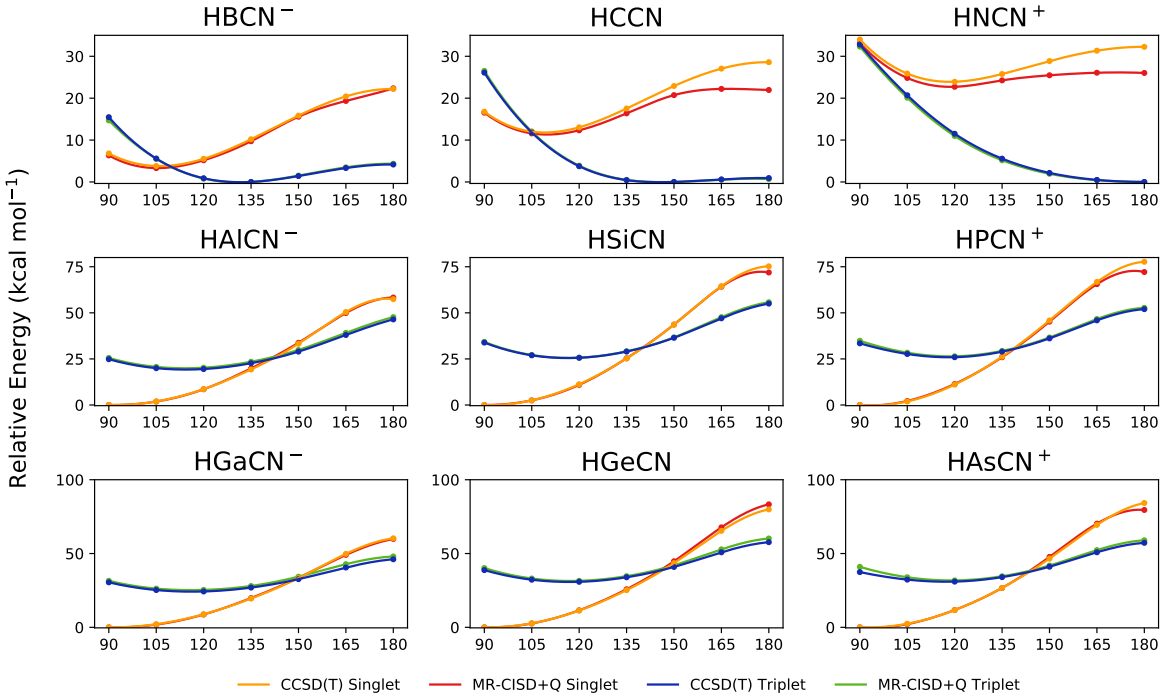

Figure 3.5: Relaxed scan of the H–X–C angles, where X is the central divalent atom. The energies were obtained by constrained optimizations at the CCSD(T) level of theory. MR-CISD+Q single point energies were computed following CASSCF (10o, 8e$^-$) constrained geometry optimizations. The basis sets used are described in the methods section. Relative energies (kcal mol^{-1}) of the singlet and triplet states at both CCSD(T) and MR-CISD+Q levels of theory are shown against the H–X–C angle in degrees. The reference energy for each level of theory is the lowest energy point out of both the singlet and triplet states for each species.

Barrier to Linearity

As mentioned in the introduction, the barrier to linearity for triplet HCCN is well studied both computationally and experimentally as evidence of its bent but quasilinear character. In Figure 3.5 we compare the barrier to linearity across the singlet and triplet states of each HXCN species, showing how the relative energies vary as the H–X–C angle increases from 90° to 180°. Both single reference and multireference methods are used, although for the most part, the relative energies between the singlet and triplet states of a given species does not change much between the two methods. The only significant difference is in the singlet state of HCCN and HNCN$^+$, in which MR-CISD+Q shows a lower energy gap relative to the triplet compared to CCSD(T) as the H–X–C angle approaches linearity.

We can make several observations:

- Compared to triplet HCCN, the rest of the HXCN species studied are less floppy near their ground state geometries, with triplet $HNCN^+$ being decidedly linear and the others decidedly bent.

- Moving down the column, the barrier to linearity increases dramatically for the singlet and somewhat less so for the triplet. A smaller H–X–C angle at the optimized geometry leads to a much larger barrier to linearity.

- The angle at which the state ordering is switched increases down the column. In the first row, the switch in state ordering happens at around 105° for $HBCN^-$ and HCCN and less than 90° for $HNCN^+$. Meanwhile, the switch happens at around 140° for the species in the second row and 150° for the species in the third row.

- When the H–X–C angle is at 180°, the triplet state is always lower than the singlet state while when H–X–C angle is at 90° the singlet is always lower than the triplet state (except for $HNCN^+$.). This is expected given that as a linear geometry is reached the energy gap between the a' and a'' orbitals on the divalent atoms decreases, and the triplet state is increasingly favored.

- While the ST gap in the optimized geometries varies wildly as discussed in previous sections, the gap between the states at a linear geometry falls within a narrower range between the species. Moreover, as the previous point mentions, the ST gaps are all positive because the triplet state is always lower than the singlet state. The first row species $HBCN^-$ (18.04 kcal mol^{-1}), HCCN (27.65 kcal mol^{-1}), and $HNCN^+$ (32.25 kcal mol^{-1}) display the largest gaps, while the second row species $HAlCN^-$ (10.96 kcal mol^{-1}), HSiCN (20.24 kcal mol^{-1}), $HPCN^+$ (25.70 kcal mol^{-1}) display the smallest gaps. The third row species $HGaCN^-$ (14.18 kcal mol^{-1}), HGeCN (22.22 kcal mol^{-1}), and $HAsCN^+$ (27.00 kcal mol^{-1}) fall in between the two. Across a row, the energy gap at the linear geometry increases on average from anions to neutral species to cations.

Comparing the ST gap of HXCN and XH_2

We also wanted to see how the substitution of a cyano (CN) group influences the ST gap of the HXCN compared to its parent derivative XH_2. Table 3.3 compares the focal-pointed ST gap of HXCN and XH_2 at the CCSD(T)/CBS level of theory using the optimized CCSD(T) geometries with either a TZ or QZ basis set. The difference between using a TZ or QZ geometry makes a difference of at most 0.09 kcal mol^{-1} in the ST gap for the HXCN species and 0.07 kcal mol^{-1} for the XH_2 species. Moving forward we will focus on the CCSD(T)/CBS gap at the QZ geometry.

The trends for the ST gap of the XH_2 species shown in Table 3.3 are similar to those obtained by Cramer and coworkers.[124] For BH_2^-, they predict a triplet ground state (ΔE_{ST} = 0.1 kcal mol^{-1}), while we predict a singlet ground state (ΔE_{ST} = –0.23 kcal mol^{-1}). For X=Si and onwards, our ST gaps are consistently 2–4 kcal mol^{-1} larger in magnitude compared to their MR-CISD+Q values. Cramer and coworkers give an excellent discussion regarding the periodic trends seen in the ST gaps of XH_2, which will be summarized here. In the first row, we start with BH_2^-, whose lowest singlet and triplet states are

Table 3.3: Singlet-Triplet Energy Differences for XH_2 and HXCN

Species	XH_2		HXCN	
X	TZ	QZ	TZ	QZ
B^-	−0.30	−0.23	3.47	3.49
C	8.95	8.99	11.81	11.86
N^+	29.14	29.18	22.66	22.64
Al^-	−14.64	−14.64	−20.43	−20.34
Si	−21.52	−21.51	−26.64	−26.63
P^+	−18.68	−18.65	−26.74	−26.73
Ga^-	−18.59	−18.59	−26.76	−26.78
Ge	−26.18	−26.20	−33.49	−33.46
As^+	−24.33	−24.28	−33.11	−33.08

Singlet-Triplet (ST) gaps defined as $\Delta E_{ST} = E_{\text{singlet}} - E_{\text{triplet}}$ and shown in kcal mol^{-1}. The ST gaps shown were extrapolated to CCSDT(Q)/CBS on a CCSD(T)/TZ (TZ columns) or CCSD(T)/QZ (QZ columns) geometry. The ST gaps for the $X=N^+$ species were extrapolated to the full CCSDTQ/CBS level of theory.

nearly degenerate. As we move across the row, the increase in nuclear charge leads to a contraction of all the orbitals and subsequently increased Coulombic repulsion. The triplet state is thus increasingly favored, because the promotion of an electron from the lone pair orbital on the divalent atom relieves this Coulombic repulsion. For later rows, this effect is not as important because the valence orbitals of higher quantum levels are more diffuse, and the need to relieve Coulombic repulsion is not as large. Instead, the dominant effect becomes the Coulombic attraction between electrons and the nucleus . Electrons in the in-plane lone pair orbital are closer to the nucleus than the out-of-plane lone pair orbital, and thus the singlet state is increasingly favored as nuclear charge (and the strength of Coulombic attraction) increases going across the row. The general trend for the ST gap in XH_2 holds when we replace H with CN, and the same periodic arguments apply. We will turn our attention now to the differences between XH_2 and HXCN.

The differences between the ST gap of HXCN and XH_2 are the most varied in the first row. The $X=B^-$ species is the only one that switches the state ordering between BH_2^- and $HBCN^-$, moving from a singlet ground state (albeit only 0.23 kcal mol^{-1} lower) to a triplet ground state. In the X=C species the ST gap also increases going from CH_2 to HCCN, showing that a triplet ground state is more favorable with the substitution of the cyano group. In the $X=N^+$ species, the ST gap actually decreases going from NH_2^+ to $HNCN^+$. Although the triplet is still the ground state in $HNCN^+$, its energy separation from

the singlet state is 6.5 kcal mol^{-1} less than in NH_2^+. For X=Al^-, Si, P^+ and X=Ga^-, Ge, As^+, the singlet ground state is maintained between XH_2 and HXCN, and the energy separation between the singlet and triplet states is consistently enlarged. The electronic ground state is further stabilized with the substitution of the cyano group.

Isodesmic Analysis

In order to gain insight into the origin of the differences between the ST gaps of HXCN and XH_2, we can isolate how the substitution of the cyano group separately affects the singlet and triplet states. To this end, we can study an isodesmic bond separation reaction originally introduced by Pople and coworkers[242] and used for carbenes by Schleyer and coworkers[113] and more recently by Eckhardt and Schreiner.[243] In an isodesmic reaction, the number of a given type of bond stays the same between products and reactants, but their connectivity changes. Equations 1 and 2 respectively give the bond separation reaction for the singlet and triplet states of the cyano-species:

$$H_3XCN + XH_2(\text{singlet}) \rightarrow XH_4 + HXCN(\text{singlet}) \tag{3.1}$$

$$H_3XCN + XH_2(\text{triplet}) \rightarrow XH_4 + HXCN(\text{triplet}) \tag{3.2}$$

The bond separation energy for these reactions measures the effect of replacing a hydrogen with a cyano group in XH_2. A negative bond separation energy indicates that the substitution of the cyano group in XH_2 provides a stabilizing interaction, while a positive bond separation energy indicates that the cyano group provides a destabilizing interaction.

Table 3.4 gives the bond separation energies for the singlet and triplet states. We note that for each species, the difference between the singlet and triplet bond separation energies mirrors the difference between the ST gaps of XH_2 and HXCN (Table 3.3). Thus the isodesmic analysis is a way to decompose the difference between the ST gaps of HXCN and XH_2 into separate contributions from the singlet and triplet states. There are three types of patterns of decomposition seen across the nine species. First, the bond separation energy is positive for the singlet state and negative for the triplet state. This pattern only applies for the case of X=B^-. Since the singlet and triplet states in BH_2^- are nearly degenerate, the stabilizing interaction provided by the cyano group for the triplet state results in a net preference for the triplet state in $HBCN^-$. Second, the bond separation energy is negative for the singlet state and positive for the triplet state. This case applies to both X=Al^-, Ga^- where the magnitude of the singlet state bond separation energy is less than the magnitude of the triplet state bond separation energy, and X=Si, Ge where the magnitude of the singlet state bond separation energy is more than the magnitude of the triplet state bond separation energy. Third, both states have a negative bond separation energy. In X=C, the bond separation energy is greater in magnitude in the triplet state than the singlet state while in X=N^+, P^+, and As^+, the opposite is true. This explains the decrease in preference for the triplet state in $HNCN^+$ as compared to NH_2^+.

Table 3.4: Isodesmic Bond Separation Energies

	B^-	C	N^+
singlet	0.89	−11.46	−64.53
triplet	−2.88	−14.32	−58.04
	Al^-	Si	P^+
singlet	−0.77	−3.76	−12.94
triplet	5.02	1.36	−4.88
	Ga^-	Ge	As^+
singlet	−2.82	−5.33	−11.52
triplet	5.35	1.97	−2.74

Bond separation energies are given in kcal mol^{-1} and are obtained with Equation 1 and 2 in the main text for the singlet and triplet states, respectively. Each species was optimized at the CCSD(T)/TZ level of theory. Focal point extrapolation and additional corrections were performed for all molecules in the same manner as HXCN, described in the methods section to give CCSDT(Q)/CBS energies for every species except $X=N^+$ (CCSDTQ/CBS energies) Table 3.3 shows that using CCSD(T)/TZ optmized geometries gives similar ST gaps as compared extrapolations completed using a CCSD(T)/QZ geometry.

NBO Analysis

NBO analysis can be used to help explain how interactions between X and CN give rise to the isodesmic bond separation energies seen in the previous section (Figure 3.6). In the singlet state, the natural bond order of X–C increases going across a row and decreases down a column while the bond order of C–N shows the opposite trend. The change in bond order transitioning from the neutral species to cations in every row is much more dramatic than going from anions to the neutral species. Weinhold's second order NBO analysis[238] points to two orbital interactions responsible for this trend (The values from the second order NBO analysis are shown in the Supporting Information): 1. Donation of electron density from the in-plane lone pair (LP) a' orbital of X into the anti-bonding orbitals of C–N (LP $\rightarrow$ C–N*); 2. Donation of electron density from the C–N bonding orbital into the out-of-plane a'' orbital of X (C–N $\rightarrow$ LP*). For this second interaction, the roles of the nitrogens in $HNCN^+$ are switched: the nitrogen of the donor C–N orbital is the divalent nitrogen cation and the acceptor orbital is on the terminal nitrogen. Both these interactions help to increase the bond order of X–C and decrease the bond order of C–N. Furthermore, the C–N $\rightarrow$ LP* interaction energy is significantly larger for the cations than the other species, and we can point to this as the reason why the bond order changes most dramatically for cations.

The increase in the natural bond order of X–C and a decrease in the natural bond order of C–N leads us to consider the importance of a resonance structure which has a X-C double bond and a C–N double bond (X=C=N), instead of a X–C single bond and a C–N triplet bond (X–C≡N). For example, $HNCN^+$ has almost equal bond orders for X–C (1.91) and C–N (2.09) and as expected, the X=C=N resonance structure in $HNCN^+$ has a significant contribution (20%) to the final geometry. The resonance structure is also has large contributions for the other cations in the series (12% for $HPCN^+$ and 10% for $HAsCN^+$). HCCN displays significant weight for the X=C=N resonance structure (11%), but the other neutral HXCN species do not (2.6% for HSiCN and 2.3% for HGeCN). The X=C=N resonance has little weight in the overall structure for all of the anions (2.4%, 0.5%, and 0.5% respectively for $HBCN^-$, $HAlCN^-$, and $HGaCN^-$).

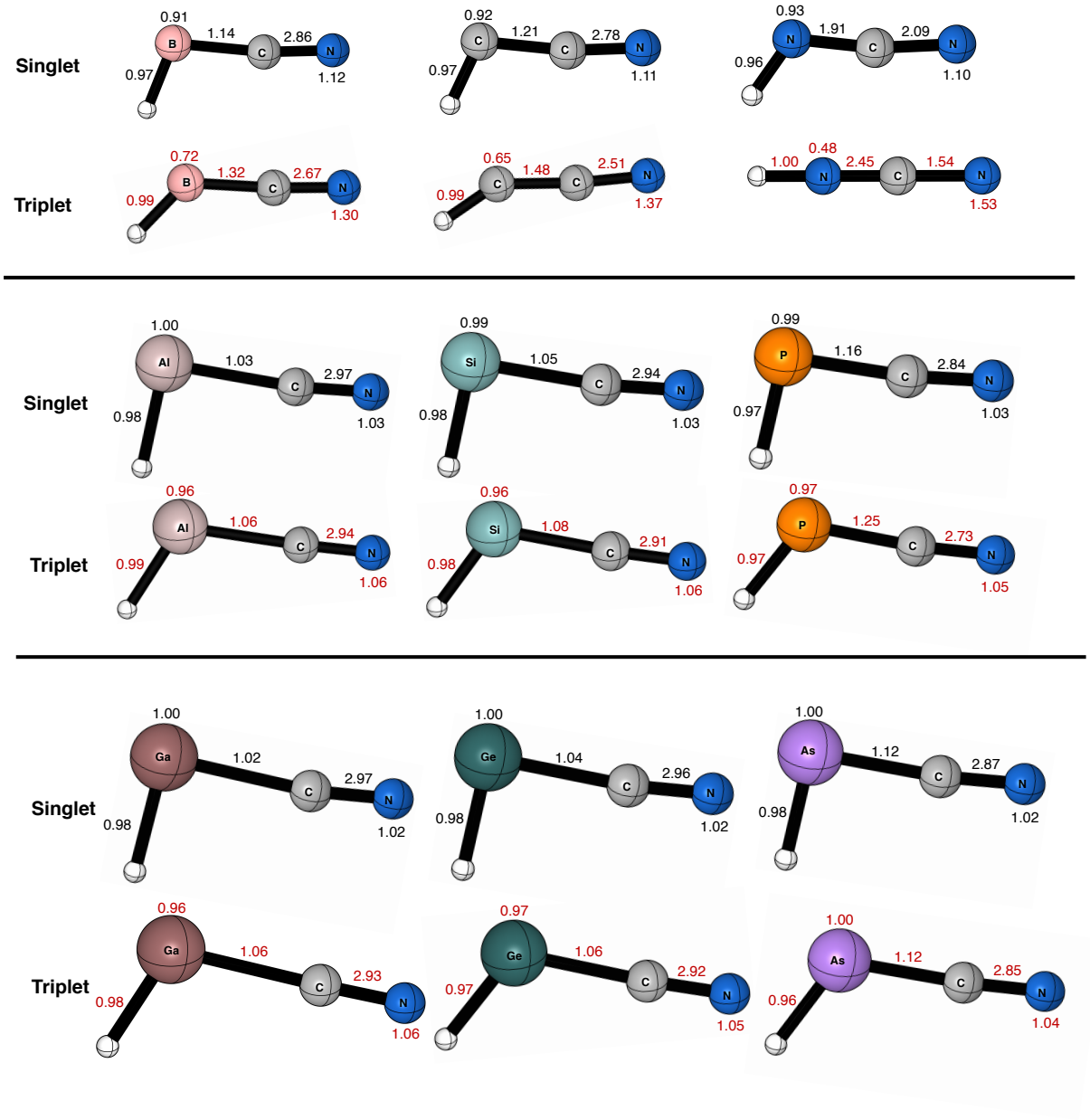

Figure 3.6: Natural bond order of the singlet (black) and triplet (red) states of the nine HXCN species. Values over atoms are lone pair occupancies. Values obtained with NBO computations performed at the B3LYP/DZ level of theory.

We conclude that in the singlet state, the cyano group provides stabilizing interactions by the pulling of electron density away from the a' in-plane LP orbital and consequently giving rise to additional resonance structures (X=C=N). The bond separation energies for the singlet states from the isodesmic analysis are well correlated with the X–C bond order and the weight of the resonance structures, not only in the trend but also the magnitude. The most dramatic example of this is seen going from HCCN to $HNCN^+$. The singlet isodesmic bond separation energy goes from −11.5 kcal mol^{-1} in HCCN to −64.5 kcal mol^{-1} in $HNCN^+$ as the X–C bond order goes from 1.21 to 1.91 and the weight of the X=C=N resonance structure increases from 11% to 20%.

Several major differences can be seen when comparing the bond orders of the singlet and triplet states. First, we notice a significant decrease on lone pair (LP) occupancy of X in the triplet states, compared to the singlet states for the first row. While the LP occupancy of the other rows decrease going from the singlet to triplet state, the difference is much smaller. Second, the X–C bond order is consistently larger in the triplet, while the C–N bond order is consistently smaller. However, the overall trend between species is the same.

In the triplet state, the in-plane a' and out-of-plane a'' orbitals are partially filled and have roles as both donor and acceptor. There are four important donor-acceptor orbital interactions in the triplet state:

1. The donation of electron density from the out-of-plane (oop) a'' orbital to the out-of-plane anti-bonding C–N orbital ($a'' \rightarrow$ oop-CN*);

2. The donation of electron density from the in-plane (ip) a' orbital into the in-plane anti-bonding C–N orbitals ($a' \rightarrow$ ip-CN*);

3. The donation of electron density from the out-of-plane bonding C–N orbital into the out-of-plane a'' orbital (oop-CN $\rightarrow a''$);

4. The donation from the in-plane bonding C–N orbitals into the in-plane a' orbital (ip-CN $\rightarrow a'$).

In the first row, donation of electron density from the a' and a'' orbitals to the C–N anti-bonding orbitals result in the relatively large bond orders as well as the negative isodesmic bond separation energy for $HBCN^-$. The additional back donation of electron density from the C–N bonding orbitals to the a' and a'' orbitals in HCCN is what gives it a larger bonding order and a more negative isodesmic bond separation energy compared to $HBCN^-$. For $HNCN^+$, an interesting phenomenon is observed. The divalent nitrogen (N1) forms a triple bond with the carbon, leaving the terminal nitrogen (N_4) with the unpaired electrons. The bond order of triplet $HNCN^+$ reflects this: the N1–C bond order (2.45) is larger than the C–N_4 bond order (1.54). The linear geometry allows for maximum orbital interactions and results in large donor-acceptor orbital interaction energies.

For the other two rows, the $a'' \rightarrow$ oop-CN* interaction energy is bigger than the $a' \rightarrow$ ip-CN* interaction energy, and increases across a row. The oop-CN $\rightarrow a''$ and ip-CN $\rightarrow a'$ interaction energies also increases across a row and are only significant in the cations $HPCN^+$ and $HAsCN^+$. The increase of these interactions across a row causes a subsequent increase in bond order. The isodesmic bond separation energies follow suit: triplet AlH_2^- and GaH_2^- is favored over triplet $HAlCN^-$ and $HGaCN^-$ by 5 kcal mol^{-1} , but triplet SiH_2 and GeH_2 is only favored over triplet HSiCN and HGeCN by 2 kcal mol^{-1}.

The additional orbital interactions in the cations cause the bond separation energy to switch signs to favor triplet $HPCN^+$ and $HAsCN^+$ over PH_2^+ and AsH_2^+ by 4.9 kcal mol^{-1} and 2.7 kcal mol^{-1} respectively.

In both the singlet and the triplet states, the NBO results for the anion and neutral species are more similar than the neutral and cation species. The overall increase of the strength of orbital interactions in the cation species could be due to the combination of the increase in nuclear charge and the removal of an electron for a formal positive charge. The interesting role of the terminal nitrogen atom in singlet and triplet $HNCN^+$ and the overall direction of electron density toward the X–C bond is probably related to the fact that nitrogen has a higher electronegativity than carbon.

3.5 Conclusions

We have presently performed by far the most comprehensive and reliable study to date on the lowest lying singlet and triplet states of the HXCN species and summarize several key observations:

1. $HBCN^-$, HCCN, and $HNCN^+$ have triplet ground states while the other molecules have singlet ground states. The first row shows the most dramatic change in ST gaps and geometrical parameters.

2. Orbital arguments used previously in the literature to explain the ground state preference of carbenes have been quantified. The singlet-triplet gaps of the nine HXCN species studied showed strong correlations both to the SOMO energy gaps in the triplet states and the percentages of s-character in the in-plane a′ orbital in the singlet state.

3. Compared to HCCN, there are larger barriers to linearity for the other HXCN species.

4. The singlet-triplet gaps of HXCN are almost always larger in magnitude than their parent XH_2 molecules. The $HNCN^+$ ST gap is smaller than the NH^+ ST gap. The ground state remains the same between HXCN and XH_2 for all pairs except X=B^-, switching from a singlet to a triplet from BH_2^- to $HBCN^-$.

5. Isodesmic and NBO analysis show that the differences in ST gaps between HXCN and XH_2 originate from the the strength of the interaction between the in-plane a′ orbital/out-of-plane a″ orbital on X and the bonding/anti-bonding C–N orbital.

6. $HNCN^+$ represents an outlier among the nine species. It has the most positive ST gap, and the most unusual singlet (biggest H–X–C angle) and triplet (completely linear) geometries. NBO analysis shows that electron density is pulled away from the C–N bond by the divalent nitrogen cation (N1). The situation in the triplet state is so extreme that the N1–C bond order is bigger than the C–N bond order, and the orbitals of nonbonding electrons are located at the terminal nitrogen atom instead of N1. Its peculiar behavior might stem from the fact that nitrogen is the most electronegative among the nine choices of X, and the only one that is more electronegative than carbon.

Chapter 4

A high level *ab initio* study on the isomerization of H2XY to HXYH (X, Y = O, S, Se, Te) [1]

[1] Abbott, B.Z.; Bralick, A. K.; Schaefer, H. F., to be submitted to *Phys. Chem. Chem. Phys.*

4.1 Abstract

Oxywater(H_2OO) is an intermediate in the oxidation of hydrogen peroxide (HOOH), and along with its relatives H_2SS, H_2SeSe, and H_2TeTe plays an important role in atmospheric and biochemical processes. In this work, we study the isomerization of H_2XY species to HXYH (X, Y = O, S, Se, Te) using highly rigorous *ab initio* methods. Geometries and harmonic frequencies were obtained using both a scalar relativistic X2C-1e-CCSD(T) approach and with non-relativistic CCSD(T) using ECPs on Se and Te. The two methods are compared for their ability to capture relativistic effects. A focal point approach was used to extrapolate electronic energies of CCSD(T)/aug-cc-pVTZ geometries to a CCSDT(Q)/CBS level of theory. Periodic trends in geometries and isomerization barriers were established and analyzed.

4.2 Introduction

The chalcogen elements are well known for their roles in biological and chemical processes. The smallest and simplest dichalcogen, H_2OO and its isomer HOOH acts as an intermediate in many combustion and atmospheric processes.[244–253] Diselenides are exploited for their applications in organic synthesis and pharmacology.[254,255] Disulfides are commonly found in proteins and enzymes.[256] Selenium-chalcogen bonds are of special interest due to their appearance in selenoproteins.[257] Understanding dichalcogen structures is an important step in understanding their physical and chemical properties. It is therefore of interest to gain insight into the mechanism of the isomerization of the simplest dichalcogen structure, H_2XY to HXYH.

Previous computational work has been done on several of the dichalcogen structures. The most abundant is for the isomerization of H_2OO to HOOH.[258–261] The rotational barrier of HXYH (X, Y = O, S, Se)[262] and HTeXH (X = O, S, Se, Te)[263,264] has been studied by several groups. The participation of S and Se containing species in several biologically relevant processes has been explored.[265–267] There have been a few studies that have focused on the periodic trends associated with the substitution of various chalcogens. In 2007, Kaur and coworkers computed relative energies of H_2X_2 (X = O, S, Se) at the G2MP2 level.[256] In 2013, Bickelhaupt and coworkers explored trends in electronegativity for HXXH and H_2XX structures at the ZORA-BP86/QZ4P level.[268] In 2017, Viana studied the HSeXH–H_2SeX isomerization for X = O, S, Se at the CCSD(T)//MP2 level of theory.[257]

In all these studies, we find a lack of balance between satisfactory relativistic treatment and highly rigorous energetics. There have been several studies that acknowledge and explore the relativistic effects of the dichalogen species in detail.[255,269–271] However, the proper relativistic treatment comes at the cost of rigorous theory for optimization and single point energies. It is therefore desirable to explore various ways to model the relativistic effects of the system while still obtaining high-level geometries and energies.

In the present study, the isomerization process of HXYH to H_2XY for X, Y = O, S, and Te is explored with several *ab initio* approaches to assess the validity of various relativistic treatments. Optimized geometries are reported at the CCSD(T)/aug-cc-pwCVXZ (X = T, Q) level of theory and focal point energies are reported at the CCSDT(Q)/CBS level of theory. Periodic trends are discussed.

4.3 Methods

Geometries and Vibrational Frequencies

Geometry optimizations and harmonic frequencies were obtained for the 10 HXYH, 16 H_2XY, and 16 transition state structures for X, Y = O, S, Se, Te. The scalar relativistic X2C-1e-CCSD(T) method was used. The X2C recontracted aug-cc-pVTZ basis set for H, the X2C recontracted aug-cc-pCVTZ basis set for O, S, and Se, the recontracted aug-cc-pwCVTZ basis set for Te. The X2C recontracted aug-cc-pCVTZ basis set was not available for Te. All electrons were correlated.

Additionally, optimized geometries and harmonic frequencies were obtained for HXYH and H_2XY structures using the CCSD(T)[213–215] level of theory along with the aug-cc-pVTZ basis set for H, aug-cc-pwCVTZ basis set for O and S, and the aug-cc-pwCVTZ-PP basis set for Se and Te. HXYH and H_2XY structures for X, Y = O, S, Se were also computed with the X2C recontracted aug-cc-pVQZ basis set for H and the X2C recontracted aug-cc-pCVQZ basis set for O, S, and Se. All electrons were correlated.

The Hartree–Fock density matrix elements, coupled cluster amplitudes, and lambda coefficients were converged to 10^{-9}. The RMS gradient for the geometry optimizations was converged to 10^{-8}. A restricted Hartree–Fock (RHF) reference was used. The electronic structure program CFOUR 2.0[222] was used to obtain all reported geometries and harmonic vibrational frequencies. IRC analysis was performed using Psi4 to confirm the transition state structures.

Energetics

Energetics were obtained via the focal point approach of Allen and coworkers.[225–228] This approach allows us to obtain isomerization energies at the CCSDT(Q)/CBS level of theory. Single point energies at the HF, MP2, CCSD, CCSD(T), CCSDT, and CCSDT(Q) levels of theory were performed using the CCSD(T)/aug-pwCVTZ optimized structures. CFOUR was used for all computations. HF energies were extrapolated using Feller's three-point formula[232] and correlation energies with Helgaker's two-point formula.[233] Additional corrections were added to the (ΔE_{ST}), including a harmonic zero-point vibrational energy correction (ΔE_{ZPVE}) and a scalar relativistic correction ($\Delta_{\mathrm{rel}} = E_{\mathrm{AE-CCSD(T)/SF-X2C-1e}} - E_{\mathrm{AE-CCSD(T)}}$), both with CFOUR 2.0. The uncontracted aug-cc-pwCVTZ basis set was used for the scalar relativistic corrections.

4.4 Results and Discussion

Geometries

Table 4.1 shows the geometries of the H_2XY species computed with three different treatments of relativistic effects. TZ geometries are performed at the CCSD(T) level of theory while TZ-X2C and QZ-X2C are performed with X2C-1e-CCSD(T). QZ-X2C geometries for species containing Te were not computed

due to compuational cost. For the X–Y bond, we see that X2C-1e-CCSD(T) with a TZ basis set gives slightly larger bonds lengths compared to CCSD(T). The difference between CCSD(T)/TZ and X2C-1e-CCSD(T)/TZ was no more than 0.007 angstroms. Interestingly, the species containing Se and Te showed the smallest difference between the bond lengths. The ECP treatment is comparable to the scalar relativistic treatment. In contrast, X2C-1e-CCSD(T) with a QZ basis set gives slightly smaller bond lengths compared to CCSD(T) The difference between CCSD(T)/TZ and X2C-1e-CCSD(T)/QZ was no more than 0.01 angstroms. For the H–X bond, the difference between the bond lengths from CCSD(T) and X2C-1e-CCSD(T) is at most 0.002 angstroms.

For the H–X–Y angle, X2C-1e-CCSD(T) with a TZ basis set gives slightly smaller angles compared to CCDS(T). The difference is the greatest for the H_2TeTe angle at 0.21 degrees. X2C-1e-CCSD(T) with a QZ basis mostly gives sightly larger angles compared to CCSD(T) out of the species computed. The biggest difference is 0.18 degrees for H_2OO. Thus, we see that the variations in the geometries from the methods compared are small, and the TZ level computations do not vary drastically from the QZ level computations.

Table 4.2 shows the geometries of the HXYH species at the X2C-1e-CCSD(T)/TZ level of theory. As expected, the bond lengths increase as the size of the atom increases. The H–O–X angles increase as the size of X increases. The H–O–Te angle is the biggest at 107.56 degrees. However, species containing X=S, Se, and Te but not oxygen do not show such a large angle. Instead, bond angle decreases with increasing atom size for X. For the X–Y–H bond angle, increasing atom sizes show a decreasing bond angle. The dihedral H–X–Y–H angle is largest for HOOH at 112.38 degrees, but then stays relatively constant at around 90 degrees for the other species.

Table 4.3 shows the geometries of the H_2XY species at the X2C-1e-CCSD(T)/TZ level of theory. For a fixed X, the X–Y bond length increases as Y increases in atomic number. In contrast, the H–X bond length decreases, but only slightly. The H–X–Y bond angle increases as Y increases in atomic number for X=O and Te. However, it decreases as Y increases in atomic number for X=S and Se. The H–X–Y–H dihedral angle increases as Y increases in atomic number for a fixed X.

Energies

Figure 4.1 shows the relative energies of the HXXH, transition state, and H_2XX. Compared to the HXXH structure, the isomerization energy to H_2XX decreases as we go from X = O to S to Se to Te. At the same time, the barriers also decrease. Figure 4.2 shows the relative energies of the HXYH, transition state, and H_2XY. Compared to HXYH, the isomerization energies to H_2XY decrease with increasing Y size. The barrier also decreases. In contrast, the isomerization of HXYH to H_2YX has higher barriers. Comparing the isomerization of HXYH to the H_2YX counterpart of H_2XY, H_2YX has a smaller isomerization energy when the size of X and Y are close, but a bigger isomerization energy when the difference in X and Y sizes increases. The transition states for H_2XY in cases where X is the same size or bigger than Y more closely resembles the H_2XY structure, while the transition states for H_2XY where X is smaller than Y more closely resembles the HXYH counterpart.

Table 4.1: A comparison of geometric parameters for H_2XY.

X	Y	$r_{X\text{-}Y}$			$r_{H\text{-}X}$			∠H-X-Y		
		TZ	TZ-X2C	QZ-X2C	TZ	TZ-X2C	QZ-X2C	TZ	TZ-X2C	QZ-X2C
O	O	1.540	1.540	1.532	0.968	0.969	0.966	100.37	100.28	100.55
O	S	1.912	1.916	1.905	0.966	0.967	0.965	104.19	104.04	104.15
O	Se	2.063	2.065	2.056	0.966	0.966	0.964	104.31	104.21	104.31
O	Te	2.269	2.268		0.965	0.965		106.22	106.14	
S	O	1.483	1.489	1.479	1.370	1.372	1.370	108.62	108.52	108.59
S	S	1.986	1.992	1.976	1.354	1.356	1.355	107.68	107.52	107.80
S	Se	2.165	2.172	2.156	1.349	1.351	1.350	106.60	106.42	106.68
S	Te	2.431	2.433		1.344	1.344		105.12	104.98	
Se	O	1.645	1.649	1.643	1.501	1.502	1.501	105.77	105.74	105.73
Se	S	2.117	2.121	2.109	1.484	1.486	1.485	105.44	105.41	105.55
Se	Se	2.279	2.284	2.271	1.478	1.480	1.479	104.81	104.76	104.92
Se	Te	2.530	2.532		1.472	1.474		103.80	103.81	
Te	O	1.811	1.815		1.687	1.688		103.67	103.48	
Te	S	2.260	2.263		1.677	1.678		104.42	104.29	
Te	Se	2.405	2.410		1.673	1.674		104.41	104.24	
Te	Te	2.638	2.640		1.668	1.669		104.24	104.03	

Bond lengths are in angstroms and angles are in degrees. TZ values are computed at CCSD(T) with the aug-cc-pVTZ basis set for H, aug-cc-pwCVTZ basis set for O and S, and aug-cc-pwCVTZ-PP basis set for Se and Te. TZ-X2C values are computed at X2C-1e-CCSD(T) with the X2C recontracted aug-cc-pVTZ basis set for H, the X2C recontracted aug-cc-pCVTZ basis set for O, S, and Se, and the recontracted aug-cc-pwCVTZ basis set for Te. The QZ-X2C values are computed at X2C-1e-CCSD(T) with the recontracted aug-cc-pCVQZ basis set for O, S, and Se.

Table 4.2: Geometric parameters for HXYH.

X	Y	$r_{X\text{-}Y}$	$r_{H\text{-}X}$	$r_{H\text{-}Y}$	∠H-X-Y	∠X-Y-H	∠H-X-Y-H
O	O	1.459	0.965	0.965	99.86	99.86	112.38
O	S	1.673	0.963	1.344	106.89	98.18	91.42
O	Se	1.812	0.963	1.469	106.07	96.38	90.27
O	Te	1.970	0.970	1.656	107.56	95.09	89.00
S	S	2.070	1.341	1.341	97.82	97.82	90.69
S	Se	2.209	1.340	1.466	97.09	96.27	90.40
S	Te	2.391	1.338	1.654	96.55	95.33	90.04
Se	Se	2.341	1.465	1.465	95.79	95.79	90.24
Se	Te	2.520	1.464	1.653	95.36	95.14	90.01
Te	Te	2.702	1.652	1.652	94.97	94.97	90.25

Bond lengths are in angstroms and angles are in degrees. Geometries are computed at the X2C-1e-CCSD(T) level of theory with the X2C recontracted aug-cc-pVTZ basis set for H, the X2C recontracted aug-cc-pCVTZ basis set for O, S, and Se, and the recontracted aug-cc-pwCVTZ basis set for Te.

4.5 Conclusions

HXYH and H_2XY structures for X, Y = O, S, Se, and Te were computed along with their interconversion transition states. Three relativistic treatments were compared, and no significant difference in the geometries were found. In general, the bond lengths of X–Y increased as X and Y increased in size, as expected. The H–X–Y angles for the HXYH species increased dramatically with X=O, but decreased slightly with an increase in the atomic number of Y for X=S, Se, and Te. The H–X–Y angles for the H_2XY species were all between 100 and 109 degrees. The isomerization energy from H_2XY to HXYH for a fixed X decreased with increasing Y atomic radii. The transition state structure more closely resembles H_2XY in cases where X is the same size or bigger than Y, and more closely resembles the HXYH structure in cases where X is smaller than Y.

Table 4.3: Geometric parameters for H_2XY.

X	Y	$r_{X\text{-}Y}$	$r_{H\text{-}X}$	∠H-X-Y	∠H-X-Y-H
O	O	1.540	0.969	100.28	109.11
O	S	1.916	0.967	104.04	111.51
O	Se	2.065	0.966	104.21	111.27
O	Te	2.268	0.965	106.14	112.62
S	O	1.489	1.372	108.52	93.64
S	S	1.992	1.356	107.52	95.64
S	Se	2.172	1.351	106.42	95.70
S	Te	2.433	1.344	104.98	95.70
Se	O	1.649	1.502	105.74	91.32
Se	S	2.121	1.486	105.41	92.63
Se	Se	2.284	1.480	104.76	92.84
Se	Te	2.532	1.474	103.81	93.07
Te	O	1.815	1.688	103.48	89.76
Te	S	2.263	1.678	104.29	90.51
Te	Se	2.410	1.674	104.24	90.71
Te	Te	2.640	1.669	104.03	91.05

Bond lengths are in angstroms and angles are in degrees. Geometries are computed at the X2C-1e-CCSD(T) level of theory with the X2C recontracted aug-cc-pVTZ basis set for H, the X2C recontracted aug-cc-pCVTZ basis set for O, S, and Se, and the recontracted aug-cc-pwCVTZ basis set for Te.

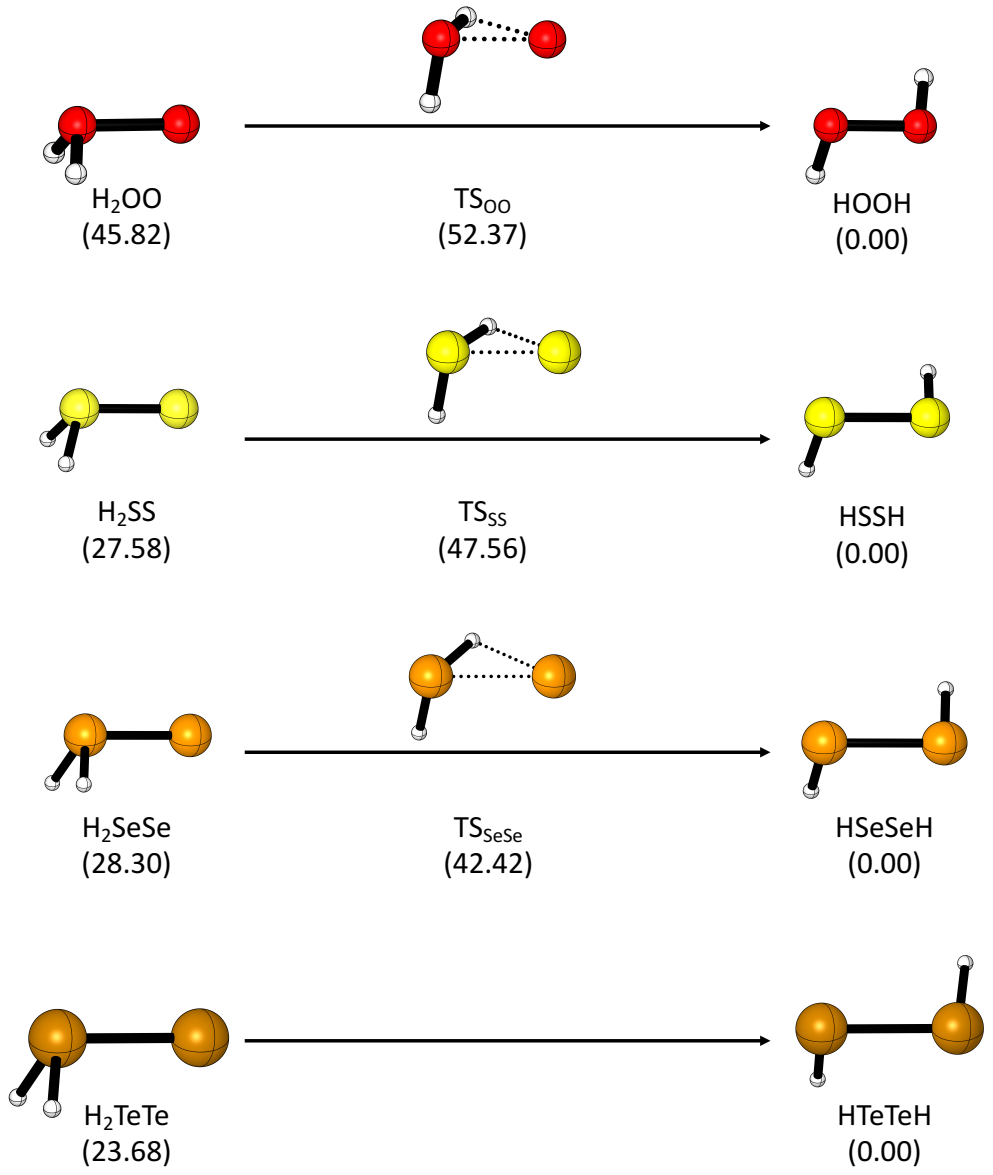

Figure 4.1: Relative energies of the XX species. Energies are computed at the X2C-1e-CCSD(T)/TZ level of theory.

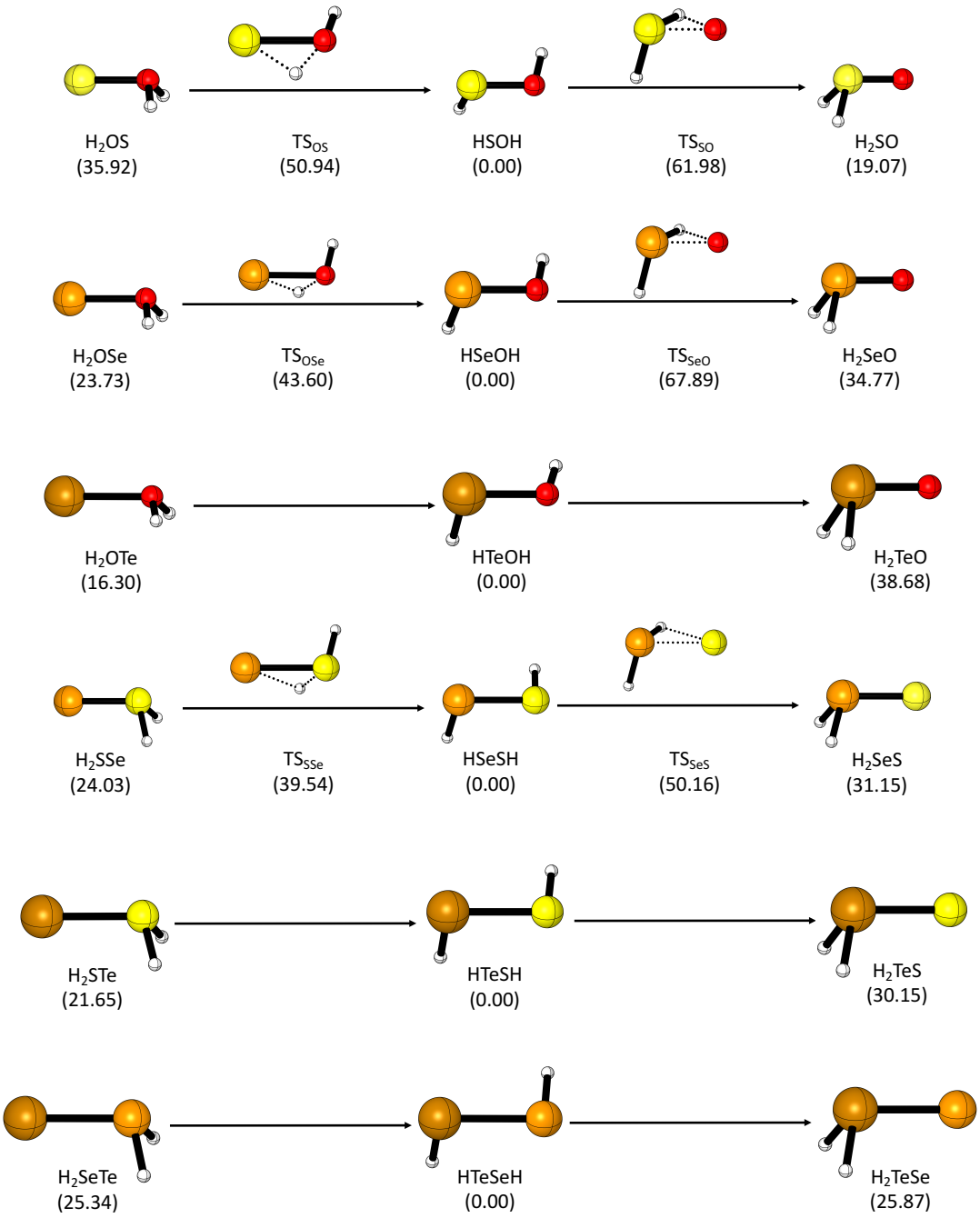

Figure 4.2: Relative energies of the XY species. Energies are computed at the X2C-1e-CCSD(T)/TZ level of theory.

Chapter 5

Janus: An Extensible Open-Source Software Package for Adaptive QM/MM Methods[1]

[1]Reproduced with permission from Zhang, B; Altarawy, D; Barnes, T; Turney, J. M.; Schaefer, H.F. *J. Chem. Theory Comput.* 2019, 15, 4362–4373, Copyright 2019 American Chemical Society. Download available via ACS Articles on Request at http://pubs.acs.org/articlesonrequest/AOR-SErZCdrpC9uCaID8v2xx.

5.1 Abstract

Adaptive quantum mechanics/molecular mechanics (QM/MM) approaches are able to treat systems with dynamic or non-localized active centers by allowing for on-the-fly reassignment of the QM region. Although these approaches have been in active development, the inaccessibility of current software has caused slow adoption and limited applications. JANUS seeks to remedy the limitations of current software by providing a free and open-source Python library for adaptive methods that is modular and extensible. Our software has implementations of many existing adaptive methods and a user-friendly input structure that removes the hindrance of complicated set-up procedures. A Python API is made available to customize JANUS's capabilities and implement novel adaptive approaches. JANUS currently interfaces with PSI4 and OPENMM, but its modular infrastructure enables easy extensibility to other molecular codes without major modifications to either code. The software is freely available at https://github.com/CCQC/janus. Our goal is that JANUS will serve as a user-driven platform for adaptive QM/MM methods.

5.2 Introduction

Background

The use of combined quantum mechanical and molecular mechanical (QM/MM) methods to treat problems otherwise intractable by either QM or MM approaches alone has long been established.[272,273] However, traditional QM/MM methods are only appropriate for systems with fixed QM regions. For systems with non-localized active centers, such as ion transport and solvent diffusion, an adaptive QM/MM approach that allows on-the-fly reassignment of the QM region during a molecular dynamics (MD) simulation should be used.[274,275] These adaptive methods have been in active development, and utilized mostly in studies on the solvation properties of ions[276–285] and select organic reactions.[286,287] Recent achievements, such as studying the exchange of molecules in protein binding sites,[288] the tracking of proton hopping in bulk water,[289,290] the determination of explicit solvation effects on nucleophilic addition to carbonyl carbons,[291] and proton transfer in a protein channel,[292] demonstrate the ability of adaptive QM/MM to address problems previously unreachable by traditional QM/MM.

The smoothing of energy and forces between steps in an MD simulation is a central problem in adaptive QM/MM. In order to have a dynamic QM region, an atom's designation to be treated as a QM or MM particle can change between one step of a MD simulation and the next. This may cause a discontinuity in the energy and forces between the two steps that must be alleviated. In most adaptive QM/MM algorithms, this smoothing is achieved by defining a buffer zone between the QM and MM regions; the particles that fall within the buffer zone will be referred to as buffer groups. Various QM/MM configurations are determined from a method-dependent partitioning of each atom in the buffer group as either a QM or MM particle. A smoothing function is then applied to interpolate the various QM/MM

partitions. Buffer groups effectively have dual QM and MM character, enabling smoothing between time steps.

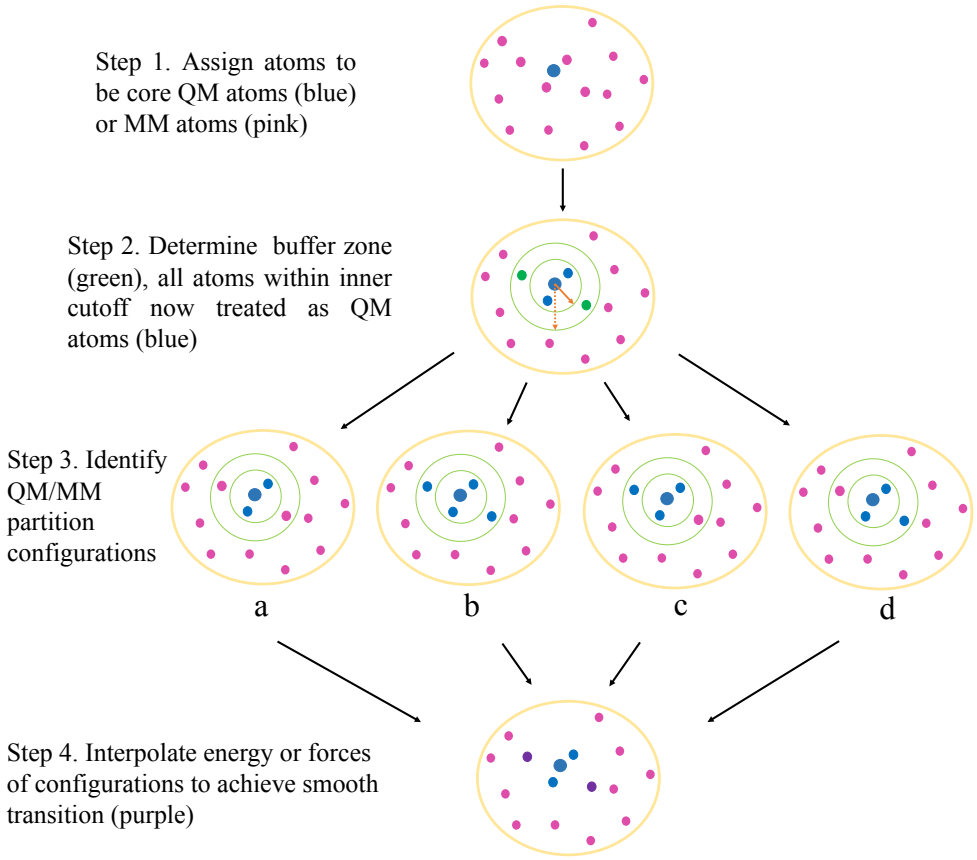

Figure 5.1: An example of a Permuted Adaptive Partitioning (PAP) scheme for a buffer zone with two groups.

Figure 5.1 shows an illustrative example of this process, and can be used to further understand the various parts of an adaptive QM/MM approach that need to be considered. Each step of Figure 5.1 is described below:

Step 1: Initial QM/MM designation. For a given system, atoms can be initially designated as a core QM atoms (blue) or a MM atom (pink). The core QM atoms will always be treated as QM throughout the course of the simulation, and are effectively what is being tracked or followed.

Step 2: Determination of buffer groups. Various approaches have been proposed for how to define the buffer zone; these include distance-,[276] number-,[293,294] and density-based[295] partitioning schemes. The distance-based scheme shown in Figure 5.1 is the most common, and involves the user defining

a R_{min} (solid orange arrow) and R_{max} (dotted orange arrow) calculated from the core QM atoms to serve as the lower and upper radial boundary for the buffer zone. Everything within R_{min} now become designated as a QM atom (blue), while everything between R_{min} and R_{max} is a buffer group (green).

Step 3: Identification of QM/MM configurations. The groups within the buffer zone can be treated as either QM or MM, and QM/MM configurations may be identified based on this method-dependent decision. Figure 5.1 shows the scheme used by the Permuted Adaptive Partitioning (PAP) method, in which all combinations of how to assign buffer group identities are considered. For the two buffer groups in question, we can designate them to be: *a*. both MM atoms; *b*. both QM atoms; *c*. 1 QM atom and 1 MM atom; and *d*. 1 MM atom and 1 QM atom. A separate QM/MM computation is performed on the four resulting QM/MM configurations.

Step 4: Interpolation scheme. The separate QM/MM configurations can then be combined as a linear combination of either energies or forces with weights determined by the smoothing function. The smoothing function depends on the relative distances of the buffer groups to the QM center; if a buffer group is close to the QM region then the QM/MM configuration in which that buffer group is treated with QM is given more weight than the QM/MM configuration in which it is treated with MM.

Algorithms with energy-interpolation [e.g., ONIOM-XS,[296] DAS,[297] and the adaptive partitioning (AP) family of methods[289,298,299]] conserve energy and momentum for the most part, but the presence of the smoothing function in the energy expression gives rise to extraneous forces due to the gradient of the smoothing function. Algorithms with force-interpolation (e.g., Buffered Force,[300] Hot-Spot,[278] and Size-Consistent Multipartitioning[301]) often do not conserve energy or have a meaningful energy expression, but there are no extraneous forces.[274] A number of reviews has been written on adaptive QM/MM methods, and we refer the interested reader to these for a more detailed discussion of these approaches.[274,275,302,303]

Current limitations to adaptive QM/MM

The development of new adaptive QM/MM methods is rapidly gaining traction, as more than half of all available methods were developed within the last six years.[275] Table 5.1 shows current methods as well as the software packages in which they are available. Despite this progress, the number of published studies using adaptive QM/MM methods is far fewer than one would expect, considering the number of different algorithms available. There are many untapped areas of potential research that could benefit from the potential insights gained with an adaptive approach. The primary cause for the lack of usage stems from the limitations of existing software,[275,303] which include:

1. Restrictive licensing. Most current software packages that have adaptive QM/MM capabilities offer restrictive licensing and/or is not free to use. Moreover, the QM or MM codes with which QM/MM software packages interface may also have restrictive licensing and/or are not free to use. This causes a barrier to researchers.

Table 5.1: Time evolution of adaptive approaches and their availability in various software packages

Year[a]	Method	Software Package
1996	Hot-Spot[276, 278, 279, 304]	FlexMD[305, 306],QMMM[307], Gaussian[308],Janus
2002	ONIOM-XS[296]	FlexMD, QMMM, Janus
2007	Permuted Adaptive Partitioning (PAP)[288, 289, 298, 299, 309]	QMMM, Janus
2007	Sorted Adaptive Partitioning (SAP)[288, 289, 298, 299]	QMMM, Janus
2009	Difference-based Adaptive Solvation (DAS)[297]	FlexMD, Janus
2012	Buffered Force (BF)[300, 310]	Flex-MD, CP2K[311], Amber[312], Janus
2012	Number-Adaptive[293]	AG-IF[313]
2014	Density-Based Adaptive (DBA)[295, 314]	Yoink[315]
2014	Size-Consistent Multipartitioning (SCMP)[301, 316],	GROMACS[317], *b*
2015	Time-Adaptive[318]	CPMD[319]
2016	Hamiltonian Adaptive Many-Body Correction (HAMBC) [320]	FlexMD, *b*
2017	Scaled Interaction Single Partition Adaptive (SISPA)[321]	pDynamo[322]

a. Methods are dated by the year in which they were first published, and subsequent modifications are cited. *b*. Method will be implemented in Janus in the near future.

2. Lack of comprehensive method selection Some codes used for traditional QM/MM, such as ChemShell,[323,324] do not have adaptive QM/MM capabilities; others fail to offer a comprehensive selection of all available methods found in Table 5.1. Comparisons between selected methods require the use of different packages, which may all have their own approach to implementation. Reproducibility of results using various codes is thus difficult to achieve, and the advantages and disadvantages of energy or force-based approaches are still debated.

3. Developer-focused software. Most implementations are developer-focused, requiring additional programming skills (FLEX-MD,[305,306] PDYNAMO[322]) or complicated system set-ups (QMMM[307]). Such need of this prior knowledge is a significant hinderance to users. In addition, the actual implementation details are not always made clear by developers. As a result, existing methods have not been used widely enough to test their robustness and performance, and newer methods have seen little application outside of benchmarking on small systems.[275,285,320] For example, the first application of the SCMP method was published just this year[285] and the HAMBC method has not been used outside of a benchmarking context. Increased usage will allow for assessment of current methods and lead to the development of a standard adaptive QM/MM methodology.

The inaccessibility of adaptive methods has been the primary barrier to their adoption; as such, there have been repeated calls in the literature for user-friendly adaptive QM/MM software.[275,302,303] The goals of such software would be as follows:

1. open-source code that interfaces with other open-source packages for easy access and community contribution;

2. user focused code with detailed documentation and tutorials for easy setup, with no prior programming knowledge required;

3. importable capabilities for flexible use by developers and advanced users;

4. the availability of most or all adaptive methods in one place for maximum choice and testing.

Several open-source programs exist for adaptive methods. The pDynamo program of Field[322] is a Python-based library for general QM/MM which can be used to implement adaptive methods. However, knowledge of the Python programming language is needed to run this program and adaptive methods are not readily available. Q-reg[325] is a library written in C++ designed for running adaptive QM/MM specifically for electrochemistry. Standard adaptive methods are not implemented and would require extensive programming knowledge on the user's part. Furthermore, neither pDynamo nor Q-reg currently interface with open-source codes for QM and MM computations.

Herein, we introduce Janus, an open-source Python library that seeks to remedy the limitations of current programs by providing a unified platform for adaptive methods. (The name "Janus" is a reference to the Roman god of transition and duality, and has been used in the literature to describe QM/MM boundary atoms.[326]) Janus lowers the barrier to using adaptive QM/MM methods by eliminating prior programing knowledge and providing a simple set-up scheme. For more advanced users and developers, the full set of the software's capabilities is available through a Python API to provide flexibility in usage. The modular design makes it possible for easy expansion of the code: only minimum modification to the code is required for implementing new methods and adding new interfaces. The key features of Janus, as well as some illustrative benchmarking results are discussed in the following sections.

5.3 Software Features

Ease of use

Janus does not require programming knowledge but instead supports input file submission through a simple command line interface (CLI). Figure 5.2 shows an example of a minimal input file that runs an adaptive QM/MM/MD simulation. The input has a YAML format with separate sections that specify the job instructions ("`system`"), any adaptive QM/MM specific keywords ("`aqmmm`"), any MD simulation specific keywords ("`md`"), as well as any parameters specifically related to the QM ("`hl`") or MM ("`ll`") computations. Although there are many potential keywords for each section, there are sensible defaults for nearly all of them. Thus, the user only has to change the keywords they desire by including them in the appropriate section, resulting in a very simple input. Instructions on installation, how to set up an input file and run the program, along with the keywords of each section and their descriptions are provided in the manual (https://ccqc.github.io/janus/) to ensure that the learning barrier is as low as possible. For more advanced users, Janus is available as an importable Python library.

Janus offers a comprehensive selection of methods (see Table 5.1) for the user to choose from. Because the different types of adaptive QM/MM approaches can all be run in the same place, the methods can

```
system:
    system_info: water.pdb
    run_aqmmm: True
    run_md: True
    aqmmm_scheme: ONIOM-XS
    ll_program: OpenMM
    hl_program: Psi4

aqmmm:
    Rmin: 3.5
    Rmax: 4.0
    qm_center: [0,1,2]

md:
    start_qmmm: 30000
    end_qmmm: 70000
    md_steps: 70000
    md_ensemble: NVE
    step_size: 0.5
ll:
    rigid_water: False
    fric_coeff: 91
    temp: 298
hl:
    method: scf
    basis: cc-pvdz
```

Figure 5.2: An example of a minimal input file for running JANUS. The presented input tells the software to run adaptive QM/MM ("`run_aqmmm`") with an ONIOM-XS scheme ("`aqmmm_scheme`") , using PSI4 for the QM computations ("`hl_program`") and OPENMM for the MM computations ("`ll_program`") and MD time-step integration ("`run_md`"). The QM center ("`qm_center`") is defined as the atoms with indices 0, 1, and 2. 30000 MD steps with pure MM will run, after which the adaptive QM/MM computation will start ("`start_qmmm`"). 40000 steps will be taken with an adaptive QM/MM computation at every step ("`end_qmmm`"), resulting a total of 70000 MD steps ("`md_steps`") . The microcanonical (NVE) ensemble ("`md_ensemble`") is specified. Additional MM ("`ll`") and QM ("`hl`") parameters are also given. No reinitialization is required.

be compared on equal footing. This will allow for consistent benchmarking across different approaches, as well as easy comparisons between them to establish general adaptive procedures. Information on each method is provided in the manual, so the user can make an informed choice that is appropriate for their specific system.

Rapid testing using an API

In addition to input file submission, JANUS provides its own Python API as an alternative way to interact with the software, which is useful both to developers and general users. For developers, the API makes JANUS's capabilities both accessible and customizable. Many of JANUS's functionalities can be used independently as a starting point for new approaches. In addition, modifications to a currently existing method in JANUS can quickly be tested.

The following section of code is an example of how a general user might use the API:

```
from janus import qm_wrapper, mm_wrapper, qmmm

# instantiate a Psi4Wrapper object as the high level wrapper
hl_wrapper = qm_wrapper.Psi4Wrapper()

# instantiate an OpenMMWrapper object as the low level wrapper
ll_wrapper = mm_wrapper.OpenMMWrapper(sys_info="water.pdb")

# instantiate a Permuted Adaptive Partitioning (PAP) object,
# varying Rmin and Rmax (Angstroms)
p1 = qmmm.PAP(hl_wrapper, ll_wrapper, sys_info="water.pdb", Rmin=5.0, Rmax=5.5)
p2 = qmmm.PAP(hl_wrapper, ll_wrapper, sys_info="water.pdb", Rmin=5.0, Rmax=6.0)

# partition the QM and buffer zone atoms
p1.find_buffer_zone()
p2.find_buffer_zone()

# find QM/MM configurations that arise from buffer zone partitioning
p1.find_configurations()
p2.find_configurations()

#print Rmin, Rmax, number of qm groups, buffer groups, and QM/MM configurations
print(p1.Rmin, p1.Rmax, p1.n_qm_groups, p1.n_buffer_groups, p1.n_configs)
print(p2.Rmin, p2.Rmax, p2.n_qm_groups, p2.n_buffer_groups, p2.n_configs)
```

In this case, the API is used as a way to test how different values of `Rmin` and `Rmax` (using a distance partitioning scheme) affect the number of buffer groups that are designated and the number of QM/MM configurations. For a given system, it might not be immediately clear to the user what are reasonable `Rmin` and `Rmax` values to set. The value of `Rmin` and `Rmax` directly affect the number of atoms in the QM region and buffer zone, and thus also determine the number of QM/MM configurations that arise (see Figure 5.1). If `Rmin` is too large, the QM region might be too large and not feasible for some QM methods. If the distance between `Rmin` and `Rmax` is too large, there might be too many buffer groups and thus too many QM/MM configurations. Therefore, it is valuable to easily test different values of `Rmin` and `Rmax` to determine what is appropriate for a specific system before starting a longer simulation job. With just a little programming knowledge, a user can take advantage of the API to run jobs more efficiently.

Minimum interface overhead with external software

Major disadvantages of current interface-based QM/MM packages such as QMMM,[307] PDYNAMO,[322] and FLEX-MD,[305,306] include the need to reinitialize the QM and MM codes at each time-step, as well as the number of file transfers needed for communication between different software. JANUS currently

interfaces with OpenMM[327,328] and Psi4,[329,330] both of which have their own application program interface (API). By calling Psi4 and OpenMM's functionalities through APIs, Janus minimizes the reinitialization of external code and file-transfer mechanisms to achieve more efficient implementations of workflow. In sample computations performed, only a trivial amount of the time is spent as overhead ($< 2\%$). The software also uses MDTraj[331] to assist in intermediate trajectory storing and manipulation. Although not all molecular software packages have an API, there has been an increasing push to develop APIs for established software (e.g. Amber,[312] NWChem,[332] GROMACS[317]), which Janus can take advantage of. Furthermore, initiatives from the Molecular Software Sciences Institute (MolSSI), such as the QCEngine package and the MolSSI Driver Interface Project, are encouraging API based interfaces for a variety of molecular software packages.[333,334] These developments are expected to provide a way for Janus to interface with more codes without the use of a file-based communication mechanism.

Easy software expansion through modular design

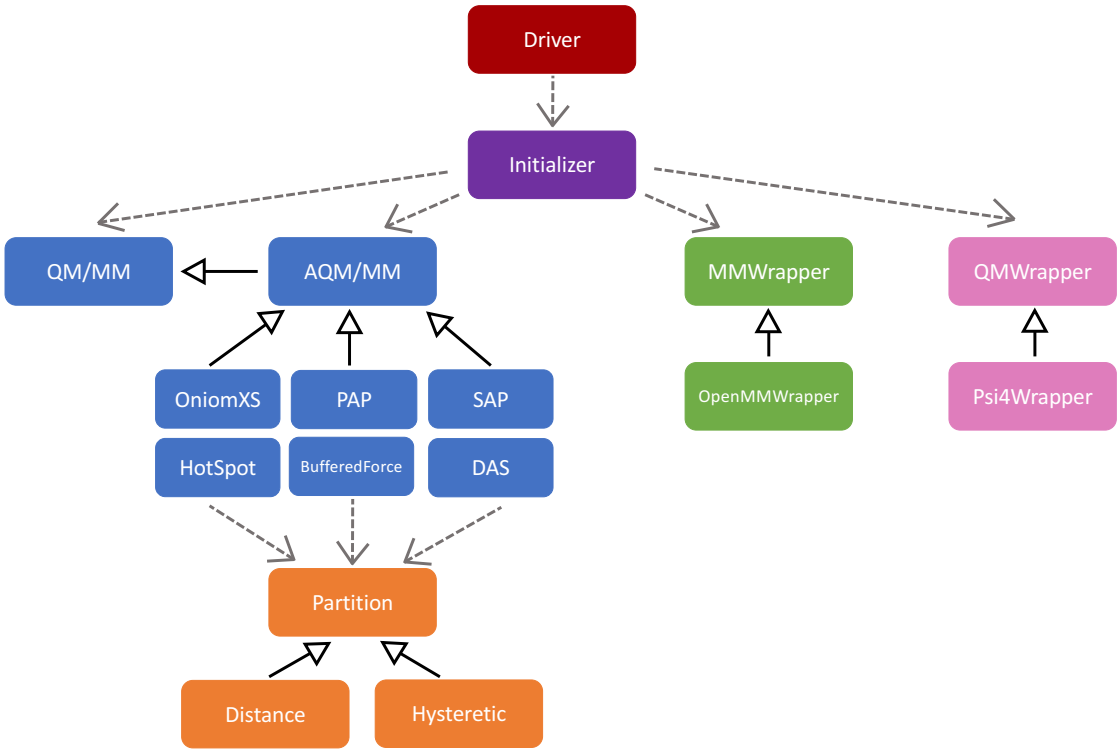

Figure 5.3: An overview of the class structure of Janus. Dashed gray arrows from class *A* to *B* indicate that *A* calls functions from *B*. Solid black arrows from class *A* to *B* indicates that *A* is a subclass of *B* and inherits from *B*.

Janus is designed with a modular approach to allow for easy method implementation and package expansion. Figure 5.3 shows the relationship between the four main modules in Janus. The MM

wrapper module (green) contains the interface to any molecular mechanics code, while the QM wrapper module (pink) contains the interface to any quantum mechanics code. As Figure 5.3 shows, each specific QM or MM software will be interfaced through its own subclass within the module. The QM/MM module (blue) includes the QMMM class as well as the AQMMM class, and contains all functionalities for traditional and adaptive QM/MM. Each adaptive method is a separate subclass that has its own set of method-dependent functions in addition to the functions inherited from AQMMM. The Partition module contains schemes for defining the buffer zone, as described in Step 2 of Figure 5.1. Currently, two schemes are implemented: the distance-based scheme first pioneered by Rode and coworkers[276] and the hysteretic scheme of Bernstein and coworkers.[300] Distance partitioning has been described previously in Figure 5.1. Hysteretic partitioning still involves defining a radial boundary for a QM and buffer region, but contains an additional set of boundaries (Figure 5.4). As the name suggests, the scheme uses information from the previous step to temper drastic changes in the number of QM and buffer atoms during a simulation.

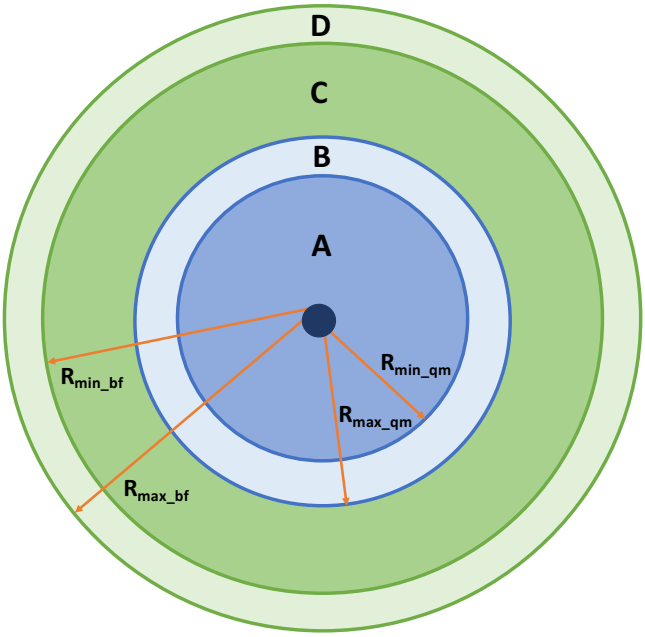

Figure 5.4: The hysteretic partitioning scheme. All molecules in region A are designated as QM molecules. Molecules in region B are designated as QM molecules if they were QM molecules in the previous time-step, and buffer atoms otherwise. Molecules in region C are designated as buffer atoms. Molecules in region D are designated as buffer atoms if they were buffer atoms in the previous time-step, and MM molecules otherwise.

The independence of separate modules allows for great flexibility both in application and implementation, and makes JANUS a great tool for testing new adaptive QM/MM approaches. In order to add a new

adaptive method, add a new partitioning method, or to interface to a new package, one simply has to create a subclass that takes advantage of the existing infrastructure without changing the rest of the code. This allows quick implementation of currently existing methods and relatively easy package expansion with other software. In addition, the separation of the Partition module from the QM/MM module allows users to test different combinations of partitioning schemes and adaptive methods to develop new approaches. For example, adaptive approaches such as ONIOM-XS[296] and the adaptive partitioning family (PAP, SAP)[298] are traditionally implemented with distance partitioning. However, with the modular implementation of different partitioning schemes one can easily test the use of hysteretic partitioning in these methods. Thus, JANUS can be used both to develop new methods as well as test untried combinations of existing approaches.

5.4 Sample Applications

A primary application for adaptive QM/MM methods is to study the coordination number of a system in explicit solvent. A radial distribution function (RDF) is often used to quantify this property, and gives the probability of finding a molecule within a certain distance of another molecule. Peaks in the RDF correspond to the solvation shells around the center system.

To demonstrate this usage in JANUS, we performed test simulations on a cluster of 1099 water molecules. The flexible TIP3P[335] forcefield was used. No cutoff was used for nonbonded interactions. In all cases, the water box was equilibrated for 25 ps in the canonical (NVT) ensemble using a Langevin integrator at a temperature of 298 K, a friction coefficient of 1 ps^{-1}, and a step size of 0.5 fs. A production run in the canonical (NVT) ensemble was then performed. The pure MD simulation ran for 30 ps with a step size of 0.5 fs. For the QM/MM and adaptive QM/MM runs, an MD simulation using molecular mechanics forces only was run for 25 ps in the NVT ensemble before starting the QM/MM or adaptive QM/MM portion for 5 ps. For the QM/MM run, the QM region was set to be one water molecule. For the adaptive schemes, the same water molecule from QM/MM was set to be the QM center. Mechanical embedding was used. While electrostatic embedding is available for traditional QM/MM, electrostatic embedding for adaptive QM/MM is not yet implemented. The QM computations were performed using Hartree–Fock theory (using a RHF reference), along with either the 6-31+G*[336] or Dunning's aug-cc-pVDZ[337] basis set. Other parameters used defaults set in JANUS, and can be found on the JANUS website (https://ccqc.github.io/janus/). Distance partitioning was used for adaptive test runs.

All computations were run using JANUS, calling PSI4 for the QM computations and OPENMM for the MM computations and time-step integration. Three separate simulations were run with the conditions above and averaged to obtain the RDF. RDFs were generated with PYTRAJ[338,339] and used the distance between the central water molecule and the other water molecules. Figure 5.5 shows the Oc–O, Oc–H, and Hc–H RDF of the water box averaged over the last 5 ps of the NVT simulation of three separate simulations for each test case. We demonstrate that JANUS not only runs adaptive QM/MM but can also perform traditional QM/MM and MD simulations to use for comparison purposes.

The results obtained are qualitatively in line with experiment and other computational RDFs previously published.[340] We note that for adaptive runs, a keyword change in the input is all that is necessary to redefine the buffer zone or QM approach.

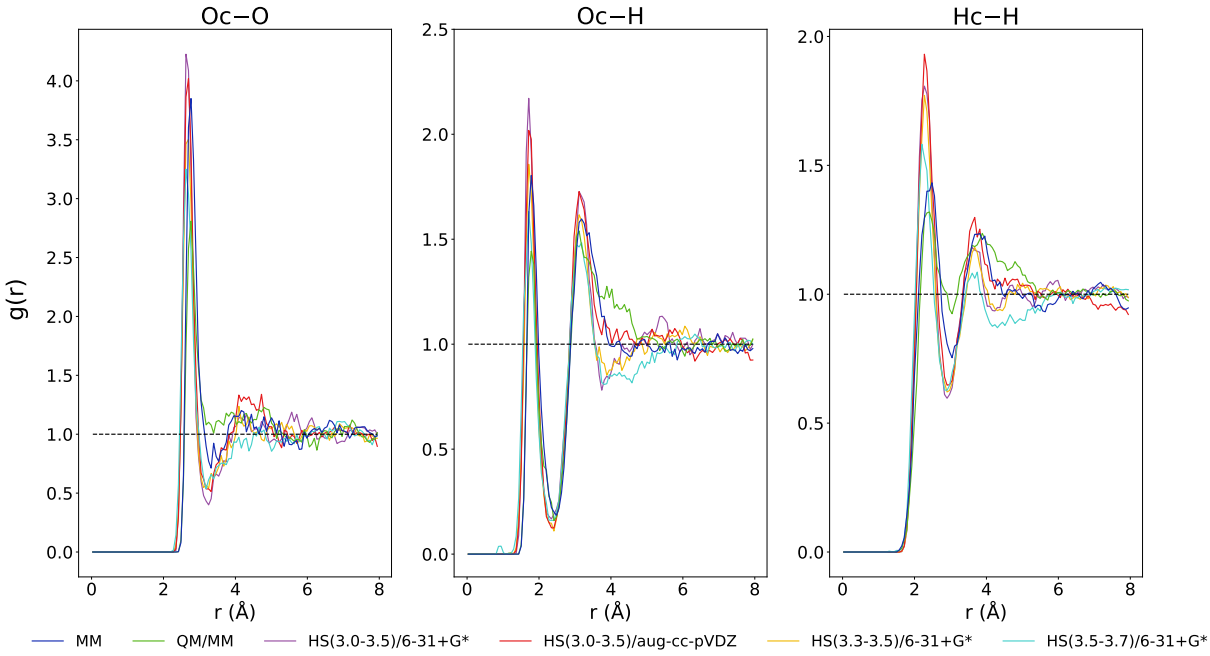

Figure 5.5: Radial distribution function for water. Oc–O was computed with the oxygen of the central water vs. all other water oxygens. Oc–H was computed with the oxygen of the central water vs. all other water hydrogens. Hc–H was computed with the hydrogens of the central water vs. all other water hydrogens. The blue curve corresponds to a pure MM simulation. The green curve corresponds to a non-adaptive QM/MM simulation. The purple curve corresponds to a Hot-Spot adaptive simulation with the buffer zone defined as R_{min} = 3.0 Å and R_{max} = 3.5 Å, and using the 6-31+G* basis set for QM computations. The red curve corresponds to a Hot-Spot adaptive simulation with the buffer zone defined as R_{min} = 3.0 Å and R_{max} = 3.5 Å, and using the aug-cc-pVDZ basis set. The yellow curve corresponds to a Hot-Spot adaptive simulation with the buffer zone defined as R_{min} = 3.3 Å and R_{max} = 3.5 Å, and using the 6-31+G* basis set. The cyan curve corresponds to a Hot-Spot adaptive simulation with the buffer zone defined as R_{min} = 3.5 Å and R_{max} = 3.7 Å, and using the 6-31+G* basis set.

Figure 5.6 shows how the QM region is changing for a 10 ps NVT simulation of the Hot-Spot adaptive approach with the SCF/6-31+G* QM method. The top graph shows the number of water molecules in a QM region (including buffer zone) with a radius of 3.5 Å. The graph shows the number of water molecules in a QM region (including buffer zone) with a radius of 3.7 Å. As the figure shows, throughout the simulation the number of waters in the smaller QM region varies from four to ten, while the number of molecules in the bigger QM region varies from four to twelve. Thus we see the adaptive approach correctly capture the movement of the water cluster and update the QM region as needed.

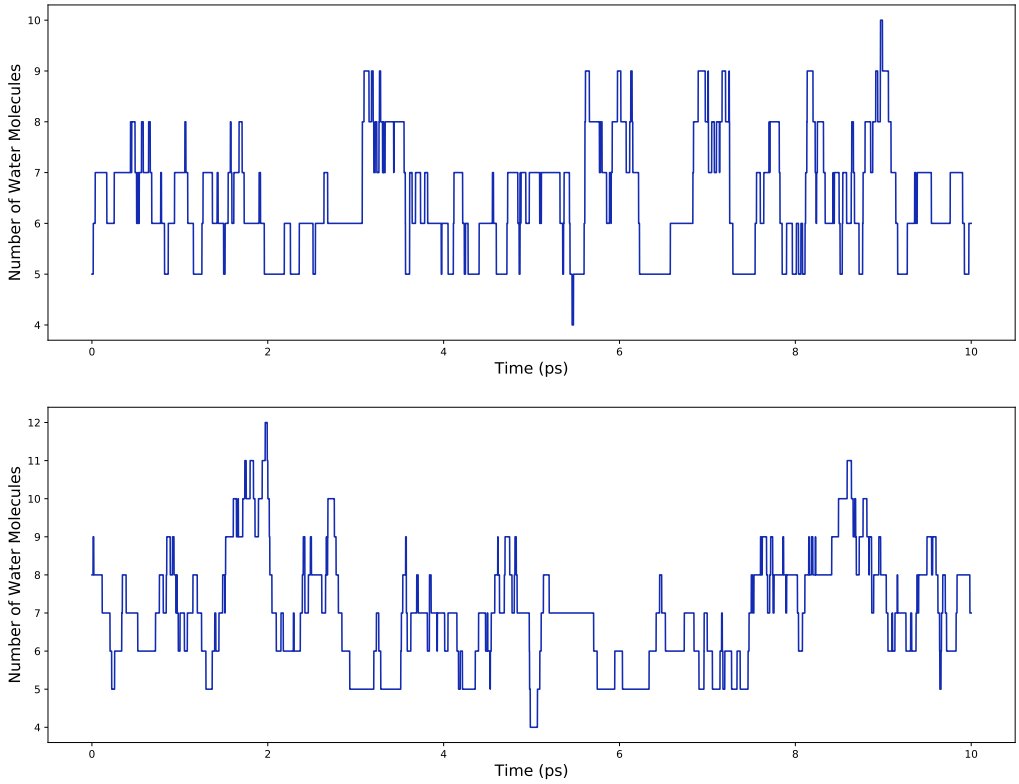

Figure 5.6: Number of water molecules in the QM region. The top corresponds to a Hot-Spot adaptive simulation with the buffer zone defined as R_{min} = 3.3 Å and R_{max} = 3.5 Å, and using the 6-31+G* basis set for QM computations. The bottom corresponds to a Hot-Spot adaptive simulation with the buffer zone defined as R_{min} = 3.5 Å and R_{max} = 3.7 Å, and using the 6-31+G* basis set for QM computations.

We also used JANUS to study the structure and solvation of N-methylacetamide (NMA) in a water cluster of 576 water molecules. NMA is the simplest model for a peptide bond and as a result has been subject to numerous experimental and theoretical studies.[341–353] The flexible TIP3P[335] forcefield was used for the water molecules and the Amber ff99SB[354,355] protein forcefield was used for NMA. No cutoff was used for nonbonded interactions. The system was equilibrated for 25 ps in the canonical (NVT) ensemble using a Langevin integrator at a temperature of 300 K, a friction coefficient of 1 ps^{-1}, and a step size of 0.5 fs. A production run in the microcanonical (NVE) ensemble was then performed for 35 ps total. The use of QM/MM forces started after 15 ps. The QM/MM run treated NMA with QM and the rest with MM. Mechanical embedding was used. The QM portion was performed using both Hartree–Fock

theory and density-fitted second-order Møller-Plesset (MP2) theory with an RHF reference, along with the STO-3G[356] and Dunning cc-pVDZ (DZ)[337] basis sets. For clarity purposes, QM/MM computations will be referenced using the notation [QM method] [QM basis set]/[MM] (e.g., MP2 STO-3G/MM).

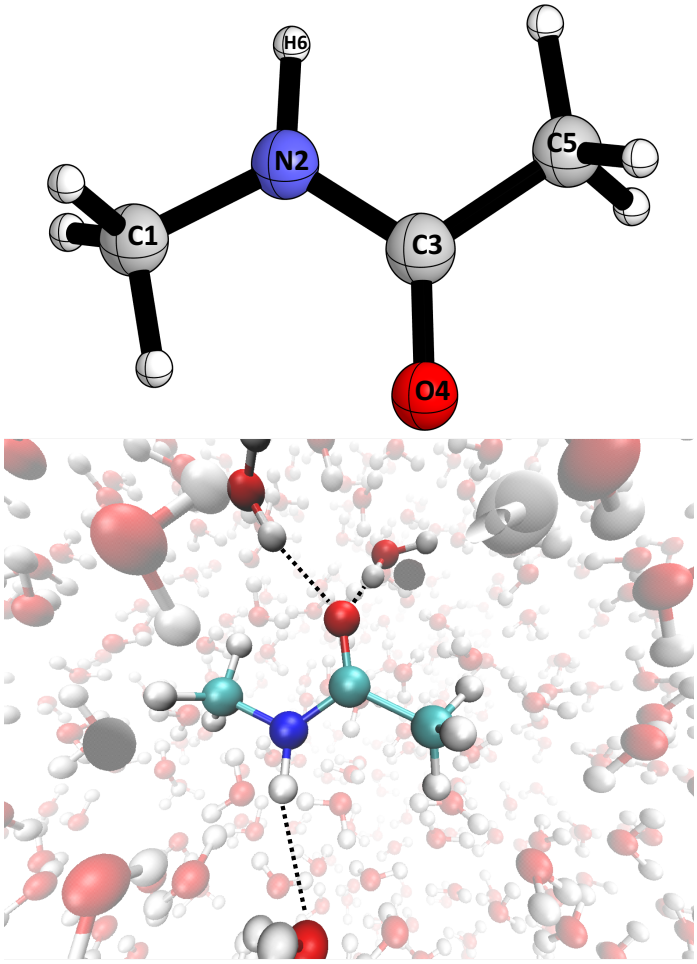

Figure 5.7: The NMA molecule and a snapshot of a hydrogen-bonded NMA–water complex during the MP2 cc-pVDZ/MM simulation generated with VMD.[357,358]

Figure 5.7 shows the labeled NMA molecule (used by Tables 5.2 and 5.3 and Figure 5.8) along with a snapshot of the MP2 DZ/MM simulation. The snapshot reveals two water molecules acting as hydrogen bond donors to the carbonyl oxygen and one water molecule acting as a hydrogen bond acceptor to the amide hydrogen. This is representative of the most common type of hydrogen bonded complex seen for all the simulations. The average numbers of water molecules hydrogen bonded with the carbonyl carbon during the 20 ps simulation are: 1.5 (MM), 1.7 (HF STO-3G/MM), 1.9 (HF DZ/MM), 1.7 (MP2 STO-3G/MM), and 1.7 (MP2 DZ/MM). For the amide hydrogen, the average numbers are: 1.0 (MM), 1.1 (HF STO-3G/MM), 1.0 (HF DZ/MM), 1.1 (MP2 STO-3G/MM), and 1.1 (MP2 DZ/MM). Average bonds were measure by tabulating the number of hydrogen bonds present at each time step and taking the mean.

Hydrogen bonds were identified using the Baker–Hubbard[359] scheme as implemented in MDTraj[331] with the criteria $r_{H...Acceptor} < 2.5$ Å.

Table 5.2 shows the geometry parameters obtained by our simulations as compared to *ab initio* MD and experiment. We note that although most parameters are comparable with one another, the ∠H–N–C–O torsion angle is highly dependent on the basis set. Table 5.3 shows the hydrogen bonding geometric parameters as compared with other QM/MM results in the literature. The average angles obtained by our results are lower than the ones in previous literature. The large variation in the standard deviation of the angles is partly due to the fact that we are averaging over all the waters that form hydrogen bonding interactions with the carbonyl carbon and amide hydrogen, as discussed above.

Figure 5.8 shows the RDF of various parts of the NMA molecule and the waters. Results obtained agree with those obtained by previous studies.[341–345] Traditional[341,345] and *ab initio* MD[342] results qualitatively match our results for the amide hydrogen and water RDFs as well as the carbonyl oxygen and water RDFs. Our QM/MM results also qualitatively match other QM/MM studies.[343,344] The O4–H_{wat} RDF has a more pronounced peak than the one obtained with ABEEM/MM,[343] but is in agreement with a HF 3-21G/MM study.[344]

Table 5.2: Geometry parameters of NMA in a water cluster

	MM	HF STO-3G/MM	HF DZ/MM	MP2 STO-3G/MM	MP2 DZ/MM	Ref 70[a]	Ref 76[b]	Ref 73[c]
				Bond Lengths (Å)				
C1–N2	1.462 ± 0.034	1.478 ± 0.026	1.448 ± 0.024	1.5074 ± 0.029	1.4542 ± 0.030	1.477 ± 0.031	1.465(13)	1.458
N2–C3	1.333 ± 0.022	1.442 ± 0.029	1.356 ± 0.024	1.4793 ± 0.034	1.3744 ± 0.026	1.351 ± 0.028	1.290(13)	1.351
C3–C5	1.510 ± 0.031	1.543 ± 0.028	1.512 ± 0.029	1.5666 ± 0.031	1.5185 ± 0.031	1.520 ± 0.035	1.536(16)	1.515
C3–O4	1.230 ± 0.020	1.223 ± 0.020	1.205 ± 0.023	1.2658 ± 0.020	1.2351 ± 0.027	1.268 ± 0.022	1.236(12)	1.243
N2–H6	1.013 ± 0.027	1.036 ± 0.040	1.004 ± 0.028	1.0614 ± 0.030	1.0217 ± 0.029	1.031 ± 0.032		1.010
				Bond Angles (°)				
∠C1–N2–C3	124.85 ± 3.23	113.84 ± 3.52	119.22 ± 3.16	110.59 ± 3.61	117.58 ± 3.63		120.5	122.4
∠N2–C3–O4	123.32 ± 2.80	120.38 ± 2.61	121.06 ± 2.44	120.45 ± 2.82	120.76 ± 2.84		123.0	121.7
∠N2–C3–C5	116.70 ± 2.93	115.58 ± 2.91	117.18 ± 2.78	114.57 ± 3.09	116.57 ± 3.11		116.5	116.9
∠O4–C3–C5	119.68 ± 3.02	123.64 ± 2.86	121.60 ± 2.66	124.31 ± 3.03	122.34 ± 3.10		120.5	
∠H6–N2–C3	116.98 ± 3.31	113.67 ± 4.65	119.73 ± 3.82	110.35 ± 5.05	119.68 ± 4.53		119.5	118.9
				Torsion Angles (°)				
∠H–N–C–O	172.40 ± 5.74	147.44 ± 12.61	168.31 ± 8.40	141.12 ± 11.67	165.81 ± 9.68	173.98 ± 10.75		180.00
∠C–N–C–C	171.93 ± 6.06	166.28 ± 8.86	171.42 ± 6.16	166.11 ± 8.71	171.01 ± 6.72	178.46 ± 10.70		180.00

Number after ± is the standard deviation.
a. Parameters determined by *ab initio* MD.
b. Parameters determined by X-ray crystal diffraction.
c. Paramters determined by B3LYP/6-31+G(d,p) in the IEF continuum.

5.5 Conclusions

Several disadvantages exist in currently available software for adaptive QM/MM, including licensing restrictions, limited selection of methods, and developer-focused usage. The growing field of adaptive

Table 5.3: Hydrogen bond parameters of NMA in a water cluster

	$O4\cdots H_{wat}$(Å)[a]	$O4–H_{wat}–O_{wat}$ (°)	$C3–O4–H_{wat}$ (°)	$H6\cdots O_{wat}$ (Å)	$N2–H_{wat}–O_{wat}$ (°)
MM	1.933 ± 0.220	156.03 ± 13.28	125.16 ± 19.49	2.141 ± 0.181	153.64 ± 13.36
HF STO-3G/MM	1.932 ± 0.223	156.02 ± 13.49	128.11 ± 18.78	2.134 ± 0.186	153.15 ± 12.98
HF DZ/MM	1.959 ± 0.212	155.29 ± 13.13	124.37 ± 19.61	2.198 ± 0.169	152.32 ± 13.79
MP2 STO-3G/MM	1.930 ± 0.218	157.66 ± 12.30	126.77 ± 19.42	2.173 ± 0.188	149.56 ± 14.01
MP2 DZ/MM	1.919 ± 0.220	157.34 ± 12.64	126.82 ± 18.34	2.195 ± 0.182	150.70 ± 13.86
OPLS[344]	1.78		141	1.94	175
AM1/MM[344]	1.70		144	1.77	176
HF 3-21G/MM[344]	1.99		143	2.10	168
[b]B3LYP/6-31+G(d,p)[345]	1.83			1.90	

Number after $\pm$ is the standard deviation.
a. $O4–H_{wat}$ related parameters are averaged over any hydrogen bond interactions with O4 at a given time.
b. Performed in IEF continuum

QM/MM can greatly benefit from a user-friendly platform with efficiently implemented methods, both to explore untested applications and to promote novel, problem-driven approaches. In the present paper, we describe Janus, an adaptive QM/MM Python library that seeks emphasizes usability and promotes method development. Janus is freely available to download and install at https://github.com/CCQC/janus and is made open-source. Contributions from the community are highly welcome and encouraged, facilitated by rigid automated testing and continuous integration. Simple input file structures and detailed documentation make it accessible to non-developers while advanced users can take advantage of the full capabilities of the Python library. An API is offered to give users great flexibility with how to use the code and allows rapid testing of various adaptive approaches. The modular design of Janus makes it easy to either implement an existing method or test new methods. In addition, interfaces to other software packages can be added on with little modification to existing code. It is our hope that Janus will promote the use of adaptive QM/MM methods and that the growth of the software will be sustained and guided by user needs.

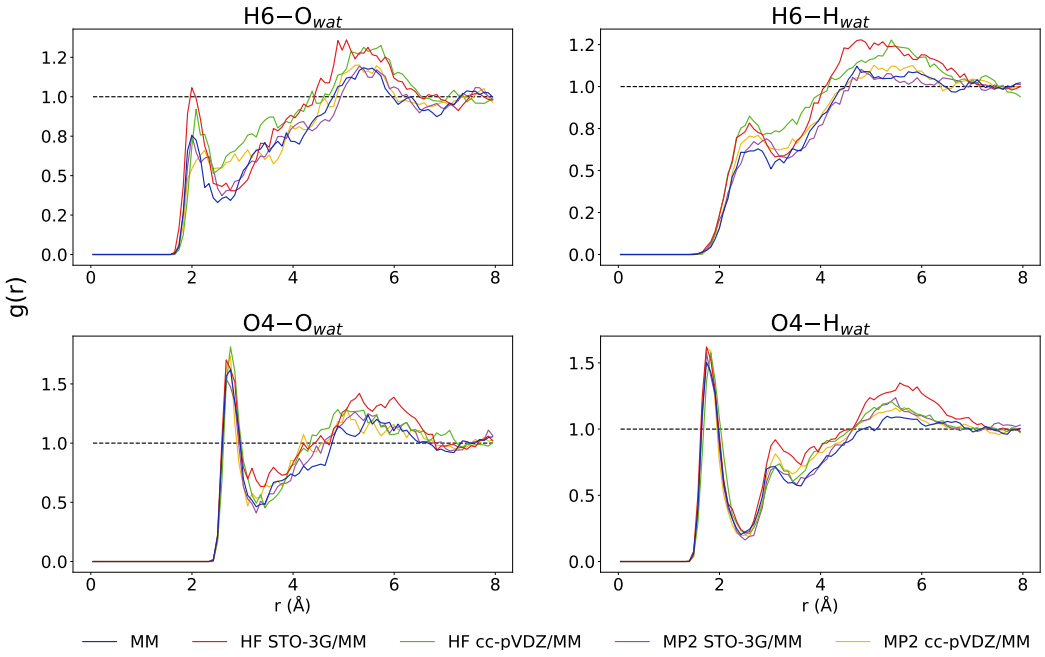

Figure 5.8: Radial distribution function of NMA and water. The H6−O_{wat} RDF was computed with the amide hydrogen and water oxygen atoms. The H6−H_{wat} RDF was computed with the amide hydrogen and water hydrogen atoms. The O4−O_{wat} RDF was computed with the carbonyl oxygen and water oxygen atoms. The O4−H_{wat} RDF was computed with the carbonyl oxygen and water hydrogen atoms. The blue curve corresponds to a pure MM simulation. The red, green, purple, and yellow curves correspond to QM/MM simulations at various levels of QM theory.

Chapter 6

Conclusion

Spin-orbit splittings of p-block elements B–I were computed via the fully relativistic four-component methods CASSCF, CASPT2, and MR-CISD+Q. The computations were within 15% of the experimental results across the board, but no method consistently outperformed the others across the periodic table. As expected, the absolute errors increased for later rows. In general, increasing the basis set size yielded more accurate results for elements B–F across all methods; however, no such trend was observed for the later rows. The 4c-CASSCF method is recommended for all cases beyond fluorine with the exception of group 13 atoms. The 4c-CASPT2 method is recommended for group 13 atoms. Although 4c-MR-CISD+Q is a more rigorous method, a frozen-core treatment with a valence-only basis set will not outperform 4c-CASSCF. In general, fully relativistic 4c methods give more accurate splittings than a perturbative treatment of SOC and are comparable with the DC Hamiltonian approach of Fleig.

Out of the 9 HXCN species studied, $HBCN^{-}$, HCCN, and $HNCN^{+}$ have triplet ground states while the other molecules have singlet ground states. They also show the most dramatic change in ST gaps and geometrical parameters. HCCN has the smallest barrier to linearity. The correlation of SOMO energy gaps in the triplet states and the percentages of s–character in the a' orbital in the singlet state to the ST energy gap was quantified and found to have a strong correlation. Isodesmic and NBO analysis show that the differences in ST gaps between HXCN and XH_2 originate from the the strength of the interaction between the a' orbital/a" orbital on X and the bonding/anti-bonding C–N orbital $HNCN^{+}$ is an outlier among the species studied, and has a fundamentally different electronic structure that is not carbene-like.

The isomerization of H_2XY species to HXYH (X, Y = O, S, Se, Te) species was studied using highly rigorous *ab initio* methods. Geometries and harmonic frequencies were obtained using the X2C-1e-CCSD(T) method with a X2C recontracted aug-cc-pCVXZ (X= T, Q) basis set. A nonrelativistic CCSD(T) was also used with an aug-cc-pwCVTZ basis set on H, O, S and an aug-cc-pwCVTZ-PP basis set on Se and Te. No significant difference in the geometries were found when comparing the various approaches. In general, the bond lengths of X–Y increased as X and Y increased in size, as expected. The H–X–Y angles for the HXYH species increased dramatically with X=O, but decreased slightly with an increase in the atomic number of Y for X=S, Se, and Te. The H–X–Y angles for the H_2XY increased for X=O as Y varied from O, S, Se and Te, but decreased with increasing atomic radii in Y for X=S, Se, and Te. The isomerization

energy from H_2XY to HXYH for a fixed X decreased with increasing Y atomic radii. The transition state structure more closely resembles H_2XY in cases where X is the same size or bigger than Y, and more closely resembles the HXYH structure in caess where X is smaller than Y.

An adaptive QM/MM Python library, JANUS, was developed to provide a user-friendly platform with efficiently implemented methods to explore untested applications and to promote novel, problem-driven approaches. Simple input file structures and detailed documentation make the adaptive QM/MM method more accessible to non-developers. The modular design of JANUS makes it easy to either implement an existing method or test new adaptive approaches. Interfaces to other software packages can be added on with little modification to existing code. JANUS is freely available to download and install at https://github.com/CCQC/janus.

Appendix A

Links for Supporting Information

The Supporting Information for Chapter 2 can be found at
https://pubs.acs.org/doi/abs/10.1021/acs.jctc.7b00989.

The Supporting Information for Chapter 3 can be found at
http://www.rsc.org/suppdata/c9/cp/c9cp05777c/c9cp05777c1.pdf.

Bibliography

(1) Pyykkö, P. *Annu. Rev. Phys. Chem.* 2012, *63*, 45–64.

(2) Fedorov, D. G.; Koseki, S.; Schmidt, M. W.; Gordon, M. S. *Int. Rev. Phys. Chem.* 2003, *22*, 551–592.

(3) Laguna, G. A.; Beattie, W. H. *Chem. Phys. Lett.* May 1982, *88*, 439–440.

(4) Minnhagen, L. *Arkiv för Fysik* 1962, *21*, 415–478.

(5) Saiz-Lopez, A.; Mahajan, A. S.; Salmon, R. A.; Bauguitte, S. J.-B.; Jones, A. E.; Roscoe, H. K.; Plane, J. M. C. *Science* July 2007, *317*, 348–351.

(6) Neufeld, D. A.; Wolfire, M. G. *Astrophys. J.* Dec. 2009, *706*, 1594–1604.

(7) Hao, Y.; Gu, J.; Guo, Y.; Zhang, M.; Xie, Y.; Schaefer, H. F. *Phys. Chem. Chem. Phys.* 2014, *16*, 2641–2646.

(8) Czakó, G.; Csaszar, A. G.; Schaefer, H. F. *J. Phys. Chem. A* 2014, *118*, 11956–11961.

(9) Wang, H.; Li, G.; Li, Q.-S.; Xie, Y.; Schaefer, H. F. *J. Phys. Chem. B* 2016, *120*, 1743–1748.

(10) Saue, T. *ChemPhysChem* 2011, *12*, 3077–3094.

(11) Autschbach, J. *J. Chem. Phys.* 2012, *136*, 150902.

(12) Kutzelnigg, W. *Chem. Phys.* 2012, *395*, 16–34.

(13) Liu, W. *Natl. Sci. Rev.* June 2016, *3*, 204–221.

(14) Quiney, H. M.; Skaane, H.; Grant, I. P. *Adv. Quantum Chem.* 1998, *32*, 1–49.

(15) Blume, M.; Watson, R. E. *Proc. R. Soc. A* 1963, *271*, 565–578.

(16) Malli, G. *J. Chem. Phys.* 1968, *48*, 1088–1091.

(17) Malli, G. *J. Chem. Phys.* 1968, *48*, 1092–1094.

(18) Jones, M. *J. Phys. B: At. Mol. Phys.* 1971, *4*, 1422–1439.

(19) Detrich, J. *Phys. Rev. A* 1972, *5*, 2014–2027.

(20) Pasternak, A.; Goldschmidt, Z. B. *Phys. Rev. A* 1972, *6*, 55–68.

(21) Veseth, L. *J. Phys. B: At. Mol. Phys.* 1981, *14*, 795–802.

(22) Pyper, N. C.; Rose, S. J.; Grant, I. P. *J. Phys. B: At. Mol. Phys.* 1982, *15*, 1319–1331.

(23) Stevens, W. J.; Krauss, M. *Chem. Phys. Lett.* 1982, *86*, 320–324.

(24) Cooper, D. L.; Wilson, S. *J. Phys. B: At. Mol. Phys.* 1982, *15*, 493–501.

(25) Teichteil, C.; Pelissier, M.; Spiegelmann, F. *Chem. Phys.* 1983, *81*, 273–282.

(26) Pyper, N. C. *J. Phys. B: At. Mol. Phys.* 1983, *16*, L211–L215.

(27) Matsushita, T.; Marian, C. M.; Klotz, R.; Hess, B. A.; Peyerimhoff, S. D. *Chem. Phys.* 1985, *96*, 371–379.

(28) Ross, R. B.; Ermler, W. C.; Christiansen, P. A. *J. Chem. Phys.* 1986, *84*, 3297–3300.

(29) Samzow, R.; Hess, B. A. *Chem. Phys. Lett.* 1991, *184*, 491–496.

(30) Ilyabaev, E.; Kaldor, U. *Phys. Rev. A* 1993, *47*, 137–142.

(31) Eliav (Ilyabaev), E.; Kaldor, U.; Ishikawa, Y. *Chem. Phys. Lett.* 1994, *222*, 82–87.

(32) Figgen, D.; Wedig, A.; Stoll, H.; Dolg, M.; Eliav, E.; Kaldor, U. *J. Chem. Phys.* 2008, *128*, 024106.

(33) Kaldor, U.; Eliav, E. *Adv. Quantum Chem.* 1998, *31*, 313–336.

(34) Seijo, L. *J. Chem. Phys.* 1995, *102*, 8078–8088.

(35) Heinemann, C.; Koch, W.; Schwarz, H. *Chem. Phys. Lett.* 1995, *245*, 509–518.

(36) Leininger, T.; Berning, A.; Nicklass, A.; Stoll, H.; Werner, H.-J.; Flad, H.-J. *Chem. Phys.* 1997, *217*, 19–27.

(37) Casarrubios, M.; Seijo, L. *J. Mol. Struct.: THEOCHEM* 1998, *426*, 59–74.

(38) Pacios, L. F.; Christiansen, P. A. *J. Chem. Phys.* 1985, *82*, 2664–2671.

(39) Hurley, M. M.; Pacios, L. F.; Christiansen, P. A.; Ross, R. B.; Ermler, W. C. *J. Chem. Phys.* 1986, *84*, 6840–6853.

(40) LaJohn, L. A.; Christiansen, P. A.; Ross, R. B.; Atashroo, T.; Ermler, W. C. *J. Chem. Phys.* 1987, *87*, 2812–2824.

(41) Ross, R. B.; Powers, J. M.; Atashroo, T.; Ermler, W. C.; LaJohn, L. A.; Christiansen, P. A. *J. Chem. Phys.* 1990, *93*, 6654–6670.

(42) Ross, R. B.; Gayen, S.; Ermler, W. C. *J. Chem. Phys.* 1994, *100*, 8145–8155.

(43) Ermler, W. C.; Ross, R. B.; Christiansen, P. A. *Int. J. Quant. Chem.* 1991, *40*, 829–846.

(44) Nash, C. S.; Bursten, B. E.; Ermler, W. C. *J. Chem. Phys.* 1997, *106*, 5133–5142.

(45) Boettger, J. C. *Phys. Rev. B.* 2000, *62*, 7809–7815.

(46) Nicklass, A.; Peterson, K. A.; Berning, A.; Werner, H.-J.; Knowles, P. J. *J. Chem. Phys.* 2000, *112*, 5624–5632.

(47) Visscher, L.; Saue, T. *The Journal of Chemical Physics* 2000, *113*, 3996.

(48) Koseki, S.; Fedorov, D. G.; Schmidt, M. W.; Gordon, M. S. *J. Phys. Chem. A* 2001, *105*, 8262–8268.

(49) Gagliardi, L.; Schimmelpfennig, B.; Maron, L.; Wahlgren, U.; Willetts, A. *Chem. Phys. Lett.* 2001, *344*, 207–212.

(50) Malmqvist, P. Å.; Roos, B. O.; Schimmelpfennig, B. *Chem. Phys. Lett.* 2002, *357*, 230–240.

(51) Roos, B. O.; Lindh, R.; Malmqvist, P. Å.; Veryazov, V.; Widmark, P.-O. *J. Phys. Chem. A* 2004, *108*, 2851–2858.

(52) Roos, B. O.; Malmqvist, P. *Phys. Chem. Chem. Phys.* 2004, *6*, 2919.

(53) Fleig, T.; Olsen, J.; Marian, C. M. *J. Chem. Phys.* 2001, *114*, 4775–4790.

(54) Fleig, T.; Olsen, J.; Visscher, L. *J. Chem. Phys.* 2003, *119*, 2963–2971.

(55) Fleig, T. *Phys. Rev. A* 2005, *72*, 052506.

(56) Dal Corso, A.; Conte, A. M. *Phys. Rev. B.* 2005, *71*, 115106.

(57) Fromager, E.; Teichteil, C.; Maron, L. *J. Chem. Phys.* 2005, *123*, 034106.

(58) Jones, M. D.; Albers, R. C. *Phys. Rev. B.* 2009, *79*, 045107.

(59) Zeng, T.; Fedorov, D. G.; Klobukowski, M. *J. Chem. Phys.* 2009, *131*, 124109.

(60) Matsuoka, T.; Someno, S.; Hada, M. *J. Comput. Chem. Jpn.* 2011, *10*, 11–17.

(61) Yang, D.-D.; Wang, F.; Guo, J. *Chem. Phys. Lett.* 2012, *531*, 236–241.

(62) Tu, Z.; Wang, F.; Li, X. *J. Chem. Phys.* 2012, *136*, 174102.

(63) Filatov, M.; Zou, W.; Cremer, D. *J. Chem. Phys.* 2013, *139*, 014106.

(64) Kim, I.; Lee, Y. S. *J. Chem. Phys.* 2013, *139*, 134115.

(65) Li, Z.; Suo, B.; Zhang, Y.; Xiao, Y.; Liu, W. *Mol. Phys.* 2013, *111*, 3741–3755.

(66) Wang, Z.; Tu, Z.; Wang, F. *J. Chem. Theory Comput.* 2014, *10*, 5567–5576.

(67) Pathak, H.; Sasmal, S.; Nayak, M. K.; Vaval, N.; Pal, S. *Comput. Theor. Chem.* 2016, *1076*, 94–100.

(68) Cao, Z.; Li, Z.; Wang, F.; Liu, W. *Phys. Chem. Chem. Phys.* 2017, *19*, 3713–3721.

(69) Thyssen, J.; Fleig, T.; Jensen, H. J. A. *J. Chem. Phys.* 2008, *129*, 034109.

(70) Jensen, H. J. A.; Dyall, K. G.; Saue, T.; Fægri, K. *J. Chem. Phys.* 1996, *104*, 4083–4097.

(71) Bates, J. E.; Shiozaki, T. *J. Chem. Phys.* 2015, *142*, 044112.

(72) Kelley, M. S.; Shiozaki, T. *J. Chem. Phys.* 2013, *138*, 204113.

(73) Shiozaki, T. *J. Chem. Phys.* 2013, *138*, 111101.

(74) Werner, H.-J.; Reinsch, E.-A. *J. Chem. Phys.* 1982, *76*, 3144–3156.

(75) Werner, H.; Knowles, P. J. *J. Chem. Phys.* 1988, *89*, 5803–5814.

(76) Shiozaki, T.; Mizukami, W. *J. Chem. Theory Comput.* 2015, *11*, 4733–4739.

(77) MacLeod, M. K.; Shiozaki, T. *J. Chem. Phys.* 2015, *142*, 051103.

(78) Granovsky, A. A. *J. Chem. Phys.* 2011, *134*, 214113.

(79) Vlaisavljevich, B.; Shiozaki, T. *J. Chem. Theory Comput.* 2016, *12*, 3781–3787.

(80) BAGEL, Brilliantly Advanced General Electronic-Structure Library. http://www.nubakery.org, under the GNU General Public License.

(81) Shiozaki, T. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 2017, e1331.

(82) Dunning, T. H. *J. Chem. Phys.* 1989, *90*, 1007–1023.

(83) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* June 1993, *98*, 1358–1371.

(84) Wilson, A. K.; Woon, D. E.; Peterson, K. A.; Dunning, T. H. *J. Chem. Phys.* 1999, *110*, 7667–7676.

(85) Bross, D. H.; Peterson, K. A. *Theor. Chem. Acc.* 2014, *133*, 1434.

(86) Dunning, T. H.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* 2001, *110*, 9244–9253.

(87) Dyall, K. G. *Theor. Chem. Acc.* 2002, *108*, 335–340.

(88) Dyall, K. G. *Theor. Chem. Acc.* 2003, *109*, 284–284.

(89) Dyall, K. G. *Theor. Chem. Acc.* 2006, *115*, 441–447.

(90) Dyall, K. G. *Theor. Chem. Acc.* 1998, *99*, 366–371.

(91) Dyall, K. G. *Theor. Chem. Acc.* 2002, *108*, 365–365.

(92) Stoychev, G. L.; Auer, A. A.; Neese, F. *J. Chem. Theory Comput.* 2017, *13*, 554–562.

(93) Neese, F. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 2012, *2*, 73–78.

(94) A. Kramida; Yu. Ralchenko; J. Reader; and NIST ASD Team, *NIST Atomic Spectra Database* (ver. 5.5.1), [Online]. Available: `https://physics.nist.gov/asd` [2017, November 10]. National Institute of Standards and Technology, Gaithersburg, MD. 2017.

(95) Visscher, L.; Dyall, K. G. *J. Chem. Phys.* 1996, *104*, 9040–9046.

(96) Bender, C. F.; Schaefer, H. F. *Journal of the American Chemical Society* 1970, *92*, 4984–4985.

(97) Bender, C. F.; Schaefer, H. F.; Franceschetti, D. R.; Allen, L. C. *Journal of the American Chemical Society* 1972, *94*, 6888–6893.

(98) Wentrup, C. *Angew. Chem. Int. Ed.* 2018, *57*, 11508–11521.

(99) Shainyan, B. A.; Kuzmin, A. V.; Moskalik, M. Y. *Comput. Theor. Chem.* 2013, *1006*, 52–61.

(100) Skell, P. S.; Garner, A. Y. *J. Am. Chem. Soc.* 1956, *78*, 5430–5433.

(101) Hine, J. S., *Divalent Carbon*; Ronald Press Co.: New York, 1964.

(102) Kirmse, W., *Carbene Chemistry*, 2nd ed.; Academic Press: New York, 1971; Vol. 1.

(103) Bourissou, D.; Guerret, O.; Gabbaï, F. P.; Bertrand, G. *Chem. Rev.* 2000, *100*, 39–92.

(104) Bertrand, G., *Carbene chemistry: from fleeting intermediates to powerful reagents*; Marcel Dekker, Inc.: New York, 2002.

(105) Hirai, K.; Itoh, T.; Tomioka, H. *Chem. Rev.* 2009, *109*, 3275–3332.

(106) De Frémont, P.; Marion, N.; Nolan, S. P. *Coord. Chem. Rev.* 2009, *253*, 862–892.

(107) Moss, R. A.; Doyle, M. P., *Contemporary Carbene Chemistry*; John Wiley & Sons: Hoboken, 2013.

(108) Hopkinson, M. N.; Richter, C.; Schedler, M.; Glorius, F. *Nature* 2014, *510*, 485–496.

(109) Sheridan, R. S. *Chem. Rev.* 2013, *113*, 7179–7208.

(110) Gerbig, D.; Ley, D. en *WIREs Comput. Mol. Sci.* 2013, *3*, 242–272.

(111) Schaefer, H. F. *Science* 1986, *231*, 1100–1107.

(112) Luke, B.; Pople, J.; Krogh-Jespersen, M.; Apeloig, Y.; Chandrasekhar, J.; Schleyer, P. v. R. *J. Am. Chem. Soc.* 1986, *108*, 260–269.

(113) Luke, B.; Pople, J.; Krogh-Jespersen, M.; Apeloig, Y.; Karni, M.; Chandrasekhar, J.; Schleyer, P. v. R. *J. Am. Chem. Soc.* 1986, *108*, 270–284.

(114) Neumann, W. P. *Chem. Rev.* 1991, *91*, 311–334.

(115) Heinemann, C.; Herrmann, W. A.; Thiel, W. *J. Organomet. Chem.* 1994, *475*, 73–84.

(116) Haaf, M.; Schmedake, T. A.; West, R. *Acc. Chem. Res.* 2000, *33*, 704–714.

(117) Mizuhata, Y.; Sasamori, T.; Tokitoh, N. *Chem. Rev.* 2009, *109*, 3479–3511.

(118) Wang, Y.; Ma, J. *J. Organomet. Chem.* 2009, *694*, 2567–2575.

(119) Falvey, D. E. *J. Phys. Org. Chem.* 1999, *12*, 589–596.

(120) Borodkin, G. I.; Shubin, V. G. *Russ. Chem. Rev.* 2008, *77*, 395.

(121) Braunschweig, H.; Chiu, C.-W.; Radacki, K.; Kupfer, T. *Angew. Chem. Int. Ed.* 2010, *49*, 2041–2044.

(122) Hicks, J.; Vasko, P.; Goicoechea, J. M.; Aldridge, S. *Nature* 2018, *557*, 92.

(123) Hinz, A.; Breher, F. *Angew. Chem. Int. Ed.* 2018, *57*, 8818–8820.

(124) Cramer, C. J.; Dulles, F. J.; Storer, J. W.; Worthington, S. E. *Chem. Phys. letters* 1994, *218*, 387–394.

(125) Hyatt, I. D.; Meza-Aviña, M. E.; Croatt, M. P. *Synlett* 2012, *23*, 2869.

(126) Kunde, V.; Aikin, A.; Hanel, R.; Jennings, D.; Maguire, W.; Samuelson, R. *Nature* 1981, *292*, 686.

(127) Morris, V.; Bhatia, S.; Stelson, A.; Hall Jr, J. *Energy & Fuels* 1991, *5*, 126–133.

(128) Francisco, J. S. *Chem. Phys. Lett.* 1994, *230*, 372–376.

(129) Wang, D.; Liu, H.; Huang, X.; Li, Y.; Geng, C.; Zhan, J.; Sun, C. *Eur. Phys. J. D* 2008, *48*, 187–196.

(130) Adamson, J.; DeSain, J.; Curl, R.; Glass, G. *J. Phys. Chem. A* 1997, *101*, 864–870.

(131) Osamura, Y.; Petrie, S. *J. Phys. Chem. A* 2004, *108*, 3615–3622.

(132) Chen, H.-L.; Chao, W.-C. *J. Phys. Chem. A* 2011, *115*, 1133–1142.

(133) Turner, B.; Pirogov, L.; Minh, Y. *Astrophys. J.* 1997, *483*, 235.

(134) Pratap, P.; Dickens, J.; Snell, R. L.; Miralles, M.; Bergin, E.; Irvine, W. M.; Schloerb, F. *Astrophys. J.* 1997, *486*, 862.

(135) Hirota, T.; Yamamoto, S.; Mikami, H.; Ohishi, M. *Astrophys. J.* 1998, *503*, 717.

(136) Liszt, H.; Lucas, R. *Astron. Astrophys.* 2001, *370*, 576–585.

(137) Mebel, A.; Kaiser, R. *Astrophys. J* 2002, *564*, 787.

(138) Takahashi, K.; Takayanagi, T. *Chem. Phys. letters* 2006, *429*, 399–404.

(139) Loison, J.-C.; Wakelam, V.; Hickson, K. M. *Mon. Notices Royal Astron. Soc.* 2014, *443*, 398–410.

(140) Loison, J.; Hickson, K. *Chem. Phys. Lett.* 2015, *635*, 174–179.

(141) Shivani, A. M.; Shivani, P. T. *Orig. Life Evol. Biospheres* 2014, *44*, 143–157.

(142) Lanzisera, D. V.; Andrews, L. *J. Phys. Chem. A* 1997, *101*, 9660–9665.

(143) Lanzisera, D. V.; Andrews, L.; Taylor, P. R. *J. Phys. Chem. A* 1997, *101*, 7134–7140.

(144) Merritt, J. M.; Douberly, G. E.; Stiles, P. L.; Miller, R. E. *J. Phys. Chem. A* 2007, *111*, 12304–12316.

(145) Liang, T.; Douberly, G. E. *Chem. Phys. Lett.* 2012, *551*, 54–59.

(146) Pappova, A.; Deakyneoe, C. A.; Skancke, A.; Černušák, I.; Liebman, J. F. *Mol. Phys.* 1996, *89*, 247–265.

(147) Liebman, J. F.; Černušák, I.; Miková, A. *Int. J. Quant. Chem* 2001, *84*, 140–148.

(148) Bernheim, R.; Kempf, R.; Humer, P.; Skell, P. *J. Chem. Phys.* 1964, *41*, 1156–1157.

(149) Bernheim, R.; Kempf, R.; Gramas, J.; Skell, P. *J. Chem. Phys.* 1965, *43*, 196–200.

(150) Bernheim, R.; Kempf, R.; Reichenbecher, E. *J Magn. Reson.* 1970, *3*, 5–9.

(151) Wasserman, E.; Yager, W.; Kuck, V. *Chem. Phys. Lett.* 1970, *7*, 409–413.

(152) Dendramis, A.; Leroi, G. *J. Chem. Phys.* 1977, *66*, 4334–4340.

(153) Saito, S.; Endo, Y.; Hirota, E. *J. Chem. Phys.* 1984, *80*, 1427–1430.

(154) Malmquist, P.-Å.; Lindh, R.; Roos, B. O.; Ross, S. *Theor. Chim. Acta.* 1988, *73*, 155–171.

(155) Brown, F.; Saito, S.; Yamamoto, S. *J. Mol. Spect.* 1990, *143*, 203–208.

(156) Endo, Y.; Yasuhiro, O. *J. Chem. Phys.* 1993, *98*, 6618–6623.

(157) Miller, C.; Eckhoff, W.; Curl, R. *J. Mol. Struct.* 1995, *352*, 435–446.

(158) McCarthy, M.; Gottlieb, C.; Cooksy, A.; Thaddeus, P. *J. Chem. Phys.* 1995, *103*, 7779–7787.

(159) Maier, G.; Reisenauer, H. P.; Rademacher, K. *Chem.:Eur. J.* 1998, *4*, 1957–1963.

(160) Sun, F.; Kosterev, A.; Scott, G.; Litosh, V.; Curl, R. *J. Chem. Phys.* 1998, *109*, 8851–8856.

(161) Han, J.-x.; Hung, P.; DeSain, J.; Jones, W.; Curl, R. *J. Mol. Spect.* 1999, *198*, 421–428.

(162) Allen, M. D.; Evenson, K. M.; Brown, J. M. *J. Mol. Spect.* 2001, *209*, 143–164.

(163) Hung, P.; Sun, F.; Hunt, N.; Burns, L.; Curl, R. *J. Chem. Phys.* 2001, *115*, 9331–9339.

(164) Poutsma, J. C.; Upshaw, S. D.; Squires, R. R.; Wenthold, P. G. *J. Phys. Chem. A* 2002, *106*, 1067–1073.

(165) Nimlos, M. R.; Davico, G.; Geise, C. M.; Wenthold, P. G.; Lineberger, W. C.; Blanksby, S. J.; Hadad, C. M.; Petersson, G. A.; Ellison, G. B. *J. Chem. Phys.* 2002, *117*, 4323–4339.

(166) Nakajima, M.; Toyoshima, H.; Sato, S.; Tanaka, K.; Hoshina, K.; Kohguchi, H.; Sumiyoshi, Y.; Ohshima, Y.; Endo, Y. *J. Chem. Phys.* 2013, *138*, 164309.

(167) Garcia, G. A.; Krüger, J.; Gans, B.; Falvo, C.; Coudert, L. H.; Loison, J.-C. *J. Chem. Phys.* 2017, *147*, 013908.

(168) Harrison, J.; Dendramis, A.; Leroi, G. *J. Am. Chem. Soc.* 1978, *100*, 4352–4356.

(169) Zandler, M. E.; Goddard, J. D.; Schaefer, H. F. *J. Am. Chem. Soc.* 1979, *101*, 1072–1076.

(170) Kim, K. S.; Schaefer, H. F.; Radom, L.; Pople, J. A.; Binkley, J. S. *J. Am. Chem. Soc.* 1983, *105*, 4148–4154.

(171) Rice, J. E.; Schaefer, H. F. *J. Chem. Phys.* 1987, *86*, 7051–7053.

(172) Seidl, E. T.; Schaefer, H. F. *J. Chem. Phys.* 1992, *96*, 4449–4452.

(173) Aoki, K.; Fueno, H.; Ikuta, S.; Nomura, O. *Chem. Phys. letters* 1993, *202*, 33–36.

(174) Aoki, K.; Ikuta, S.; Nomura, O. *J. Chem. Phys.* 1993, *99*, 3809–3814.

(175) Kellogg, C. B.; Galbraith, J. M.; Fowler, J. E.; Schaefer, H. F. *J. Chem. Phys.* 1994, *101*, 430–435.

(176) Morter, C.; Farhat, S.; Curl, R. *Chem. Phys. letters* 1993, *207*, 153–158.

(177) Goldberg, N.; Fiedler, A.; Schwarz, H. *J. Phys. Chem.* 1995, *99*, 15327–15334.

(178) Koput, J. *J. Phys. Chem. A* 2002, *106*, 6183–6188.

(179) Koput, J. *J. Phys. Chem. A* 2003, *107*, 4717–4723.

(180) Ionescu, E.; Reid, S. A. *J. Mol. Struct.: THEOCHEM* 2005, *725*, 45–53.

(181) Kassaee, M.; Musavi, S.; Buazar, F. *J. Mol. Struct.: THEOCHEM* 2005, *728*, 15–24.

(182) Mladenović, M.; Botschwina, P.; Puzzarini, C. *J. Phys. Chem. A* 2006, *110*, 5520–5529.

(183) Kassaee, M.; Ghambarian, M.; Musavi, S. *Heteroatom Chemistry* 2008, *19*, 377–388.

(184) Zhao, Z.-X.; Zhang, H.-X.; Sun, C.-C. *J. Phys. Chem. A* 2008, *112*, 12125–12131.

(185) Inostroza, N.; Huang, X.; Lee, T. J. *J. Chem. Phys.* 2011, *135*, 244310.

(186) Cernicharo, J.; Gottlieb, C.; Guelin, M.; Thaddeus, P.; Vrtilek, J. *Astrophys. J.* 1989, *341*, L25–L28.

(187) Guélin, M.; Muller, S.; Cernicharo, J.; Apponi, A.; McCarthy, M.; Gottlieb, C.; Thaddeus, P. *Astronomy and Astrophysics* 2000, *363*, L9–L12.

(188) Guélin, M.; Muller, S.; Cernicharo, J.; McCarthy, M.; Thaddeus, P. *Astronomy & Astrophysics* 2004, *426*, L49–L52.

(189) Apponi, A.; McCarthy, M.; Gottlieb, C.; Thaddeus, P. *Astrophys. J. Lett.* 1999, *516*, L103.

(190) Sanz, M. E.; McCarthy, M. C.; Thaddeus, P. *Astrophys. J. Lett.* 2002, *577*, L71.

(191) Zhao, Z.-X.; Hou, C.-Y.; Shu, X.; Zhang, H.-X.; Sun, C.-c. *Theor. Chem. Acc.* 2009, *124*, 85–93.

(192) Flores, J.; Perez-Juste, I.; Carballeira, L. *Chem. Phys.* 2005, *313*, 1–15.

(193) Wang, Q.; Ding, Y.-h.; Sun, C.-c. *Chem. Phys.* 2006, *323*, 413–428.

(194) Kalcher, J. *J. Phys. Chem. A* 2005, *109*, 11437–11442.

(195) Kalcher, J.; Skurski, P.; Simons, J. *J. Phys. Chem. A* 2007, *111*, 401–410.

(196) Thorwirth, S.; Harding, M. E. *J. Chem. Phys.* 2009, *130*, 214303.

(197) Yamamoto, J.-i.; Okabe, Y. *Comput. Theor. Chem.* 2011, *963*, 24–33.

(198) Onyszchuk, M.; Castel, A.; Riviere, P.; Satge, J. *J. Organomet. Chem.* 1986, *317*, C35–C37.

(199) Hihara, G.; Hynes, R. C.; Lebuis, A.-M.; Rivière-Baudet, M.; Wharf, I.; Onyszchuk, M. *J. Organomet. Chem.* 2000, *598*, 276–285.

(200) Brown, Z. D.; Vasko, P.; Fettinger, J. C.; Tuononen, H. M.; Power, P. P. *J. Am. Chem. Soc.* 2012, *134*, 4045–4048.

(201) Huang, M.; Su, M.-D. *J. Organomet. Chem.* 2002, *659*, 121–124.

(202) Wang, Q.; Ding, Y.-h.; Sun, C.-c. *J. Chem. Phys.* 2005, *122*, 204305.

(203) Bundhun, A.; Abdallah, H. H.; Ramasami, P.; Gaspar, P. P.; Schaefer, H. F. *Inorg. Chem.* 2012, *51*, 12152–12164.

(204) Kassaee, M.; Ghambarian, M.; Musavi, S. *J. Organomet. Chem.* 2005, *690*, 4692–4703.

(205) Xu, S.; Lin, M. *J. Phys. Chem. A* 2007, *111*, 6730–6740.

(206) Jian, R.-C.; Tsai, C.; Hsu, L.-C.; Chen, H.-L. *J. Phys. Chem. A* 2010, *114*, 4655–4663.

(207) Chen, H.-L.; Wu, S.-k.; Lu, Y.-H. *J. Phys. Chem. A* 2012, *116*, 3267–3273.

(208) Wu, N.-N.; He, C.-Z.; Duan, X.-M.; Liu, J.-Y. *Journal of computational chemistry* 2011, *32*, 1449–1455.

(209) Berman, M. R.; Tsuchiya, T.; Gregušová, A.; Perera, S. A.; Bartlett, R. J. *J. Phys. Chem. A* 2007, *111*, 6894–6899.

(210) Puzzarini, C.; Gambi, A. *J. Chem. Phys.* 2005, *122*, 064316.

(211) Antoniotti, P.; Borocci, S.; Bronzolino, N.; Grandinetti, F. *ChemPhysChem* 2004, *5*, 1345–1351.

(212) Betterton, S. A.; Berka, A. S.; Fleming, P. E. *J. Theor. Comput. Chem.* 2010, *9*, 189–200.

(213) Staton, J. *Chem. Phys. Lett.* 1997, *281*, 130–134.

(214) Bartlett, R. J.; Watts, J. D.; Kucharski, S.; Noga, J. *Chem. Phys. Lett.* 1990, *165*, 513–522.

(215) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* 2013, *589*, 37–40.

(216) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. en *J. Chem. Phys.* 1992, *96*, 6796–6806.

(217) Dunning Jr, T. H. *J. Chem. Phys.* 1989, *90*, 1007–1023.

(218) Woon, D. E.; Dunning Jr, T. H. *J. Chem. Phys.* 1993, *98*, 1358–1371.

(219) Dunning Jr, T. H.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* 2001, *114*, 9244–9253.

(220) DeYonker, N. J.; Peterson, K. A.; Wilson, A. K. *J. Phys. Chem. A* 2007, *111*, 11383–11393.

(221) Wilson, A. K.; Woon, D. E.; Peterson, K. A.; Dunning Jr, T. H. *J. Chem. Phys.* 1999, *110*, 7667–7676.

(222) Stanton, J. F.; Gauss, J.; Cheng, L.; Harding, M. E.; Matthews, D. A.; Szalay, P. G. CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package., With contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, O. Christiansen, F. Engel, R. Faber, M. Heckert, O. Heun, M. Hilgenberg, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, T. Kirsch, K. Klein, W.J. Lauderdale, F. Lipparini, T. Metzroth, L.A. Mück, D.P. O'Neill, D.R. Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J.D. Watts and the integral packages MOLECULE (J. Almlöf and P.R. Taylor), PROPS (P.R. Taylor), ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see http://www.cfour.de.

(223) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. *WIREs Comput. Mol. Sci.* 2012, *2*, 242–253.

(224) Werner, H.-J. et al. MOLPRO, version 2015.1, a package of ab initio programs., see http://www.molpro.net, Cardiff, UK, 2015.

(225) Schuurman, M. S.; Muir, S. R.; Allen, W. D.; Schaefer, H. F. *J. Chem. Phys.* 2004, *120*, 11586–11599.

(226) Gonzales, J. M.; Pak, C.; Cox, R. S.; Allen, W. D.; Schaefer, H. F.; Császár, A. G.; Tarczay, G. *Chem. Eur. J.* 2003, *9*, 2173–2192.

(227) Császár, A. G.; Allen, W. D.; Schaefer, H. F. *J. Chem. Phys.* 1998, *108*, 9751–9764.

(228) East, A. L. L.; Allen, W. D. *J. Chem. Phys.* 1993, *99*, 4638–4650.

(229) Harding, M. E.; Metzroth, T.; Gauss, J.; Auer, A. A. *Journal of Chemical Theory and Computation* 2008, *4*, 64–74.

(230) MRCC, a quantum chemical program suite written by M. Kállay, Z. Rolik, J. Csontos, P. Nagy, G. Samu, D. Mester, J. Csóka, B. Szabó, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, and B. Hégely. www.mrcc.hu.

(231) Kállay, M.; Gauss, J. *J. Chem. Phys.* 2008, *129*, 144101.

(232) Feller, D.; Peterson, K. A.; Crawford, T. D. *J. Chem. Phys.* 2006, *124*, 054107.

(233) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* 1997, *106*, 9639–9646.

(234) Rolik, Z.; Szegedy, L.; Ladjánszki, I.; Ladóczki, B.; Kállay, M. *J. Chem. Phys.* 2013, *139*, 094105.

(235) Woon, D. E.; Dunning Jr, T. H. *J. Chem. Phys.* 1995, *103*, 4572–4585.

(236) Peterson, K. A.; Dunning Jr, T. H. *J. Chem. Phys.* 2002, *117*, 10548–10560.

(237) Cheng, L.; Gauss, J. *J. Chem. Phys.* 2011, *135*, 084114.

(238) Glendening, E. D.; Landis, C. R.; Weinhold, F. en *WIREs Comput. Mol. Sci.* 2012, *2*, 1–42.

(239) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. TB; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X. et al. *Mol. Phys.* 2015, *113*, 184–215.

(240) Lee, C.; Yang, W.; Parr, R. G. *Physical review B* 1988, *37*, 785.

(241) CYLview, 1.0b, Legault, C. Y., Université de Sherbrooke, 2009 (http://www.cylview.org).

(242) Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. *J. Am. Chem. Soc.* 1970, *92*, 4796–4801.

(243) Eckhardt, A. K.; Schreiner, P. R. *Angew. Chem. Int. Ed.* 2018, *57*, 5248–5252.

(244) Schröder, D.; Schalley, C. A.; Schwarz, H.; Goldberg, N.; Hrûsák, J. *Chem. Euro. J.* 1996, *2*, 1235–1242.

(245) Matsika, S.; Yarkony, D. R. *J. Chem. Phys.* 2002, *117*, 3733–3740.

(246) Sennikov, P.; Ignatov, S.; Schrems, O. *ChemPhysChem* 2005, *6*, 392–412.

(247) Braunstein, M.; Panfili, R.; Shroll, R.; Bernstein, L. *J. Chem. Phys.* 2005, *122*, 184307.

(248) Chattaraj, P. K.; Roy, D. R. *J. Phys. Chem. A* 2006, *110*, 11401–11403.

(249) Pehkonen, S.; Marushkevich, K.; Khriachtchev, L.; Räsänen, M.; Grigorenko, B. L.; Nemukhin, A. V. *J. Phys. Chem. A* 2007, *111*, 11444–11449.

(250) Franz, J.; Francisco, J. S.; Peyerimhoff, S. D. *J. Chem. Phys.* 2009, *130*, 084304.

(251) Conforti, P. F.; Braunstein, M.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* 2010, *133*, 164312.

(252) Tozihi, M.; Vahedpour, M.; Nazari, F. *J. Iranian Chem. Soc.* 2010, *7*, 585–596.

(253) Oba, Y.; Watanabe, N.; Hama, T.; Kuwahata, K.; Hidaka, H.; Kouchi, A. *Astrophys. J.* 2012, *749*, 67.

(254) Gámez, J. A.; Yáñez, M. *J. Chem. Theor. Comput.* 2011, *7*, 1726–1735.

(255) Zaccaria, F.; Wolters, L. P.; Fonseca Guerra, C.; Orian, L. *J. Comput. Chem.* 2016, *37*, 1672–1680.

(256) Kaur, D.; Sharma, P.; Bharatam, P. V. *J. Mol. Struct.: THEOCHEM* 2007, *810*, 31–37.

(257) Viana, R. B. *Mol. Phys.* 2017, *115*, 1004–1013.

(258) Hrušák, J.; Friedrichs, H.; Schwarz, H.; Razafinjanahary, H.; Chermette, H. *J. Phys. Chem.* 1996, *100*, 100–110.

(259) Huang, H. H.; Xie, Y.; Schaefer, H. F. *J. Phys. Chem.* 1996, *100*, 6076–6080.

(260) Jursic, B. S. *J. Mol. Struct.: THEOCHEM* 1997, *417*, 81–88.

(261) Hoy, E. P.; Schwerdtfeger, C. A.; Mazziotti, D. A. *Mol. Phys.* 2012, *110*, 765–773.

(262) Laitinen, R.; Pakkanen, T. *J. Mol. Struct.* 1982, *91*, 337–352.

(263) BelBruno, J. J. *Heteroat. Chem.* 1997, *8*, 199–202.

(264) Sánchez-Sanz, G.; Trujillo, C.; Alkorta, I.; Elguero, J. *ChemPhysChem* 2012, *13*, 496.

(265) Hinchliffe, A. *J. Mol. Struct.* 1979, *55*, 127–134.

(266) Suontamo, R. J.; Laitinen, R. S. *Main Group Chem.* 1996, *1*, 241–251.

(267) Goodarzi, M.; Vahedpour, M.; Karimi, L. *Comput. Theor. Chem.* 2012, *980*, 101–107.

(268) El-Hamdi, M.; Poater, J.; Bickelhaupt, F. M.; Sola, M. *Inorg. Chem.* 2013, *52*, 2458–2465.

(269) Laerdahl, J. K.; Schwerdtfeger, P. *Phys. Rev. A* 1999, *60*, 4439.

(270) Pagola, G. I.; Larsen, M. A.; Ferraro, M.; Sauer, S. P. *J. Comput. Chem.* 2018, *39*, 2589–2600.

(271) Senami, M.; Ito, K. *Phys. Rev. A* 2019, *99*, 012509.

(272) Warshel, A.; Karplus, M. *J. Am. Chem. Soc.* 1972, *94*, 5612–5625.

(273) Warshel, A.; Levitt, M. *J. Mol. Biol.* 1976, *103*, 227–249.

(274) Pezeshki, S.; Lin, H. *Mol. Sim.* 2015, *41*, 168–189.

(275) Duster, A. W.; Wang, C. H.; Garza, C. M.; Miller, D. E.; Lin, H. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 2017, *7*, 1–21.

(276) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. *Chem. Phys.* 1996, *211*, 313–323.

(277) Schwenk, C. F.; Loeffler, H. H.; Rode, B. M. *J. Am. Chem. Soc.* 2003, *125*, 1618–1624.

(278) Hofer, T. S.; Pribil, A. B.; Randolf, B. R.; Rode, B. M. *J. Am. Chem. Soc.* 2005, *127*, 14231–14238.

(279) Rode, B. M.; Hofer, T. S.; Randolf, B. R.; Schwenk, C. F.; Xenides, D.; Vchirawongkwin, V. *Theor. Chem. Acc.* 2006, *115*, 77–85.

(280) Azam, S. S.; Hofer, T. S.; Randolf, B. R.; Rode, B. M. *J. Phys. Chem. A* 2009, *113*, 1827–1834.

(281) Rowley, C. N.; Roux, B. *J. Chem. Theory Comput.* 2012, *8*, 3526–3535.

(282) Lev, B.; Roux, B.; Noskov, S. Y. *J. Chem. Theory Comput.* 2013, *9*, 4165–4175.

(283) Kabbalee, P.; Sripa, P.; Tongraar, A.; Kerdcharoen, T. *Chem. Phys. Lett.* 2015, *633*, 152–157.

(284) Kabbalee, P.; Tongraar, A.; Kerdcharoen, T. *Chem. Phys.* 2015, *446*, 70–75.

(285) Watanabe, H. C.; Kubillus, M.; Kubař, T.; Stach, R.; Mizaikoff, B.; Ishikita, H. *Phys. Chem. Chem. Phys.* 2017, *19*, 17985–17997.

(286) Park, K.; Götz, A. W.; Walker, R. C.; Paesani, F. *J. Chem. Theory Comput.* 2012, *8*, 2868–2877.

(287) Várnai, C.; Bernstein, N.; Mones, L.; Csányi, G. *J. Phys. Chem. B* 2013, *117*, 12202–12211.

(288) Pezeshki, S.; Davis, C.; Heyden, A.; Lin, H. *J. Chem. Theory Comput.* 2014, *10*, 4765–4776.

(289) Pezeshki, S.; Lin, H. *J. Chem. Theory Comput.* 2015, *11*, 2398–2411.

(290) Jiang, T.; Boereboom, J. M.; Michel, C.; Fleurat-Lessard, P.; Bulo, R. E. In *Quantum Modeling of Complex Molecular Systems*, Rivail, J.-L., Ruiz-Lopez, M., Assfeld, X., Eds.; Springer: Cham, 2015; Vol. 21, pp 51–91.

(291) Boereboom, J. M.; Fleurat-Lessard, P.; Bulo, R. E. *J. Chem. Theory Comput.* 2018, *14*, 1841–1852.

(292) Duster, A. W.; Garza, C. M.; Aydintug, B. O.; Negussie, M. B.; Lin, H. *J. Chem. Theory Comput.* 2019, DOI: 10.1021/acs.jctc.8b01128.

(293) Takenaka, N.; Kitamura, Y.; Koyano, Y.; Nagaoka, M. *Chem. Phys. Lett.* 2012, *524*, 56–61.

(294) Takenaka, N.; Kitamura, Y.; Koyano, Y.; Nagaoka, M. *J. Chem. Phys.* 2012, *137*, 024501.

(295) Waller, M. P.; Kumbhar, S.; Yang, J. *ChemPhysChem* 2014, *15*, 3218–3225.

(296) Kerdcharoen, T.; Morokuma, K. *Chem. Phys. Lett.* 2002, *355*, 257–262.

(297) Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. *J. Chem. Theory Comput.* 2009, *5*, 2212–2221.

(298) Heyden, A.; Lin, H.; Truhlar, D. G. *J. Phys. Chem. B* 2007, *111*, 2231–2241.

(299) Pezeshki, S.; Lin, H. *J. Chem. Theory Comput.* 2011, *7*, 3625–3634.

(300) Bernstein, N.; Várnai, C.; Solt, I.; Winfield, S. A.; Payne, M. C.; Simon, I.; Fuxreiter, M.; Csányi, G. *Phys. Chem. Chem. Phys.* 2012, *14*, 646–656.

(301) Watanabe, H. C.; Kubar, T.; Elstner, M. *J. Chem. Theory Comput.* 2014, *10*, 4242–4252.

(302) Bulo, R. E.; Michel, C.; Fleurat-Lessard, P.; Sautet, P. *J. Chem. Theory Comput.* 2013, *9*, 5567–5577.

(303) Zheng, M.; Waller, M. P. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 2016, *6*, 369–385.

(304) Csányi, G.; Albaret, T.; Payne, M. C.; De Vita, A. *Phys. Rev. Lett.* 2004, *93*, 1–4.

(305) Jacob, C. R.; Beyhan, S. M.; Bulo, R. E.; Gomes, A. S. P.; Götz, A. W.; Kiewisch, K.; Sikkema, J.; Visscher, L. *J. Comput. Chem.* 2011, *32*, 2328–2338.

(306) Te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* 2001, *22*, 931–967.

(307) Lin, H.; Zhang, Y.; Pezeshki, S.; Wang, B.; Truhlar, D. G., *QMMM2015*; University of Minnesota: Minneapolis, 2015.

(308) Gaussian16, Available at: http://www.gaussian.com.

(309) Duster, A.; Wang, C.-H.; Lin, H. *Molecules* 2018, *23*, 2170–2186.

(310) Mones, L.; Jones, A.; Götz, A. W.; Laino, T.; Walker, R. C.; Leimkuhler, B.; Csányi, G.; Bernstein, N. *J. Comput. Chem.* 2015, *36*, 633–648.

(311) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; Vandevondele, J. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 2014, *4*, 15–25.

(312) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 2013, *3*, 198–210.

(313) Okamoto, T.; Yamada, K.; Koyano, Y.; Asada, T.; Koga, N.; Nagaoka, M. *J. Comput. Chem.* 2011, *32*, 932–942.

(314) Zheng, M.; Kuriappan, J. A.; Waller, M. P. *Int. J. Quant. Chem.* 2017, *117*, e25336.

(315) Zheng, M.; Waller, M. P. *J. Comput. Chem.* 2018, *39*, 799–806.

(316) Watanabe, H. *Molecules* 2018, *23*, 1882–1897.

(317) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* 2015, *1-2*, 19–25.

(318) Böckmann, M.; Doltsinis, N. L.; Marx, D. *J. Chem. Theory Comput.* 2015, *11*, 2429–2439.

(319) CPMD, Available at: http://www.cpmd.org.

(320) Boereboom, J. M.; Potestio, R.; Donadio, D.; Bulo, R. E. *J. Chem. Theory Comput.* 2016, *12*, 3441–3448.

(321) Field, M. J. *J. Chem. Theory Comput.* 2017, *13*, 2342–2351.

(322) Field, M. J. *J. Chem. Theory Comput.* 2008, *4*, 1151–1161.

(323) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 2014, *4*, 101–110.

(324) Lu, Y.; Farrow, M. R.; Fayon, P.; Logsdail, A. J.; Sokol, A. A.; Catlow, C. R. A.; Sherwood, P.; Keal, T. W. *J. Chem. Theory Comput.* 2018, DOI: 10.1021/acs.jctc.8b01036.

(325) Dohm, S.; Spohr, E.; Korth, M. *J. Comput. Chem.* 2017, *38*, 51–58.

(326) Senn, H. M.; Thiel, W. In *Atomistic approaches in modern biology*, Reiher, M., Ed.; Springer: Berlin, 2006; Vol. 268, pp 173–290.

(327) Eastman, P. et al. *J. Chem. Theory Comput.* 2013, *9*, 461–469.

(328) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. *PLOS Comput Biol* 2017, *13*, 1–17.

(329) Turney, J. M. et al. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 2012, *2*, 556–565.

(330) Parrish, R. M. et al. *J. Chem. Theory Comput.* 2017, *13*, 3185–3197.

(331) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. *Biophys. J.* 2015, *109*, 1528–1532.

(332) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; De Jong, W. A. *Comput. Phys. Commun.* 2010, *181*, 1477–1489.

(333) Wilkins-Diehr, N.; Crawford, T. D. *Comput. Sci. Eng.* 2018, *20*, 26–38.

(334) Krylov, A.; Windus, T. L.; Barnes, T.; Marin-Rimoldi, E.; Nash, J. A.; Pritchard, B.; Smith, D. G. A.; Altarawy, D.; Saxe, P.; Clementi, C.; Crawford, T. D.; Harrison, R. J.; Jha, S.; Pande, V. S.; Head-Gordon, T. *J. Chem. Phys.* 2018, *149*, 180901-1–180901-11.

(335) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* 1983, *79*, 926–935.

(336) Petersson, G.; Al-Laham, M. A. *J. Chem. Phys.* 1991, *94*, 6081–6090.

(337) Dunning Jr, T. H. *J. Chem. Phys.* 1989, *90*, 1007–1023.

(338) Roe, D. R.; Cheatham III, T. E. *J. Chem. Theory Comput.* 2013, *9*, 3084–3095.

(339) Nguyen, H.; Roe, D.; Swails, J.; Case, D. PYTRAJ: Interactive data analysis for molecular dynamics simulations, Available at: https://github.com/Amber-MD/pytraj.

(340) Head-Gordon, T.; Hura, G. *Chem. Rev.* 2002, *102*, 2651–2670.

(341) Yu, H. A.; Karplus, M.; Pettitt, B. M. *J. Am. Chem. Soc.* 1991, *113*, 2425–2434.

(342) Gaigeot, M. P.; Vuilleumier, R.; Sprik, M.; Borgis, D. *J. Chem. Theory Comput.* 2005, *1*, 772–789.

(343) Qian, P.; Lu, L.-N.; Yang, Z.-Z. *Can. J. Chem.* 2009, *87*, 1738–1746.

(344) Gao, J.; Freindorf, M. *J. Phys. Chem. A* 1997, *101*, 3182–3188.

(345) Mennucci, B.; Martınez, J. M. *J. Phys. Chem. B* 2005, *109*, 9818–9829.

(346) Guo, H.; Karplus, M. *J. Phys. Chem.* 1992, *96*, 7273–7287.

(347) Han, W.-G.; Suhai, S. *J. Phys. Chem.* 1996, *100*, 3942–3949.

(348) Katz, J. L.; Post, B. *Acta Crystallogr.* 1960, *13*, 624–628.

(349) Kitano, M.; Fukuyama, T.; Kuchitsu, K. *Bull. Chem. Soc. Jpn.* 1973, *46*, 384–387.

(350) Minami, H.; Iwahashi, M. *Int. J. Spectrosc.* 2011, *2011*.

(351) Mirkin, N. G.; Krimm, S. *J. Mol. Struct.: THEOCHEM* 1995, *334*, 1–6.

(352) Buck, M.; Karplus, M. *J. Phys. Chem. B* 2001, *105*, 11000–11015.

(353) Xiao, X.; Tan, Y.; Zhu, L.; Guo, Y.; Wen, Z.; Li, M.; Pu, X.; Tian, A. *J. Mol. Model.* 2012, *18*, 1389–1399.

(354) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* 2006, *65*, 712–725.

(355) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Natl. Acad. Sci.* 2006, *103*, 915–920.

(356) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* 1969, *51*, 2657–2664.

(357) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* 1996, *14*, 33–38.

(358) VMD, Available at: http://www.ks.uiuc.edu/Research/vmd/.

(359) Baker, E.; Hubbard, R. *Prog. Biophys. Mol. Biol.* 1984, *44*, 97–179.