University of Mississippi

eGrove

Electronic Theses and Dissertations | Graduate School

2015

Accurate Computation Of Molecular Properties From Novel Applications Of Quantum Mechanical Wavefunction Methods

James Coleman Howard
University of Mississippi

Follow this and additional works at: https://egrove.olemiss.edu/etd

 Part of the Chemistry Commons

Recommended Citation

Howard, James Coleman, "Accurate Computation Of Molecular Properties From Novel Applications Of Quantum Mechanical Wavefunction Methods" (2015). *Electronic Theses and Dissertations*. 924.
https://egrove.olemiss.edu/etd/924

This Dissertation is brought to you for free and open access by the Graduate School at eGrove. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of eGrove. For more information, please contact egrove@olemiss.edu.

ACCURATE COMPUTATION OF MOLECULAR PROPERTIES FROM NOVEL APPLICATIONS OF QUANTUM MECHANICAL WAVEFUNCTION METHODS

A Dissertation
presented in partial fulfillment of requirements
for the degree of Doctor of Philosophy
in the Department of Chemistry and Biochemistry
The University of Mississippi

by

J. Coleman Howard

December 2015

Copyright J. Coleman Howard 2015
ALL RIGHTS RESERVED

ABSTRACT

High-accuracy quantum mechanical (QM) wavefunction methods have been applied to compute molecular properties of weakly-bound clusters. This work focuses on both extending the applicability of robust theoretical methods to larger systems and also determining the inherent accuracy of *ab initio* methods when compared to experimentally measured properties. Described here is the development of an efficient many-body approach that offers the ability to reduce both the time and the computational resources normally required to compute these properties reliably. The *N*-body:Many-body QM:QM technique has been extended to compute harmonic vibrational frequencies of clusters. Applying this methodology to small hydrogen-bonded clusters demonstrates that this approach yields both optimized geometries and harmonic vibrational frequencies in excellent agreement with the "gold standard" of correlated wavefunction methods, the CCSD(T) method, but with much greater efficiency. In addition, this work includes careful calibration studies to examine the basis set convergence of harmonic frequencies to determine which basis sets can be employed to obtain *ab initio* frequencies lying near the complete basis set (CBS) limit. These benchmark values are used to calibrate more efficient methods including *ab initio* methods, various density functional approximations and water potentials. Lastly, anharmonic vibrational frequencies and dissociation energies have been computed for small hydrogen-bonded dimers, allowing for a direct comparison to experiment.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation under Grant numbers EPS-0903787, CHE-0957317, CHE-1156713 and CHE-1338056. I have also been supported by the University of Mississippi Department of Chemistry and Biochemistry and the University of Mississippi Graduate School. I would also like to thank my advisor Dr. Gregory S. Tschumper and the entire Tschumper Research Group, as well as my committee members Dr. Steven Davis, Dr. Jared Delcamp, Dr. Robert Doerksen and Dr. Nathan Hammer. Lastly, I would like to thank the University of Missisippi Department of Chemistry and Biochemistry, Chairman Dr. Charles Hussey and Graduate Program Coordinator Dr. Walter Cleland.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Theoretical Background

Noncovalent interactions are the relatively weak interactions existing between molecules. These interactions are considered "weak" insofar as the energy associated with such an interaction is typically on the order of a few kcal mol^{-1}, much smaller than the energy associated with the covalent bond that binds together an O_2 molecule, for example. Nevertheless, it is difficult to overstate the importance of these interactions. From the structure of DNA to the unique properties of water and even to the very existence of condensed phases, noncovalent interactions play a variety of critical roles in the natural world. The origin of these interactions differs depending on the context. For example, the attractive interaction between stacked DNA bases is very different from the binding of two water molecules. Whether we speak of a stacking interaction in the case of DNA bases or a hydrogen bond formation in the case of the water molecules, the ability to accurately predict the details of either phenomenon requires the tools of *ab initio* quantum mechanical wavefunction methods.

1.1.1 The Hartree-Fock Approximation

The central problem of quantum chemistry is finding approximate solutions to the nonrelativistic time-independent electronic Schrödinger equation for a collection of nuclei

Portions of this chapter are reproduced with permission from "Wavefunction methods for the accurate characterization of water clusters," J. Coleman Howard and Gregory S. Tschumper, *WIRES: Comp. Mol Sci. 4*(3), Copyright © 2013, Wiley Periodicals, Inc.

and electrons.[1]

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

Here, $\hat{H}$ is the Hamiltonian operator acting on the wavefunction Ψ to yield the total energy E. For a collection of N electrons and M nuclei, the Hamiltonian (in atomic units) takes the form of Equation 1.2.

$$\hat{H} = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{1.2}$$

The terms in Equation 1.2 represent, from left to right, the kinetic energy of the electrons, the kinetic energy of the nuclei, the attraction between electrons and nuclei, electron-electron repulsion and nuclear-nuclear repulsion. In quantum chemistry, a (very) common assumption is that the nuclei are fixed with respect to the much smaller, faster electrons. By invoking this Born-Oppenheimer approximation, the kinetic energy of the nuclei (second term in Equation 1.2) becomes zero, and the nuclear-nuclear repulsion (last term in Equation 1.2) becomes a constant. In this picture, nuclei move in a potential created by the electrons, leading to a concept central to many areas of chemistry, the potential energy surface (PES).

As the foundation of many *ab initio* methods, the wavefunction of an N-electron system is approximated as an $N \times N$ Slater determinant, shown in Equation 1.3.

$$\Psi_0 = |\phi_{P_1}\phi_{P_2}\dots\phi_{P_N}| = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_{P_1}(\mathbf{x_1}) & \phi_{P_2}(\mathbf{x_1}) & \dots & \phi_{P_N}(\mathbf{x_1}) \\ \phi_{P_1}(\mathbf{x_2}) & \phi_{P_2}(\mathbf{x_2}) & \dots & \phi_{P_N}(\mathbf{x_2}) \\ \vdots & \vdots & & \vdots \\ \phi_{P_1}(\mathbf{x_N}) & \phi_{P_2}(\mathbf{x_N}) & \dots & \phi_{P_N}(\mathbf{x_N}) \end{vmatrix} \tag{1.3}$$

Here, the ϕ_P represent occupied spin molecular orbitals, denoted by subscripts and organized into columns, shown in the more explicit representation on the right-hand side of the equation. The spin molecular orbitals are assumed to be orthonormal. Because these are spin orbitals, the $\mathbf{x}$ coordinates contain both spatial and spin coordinate information, and the

subscripts on these denote a particular electron. Interchanging two rows of this determinant would be equivalent to exchanging the coordinates of the two electrons labeled by those two rows. Since swapping rows of a determinant changes its sign, we can see that this form of the wavefunction satisfies the antisymmetry requirement of fermions. Now, consider the effect of two of the ϕ_P being identical. This is analogous to having two electrons in the same spin orbital. The properties of determinants also dictate that if two columns are identical, then the determinant is zero, satisfying the Pauli exclusion principle.[2]

Setting aside the constant nuclear repulsion term, it will be convenient to write the Born-Oppenheimer Hamiltonian in a compact form (Equation 1.4).

$$\hat{H} = \sum_{i=1}^{N} \hat{h}(i) + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}} \tag{1.4}$$

Here, the electron kinetic energy and the electron-nuclear attraction have been combined into the one-electron operator $\hat{h}(i)$, implying that the operator depends on the spatial coordinates of electron i. The expectation value of the electronic energy is then given by Equation 1.5.[2]

$$\begin{aligned} E_0 &= \int\int \dots \int \Psi_0^* \ \hat{H} \ \Psi_0 \ d\mathbf{x_1} \dots d\mathbf{x_{N-1}} d\mathbf{x_N} = \left\langle \Psi \,\middle|\, \hat{H} \,\middle|\, \Psi \right\rangle \\ &= \sum_{a=1}^{N} \left\langle \phi_{P_a}(\mathbf{x_1}) \,\middle|\, \hat{h}(1) \,\middle|\, \phi_{P_a}(\mathbf{x_1}) \right\rangle + \frac{1}{2} \sum_{a=1}^{N} \sum_{b=1}^{N} \left\langle \phi_{P_a}(\mathbf{x_1}) \phi_{P_b}(\mathbf{x_2}) \left| \frac{1}{r_{12}} \right| \phi_{P_a}(\mathbf{x_1}) \phi_{P_b}(\mathbf{x_2}) \right\rangle \\ &\quad - \frac{1}{2} \sum_{a=1}^{N} \sum_{b=1}^{N} \left\langle \phi_{P_a}(\mathbf{x_1}) \phi_{P_b}(\mathbf{x_2}) \left| \frac{1}{r_{12}} \right| \phi_{P_a}(\mathbf{x_2}) \phi_{P_b}(\mathbf{x_1}) \right\rangle \end{aligned} \tag{1.5}$$

Equation 1.5 introduces the bra-ket notation that implies integration over the coordinates of all electrons. The summations here do not index electrons, but instead occupied spin molecular orbitals. The symmetry of the Slater determinant and the fact that the Hamiltonian only contains one- and two-electron operators allow us to arbitrarily label each electron in the one-electron terms as electron 1 and likewise the electrons in the two-electron terms as

electrons 1 and 2. By assuming the order of the electronic coordinates in the integrals above and replacing the two electron repulsion integrals with a single antisymmetrized integral, Equation 1.5 can be expressed in an even more compact form (Equation 1.6).

$$E_0 = \sum_{a=1}^{N} \left\langle \phi_{P_a} \left| \hat{h} \right| \phi_{P_a} \right\rangle + \frac{1}{2} \sum_{a=1}^{N} \sum_{b=1}^{N} \left\langle \phi_{P_a} \phi_{P_b} \left| \right| \phi_{P_a} \phi_{P_b} \right\rangle \tag{1.6}$$

To find the spin orbitals needed to compute the ground state energy in Equation 1.6, the energy is minimized with respect to the orbitals. Those orbitals which minimize Equation 1.6 are the Hartree-Fock (HF) spin orbitals, and the process of finding those orbitals is the Hartree-Fock Self-Consistent Field (HF-SCF) procedure. The basic idea of the HF-SCF procedure is to choose a set of basis functions (a basis set) from which to expand the molecular orbitals as linear combinations of these basis functions. In practice, the most common choice is to use functions that mimic the hydrogen atomic orbitals, utilizing multiple functions of differing angular momenta (i.e, s, p, d, f, etc.) depending on the chemical system under investigation. Quoting the text "Modern Quantum Chemistry" by Szabo and Ostlund,[2] "the choice of a basis set is more of an art than a science." Fortunately, the variational nature of the HF-SCF procedure defines a convergent method, where the HF molecular orbitals become closer to exact as the basis set becomes more complete. If the molecular orbitals are expanded from a set of K atomic orbital basis functions, then the HF-SCF procedure yields a total of $2K$ molecular orbitals. The N lowest-energy molecular orbitals will then be occupied by an electron, and the $2K - N$ other orbitals are called unoccupied or virtual orbitals. The number of electron repulsion integrals to compute grows as the fourth power of the number of basis functions so the choice of basis set is limited by time and computational considerations. Often, it is the decades of existing computational chemistry literature that provide the best motivation for choosing a particular basis set. Section 1.2.2 expands on this topic of choosing an appropriate basis set.

1.1.2 Beyond the HF approximation

Even with a complete set of basis functions, the HF-SCF method has a serious deficiency in its treatment of electron-electron interactions, which can be seen by examining the Hartree-Fock equation for the energy (ϵ_a) of a particular spin orbital (ϕ_{P_a}).

$$\hat{h}\phi_{P_a}(1) + \sum_{b\neq a}^{N}\left(\int d\mathbf{x_2}\,|\phi_{P_b}(2)|^2\,\frac{1}{r_{12}}\right)\phi_{P_a}(1) - \sum_{b\neq a}^{N}\left(\int d\mathbf{x_2}\phi^*_{P_b}(2)\phi_{P_a}(2)\frac{1}{r_{12}}\right)\phi_{P_b}(1) = \epsilon_a\phi_{P_a}(1) \quad (1.7)$$

In the second term of Equation 1.7, the parenthetical part can be identified as a Coulomb operator. Recognizing the probability factor for electron 2 immediately inside the integrand, this Coulomb term may be interpreted as an average repulsion experienced by electron 1, and summing over all other molecular orbitals effectively means that the energy of each orbital only includes the average repulsion from the other electrons.[2] For this reason, Hartree-Fock theory is often called a mean-field theory.

The HF energy usually accounts for more than 99% of the total electronic energy.[1] However, to accurately predict chemical phenomena, this is often not enough. This missing energy is necessarily negative because of the variational nature of HF theory, and it is referred to as "correlation energy."[a] If the set of spin orbitals $\{\phi_P\}$ is complete,[b] then the exact N-electron wavefunction can be expressed as a linear combination of all the possible Slater determinants that can be formed from these spin orbitals. Besides the ground state wavefunction, additional Slater determinants are formed by substituting virtual orbitals for occupied orbitals. These replacements can be viewed as electron excitations, producing

[a] Technically, this correlation energy is "dynamical correlation energy", referring to the energy lowering by correlating the motions of the electrons, leading to decreased repulsion. This is distinct from so-called "static" correlation, which results when the HF wave function is not a good zeroth order approximation. Static correlation is not considered in this work.

[b] A complete set of spin orbitals is unobtainable in practice, and the wavefunction will only be exact within the space spanned by the one-electron expansion functions.

excited configurations.

$$|\Phi_0\rangle = |\Psi_0\rangle + \left(\frac{1}{2}\right)^2 \sum_{a,r} c_a^r \, |\Psi_a^r\rangle + \left(\frac{1}{6}\right)^2 \sum_{ab,rs} |\Psi_{ab}^{rs}\rangle + \left(\frac{1}{24}\right)^2 \sum_{abc,rst} |\Psi_{abc}^{rst}\rangle + \cdots \tag{1.8}$$

In Equation 1.8, the exact wave function $|\Phi_0\rangle$ is expanded as the sum of the HF reference wavefunction $|\Psi_0\rangle$ and all possible configurations including single excitations (second term), double excitations (third term), triple excitations (fourth term), etc. up through N-tuple excitations. The vector $|\Psi_a^r\rangle$ represents a Slater determinant formed by replacing Φ_{P_a} in $|\Psi_0\rangle$ with Φ_{P_r}. Each term has a coefficient associated with it that can be viewed as a relative weight to the determinant. Minimizing the total energy with respect to these coefficients is equivalent to diagonalizing the Hamiltonian matrix and yields the best possible approximation to the ground state energy within the space spanned by the Slater determinants. This method is referred to as configuration interaction (CI), and including all possible excitations defines the Full CI (FCI) method. In practice, the FCI matrix is far too large, and truncated CI methods are used instead, where determinants are, for example, limited to singly and doubly excited configurations (CISD). Unfortunately, the truncated CI models are not very useful either, as they lack an important quality called size consistency.[2] In a size-consistent method, the total energy of a collection of N particles becomes proportional to N as $N \to \infty$. The CI method is the simplest correlation method conceptually, though, and it demonstrates how the deficiencies in the HF wavefunction can be remedied by including excited determinants to capture electron correlation energy. In Section 1.3.1, some size-consistent correlation methods that are more commonly used are discussed.

1.2 Hydrogen Bonding

The focus of the research presented here is hydrogen-bonded systems. While perhaps best known as being the stabilizing interaction between water molecules, hydrogen bonding also has important roles in crystal packing, protein folding and DNA base pairing, for ex-

ample.[3] The concept of the hydrogen bond has existed for more than a century now, but a precise definition has proven difficult. As more theoretical and experimental evidence has become available, the accepted ideas on what constitutes hydrogen bonding have shifted.[4] The International Union of Pure and Applied Chemistry (IUPAC) established a task force in 2005 to determine a definition for the hydrogen bond.[4] Their recommendation was the following: "The hydrogen bond is an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X-H in which X is more electronegative than H, and an atom or a group of atoms in the same or different molecule, in which there is evidence of bond formation." A hydrogen bond is typically written X-H$\cdots$Y, where the X-H group is referred to as the "donor." Under the IUPAC definition, the only requirement for the "acceptor" Y is that it contain an electron-rich region, such as a lone pair of electrons in the case of the O-H$\cdots$O hydrogen bonds in water. This IUPAC definition does not require any one particular criterion to be met for there to exist evidence of a hydrogen bond, but instead offers a list of criteria that can be used to gauge the reliability in considering a particular interaction to be considered a hydrogen bond.

While the IUPAC definition of a hydrogen bond is sufficiently broad to include many of the more ambiguous cases, we will focus here on the common qualities of the stronger variety of hydrogen bond involving highly electronegative atoms, such as oxygen or fluorine. The stability of the hydrogen bond in these cases is largely electrostatic in origin.[4–6] For example, in a single water molecule, the electronegative oxygen atom causes a partial negative charge to accumulate on the oxygen and a partial positive charge on the hydrogen atoms. When multiple water molecules are present, the partially positive hydrogen atoms are attracted to the lone pair of a nearby oxygen, leading to hydrogen bond formation. Hydrogen bonding interactions are directional, and an X-H$\cdots$Y angle near 180° is typically associated with a stronger hydrogen bond.[4]

Experimentally, NMR spectra can be useful to establish evidence of hydrogen bond formation, as the proton deshielding is enhanced in a donor molecule.[4] Important to the

work presented here is the effect of hydrogen bonding on a vibrational spectrum. Upon hydrogen bond formation, the X-H bond length in the donor molecule is lengthened, and the X-H vibrational frequency shifts to lower energy, commonly called a "red shift."[4,6] This can be a sensitive indicator of the hydrogen bonding environment and serve as a spectral "signature" for the hydrogen bond.

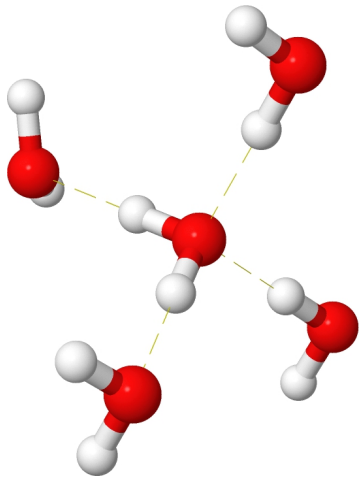


Figure 1.1. Depiction of a water molecule (center) accepting two and donating two hydrogen bonds denoted by the dashed lines O-H$\cdots$O.

1.3 Computational Considerations

Ab initio wavefunction-based methods for intermolecular interactions fall into two categories: perturbation approaches and supermolecular approaches. Symmetry adapted perturbation theory (SAPT) is a powerful technique that falls into the former classification. SAPT and its variants directly calculate the interaction energy and components thereof (electrostatic, induction, dispersion, and exchange repulsion) between two fragments (usually monomers) as a perturbation to non-interacting monomer wavefunctions or Kohn-Sham densities.[7–11] The procedure can be extended to larger clusters in principle. In practice, however, implementing the perturbation approach for a molecular cluster with more than two fragments is not straightforward. Three-body versions of SAPT and SAPT (DFT) exist,[12–17] but we are not aware of higher-order implementations at this time. Nevertheless,

the technique has provided a wealth of insight into noncovalent complexes because the lowest-order interactions tend to be the most important.

In water clusters and other hydrogen bonded systems, cooperative effects are significant, and higher-order interactions cannot be ignored. In fact, cooperative effects can account for more than 30% of the binding energy in $(H_2O)_n$ clusters (n>4).[18] The supermolecular approach includes these many-body interactions by determining the binding energy (E_{bind}) of a noncovalent cluster from the difference between the energies of the individual fragments ($f_1, f_2, \ldots, f_n$) and that of the cluster.

$$E_{bind} = E[\text{cluster}] - \sum_{i=1}^{n} E[f_i], \tag{1.9}$$

$E[X]$ represents the total electronic energy of species X computed with a size consistent method.[19,20] (Reference 21 includes a tutorial on the relationship between size consistency and the supermolecular approach.)

If the internal coordinates of the fragments do not change as the complex forms (i.e., the rigid or frozen monomer approximation), then E_{bind} is formally equivalent to the interaction energy (E_{int}), the latter requiring that the internal coordinates of each fragment be the same in both the complex and fragment computations.[22] Automated geometry optimization procedures are available for a wide range of DFT, semi-empirical and wavefunction methods, and as such, it is now common practice in supermolecular electronic structure studies to use fully optimized internal coordinates when computing E[cluster] and each $E[f_i]$. The quantity computed in this manner via Equation 1.9 formally corresponds to the equilibrium binding energy (simply denoted E_{bind} here), differing from E_{int} by the energy required to distort each fragment from their optimized geometries to the ones adopted in the complex. This distortion can also be viewed as the monomer relaxation energy, a quantity with the same magnitude but opposite sign as the distortion energy.

Consequently, E_{bind} and E_{int} can be thought of as adiabatic and vertical quantities,

respectively, representing the energetic stabilization that results as n non-interacting fragments associate with each other to form a cluster, and their values are necessarily negative for a bound complex. In practice, less rigorous definitions are commonly employed for E_{int}, E_{bind} and closely related terms (e.g., stabilization energy, association energy, complexation energy and dissociation energy). Therefore, a great deal of confusion can be avoided if authors clearly define the terminology adopted in a particular study.

If the isolated fragments are identical (e.g., water molecules), then the summation can be eliminated because each $E[f_i]$ has the same value. For example, E_{bind} for a water cluster composed of n monomers reduces to the following expression.

$$E_{bind} = E[(H_2O)_n] - nE[H_2O] \tag{1.10}$$

The dissociation energy (D_e) is closely related to the binding energy, having the same magnitude but opposite sign because it refers to the reverse process. The term D_e is typically reserved for minima on a potential energy surface (PES).

Often it is convenient to use a reference other than the separated monomers. For example, the energy of a tetramer could be compared to the energy of two dimers rather than four monomers. When studying the sequential growth of water clusters, it may be useful to examine the energy associated with the addition of each water.

$$\Delta E_n^{n+1} = E[(H_2O)_{n+1}] - (E[(H_2O)_n] + E[H_2O]) \tag{1.11}$$

Potential energy surfaces of even the smallest water clusters possess a number of low-energy stationary points, and the number of low-lying minima on the PES grows with the value of n. The number of hydrogen bonding topologies increases exponentially with cluster size.[23] Consequently, the relative energy of the stationary points can be a more useful quantity than E_{bind}. Any stationary point can serve as a reference point, but the most

common choice is the structure presumed to be the global minimum, $(H_2O)_n^{g.m.}$.

$$\Delta E = E[(H_2O)_n] - E[(H_2O)_n^{g.m.}] \tag{1.12}$$

The equations presented in this section also apply to energies that include the zero-point vibrational energy (ZPVE) of each species, usually denoted by a superscript 0 (e.g., E^0_{bind}, D^0_e or ΔE^0). The simplest approaches obtain the ZPVE from the harmonic vibrational frequencies or employ suitable scaling factors[24,25] to estimate anharmonic effects, which are particularly significant in water clusters and other weakly bound systems. More rigorous procedures, such as vibrational perturbation theory, can be used to compute anharmonic vibrational frequencies and the corresponding ZPVE.[26] A recent paper by Temelso and Shields includes a superb overview of procedures for estimating vibrational anharmonicity in water clusters.[27,28] Analogous expressions can also be obtained for related thermodynamic quantities, (e.g., enthalpy, entropy and free energy) which have provided much insight into thermal and entropic effects on the formation of water clusters.[29]

1.3.1 Electron Correlation

Correlated wavefunction theory methods and sufficiently flexible atomic orbital (AO) basis sets are required to consistently compute accurate energetic quantities for Equations 1.9 – 1.12 as well as related properties (optimized geometries, harmonic vibrational frequencies, etc.). The most successful and widely utilized quantum mechanical electronic structure techniques for clusters of molecules held together by noncovalent interactions are based on many-body perturbation theory (MBPT) and coupled-cluster (CC) theory.[30] This collection includes Møller-Plesset perturbation theory (MPPT),[31] which is a particular partitioning scheme for the former, and quadratic configuration interaction (QCI) theory, which is a special case of the latter.[32–34]

Practically speaking, the level of electron correlation and the size of the basis set are limited by computational resources, because the computational demands of sufficiently

accurate wavefunction methods scale unfavorably with the size of the system (e.g., number of atoms, number electrons, number of basis functions). The hierarchy of single reference post-HF methods is defined by the categories of excited determinants or configurations included (i.e., single, double, triple, etc.) in the construction of the wavefunction.

Second-order Møller-Plesset perturbation theory (MP2) is generally considered the simplest correlated wavefunction method to provide reliable results for hydrogen bonds as long as dispersion does not make a relatively large contribution.[35] MP2 has the advantage of being a non-iterative, size-consistent technique that scales asymptotically as the fifth power with the size of the system (N^5). In practice, the scaling can be somewhat lower for systems that are not extremely large. Unfortunately, MP2 does not perform equally well for all types of noncovalent interactions, and it is known to significantly overbind dispersion-dominated interactions, particularly those between aromatic molecules that adopt face-to-face arrangements.[36–39]

More demanding post-HF methods are required to correctly describe the entire spectrum of noncovalent interactions. Although the coupled-cluster technique that includes all single and double substitutions (CCSD) tends to substantially underbind the systems for which MP2 overbinds,[40,41] reliable results can obtained by including a perturbative estimate of connected triple excitations (i.e., the CCSD(T) method). The CCSD(T) method is widely employed to generate benchmark binding energies in this context,[21,42–46] but QCISD(T) may also be an effective alternative for clusters composed of closed-shell fragments.[47] Unfortunately, the additional complexity associated with the CCSD(T) method (and the QCISD(T) method) increases the scaling to N^7. In terms of occupied (o) and virtual (v) orbitals, the CCSD(T) method actually scales as o^3v^4. So increasing the size of the AO basis set even for a small system can substantially increase the computational demands. Higher orders of Møller-Plesset perturbation theory (MP3, MP4, etc.) are not typically employed to evaluate electron correlation effects beyond the MP2 level because the series can sometimes diverge.[48–50] Quantum Monte Carlo (QMC) techniques offer an alternative approach

to computing electronic energies with high accuracy and are naturally suited to large-scale parallelization. Recent QMC results for water clusters[51–55] show great promise for further application and development.

1.3.2 Atomic Orbital Basis Sets

The AO basis set also significantly affects the accuracy of results obtained with post-HF wavefunction methods. Several popular classes of basis sets have recently been reviewed by Jensen in *WIREs Comp. Mol. Sci.*[56] In principle, it is desirable to use an extremely large basis set that lies near the complete basis set (CBS) limit. Unfortunately, the unfavorable scaling of computational demands with the number of virtual orbitals (*vide supra*) requires that truncated basis sets must be used in practice, which leads to basis set incompleteness error (BSIE). This quantity is merely the difference between a value computed with a particular finite basis set and the corresponding CBS value obtained with the same electronic structure method.

The development of Dunning's correlation consistent Gaussian basis sets[57–64] has been critical to understanding and overcoming the inherent limitations of truncated basis sets. The correlation consistent family (cc-pVXZ or simply XZ) was designed to systematically approach the CBS limit as the cardinal number (X) of the basis set increases. For water clusters, X simply corresponds to the maximum angular momentum of the basis functions. The versions augmented with diffuse functions (aug-cc-pVXZ or simply aXZ) are particularly important for hydrogen bonding and other noncovalent interactions. Diffuse functions have relatively small orbital exponents which leads to slower radial decay than their valence counterparts. They usually provide improved results for anions, excited states and long range interactions because they help describe regions farther from the nuclear centers that still have appreciable electron density. The addition of diffuse functions is generally critical for reliably modeling hydrogen bonding and noncovalent interactions.[21,65–68]

For hydrogen bonding, particularly in $(H_2O)_n$ clusters, it is often not necessary to

put diffuse functions on the H atoms, a practice that can be computationally advantageous. The electronegative atom to which H is bonded (O in this case) withdraws electron density, leaving H with a substantial partial positive charge ($\delta+$). Diffuse functions are not needed to describe the depleted electron density around this partially exposed proton. The opposite is true, however, for the heavier electronegative atom because it acquires an appreciable partial negative charge ($\delta-$). Our research group has dubbed these "heavy" aug-cc-pVXZ (or ha$X$Z) basis sets because diffuse functions are only added to the heavy (non-hydrogen) atoms. Other popular notations include aug′-cc-pVXZ and more recently jul-cc-pVXZ.[69] It is not clear, however, that the smaller jun-, may- and apr- variants of the partially augmented calendar correlation consistent basis sets[69] can be employed in CBS extrapolation schemes because the correlation consistency has been disrupted. It should be noted that these ha$X$Z style basis sets are not physically justified for all noncovalent interactions though. For example, diffuse functions on H may be necessary in dihydrogen bonding where one of the participating H atoms actually develops a partial negative charge ($\delta-$).

The smooth convergence of Dunning's basis sets has inspired several different techniques[61,70–74] for extrapolating both the HF and correlation energies to the CBS limit. Feller's exponential fit of $(H_2O)_2$ total energies calculated with a correlation consistent family of basis sets was certainly one of the earliest applications to water clusters.[70]

$$E_{\mathrm{HF}}^{X} = E_{\mathrm{HF}}^{\mathrm{CBS}} + Ae^{-bX} \tag{1.13}$$

This simple formula with 3 parameters can be expressed in a closed algebraic form for a set of energies calculated with 3 successive values of the cardinal number ($X-2$, $X-1$ and X).

$$E_{\mathrm{HF}}^{\mathrm{CBS}} = E_{\mathrm{HF}}^{X} - \frac{(E_{\mathrm{HF}}^{X} - E_{\mathrm{HF}}^{X-1})^2}{(E_{\mathrm{HF}}^{X} - E_{\mathrm{HF}}^{X-1}) - (E_{\mathrm{HF}}^{X-1} - E_{\mathrm{HF}}^{X-2})} \tag{1.14}$$

Both Equations 1.13 and 1.14 are written in terms of the HF energy because it is common to extrapolate the correlation energy separately, given their different convergence behav-

iors.[65,74–78]

It has been known for some time that the correlation energy converges asymptotically as an inverse power of the maximum angular momentum of the basis set.[79,80] Extrapolation schemes for the correlation energy are based on this observation and differ from one another in the actual exponents and number of terms in the function.[65,72,75,81,82] Martin[75] and Helgaker[65] each proposed 2-parameter formulae in this manner:

$$E_{\text{corr}}^{X} = E_{\text{corr}}^{\text{CBS}} + \frac{A}{(X+1/2)^4} \tag{1.15}$$

$$E_{\text{corr}}^{X} = E_{\text{corr}}^{\text{CBS}} + \frac{A}{X^3} \tag{1.16}$$

Martin's formula (Equation 1.15) has also been modified and tested with an additional sixth power term and a variable exponent as well.[72,75] Note that for the correlation consistent basis sets, X is equal to the maximum angular momentum for all first and second row atoms. Binding energies (E_{bind}) of water clusters have also been extrapolated using heuristic expressions based on inverse powers of angular momentum[28,83–85] and can exhibit different convergence behavior than interaction energies.[83] More recent extrapolation formulae such as that of Schwenke[82] have also been proposed, however there is still no consensus as to which procedure is the most accurate across a variety of chemical systems.[78]

1.3.3 Basis Set Superposition Error

The use of a finite AO basis set within the supermolecular approach introduces an inconsistency for any sort of dissociative (or associative) process that compares the energies of the parts (e.g., a set of fragments) to the energies of the whole (e.g., a cluster).[86–91] It applies to both the rupture (or formation) of a covalent bond as well as the dissociation (or binding) energies of noncovalent clusters such as Equations 1.10 and 1.11. This inconsistency is better known as (intermolecular) basis set superposition error (BSSE).[86,89]

BSSE arises from the fact that the fragments are allowed to "share" basis functions

when interacting (e.g., in a cluster) but not when separated from each other. Consequently, the magnitude of E_{bind} (or D_e) is artificially too large for conventional electronic structure theory computations. The most common method for correcting the BSSE is the counterpoise (CP) procedure.[87,88] CP corrections remove the inconsistency by calculating the energy of the fragments in the basis set of the cluster. This necessitates utilizing the geometries that the fragments adopt in the cluster, $E[f_i]^{\text{cluster basis}}_{\text{cluster geom}}$. If the fragment geometries are flexible rather than rigid, then the energy change associated with the fragment distortion (or relaxation, depending on your perspective) must also be computed in the fragment basis set by comparing the energy of the fragment at its optimized fragment geometry, $E[f_i]^{\text{fragment basis}}_{\text{fragment geom}}$, to that of its distorted geometry in the cluster, $E[f_i]^{\text{fragment basis}}_{\text{cluster geom}}$.

Applying the CP procedure to a dimer composed of monomers m_1 and m_2 yields the following expression for the binding energy.

$$\begin{aligned} E^{\text{CP}}_{bind} = {} & E[\text{dimer}] - E[m_1]^{\text{dimer basis}}_{\text{dimer geom}} - E[m_2]^{\text{dimer basis}}_{\text{dimer geom}} \\ & + \left(E[m_1]^{\text{monomer basis}}_{\text{dimer geom}} - E[m_1]^{\text{monomer basis}}_{\text{monomer geom}}\right) \\ & + \left(E[m_2]^{\text{monomer basis}}_{\text{dimer geom}} - E[m_2]^{\text{monomer basis}}_{\text{monomer geom}}\right) \end{aligned} \tag{1.17}$$

If the rigid monomer approximation is employed, then the terms in parentheses (distortion energies) do not need to be evaluated. Some ambiguities arise when applying the CP procedure to clusters larger than a dimer.[90,91] One straightforward option simply extends the procedure for 2 monomers in a dimer to each fragment in a cluster.

$$\begin{aligned} E^{\text{CP}}_{bind} = {} & E[\text{cluster}] - \sum_{i=1}^{n} E[f_i]^{\text{cluster basis}}_{\text{cluster geom}} \\ & + \sum_{i=1}^{n} \left(E[f_i]^{\text{fragment basis}}_{\text{cluster geom}} - E[f_i]^{\text{fragment basis}}_{\text{fragment geom}}\right) \end{aligned} \tag{1.18}$$

Again, if the rigid monomer approximation is employed, then the 2nd summation involving computations in the fragment basis set vanishes.

Regardless of the implementation, the CP procedure requires extra computational effort. The additional monomer computations in the monomer basis are relatively trivial, but those in the cluster basis set can actually be quite demanding (sometimes comparable to the computation on the whole cluster). There will be n of each type when there are no symmetrically equivalent fragments. Fortunately, these computations are completely independent and, therefore, well-suited to coarse-grained parallelization. In addition to the *a posteriori* procedure for computing E_{bind}^{CP} (e.g., Equations 1.17 and 1.18), a counterpoise-corrected PES can be defined.[90]

$$E^{\mathrm{CP}} = E[(\mathrm{H_2O})_n] + \sum_{i=1}^{n} \left(E_i[\mathrm{H_2O}]_{\text{cluster geom}}^{\text{monomer basis}} - E_i[\mathrm{H_2O}]_{\text{cluster geom}}^{\text{cluster basis}} \right) \tag{1.19}$$

A priori CP-corrected geometry optimizations and harmonic vibrational frequency computations can be obtained from the corresponding first and second geometrical derivatives of this expression (i.e., gradients and Hessians).

CP corrections are widely utilized in studies of water clusters and other noncovalent complexes. Diffuse functions have been shown to significantly reduce the BSSE in hydrogen-bonded clusters.[92] Correlation consistent basis sets that only augment the heavy (non-hydrogen) atoms are particularly attractive for water clusters because they often provide energetics closer to the CBS limit for these systems than their fully augmented versions.[21,93,94] Furthermore, uncorrected results obtained with these ha$X$Z basis sets are often closer to the CBS limit than the CP-corrected values for small water clusters.[21,93] It is not guaranteed that these trends will extend to larger $(\mathrm{H_2O})_n$ clusters, and the behavior of these ha$X$Z basis sets needs to be thoroughly evaluated for different hydrogen bonding topologies.

These results highlight two aspects of BSSE and CP corrections that are often overlooked. Specifically, large BSSE does not necessarily indicate a large error, and CP corrections are not guaranteed to move calculated values closer to the CBS limit.[21] These issues do not indicate a deficiency of the CP procedure but are, instead, a manifestation of basis

set incompleteness error.[90] BSIE, in contrast, actually does measure the error relative to the CBS limit. The two quantities (BSSE and BSIE) are not unrelated though. By definition, BSSE must vanish in the limit of a complete basis, and small BSSE indicates the results are close to the CBS limit (i.e., small BSIE).

In light of the previous paragraph, "basis set superposition *inconsistency*" (BSSI) is perhaps a more appropriate name than "basis set superposition *error*", but the former will not likely supplant the latter because the term BSSE is deeply entrenched in the literature. Fortunately, this inconsistency is less of an issue when relative energies (ΔE) are the quantity of interest rather than binding (E_{bind}) or dissociation energies (D_e). However, the CP procedure can still have an appreciable effect on ΔE for isomers with significantly different structures and/or hydrogen bonding topologies.

1.3.4 Explicitly Correlated Methods

Explicitly correlated methods provide a different route to the CBS limit without resorting to extrapolation techniques. These R12 or F12 procedures construct the wavefunction with explicit dependence on interelectronic distance and correctly model the electron-electron cusp to significantly accelerate convergence of the correlation energy. In fact, some explicitly correlated coupled-cluster calculations with small basis sets (X = D or T) have been shown to be as accurate as canonical coupled-cluster calculations performed with appreciably larger basis sets (X = T or Q).[95] Although the advantages of explicitly correlated methodologies were recognized nearly a century ago,[96] their application to larger systems has only become practical with recent advances. Several reviews detail the history and key breakthroughs, such as resolution of the identity (RI) techniques.[97–100] Today explicitly correlated MP2 and CCSD methods are available in a number of software packages implemented within various *ansatze*. While there is no consensus on the best approach for a broad range of chemical applications, most modern implementations provide more accurate energetics than the corresponding canonical wavefunction methods for almost no additional computational

effort.[101]

Explicitly correlated CCSD(T) methods deserve special attention[102,103] because direct inclusion of F12 terms in the perturbative estimate of connected triples substitutions is challenging. For example, the popular F12a and F12b implementations[104,105] available in the `Molpro`[106] software package do not directly affect the triples energy associated with the (T) part of the calculation. Scaling procedures have been introduced to estimate the effect of explicit correlation on the triples energy by assuming that it is proportional to the effect on the MP2 correlation energy.[105,107] The approximate triples energy obtained in this manner is denoted (T*).

$$\frac{E_{\mathrm{corr}}^{\mathrm{MP2-F12}}}{E_{\mathrm{corr}}^{\mathrm{MP2}}} = \frac{E_{\mathrm{corr}}^{(\mathrm{T}*)}}{E_{\mathrm{corr}}^{(\mathrm{T})}} \tag{1.20}$$

Unfortunately, CCSD(T*)-F12 approaches are not size consistent. The CCSD(T**)-F12 methods were introduced to correct this problem by employing a common scaling factor, usually that of the complex rather than the monomers.[102] The CCSD(T**)-F12a technique and the related dispersion-weighted approach have been shown to yield very accurate results for hydrogen bonded dimers.[103] Hopefully this impressive performance will also extend to larger $(H_2O)_n$ clusters.

1.4 Water Clusters

1.4.1 Accurate Computational Strategies

The evolution of quantum mechanical electronic structure studies of water clusters has followed a fairly consistent pattern since the earliest HF calculations on $(H_2O)_2$.[108–114] The water dimer is small enough that some of the most sophisticated methods available at a particular time can be applied to it. These "high-level" results for $(H_2O)_2$ can be used to calibrate less demanding "low-level" procedures. (Keep in mind that the definition of high- and low-level changes over time as computing power increases.) The expectation has been that methods yielding reliable results for the dimer will also give reliable results for the trimer. Techniques that work well for $(H_2O)_3$ can then be applied with reasonable confidence

to $(H_2O)_4$, and so forth. Eventually computational resources improve to the point that the high-level methods can be applied to larger clusters to test the accuracy of the the low-level methods. Then the cycle starts over again. Sometimes techniques that work well for $(H_2O)_2$ or $(H_2O)_3$ do not reliably transfer to larger clusters, but these setbacks often provide insight into the chemical physics of the noncovalent interactions of water clusters.

For several decades, MP2 has provided the foundation for accurate electronic structure computations on water clusters. Extensive calibration has shown that MP2 optimized structures and energies are quite reliable when obtained with sufficiently flexible basis sets (e.g., aug-cc-pVTZ for O and cc-pVTZ for H). Additionally, MP2 algorithms are efficient enough that CBS limit energetics can routinely be obtained for rather sizable $(H_2O)_n$ clusters with the extrapolation techniques or explicitly correlated methods (e.g., MP2-R12 or MP2-F12) described earlier. In fact, MP2 CBS limit binding energies were computed for $(H_2O)_{20}$ isomers in 2004.[115]

CCSD(T) computations are now feasible, but not necessarily routine, for clusters containing more than a dozen water molecules. The largest water cluster to which the method has been applied is likely $(H_2O)_{20}$,[116] an accomplishment which required significant CPU resources. With this more demanding procedure, basis sets tend to be limited to double- or triple-ζ quality, which can produce unreliable extrapolations of the CCSD(T) correlation energy. However, the difference between the MP2 and CCSD(T) correlation energies converges to the CBS limit more quickly and typically can be reliably described with such basis sets.[117,118] Consequently, the additive approach that is widely employed to generate benchmark binding energies for noncovalent dimers[42,43] can be employed to reliably estimate E_{bind} at the CCSD(T) CBS limit for $(H_2O)_n$ clusters. The most popular scheme computes a higher-order correlation correction by comparing the MP2 and CCSD(T) binding energies obtained with a modest basis set, simply denoted MBS.

$$\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}} = E_{bind}^{\mathrm{CCSD(T)/MBS}} - E_{bind}^{\mathrm{MP2/MBS}} \tag{1.21}$$

This quantity can then be combined with E_{bind} at the MP2 CBS limit to estimate the binding energy at the CCSD(T) CBS limit.

$$E_{bind}^{\mathrm{CCSD(T)/CBS}} \approx E_{bind}^{\mathrm{MP2/CBS}} + \delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}} \tag{1.22}$$

Other combinations are possible. For example, the MP2 values could be replaced by the corresponding CCSD ones, or QCISD(T) could be used instead of CCSD(T) to estimate the higher-order correlation effects. Janowski and Pulay have offered an alternate interpretation of this additive scheme.[47] By regrouping the terms, the process can be viewed as a basis set correction rather than a higher-order correlation correction.

$$E_{bind}^{\mathrm{CCSD(T)/CBS}} \approx E_{bind}^{\mathrm{CCSD(T)/MBS}} + \left(E_{bind}^{\mathrm{MP2/CBS}} - E_{bind}^{\mathrm{MP2/MBS}} \right) = E_{bind}^{\mathrm{CCSD(T)/MBS}} + \delta_{\mathrm{MBS}}^{\mathrm{CBS}} \tag{1.23}$$

Regardless of the interpretation, binding and relative energies obtained with this additive approach agree closely with available data from explicitly correlated CCSD(T) computations on water clusters and other hydrogen bonded systems. For benchmark work, it should be noted that double-ζ basis sets may not provide sufficiently reliable $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ values for some noncovalent dimers.[41,102]

The next few sections review high-accuracy studies of $(H_2O)_n$ clusters, starting with the dimer and progressing to larger values of n. Comparison of MP2 and CCSD(T) CBS limit energetics has started to reveal that not all hydrogen bonds in water clusters are created equal. The two methods yield nearly identical results for $(H_2O)_n$ systems with similar hydrogen bonding topologies, such as a collection of cyclic structures in which each water donates one hydrogen bond and accepts another (homodromic hydrogen bonding networks). If more diverse networks are considered with not only single donors and single acceptors but also double donors and/or double acceptors, then non-negligible discrepancies can arise between the MP2 and CCSD(T) methods. Other factors can lead to differences between the two methods (e.g., significant geometrical distortions about the O–H$\cdots$O hydrogen bond),

but MP2 computations still play a vital role in the study of water clusters. When studying water clusters with qualitatively different hydrogen bonding networks, a more sophisticated method, such as CCSD(T), should be used to obtain reliable results via explicitly correlated techniques, extrapolation of the CCSD(T) correlation energy or an additive approximation (e.g., in conjunction with MP2 data).

The following discussion is focused primarily on investigations that have used high-accuracy wavefunction methods to compute energies and structures of small water clusters. The excessive computational demands associated with these correlated electronic structure methods have limited their extension to computing other properties of interest, such as vibrational frequencies. The reliable computation of spectra and other properties with accurate *ab initio* quantum chemistry techniques is an area of great interest,[27,54,119–122] particularly in light of the recent advances in spectroscopic techniques for probing small neutral water clusters.[123,124]

1.5 The Water Dimer

The water dimer has attracted significant interest as the smallest water cluster and a prototype for hydrogen bonding, in general. Furthermore, the water dimer is the foundation for understanding the pairwise interactions which account for a significant portion of the binding energy from small hydrogen-bonded clusters all the way to the bulk phase. The first dimer calculations were performed by Morokuma and Pedersen in 1968 at the HF level with a minimal basis set.[108] Since these earliest *ab initio* investigations[109–114,125,126] a variety of correlated post-HF methods and a wide range of basis sets have been used to gain insight into $(H_2O)_2$ and hydrogen bonding. These dimer studies have revealed much about the applicability and convergence properties of the methods themselves. With only two non-hydrogen atoms, the water dimer is well-suited for high accuracy electronic structure computations. So it is unsurprising that the dimer seems to regularly be the target of very sophisticated theoretical methods. Our attention will be focused on these high-accuracy

studies, with an emphasis on their relationship to the computational challenges described in the previous section.

Experiment and theory have long been in agreement that the minimum-energy structure of the water dimer is of C_s symmetry with a nearly linear hydrogen bond.

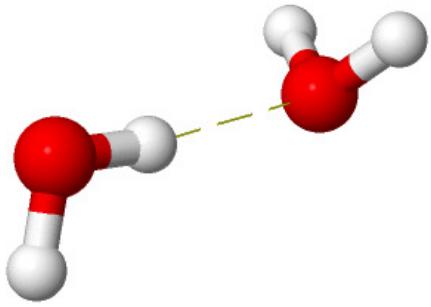


Figure 1.2. The global minimum C_s $(H_2O)_2$ structure

For this structure, a key intermolecular parameter of interest is the separation between the two oxygen atoms. Early microwave spectroscopic studies gave a vibrationally averaged O$\cdots$O distance of about 2.97 Å,[127] which is expected to be somewhat larger than the corresponding equilibrium value. Within each water monomer, the geometries remain mostly rigid upon dimer complexation. However, the covalent O-H bond participating in the O-H$\cdots$O hydrogen bond is appreciably elongated in the dimer. So the parameter R(O-H_d), where d denotes donor, and its change relative to the monomer value are of particular interest.

Throughout the 1990s, the structural properties of the global minimum were thoroughly explored at the MP2 level.[76,128–130] In particular, the efforts of Xantheas and Hobza provided much insight into the basis set dependencies of MP2 geometry optimizations on both CP-corrected and uncorrected PESs.[83,84,130,131] By including electron correlation at the MP2 level, the equilibrium distance R(O$\cdots$O) consistently decreases relative to the HF value. With the aTZ basis set, the optimal value for R(O$\cdots$O) is 2.905 Å.[130] The R(O-H_d) elongation at the MP2/aTZ level leads to a distance of 0.969 Å, nearly 0.01 Å longer than R(O-H) in the optimized monomer computed at the same level of theory.

In these works, the convergence pattern of the dimer geometry was thoroughly examined. Optimization of R(O$\cdots$O) was carried out at the MP2/aQZ and MP2/a5Z levels. As would eventually be seen for larger water clusters with homodromic hydrogen bonding networks, the MP2/aTZ geometries were found to be well converged. The MP2/a5Z optimal value for R(O$\cdots$O) was 2.905 Å,[83] a value perfectly matched with the much smaller aTZ basis set.

Although not routine because of the additional computations required, it is possible to perform CP-corrected geometry optimizations.[132] With this approach, the BSSE is removed *a priori*, rather than by removing the BSSE *a posteriori* with a series of energy calculations at the optimized geometry on the uncorrected PES. Xantheas and Hobza both evaluated the effects of CP-corrected geometry optimizations for the C_s global min.[83,130] The CP corrections consistently increase the distance between the two monomers, an observation that is consistent with its effect on the binding energy. For instance, unconstrained optimizations at the MP2/aDZ level produce R(O$\cdots$O) values of 2.921 and 2.977 Å for the standard and CP-corrected procedures, respectively.[130] On the other hand, the optimal R(O-H_d) value is unaffected by the method of optimization, as both techniques yield the same value to within 0.001 Å regardless of the basis set. Intermolecular geometrical parameters of $(H_2O)_2$ appear to converge more slowly with respect to X on the CP-corrected PES. It appears that the computation of accurate energetics for this system, however, does not require CP-corrected geometry optimizations. As noted by Hobza,[130] "Good agreement between stabilization energies resulting from CP-corrected and standard PESs gives evidence that stabilization energy, contrary to complex geometry, does not depend on the way in which the BSSE was covered (*a priori* or *a posteriori*)."

The MP2 CBS limit binding energy of the $(H_2O)_2$ global minimum was estimated in many studies through systematic improvement of basis set[73,76,83,92,130,133,134] and with explicitly-correlated calculations.[76,135] Large basis set HF calculations combined with MP2 single point energies obtained with more compact basis sets have been used to estimate an

MP2 CBS limit E_{bind} of -4.95 kcal mol^{-1}.[133] Xantheas[83] calculated the D_e of the water dimer as 4.88 kcal mol^{-1} at the MP2/a5Z level, including corrections for BSSE. Klopper performed explicitly-correlated MP2 computations on the global minimum and found that their MP2-R12 binding energy (-4.92 kcal mol^{-1})[136] was directly between the other two values. More recent estimates of E_{bind} at the MP2 CBS limit are in excellent agreement with these values, despite using slightly different geometries.[76,93,135]

As previously discussed, the size of the water dimer makes it an attractive candidate for calibrating methods and examining basis set effects. Halkier *et al.* examined the MP2 binding energy of the complex as the basis set is improved from aDZ all the way to a5Z.[73] That study also looked at the difference in using the doubly-augmented versions of Dunning's correlation consistent basis sets (daug-cc-pVXZ) and the unaugmented forms, as well. They found that employing the singly-augmented basis sets reduced the CP correction by an order of magnitude, relative to the corresponding XZ basis. The authors of that study also noted the surprising performance of aDZ without CP corrections, certainly a fortuitous result, but one that has remained useful for the study of larger homodromic cyclic clusters (*vide infra*). The doubly-augmented basis sets were not found to improve results, as the basis set errors were always larger with the daug-cc-pVXZ computed E_{bind} in comparison to MP2 CBS estimates.

In a more general theoretical study of hydrogen-bonded dimers, Halkier, Helgaker and others examined in great detail how the binding energy converges with respect to basis set for correlated calculations.[92] A major finding of that study was the unsystematic convergence of both the Hartree-Fock binding energies and of the correlation contributions to E_{bind}. For the SCF part of the binding energy, the convergence behavior was not affected by CP corrections. It was very unsystematic and even nonmonotonic for ΔE^{CP} in the water dimer. However, the authors discovered an important trend for the correlation contributions. The CP-corrected correlation contributions to E_{bind} always converged from above, suggesting that an average of the CP-corrected and standard correlation contributions to the binding energy

would be a useful approximation. Further, they found that removing the BSSE from the correlation energy, leaving only the basis set incompleteness led to a "monotonic, smooth, slow and systematic" convergence for $\Delta E_{\rm corr}^{\rm CP}$. The authors found that Helgaker's X^{-3} form of Equation 1.16 could be applicable to binding energies as well, as long as the basis set inconsistency had been removed first.

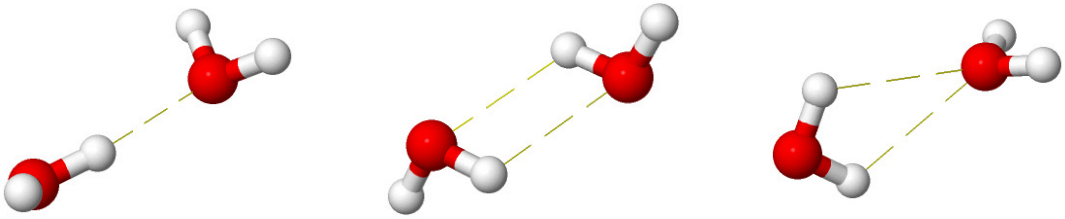
With the structural and energetic profile of the C_s global minimum well defined at the MP2 level, researchers began to explore the effects of higher-order correlation in detail. One study, for example,[93] applied a series of correlated methods and large basis sets in the spirit of the focal point analysis developed by Allen and co-workers.[117,118,137–140] The dimer global minimum structure was fully optimized at the CCSD(T) level with all electrons correlated and a triple-ζ basis set with two sets of polarization functions augmented with higher angular momentum and diffuse functions (TZ2P(f,d)+dif). The CCSD(T) optimized geometrical parameters agreed very well with prior MP2/aTZ results despite differences in the basis sets. The interoxygen separations differ by less than 0.01 Å, and R(O-H_d) values are essentially identical. The deviation between the O$\cdots$O-H_d hydrogen bond angles was found to be smaller than 0.2°.

In addition to extrapolating CCSD(T) energies to the CBS limit from basis sets as large as a6Z, the effects of connected quadruple excitations were estimated from BD(TQ) calculations.[93] CCSD(T) results slightly increase the dissociation energy, relative to MP2. The higher-order correlation effects from BD(TQ) nearly cancel the CCSD(T) effects, however. At the CBS limit, electron correlation beyond the MP2 level only increased the dissociation energy by 0.02 kcal mol^{-1}. It should be noted that the same CBS results were obtained with the ha$X$Z series of basis sets. The average CBS extrapolated D_e for the ha$X$Z series differs from the corresponding aXZ value by less than 0.01 kcal mol^{-1}. The effect of correlating the core electrons was evaluated at the MP2/aug-cc-pCV5Z level and increased the $D_e^{\rm CP}$ by 0.04 kcal mol^{-1}, relative to MP2(FC)/aug-cc-pCV5Z results. CCSD(T)/cc-pCVTZ relativistic corrections were factored into D_e and found to have a tiny destabilizing effect, on the order

of 0.002 kcal mol^{-1}. The best estimate of the CP-corrected D_e from this analysis is 5.02 kcal mol^{-1}. So the overall improvement beyond MP2 for describing the energetics of the $(H_2O)_2$ global minimum is only on the order of a few hundredths of a kcal mol^{-1}, which was consistent with other estimates of electron correlation effects beyond MP2.[76,135]

While the C_s dimer is the only minimum on the $(H_2O)_2$ PES, other low-lying stationary points have been identified and thoroughly characterized with *ab initio* wavefunction methods. In 1990, the structures of ten stationary points on the $(H_2O)_2$ PES were characterized via MP2 optimizations and frequencies as well as MP4 energies.[128] The authors identified three transitions states which effectively interconvert equivalent forms of the C_s global minimum. These are depicted in Figure 1.3.

Figure 1.3. Transition states on the $(H_2O)_2$ PES



(a) Open C_1 TS (b) Cyclic C_i TS (c) Bifurcated C_{2v} TS

The open C_1 TS in Figure 1.3a corresponds to interchanging the hydrogen atoms in the acceptor molecule and is the lowest-lying of the transition states, lying less than 0.6 kcal mol^{-1} above the C_s global minimum.[93,128] The cyclic C_i transition structure (Figure 1.3b) represents a pathway to interchanging donor and acceptor molecules. The bifurcated transition state (Figure 1.3c) "scrambles" all the hydrogens,[128] but lies at least 1.7 kcal mol^{-1} above the global minimum at MP2, MP4 and CCSD(T) levels.[93,128] The same process can be achieved via the two lower energy transition states.

Recently, Lane *et al.* have reported a series of impressive computations on the $(H_2O)_2$ global minimum, up to and including the CCSDTQ and partially augmented QZ basis

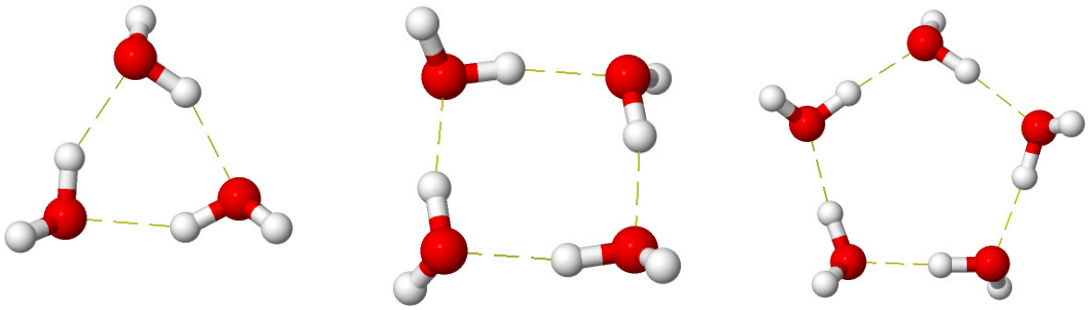
sets.[94,141] For CCSD(T)/aQZ optimizations, the R(O$\cdots$O) separation was 2.91 Å.[94] This intermolecular distance is in very good quantitative agreement with optimizations performed at lower levels of theory,[83,93,130] as are the intramolecular parameters of the monomers. The dimer geometry was also optimized with the explicitly-correlated CCSD(T)-F12x (x = a and b) methods, producing nearly identical results. Lane *et al.* calculated high-accuracy binding energies by estimating the effects from electron correlation beyond CCSD(T) and core correlation, as well as relativistic and non-Born-Oppenheimer behavior. The best estimate of the $(H_2O)_2$ D_e from that work is 5.02 kcal mol^{-1}, a value very similar to those from previous benchmark studies.[76,93,134,135]

1.6 2-Dimensional Hydrogen Bonding Networks: $(H_2O)_{n=3-5}$

1.6.1 Homodromic Global Minima

The global minima for $(H_2O)_n$ are all homodromic cyclic structures when n = 3, 4 or 5. Each water molecule functions as both the donor of and acceptor of a single hydrogen bond (Figure 1.4).

Figure 1.4. Global minima for $(H_2O)_n$ ($n = 3-5$)



(a) C_1 global min (b) S_4 global min (c) C_1 global min

The same computational strategies that were successful for the dimer system have been shown to be reliable for these 2-dimensional hydrogen bonding networks as well. Del Bene and Pople's 1970 SCF calculations predicted cyclic structures with staggered free hydrogens to be lower in energy than chain-like configurations for the trimer and larger clus-

ters. Afterwards, *ab initio* calculations[142–147] and empirical potentials[148] confirmed that these closed structures had lower electronic energies than any chain configurations. These early *ab initio* studies were limited to constrained optimizations with small basis sets in the Hartree-Fock approximation, yet they were still able to identify the structures of the global minima and provide quantitative evidence for hydrogen bonding cooperativity.

First introduced by Schütz,[149] a convenient notation for these homodromic systems defines each structure according to the orientations of the free hydrogens, relative to the pseudo-plane defined by the O atoms. For example, the C_1 trimer global minimum structure in Figure 1.4a can be called *uud*, meaning two of the free hydrogens are positioned "up" above the plane defined by the 3 O atoms and one is "down" below the ring. All of the minima from Figure 1.4 can be described by this *up* or *down* notation, but other stationary points will utilize *p* for planar orientations or *bi* to denote the presence of a bifurcated hydrogen bond.

Permuting the letter codes identifies different versions of the same minimum (e.g., *duu*). For the trimer, there are six such permutations that all specify equivalent versions of the global minimum structure. There are actually 96 equivalent versions of this trimer structure on the PES from other possible permutations.[148] For example, the directionality of the hydrogen bonding pattern in the ring can be reversed (e.g., clockwise versus counterclockwise), and each water molecule can exchange its free hydrogens with its bound hydrogens.

The early 90s saw the first correlated calculations on these clusters. Mó,[150] Xantheas[151] and Nielsen[152] are among those who examined these homodromic clusters in great detail at the MP2 level. The convergence of the trimer *uud* global minimum geometry, in particular, was established by Nielsen *et al.*[152] As with $(H_2O)_2$, the MP2-optimized structure of the global minimum is well converged with the aTZ basis set. The triangle formed by the oxygen atoms is slightly asymmetric, with O$\cdots$O separations of 2.7794 Å, 2.7815 Å and 2.7839 Å at the MP2/aQZ level.[152] These values are more than 0.1 Å shorter than the O$\cdots$O separation in the dimer. The distortion of the O-H_d bond is larger in the trimer.

At the MP2/aQZ level, the average r(O-H_d) is 0.9722 Å in the trimer, compared to 0.966 Å in the dimer[130] and 0.9589 Å in the monomer. The cyclic nature of the trimer forces the hydrogen bond angles away from a linear arrangement. The intermolecular O$\cdots$O-H angle is around 5° for the dimer[93,130] but increases to nearly 20° for all three values in the trimer.

The same convergence trends were observed for the S_4 tetramer and the C_1 pentamer. The R(O$\cdots$O) separations are all equal by symmetry, increasing to 2.731 Å at the MP2/aQZ level. The interoxygen distances only change by 0.001 Å between the aTZ and aQZ basis sets. Although the elongation of R(O-H_d) relative to the optimized monomer structure consistently increases from the dimer (+0.004 Å) to the trimer (+0.012 Å) to the tetramer (0.018 Å), this distortion is nearly identical for the $(H_2O)_4$ and $(H_2O)_5$ global minima (+0.019 Å) at the CCSD(T)/haTZ level of theory.[153] As the size of the ring grows to 4 and 5 water molecules, the angle about the hydrogen bond returns to a more linear value as seen in the dimer. The intermolecular O$\cdots$O-H_d angles are approximately 20°, 8° and 3° for the $(H_2O)_3$, $(H_2O)_4$ and $(H_2O)_5$ global minima, respectively at the CCSD(T)/haTZ level.[153]

The convergence of the cyclic homodromic minima has also been examined by calculating a series of binding energies within the aXZ hierarchy.[84] Xantheas *et al.* performed this analysis for the *uud* trimer and the *udud* tetramer. The trimer was optimized at the MP2 level with basis sets ranging from aDZ to a5Z, and high-level dissociation energies were calculated for each geometry. The difference in D_e values obtained with the a5Z basis set was only 0.002 kcal mol^{-1} when using the aTZ-optimized structure vs. the optimal a5Z geometry. The deviation associated with the MP2/aDZ optimized structure was roughly an order of magnitude larger. Similar results were obtained for the tetramer. The aTZ structure yields an MP2/a5Z D_e deviating by only 0.001 kcal mol^{-1} from that computed using the aQZ geometry, whereas the aDZ tetramer structure is in error by 0.04 kcal mol^{-1}.

In that same study,[84] aQZ and a5Z interaction energies were used to estimate MP2 CBS limit dissociation energies. D_e for the trimer was estimated to be 15.83 and 15.82 kcal mol^{-1} with and without CP corrections, respectively. D_e for $(H_2O)_4$ was found to be

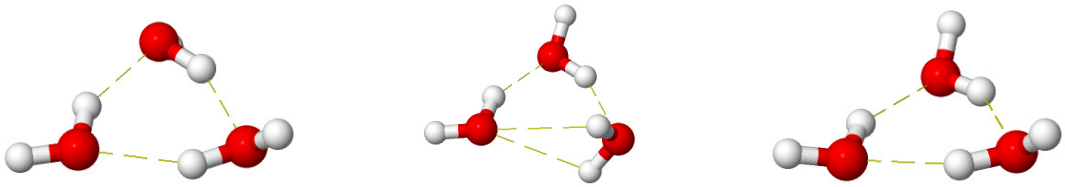
27.63 kcal mol^{-1} both with and without CP corrections. D_e increases to 36.31 kcal mol^{-1} for the pentamer when the uncorrected dissociation energies were extrapolated to the MP2 CBS limit. This value is 0.03 kcal mol^{-1} larger than the corresponding extrapolated value obtained with CP-corrected energies.

The energetics of the low-lying homodromic cyclic structures of the trimer, tetramer and pentamer are well converged at the MP2 level. Nielsen observed cancellation between the effects of correlating core electrons and improving the valence description of electron correlation with CCSD(T) calculations.[152] Another study evaluated the effects of CCSD(T) contributions with basis sets as large as aQZ and BD(TQ) contributions with the haDZ basis set.[154] Results from that work also showed that the effects of correlating the core electrons was roughly the same magnitude but opposite sign as valence correlation beyond the MP2 level. The resulting best estimate of D_e for trimer C_1 global minimum, including these effects, was identical to the MP2 CBS results of Xantheas and Nielsen at 15.82 kcal mol^{-1}. Thus, inclusion of higher-order correlation in the other homodromic global minima is not expected to change the value of D_e significantly.

1.6.2 $(H_2O)_n$ $(n = 3 - 5)$ Saddle Points

In addition to the *uud* global minimum, at least five important cyclic stationary points have been identified on the $(H_2O)_3$ PES.[149,154–159] Three of these are depicted in Figure 1.5. These trimer structures have been identified as transition states at the MP2 level.[154,160] It should be noted that many non-cyclic stationary points exist, but are predicted to be much higher in energy.[150]

Figure 1.5. Transition states on the $(H_2O)_3$ PES



(a) $C_1 \rightarrow C_1$ TS (*udp*) (b) Bifurcated TS (*upbi*) (c) $C_1 \rightarrow C_3$ TS (*uup*)

Exhaustive studies from Wales[157,160] and Schütz[149] have provided a wealth of knowledge about the $(H_2O)_3$ PES, helping to explain observed tunneling splittings in the VRT spectra.[155] The *udp* TS in Figure 1.5a interconverts equivalent versions of the *uud* C_1 global minimum by flipping the orientation of one of the free hydrogens.[148] This structure is the lowest-lying transition state on the $(H_2O)_3$ PES. This barrier has been predicted to be only 0.25 kcal mol^{-1} at the MP2,[160] CCSD[158] and CCSD(T) levels.[154] The *upbi* transition state in Figure 1.5b also connects equivalent versions of the global minimum.[157,158,160] However, this structure is over 2 kcal mol^{-1} above the global minimum.

The last transition state in Figure 1.5c (*uup*) connects the global minimum to a C_3 "Crown" structure (*uuu*).[154,160] This TS is just under 0.8 kcal mol^{-1} above the *uud* minimum according to MP2 and CCSD(T) calculations. These computations indicate the C_3 crown structure is a true minimum on the surface, but there is almost no barrier preventing it from flipping back to the C_1 global minimum. We briefly mention the fully planar *ppp* higher-order saddle point because it lies only 1.25 kcal mol^{-1} higher in energy than the global minimum.[154]

The cyclic tetramer rearrangement mechanisms have also been studied in depth by Schütz[161] and Wales.[162] The only transition state found to connect equivalent versions of the S_4 minimum is a bifurcated conformation *udbid*.[162] This stationary point, however, has not been characterized at a correlated level of theory. Two distinct but nearly isoenergetic transition states (*uudp* and *uupd*) have been characterized at the MP2/aDZ level.[161] These connect the S_4 minimum to a C_i *uudd* minimum, lying about 0.9 kcal mol^{-1} above the global minimum, according to MP2-R12 and CCSD(T) calculations.[161]

A third homodromic cyclic minimum has been confirmed at the MP2 level in which three of the free hydrogens are on the same side of the oxygen plane (*uuud*). This minimum is predicted to be very close in energy to the C_i *uudd* structure at all levels of theory.[163] Many other stationary points exist on the $(H_2O)_4$ PES, such as bicyclic structures and double hydrogen bond donor arrangements, but they have significantly higher electronic

energies.[163,164] Interestingly, the $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ corrections for this more diverse group of $(H_2O)_4$ stationary points can be quite different, depending on the hydrogen bonding environment. For example, $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ values calculated for the homodromic minima with the 6-311++G** are all 0.59 ± 0.01 kcal mol^{-1}, while for other structures the value can be as low as 0.08 kcal mol^{-1}.[163]

The pentamer has been studied less extensively, due to not only its larger size than $(H_2O)_3$ and $(H_2O)_4$, but also its similar hydrogen bonding network. Wales[159] performed a rearrangement analysis of the C_1 global minimum similar to the trimer and found analogous rearrangement processes including the single flip and bifurcation mechanisms. A number of higher-lying pentamer conformations have been identified on MP2 PESs including some 3-dimensional hydrogen-bonding networks,[164,165] and three minima have been found to lie within 1.2 kcal mol^{-1} of the global minimum according to CCSD(T)/aTZ energies.[165]

1.7 3-Dimensional Hydrogen Bonding Networks: $(H_2O)_{n\geq 6}$

1.7.1 The Water Hexamer

The water hexamer is of particular interest because it is the smallest water cluster for which a 3-dimensional hydrogen bonding network is lower in energy than a cyclic structure. In fact, several isomers of $(H_2O)_6$ have electronic energies below the cyclic homodromic motif which characterizes the trimer, tetramer and pentamer global minima.[166] The PES of $(H_2O)_6$ is much more complex than that of smaller clusters, with diverse hydrogen bonding patterns and significant populations for multiple isomers at different thermal conditions.[29] Because of the crossover from 2- to 3-dimensional hydrogen bonding networks that occurs at six water molecules, the hexamer has been referred to as the "smallest piece of ice."[167]

In early structural and energetic water hexamer investigations, a variety of water potentials were employed, and they produced different conclusions as to the identity of the lowest energy hexamer conformations.[168–171] In 1993, Kim and co-workers examined several conformations of $(H_2O)_6$ with both HF and MP2 optimizations.[172] In addition to the cyclic structure, a few stable 3-dimensional hydrogen bonding motifs were explored, including a "Prism"-type structure (9 hydrogen bonds), a "Cage" motif (8 hydrogen bonds), a "Bag" structure and a "Book" arrangement (both with 7 hydrogen bonds). All of their HF energies suggested that the quasiplanar cyclic structure was lower in energy than the 3-dimensional structures. In contrast, most of the MP2 results indicated that a Prism structure was lowest in energy by a few tenths of a kcal mol^{-1}. However, the method of optimization and the application of CP corrections changed the ordering of energies in some instances. Subsequent MP2 calculations from Kim[173] and Jordan[173,174] concluded that 3-dimensional structures had lower electronic energies than the S_6 cyclic homodromic structure. Both studies found a Prism structure and a Cage structure to be around 1 kcal mol^{-1} below the cyclic conformation. In 1996, Liu *et al.* reported in Nature[175] a VRT spectrum of $(H_2O)_6$. They observed only one isomer, and the experimental data was consistent with the calculated

rotational constants and ZPVE-inclusive relative energies for the Cage structure. Kim and co-workers[176] revisited the lowest-lying hexamer structures, applying much higher levels of calculation including MP2/TZ2P++ geometry optimizations and MP2 energy points with basis sets as large as hextuple-ζ. Regardless of the CP correction, the Prism structure had the lowest MP2 electronic energy, with the Cage about 0.1 kcal mol^{-1} higher. The Bag and cyclic Ring structure were both found to be $\approx$ 1 kcal mol^{-1} above the Prism. The electronic energy of the Book structure was suggested to be in the middle of these groups, with ΔE somewhere from 0.3 to 0.5 kcal mol^{-1}. Kim also evaluated the ZPVE for each conformation with MP2/6-311+G** vibrational frequencies. The Cage becomes the lowest energy minimum on a high quality MP2 PES when ZPVE is included (approximately 0.2 kcal mol^{-1} below the prism), and the Book drops about 0.1 kcal mol^{-1} below the Prism. Xantheas estimated MP2 CBS limit binding energies for the Prism, Cage, Book, and cyclic Ring isomers at their optimal MP2/aTZ geometries.[84] MP2/aXZ energies (X=D,T,Q) consistently found the Prism to be the global minimum but only by 0.1 kcal mol^{-1} at the CBS limit (D_e = 45.9 and 45.8 kcal mol^{-1} for the Prism and Cage, respectively). The results of Xantheas were in qualitative agreement with Kim's MP2 calculations.[176] The estimated MP2 CBS dissociation energies for the Book and Ring were 45.6 and 44.8 kcal mol^{-1}, corresponding to relative energies of 0.3 and and 1.1 kcal mol^{-1}.

In 2007, Olson[177] and co-workers calculated CCSD(T)/aTZ binding energies for the Prism, Cage, Book and Ring (or "Cyclic") structures with a newly developed parallel algorithm. In addition, a different homodromic cyclic minimum, dubbed "Boat,[178]" was investigated. Of particular interest in this study are the effects of higher-order correlation on the binding energies. CCSD(T) computations were found to increase the magnitude of E_{bind} for the Prism relative to MP2 results obtained with the same basis set. In contrast, E_{bind} did not change for the Cage and its magnitude decreased for the other isomers. Consequently, the Cage, Book, Ring and Boat conformations were destabilized relative to the Prism isomer by these higher-order correlation effects. The extent of this destabilization ($\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$)

is not consistent for the different hydrogen bonding environments. For the Cage and Book isomers, $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ is 0.2 and 0.4 kcal mol^{-1}, respectively. The Ring and Boat isomers are both destabilized by ≈ 0.6 kcal mol^{-1} relative to the Prism when the level of electron correlation is elevated to the CCSD(T) level. The results are a departure from that observed for the low-lying minima of the smaller clusters. Although this dependence of $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ on the nature of the hydrogen bonding network has been observed for $(H_2O)_4$,[163] the hexamer is the smallest $(H_2O)_n$ system for which it noticeably affects the relative energies of the low-lying minima on the PES. Their study also demonstrated that the $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ term was insensitive to the presence of diffuse functions on the H atoms. The aTZ and haTZ values differed by no more than 0.01 kcal mol^{-1}.[177]

The same higher-order correlation effects were observed by others for a similar set of hexamer isomers optimized with a larger basis set (haTZ).[179] The calculated MP2 and CCSD(T) energies were nearly identical to those from the work of Olson *et al.*,[177] and the results again demonstrated that there was a strong dependence of $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ on the hydrogen bonding network. Interestingly, the correction factor for the Bag isomer matched that of the Book to 0.01 kcal mol^{-1}, and both structures have one double hydrogen bond donor and a total of seven hydrogen bonds.

A subsequent investigation examined how higher-order correlation corrections would affect the MP2 CBS limit relative energies.[166] That study examined the Prism, Cage, Bag and Ring isomers, as well as two Book and two Boat conformations. The extra Book and Boat isomers originated from the fact that References 177 and 179 used structures with slightly different free-hydrogen orientations. The MP2 CBS limits for these structures were estimated with explicitly-correlated MP2-R12 calculations. The MP2-R12 relative energies were in excellent agreement with the extrapolated values of Xantheas[84] (e.g., 0.06 kcal mol^{-1} and 0.07 kcal mol^{-1} for the Cage, respectively). The destabilization from higher-order correlations (specifically CCSD(T) vs. MP2) is very important for the low-lying conformations. For example, $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ for the Cage is significantly larger than the magnitude of ΔE at the MP2

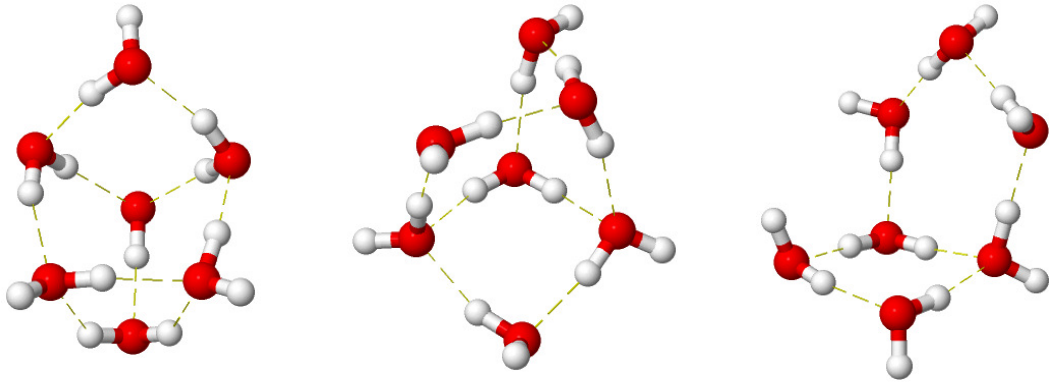
CBS limit. Applying this correction factor to the MP2 CBS values, CCSD(T) CBS relative energies were estimated via the additive scheme outline in Equation 1.23 using the haTZ basis set. At this level, the Cage is 0.25 kcal mol^{-1} above the Prism. Including higher-order correlation, the only other isomer within 1 kcal mol^{-1} of the Prism is one of the Book conformations ($\Delta E_{\mathrm{e}}^{\mathrm{CCSD(T)/CBS}} = 0.72$ kcal mol^{-1}). As with the Olson study[177] and Pérez's tetramer work,[163] the clusters with similar hydrogen bonding patterns saw comparable values for the $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ correction. The corrections for the Bag and 2 Book isomers were identical (0.39 kcal mol^{-1}). The cyclic Ring and Boat structures, each consisting of 6 hydrogen bonds in a homodromic pattern, all had $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ values of 0.58 ± 0.01 kcal mol^{-1}.

ZPVE corrections, computed at the MP2/haTZ level, have the opposite effect as the $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ corrections. ZPVE stabilizes the Cage by 0.16 kcal mol^{-1} relative to the Prism, resulting in a best estimate for ΔE^0 of 0.09 kcal mol^{-1} for the Cage at the CCSD(T) CBS limit. Both Book structures considered in that study are within 0.5 kcal mol^{-1} of the Prism after ZPVE is included, while the Bag and Ring conformations were both found to be less than 1 kcal mol^{-1} above the Prism. The cyclic Boat configurations had the highest electronic energies, with ΔE^0 around 1.25 kcal mol^{-1} at the CCSD(T) CBS limit, however the CCSD(T) CBS Gibbs free energy values computed by Shields and co-workers suggest it becomes the most stable around 200 K.[28]

Very recently, broadband rotational spectroscopy[123] and state-of-the-art quantum simulations[54] have confirmed the predictions of these estimates of the CCSD(T) CBS limit energetics. The Prism was identified in 2012 for the first time experimentally, and the relative isomer populations for the Prism, Cage and Book were estimated as 1:1:0.25.[123] Subsequently, full-dimensional diffusion Monte Carlo simulations found the Cage and Prism to have statistically indistinguishable energy differences at very low temperatures when anharmonic ZPVE was included.[54]

1.7.2 Larger Clusters

Aside from the additional computational cost incurred by applying wavefunction methods to larger water clusters, the number of possible hydrogen bonding topologies further complicates efforts to analyze the lowest-energy conformations at a highly accurate level. There are over sixty unique hydrogen bonding configurations available for a water hexamer.[23] (These are unique structures, not simply permutations of equivalent structures.) That number increases to over 2000 for a decamer.[23] Thus PES sampling techniques, such as Monte Carlo and molecular dynamics simulations, the use of parameterized water potentials and other strategies have been vital to identifying low-lying candidates.[28,168,174,178,180–189]



(a) Prism (b) Cage (c) Chair

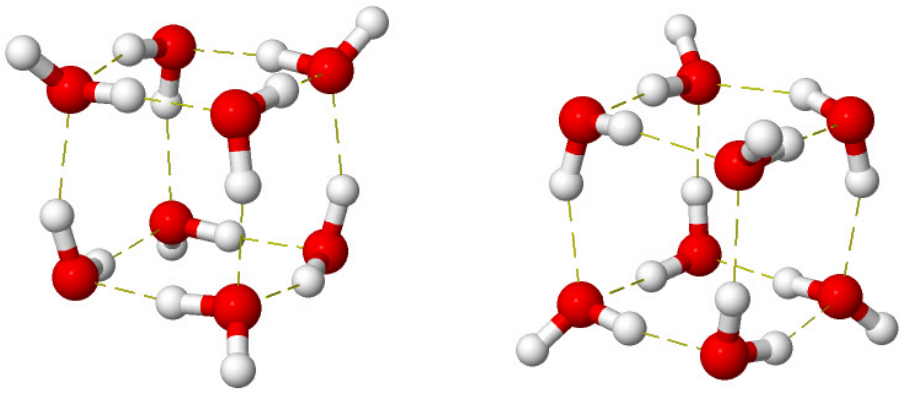
Figure 1.6. Low-lying minima for $(H_2O)_7$

The lowest energy heptamer structures exhibit hydrogen bonding motifs similar to those in the hexamer. There exist multiple low-lying isomers which can be classified as Prism, Cage and Chair conformations (Figure 1.6).[28,29,183,190] A Prism structure has been calculated as the lowest energy heptamer with conventional MP2 calculations[29,190] and with RI-MP2 extrapolations to the estimated CBS limit.[28] When higher-order correlation is included from $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ corrections, at least three prism conformations lie within 0.5 kcal mol^{-1} of the lowest energy Prism configuration ($D_e = 57.4$ kcal mol^{-1}).[28] The lowest energy Cage structure is nearly 1.7 kcal mol^{-1} above the Prism minimum. These two conformations are depicted in Figure 1.6, along with the lowest-lying Chair structure ($\Delta E \approx 3$ kcal mol^{-1}),

which is actually the most stable at 298 K when thermodynamic effects are included.[28]

The most stable octamers can be constructed from stacked tetramers, forming cubic structures. High-symmetry (D_{2d} and S_4) of the two lowest-energy octamers has facilitated MP2 computations with large basis sets on these $(H_2O)_8$ isomers depicted in Figure 1.7. They differ only in the relative directionality of hydrogen bonds in their component tetramers. Xantheas *et al.* established benchmark MP2 CBS energetics via extrapolations with the aXZ basis sets ($X = 3 - 5$). Both CP-corrected and uncorrected energies indicate that the two structures are virtually isoenergetic with the D_{2d} isomer being slightly more stable by less than 0.1 kcal mol^{-1}. The CP-corrected D_e is estimated as 72.4 kcal mol^{-1} for the D_{2d} isomer. CCSD(T) CBS limit binding energies for these structures were computed by Shields and co-workers[28] and found E_{bind} to be 72.55 kcal mol^{-1} for both isomers.

In that study,[28] CCSD(T) CBS limit binding energies for nonamer and decamer structures were estimated as well, employing RI-MP2 extrapolations and a $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ correction computed with the aDZ basis set. Similar to the octamer, the lowest energy nonamer and decamer structures are composed of stacks of smaller cyclic clusters (tetramers and pentamers). Thermodynamic corrections including anharmonicity indicated that these same isomers were also the most stable at 298 K.



(a) D_{2d} octamer (b) S_4 octamer

Figure 1.7. Low-lying minima for $(H_2O)_8$

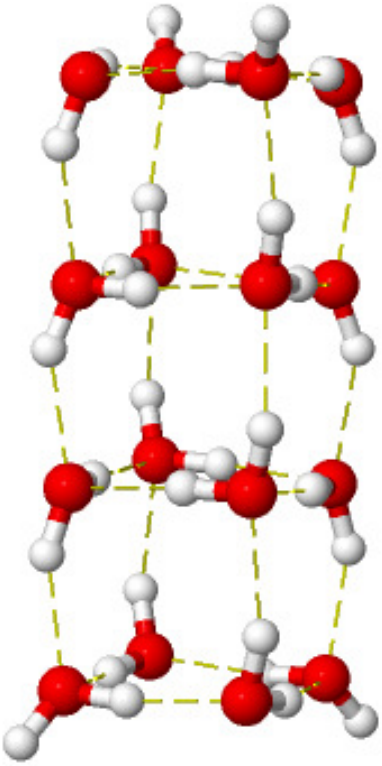
High-level *ab initio* calculations are scarce in the literature for clusters larger than

the octamer. Shields and co-workers computed CCSD(T) CBS binding energies of clusters up to the decamer with RI-MP2 extrapolations and $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ values from the aDZ basis set.[28] They noted that the $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ changed the binding energies by less than 1% but were important given the narrow energy spacings between low-lying isomers. Binding energies of $(H_2O)_{11}$ and $(H_2O)_{13}$ been computed at the MP2/a5Z level.[191] Also, MP2 CBS estimates for representative $(H_2O)_{20}$ structures have been established.[115,192] Aprà *et al.* used $(H_2O)_{20}$ structures to demonstrate the impressive scalability possible with their CCSD(T) algorithm.[116] Reference 193 reviews *ab initio* energies and structures of water clusters as large as $(H_2O)_{24}$, where computations are limited to double-ζ or triple-ζ basis sets with MP2. These cages are interesting as potential building blocks for ice clathrate structures.

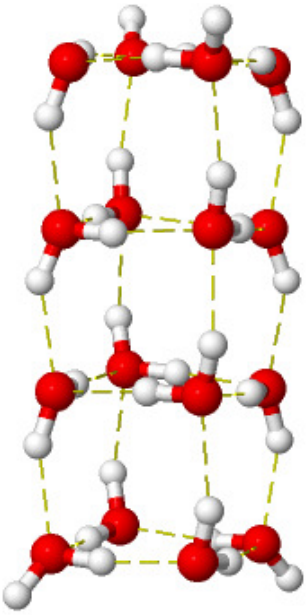
Clusters of sixteen and seventeen water molecules have been a target of high-accuracy electronic structure methods because a structural transition from "all-surface" to "internally solvated" arrangements are thought to occur in that regime.[194,195] In 2010, the binding energies for several low-lying isomers of $(H_2O)_{16}$ and $(H_2O)_{17}$ were computed at the CCSD(T)/aTZ level, representing "the current state-of-the-art in scalable high-level electronic structure theory."[196] Each CCSD(T) computation on $(H_2O)_{16}$ took roughly 3.33 hours utilizing 120,000 AMD Opteron cores on the CRAY XT5 partition at ORNL, the world's fastest supercomputer at the time.[197] The results demonstrated the importance of high-level electron correlation and large atomic orbital basis sets. The $(H_2O)_{16}$ boat-a structure (Figure 1.8a) is the lowest-energy conformation at the MP2/aTZ level, yet it lies over 0.25 kcal mol^{-1} above the 4444-a conformation (Figure 1.8b) on the CCSD(T)/aTZ PES. The MP2/aTZ energies actually suggest that the 4444-a isomer is more than 0.38 kcal mol^{-1} above the boat-a structure. The aTZ $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ higher-order correlation effectively shifts the relative energies of these two structures (boat-a and 4444-a) by more than 0.6 kcal mol^{-1}. The relative energies of the other isomers examined are also affected, but not as dramatically. Although one should not expect that the CCSD(T) energies obtained with the aTZ basis set are converged to the CBS limit, these results represent the highest-level computations that

have been performed with canonical post-HF wavefunction methods on these $(H_2O)_{16}$ and $(H_2O)_{17}$ water clusters.

The aDZ results predict a different ordering of the $(H_2O)_{16}$ isomers at both the MP2 and CCSD(T) levels. In addition, the haDZ and haTZ ΔE values can differ significantly. For example, CCSD(T) ΔE for a different $(H_2O)_{16}$ boat isomer changes from 0.94 kcal mol^{-1} with the haDZ basis set to 0.42 kcal mol^{-1} with haTZ. For the $(H_2O)_{17}$ clusters, both MP2 and CCSD(T) calculations confirmed that an internally solvated Sphere structure has a lower energy than the all-surface isomers.



(a) Boat-a



(b) 4444-a

Figure 1.8. Low-lying minima for $(H_2O)_{16}$

1.8 Strategies for Extending High-Accuracy Methods to Larger Clusters

A significant amount of research has been dedicated to calculating accurate molecular properties with greater efficiency. Fragmentation methods have facilitated the extension of QM methods to increasingly larger molecules and chemical systems.[198–210] Fragmentation methods divide a molecular system into smaller subunits and combine the results in some way to obtain an approximation of a calculation on the full system. If the partitioning scheme and the chosen theoretical methods are appropriate, the results of the fragmentation method can approach those of the canonical calculation. Reference 211 provides an extensive review of fragmentation methods.

Related to fragmentation methods are integrated or hybrid methods (e.g., QM:MM, QM:QM and multi-layered generalizations such as ONIOM).[212–216] These methods apply a high level of theory to a "chemically important" subset of the molecule and lower levels of theory to the rest of or the entire molecule. An important distinction is that these methods describe non-local phenomena with a low-level calculation on the entire system (or the majority of the entire system). It has been recognized for quite some time that extremely accurate results can be obtained very efficiently when this integrated approach to QM:QM computations is combined with the many-body expansion for noncovalent complexes.[18,112,217]

In 1998, for example, Klopper, Quack and Suhm reported CCSD(T) CBS limit dissociation energies for hydrogen fluoride clusters, $(HF)_n$.[217] For the trimer, they employed the explicitly correlated CCSD(T)-R12 method to describe the 1- and 2-body interactions and the canonical CCSD(T) method to capture the cooperative (3-body) effects. This approach was effectively a QM:QM treatment where the leading terms in the many-body expansion were computed at a very high level while the 3-body interactions were computed at a lower level. For $(HF)_4$ and $(HF)_5$, the MP2-R12 method was used to compute the 4- and 5-body contributions in what was essentially a QM:QM:QM treatment. The approach was extremely efficient because the demanding CCSD(T)-R12 computations only had to be performed on individual fragments (i.e., HF monomers) and pairs of fragments (i.e., HF dimers). Many

hybrid methods are available today that adopt a similar approach by applying 2 or more levels of theory to the many-body expansion, including Extended ONIOM (XO),[218] the Molecules in Molecules (MIM) method,[219,220] the Hybrid Many-Body Interaction (HMBI) method,[221,222] the Multilevel Fragment-Based Approach (MFBA),[223] the Stratified Approximation Many-Body Approach (SAMBA)[224] and the *N*-body:Many-body QM:QM technique developed by our group.[225–231] Reference 232 presents a general framework for and a very lucid overview of the various fragmentation procedures based on the many-body expansion.

CCSD(T) and MP2 computations on small water clusters have been used to model the dominant terms of the many-body expansion in the construction of accurate *ab initio* potential energy surfaces.[121,233–240] Potentials constructed in this way can be used both for exploring the potential energy landscapes of large water clusters and for the detailed characterization of smaller clusters with more rigorous approaches, e.g., quantum molecular dynamics simulations.[54,238,241]

Góra *et al.* applied a many-body expansion with multiple levels of theory to study water clusters as large as $(H_2O)_{16}$ and $(H_2O)_{24}$ with SAMBA. The 2-body interactions were captured at the CCSD(T) CBS limit, and the higher-order interactions were calculated up to fourth order with CCSD(T)/aDZ and MP2/aDZ. Their SAMBA calculations also corrected for BSSE with CP corrections.[224]

Relatively large water clusters have also been examined with the *N*-body:Many-body QM:QM technique.[230] The 3-body:Many-body CCSD(T):MP2 approach provided binding energies nearly identical to the canonical CCSD(T) values for more than 40 low-lying isomers of $(H_2O)_{n=3-10}$ with the haTZ basis set. The maximum deviation was only 0.07 kcal mol^{-1} for this test set. Similar differences were observed for $(H_2O)_{16}$ and $(H_2O)_{17}$. The CCSD(T):MP2 electronic energies deviated by only 0.06 kcal mol^{-1} on average for the 7 CCSD(T)/aTZ values reported by Yoo *et al.*[196]

The 1998 study of Klopper, Quack and Suhm already noted that the linear expressions for the energy associated with an integrated many-body approach facilitates the evaluation

of geometrical derivative properties. Analytic gradients, for example, have been implemented for N-body:Many-body QM:QM method and applied to water and other hydrogen bonded clusters.[229,231] In a study of more than 70 water clusters, 2-body:Many-body optimizations with MP2 as the high-level method and Hartree-Fock (HF) as the low-level method (MP2:HF) yielded structures nearly indistinguishable to those from the conventional MP2 calculation with the haTZ basis set.[231]

CHAPTER 2

N-BODY: MANY-BODY QM:QM VIBRATIONAL FREQUENCIES: APPLICATION TO SMALL HYDROGEN-BONDED CLUSTERS

We present an efficient method for reproducing CCSD(T) optimized geometries and harmonic vibrational frequencies for molecular clusters with the N-body:Many-body QM:QM technique. In this work, all 1-body through N-body interactions are obtained from CCSD(T) computations, and the higher-order interactions are captured at the MP2 level. The linear expressions from the many-body expansion facilitate a straightforward evaluation of geometrical derivative properties (e.g., gradients and Hessians). For $(H_2O)_n$ clusters ($n = 3 - 7$), optimized structures obtained with the 2-body:Many-body CCSD(T):MP2 method are virtually identical to CCSD(T) optimized geometries. Harmonic vibrational frequencies calculated with this 2-body:Many-body approach differ from CCSD(T) frequencies by at most a few cm^{-1}. These deviations can be systematically reduced by including more terms from the many-body expansion at the CCSD(T) level. Maximum deviations between CCSD(T) and 3-body:Many-body CCSD(T):MP2 frequencies are typically only a few tenths of a cm^{-1} for the H_2O clusters examined in this work. These results are obtained at a fraction of the wall time of the supermolecular CCSD(T) computation, and the approach is well-suited for parallelization on relatively modest computational hardware.

Reprinted with permission from " N-body:Many-body QM:QM Vibrational Frequencies: Application to Small Hydrogen-Bonded Clusters,"J. Coleman Howard and Gregory S. Tschumper, *J. Chem. Phys. 139*(18), 184113. http://dx.doi.org/10.1063/1.4829463 Copyright © 2013, American Institute of Physics

2.1 Introduction and theory

Ab initio quantum mechanical (QM) computations provide a reliable description of molecular systems, but the application of correlated QM methods is limited by the steep polynomial scaling of the computational demands with the size of the system being studied. A significant amount of research has been dedicated to computing accurate molecular properties with greater efficiency. Fragmentation methods, for example, have successfully extended correlated QM methods to increasingly larger molecules and chemical systems[198–210,242–245] by dividing a molecular system into smaller subunits and combining the results of many smaller calculations in some way to approximate a single canonical calculation on the full system. For carefully selected combinations of the partitioning scheme and the theoretical methods, the results of the fragmentation method can approach those of the canonical calculation. A thorough review of fragmentation methods can be found in Reference 211. In addition, a recent issue of Phys. Chem. Chem. Phys. was dedicated to the subject.[246]

Related to fragmentation methods are hybrid or integrated methods, such as IMOMM and its multi-layer generalization ONIOM.[212–216,247] These methods apply a high level of theory to a "chemically important" region of the molecule and lower levels of theory to the rest of the molecule. The ONIOM approach describes non-local interactions by performing a calculation on the entire system. The N-body:Many-body QM:QM technique[225–231] developed by our group is essentially a multi-centered ONIOM approach. When applied to clusters, this technique is equivalent to approximating a full many-body expansion[112] by computing the leading dominant terms with a high-level QM method and capturing the higher-order effects with a less demanding theoretical method. In a 1998 paper, Klopper, Quack and Suhm[217] utilized a multi-level approach such as this to calculate accurate interaction energies for $(HF)_{n=2-5}$ clusters, where the cooperative effects (>2-body) contribute more than 40% of the binding energy for the pentamer.[217] Several other hybrid methods take a similar approach, employing multiple levels of theory, including Extended ONIOM (XO),[218] Molecules-in-Molecules (MIM),[219,220] the Hybrid Many-body Interaction

method (HMBI),[221,222] the Multilevel Fragment-Based Approach (MFBA)[223] and Stratified Approximation Many-body Approach (SAMBA).[224] In Reference 232, a general framework is presented for the various many-body approaches.

The many-body expansion[112] for the total energy for a cluster of n individual fragments $[f_i f_j \dots f_n]$ takes the form of Equation 2.1.

$$E[f_i f_j \dots f_n] = E_1 + E_2 + E_3 + \cdots + E_n \tag{2.1}$$

Here, the energy of the cluster is expressed in terms of fragment, or 1-body, energies (E_1), pairwise (2-body) interactions (E_2), triad (3-body) interactions (E_3), etc.

$$E_1 = \sum_{i=1}^{n} E[f_i] \tag{2.2}$$

$$E_2 = \sum_{i=1}^{n-1} \sum_{j>i}^{n} \Big[E[f_i f_j] - (E[f_i] + E[f_j]) \Big] \tag{2.3}$$

$$E_3 = \sum_{i=1}^{n-2} \sum_{j>i}^{n-1} \sum_{k>j}^{n} \Big[E[f_i f_j f_k] - (E[f_i f_j] + E[f_i f_k] + E[f_j f_k]) + (E[f_i] + E[f_j] + E[f_k]) \Big] \tag{2.4}$$

In the context of molecular clusters, monomers are a natural and very common choice for the fragments, but Equation 2.1 is general for all types of fragments. This study will exclusively use monomers as the fragments. The full many-body expansion to nth order gives the exact energy of the cluster. However, the many-body expansion converges rather quickly so a reasonable approximation for neutral fragments/molecules can often be obtained by truncating the series at the 2nd or 3rd order.[18]

In the N-body:Many-body QM:QM approach, all 1-body through N-body interactions are captured with a high level of theory and all $(N+1)$-body through n-body interactions are computed with a less demanding method. If the higher-order (>N) effects are the

same for the chosen high and low level methods, then the QM:QM result will be identical to the full high-level result. Equation 2.5 gives the form of the 2-body:Many-body QM:QM energy.

$$\begin{aligned} E^{\text{Hi:Lo}}[f_i f_j \dots f_n] &= E_1^{\text{Hi}} + E_2^{\text{Hi}} + E_3^{\text{Lo}} + \dots + E_n^{\text{Lo}} \\ &= E^{\text{Lo}}[f_i f_j \dots f_n] + (E_1^{\text{Hi}} - E_1^{\text{Lo}}) + (E_2^{\text{Hi}} - E_2^{\text{Lo}}) \end{aligned} \tag{2.5}$$

Note that with the latter ONIOM-like form of Equation 2.5, the explicit dependence of E_3 on triad energies ($E[f_i f_j f_k]$) can be bypassed by calculating the energy of the entire cluster at a low level ($E^{\text{Lo}}[f_1 f_2 \dots f_n]$) and subtracting E_1^{Lo} and E_2^{Lo}. So the 2-body:Many-body QM:QM energy depends only on the energies of the monomers ($E[f_i]$) and pairs ($E[f_i f_j]$) computed at both high and low levels of theory, and the energy of the entire cluster ($E[f_1 f_2 \cdots f_n]$) computed at the low level. Similarly, the 3-body:Many-body QM:QM energy requires high-level computations to be performed on fragments, pairs and triads. The advantage is easier to see when the terms for the energies are explicitly written out.

$$\begin{aligned} E^{\text{2bHi:Lo}} = E^{\text{Lo}}[f_1 f_2 \dots f_n] &+ \sum_{i=1}^{n-1} \sum_{j>i}^{n} \Big[E^{\text{Hi}}[f_i f_j] - E^{\text{Lo}}[f_i f_j] \Big] \\ &- (n-2) \sum_{i=1}^{n} \Big[E^{\text{Hi}}[f_i] - E^{\text{Lo}}[f_i] \Big] \end{aligned} \tag{2.6}$$

$$\begin{aligned} E^{\text{3bHi:Lo}} = E^{\text{Lo}}[f_1 f_2 \dots f_n] &+ \sum_{i=1}^{n-2} \sum_{j>i}^{n-1} \sum_{k>j}^{n} \Big[E^{\text{Hi}}[f_i f_j f_k] - E^{\text{Lo}}[f_i f_j f_k] \Big] \\ &- (n-3) \sum_{i=1}^{n-1} \sum_{j>i}^{n} \Big[E^{\text{Hi}}[f_i f_j] - E^{\text{Lo}}[f_i f_j] \Big] \\ &+ \frac{(n-2)(n-3)}{2} \sum_{i=1}^{n} \Big[E^{\text{Hi}}[f_i] - E^{\text{Lo}}[f_i] \Big] \end{aligned} \tag{2.7}$$

From the forms of $\text{E}^{\text{2bHi:Lo}}$ and $\text{E}^{\text{3bHi:Lo}}$, one can see that the computational require-

ments for the high-level method increase quadratically and cubically with the size of the cluster. For large n, the QM:QM computations can actually be dominated by the low-level method. It is also worth noting that the calculations are completely independent so the method lends itself to efficient parallelism. In addition, the expressions are linear, which facilitates the evaluation of analytic geometrical derivative properties (gradients, Hessians). The development and implementation of 2-body:Many-body gradients have been described elsewhere.[229,231] The N-body:Many-body technique has previously been applied to hydrogen-bonded clusters including hydrogen fluoride, methanol and water.[228] In a study of more than 70 water clusters, 2-body:Many-body optimizations with MP2 as the high-level method and Hartree-Fock (HF) as the low-level method (MP2:HF) along with a triple-ζ quality basis set yielded structures nearly identical to those from the canonical high-level calculation.[231] In another study,[230] 3-body:Many-body CCSD(T):MP2 energies of $(H_2O)_n$ clusters ($n = 3 - 10$) never differed from canonical CCSD(T) total energies by more than 0.07 kcal mol^{-1}. Energies for $(H_2O)_{16}$ and $(H_2O)_{17}$ clusters calculated with the aug-cc-pVTZ basis set showed an average absolute error of 0.06 kcal mol^{-1}, relative to the full CCSD(T) total energies.[231]

Here, we expand on the N-body:Many-body QM:QM methodology for clusters in three important ways. Firstly, QM:QM gradient techniques that proved successful at the MP2 level have been extended to the CCSD(T) method. Secondly, the 2-body:Many-body gradients have been extended to 3-body:Many-body and arbitrarily high orders, although it does not appear to be necessary to go beyond 2- or 3-body in practice. Finally, QM:QM Hessians have been implemented for the calculation of harmonic vibrational frequencies for the first time with the N-body:Many-body QM:QM method. The expression for calculating a particular element of the 2-body:Many-body Hessian matrix is shown in Equation 2.8 where

q_α and q_β each represent two generic Cartesian coordinates of the full molecular cluster.

$$\begin{aligned}\frac{\partial^2}{\partial q_\alpha \partial q_\beta} E^{\text{2bHi:Lo}} &= \frac{\partial^2}{\partial q_\alpha \partial q_\beta} E^{\text{Lo}}[f_1 f_2 \dots f_n] \\ &+ \sum_{i=1}^{n-1} \sum_{j>i}^{n} \left(\frac{\partial^2}{\partial q_\alpha \partial q_\beta} E^{\text{Hi}}[f_i f_j] - \frac{\partial^2}{\partial q_\alpha \partial q_\beta} E^{\text{Lo}}[f_i f_j] \right) \\ &- (n-2) \sum_{i=1}^{n} \left(\frac{\partial^2}{\partial q_\alpha \partial q_\beta} E^{\text{Hi}}[f_i] - \frac{\partial^2}{\partial q_\alpha \partial q_\beta} E^{\text{Lo}}[f_i] \right)\end{aligned} \tag{2.8}$$

Fragmentation methods have scarcely been applied to the computation of vibrational frequencies, with some notable exceptions being via the molecular tailoring approach[208,245] and also within the embedded fragmentation method of Hirata and coworkers.[244] We apply these methods to a series of HF and H_2O clusters to examine the efficiency and accuracy of the 2- and 3-body:Many-body techniques relative to canonical CCSD(T) gradients and harmonic vibrational frequencies.

2.2 Computational Methods

Nineteen different stationary points of $(H_2O)_n$ clusters ($n = 3 - 7$) and five $(HF)_n$ ($n = 3 - 6$) clusters were optimized with the 2-body:Many-body and 3-body:Many-body methods using CCSD(T) as the high-level method and MP2 as the low-level method, referred to as CCSD(T):MP2. The geometries of these clusters were also optimized with the canonical CCSD(T) method for comparison. For optimized structures, the maximum absolute Cartesian force component on each atom was less than 2.5×10^{-4} E_h a_0^{-1}. CCSD(T):MP2 and CCSD(T) harmonic vibrational frequencies were also computed for each optimized structure. (CCSD(T) vibrational frequencies for the water heptamers were not calculated due to the excessive computational demands.)

All computations were performed with Dunning's double-ζ correlation consistent basis set,[57,58] augmented with diffuse functions on the heavy (non-hydrogen) atoms, hereafter referred to as haDZ (i.e., cc-pVDZ for H and aug-cc-pVDZ for O and F). The use of an

incomplete basis set causes an inconsistency when comparing the energies of fragments, commonly dubbed "basis set superposition error."[90] The standard procedure for removing this inconsistency is the counterpoise (CP) correction.[87,88] The computations in this work did not include CP corrections, but they could be incorporated into the N-body:Many-body QM:QM method. All CCSD(T) computations used CFOUR[248] for analytic first and second derivatives. MP2 analytic gradients were obtained from MPQC.[249] For the harmonic vibrational frequency computations, MP2 analytic Hessians were computed with Gaussian09.[250] The geometry optimization steps and the calculation of harmonic vibrational frequencies from QM:QM Hessians were performed by Gaussian09 through the external keyword.

2.3 Results and Discussion

2.3.1 Gradients Geometry Optimizations

2.3.1.1 Water Clusters

In past works, the N-body:Many-body geometry optimizations have been compared to the canonical high-level optimizations on the basis of structural differences, such as internal coordinates[229] or the minimum root-mean-square (RMS) deviation of unweighted Cartesian coordinates.[231] In this study, however, the geometries from the canonical CCSD(T) optimizations and those from the CCSD(T):MP2 optimizations are virtually identical.

To illustrate this structural similarity, consider the optimization procedures employed in this work. Due to the relative computational demands, most of the structures examined here were first optimized with the 2-body:Many-body (2b:Mb) technique with CCSD(T) as the high-level method and MP2 as the low-level method. Next, the geometries were further refined with 3-body:Many-body (3b:Mb) optimizations and finally canonical CCSD(T). Typically, the geometries obtained from the 2b:Mb optimizations were also converged structures according to the 3b:Mb and canonical CCSD(T) gradients. Note that in these circumstances the RMS deviations of the unweighted Cartesian coordinates of the optimized structures is zero. (In only a few of the most challenging cases (the larger HF clusters) a small number

of additional steps were required to converge the 3b:Mb and/or CCSD(T) optimizations.)

The individual components of the Cartesian gradients are used here to provide a more stringent comparison of the N-body:Many-body CCSD(T):MP2 forces to the corresponding supermolecular CCSD(T) values (all computed with the haDZ basis set). Table 2.1 lists the maximum absolute deviation (MaxAD) from any component of the canonical CCSD(T)/haDZ Cartesian gradient associated with various 2- and 3-body gradients for a series of water clusters (all at their 2b:Mb optimized geometries). The corresponding average absolute deviations (AvgAD provided in the Supplementary Material[251]) are roughly the same magnitude but approximately 2 to 4 times smaller.

The 2b column indicates the maximum absolute deviations associated with gradients obtained from a many-body expansion that has been truncated after the inclusion of CCSD(T) pairwise (2-body) interactions. The MaxADs range from 4.0×10^{-3} E_h a_0^{-1} for the C_1 structure of $(H_2O)_3$ to 1.6×10^{-2} E_h a_0^{-1} for the Bag isomer of $(H_2O)_6$. If the many-body expansion is extended through 3rd order at the CCSD(T) level, the deviations decrease by roughly an order of magnitude as indicated by the 3b column in Table 2.1. Far greater improvement can be realized, however, by using a low-level method to estimate the higher-order terms in the many-body expansion. The 2b:Mb technique, for example, recovers the 3-body through n-body interactions at the MP2 level, and the absolute deviations never exceed 2.4×10^{-4} E_h a_0^{-1}. The deviations can be further reduced with the 3b:Mb procedure for which MaxAD does not grow larger than 6.7×10^{-5} E_h a_0^{-1} for the PR1 structure of the water heptamer.

These results appear to be fairly insensitive to the size of the basis set and to the geometry. The deviations did not increase when a larger haTZ basis (aug-cc-pVTZ for O, cc-pVTZ for H) was used. Very similar MaxADs (and AvgADs) were obtained from computations employing the haTZ basis set for some of the smaller clusters for which the canonical CCSD(T)/haTZ gradient computations were feasible. When Hartree-Fock/haDZ optimized geometries were used instead of the 2b:Mb optimized structures, the deviations

from the components of the CCSD(T)/haDZ Cartesian gradients actually decreased slightly. (See data in Supplementary Material.[251])

The computational demands associated with the 2b:Mb and 3b:Mb gradients are significantly reduced relative to those of the CCSD(T) gradients. For the largest systems in this study (the $(H_2O)_7$ isomers), the wall time to compute a CCSD(T)/haDZ gradient on a single node was just over 38 times longer than corresponding time for a 2b:Mb gradient and 4 times longer than for a 3b:Mb gradient. These speedups can be further enhanced by parallelization due to the independent nature of the computations associated with the QM:QM schemes. For example, over 95% of the wall time for a 3b:Mb gradient for each $(H_2O)_7$ isomer is dedicated to 56 independent CCSD(T) gradient computations on 21 unique pairs of fragments and 35 unique triads. For these relatively small water clusters, nearly perfect scaling can be realized with 10 nodes, and the speedup for the 3b:Mb gradients would increase from about 4 to nearly a factor of 40.

2.3.1.2 Hydrogen Fluoride Clusters

Because they exhibit significant cooperative effects, $(HF)_n$ $(n = 3 - 6)$ clusters have been analyzed in a similar manner. Table 2.2 compares the N-body:Many-body CCSD(T):MP2 gradients to the CCSD(T) results. Here, we have also extended the methodology to compute 4-body and 4-body:Many-body gradients to demonstrate the systematic convergence of the QM:QM approach even for strongly coupled molecular clusters.

The 2b and 3b columns of Table 2.2 show the MaxADs in the gradients computed with the many-body expansion terminated at 2nd and 3rd order. These MaxADs are consistently larger than those for water clusters, indicative of the high degree of cooperativity in these HF clusters. Even with the inclusion of CCSD(T) 4-body effects, the MaxADs are still larger than 3×10^{-4} E_h a_0^{-1} for the $(HF)_5$ and $(HF)_6$ structures. The MaxADs of the 2b:Mb gradients are actually slightly smaller than the truncated 4b CCSD(T) gradients, but they can still exceed 2.5×10^{-4} E_h a_0^{-1} for HF clusters larger than the tetramer. The deviations can be systematically decreased by computing 3- and 4-body effects at the high level via the

3b:Mb and 4b:Mb methods, respectively.

2.3.2 Harmonic Vibrational Frequencies

2.3.2.1 Water Clusters

Harmonic vibrational frequencies for water clusters were computed with the 2-body, 3-body, 2b:Mb and 3b:Mb methods. Truncated 2-body and 3-body vibrational frequencies were computed for some of the smaller clusters, and the agreement with CCSD(T) was quite poor. For the C_1 trimer, the 2-body average absolute deviation from CCSD(T) was over 40 cm^{-1}, and the maximum absolute deviation was over 100 cm^{-1}. For the S_4 tetramer, the 2-body maximum deviation was greater than 200 cm^{-1}, and the 3-body maximum deviation was more than 40 cm^{-1}. In one case, the Hessian index (the number of negative eigenvalues corresponding to vibrational motion) was incorrect relative to CCSD(T). Due to this unsatisfactory performance for even the smallest water clusters in this study, the vibrational frequencies of the other clusters were not computed with these methods.

In Table 2.3, harmonic vibrational frequencies calculated with the 2b:Mb and 3b:Mb methods are compared to frequencies from a CCSD(T) supermolecular computation for $(H_2O)_n$ $(n = 3 - 6)$. The first column of data gives the MaxAD in cm^{-1} for each $(H_2O)_n$ isomer using the 2b:Mb method. The MaxAD values are on the order of a few cm^{-1} for clusters smaller than the hexamer. For the $(H_2O)_6$ structures, the largest MaxAD is 6 cm^{-1} for the Bag isomer. For every water structure larger than $(H_2O)_3$, the 2b:Mb MaxAD corresponds to a bound OH stretch, where the hydrogen is participating as a hydrogen bond donor. The "Speedup" column in Table 2.3 under 2b:Mb is simply a ratio of the CCSD(T) wall time to the 2b:Mb wall time (both using the same single node for all computations). For $(H_2O)_6$ isomers, the speedup reaches a factor of 175 for the Boat1 isomer. The 2b:Mb computations could be trivially parallelized and would be over a thousand times faster than the CCSD(T) computation with 10 dedicated nodes.

The 3b:Mb deviations demonstrate significant improvement over the 2b:Mb (3rd col-

umn of data in Table 2.3). In fact, the MaxAD of 1.3 cm^{-1} for Boat1 is the only vibrational mode in any of the $(H_2O)_n$ isomers examined to exceed 1 cm^{-1}. The 3b:Mb speedup factors are much smaller than the 2b:Mb speedups, reflecting the large number of relatively time-consuming CCSD(T) Hessian computations on triads of fragments (i.e., trimers) required for the 3b:Mb approach. Again, parallelization can significantly enhance the performance of the 3b:Mb method.

A more detailed analysis of the 2b:Mb and 3b:Mb deviations is presented in Table 2.4. In Table 2.4, the maximum absolute deviations are separated into free OH stretches (OH_f above 3850 cm^{-1}), bound OH stretches (OH_b between 3300 and 3820 cm^{-1}), intramolecular HOH bends (between 1650 and 1750 cm^{-1}) and low frequency intermolecular modes (below 1100 cm^{-1}).(See Table S4 of the Supporting Information for the average absolute deviations of these modes.) Table 2.4 reveals that 2b:Mb frequencies for the bends and free OH stretches never deviate by more than 0.5 cm^{-1} from the CCSD(T) values. The largest deviations are always due to the bound OH stretches and low frequency intermolecular modes. The 3b:Mb approach reduces the deviations associated with the bound OH stretches to less than 1 cm^{-1} for all of the water clusters examined in this work, and again the MaxAD values associated with the free OH stretches and intramolecular bends are significantly smaller, typically on the order of 0.1 cm^{-1}. Only for the low frequency and bound OH stretching modes do the deviations approach (or exceed in the case of the Boat1 hexamer) 1 cm^{-1}.

Finally, we have added a discussion of closely related work that was published while the current paper was being reviewed. Miliordos, Aprà and Xantheas have reported CCSD(T) frequencies for some of the same water clusters examined here (C_1 trimer, S_4 tetramer, C_1 pentamer along with the Prism, Cage, Book1 and Ring hexamers).[252] Their optimized structures and vibrational frequencies were obtained with the aug-cc-pVDZ basis set via finite difference procedures. To provide a direct comparison to that work, we carried out another set of analytical 2b:Mb and 3b:Mb frequency computations with the aug-cc-pVDZ basis set for their optimized structures. In general, the agreement is excellent. For the cyclic

tetramer, pentamer and hexamer, the average absolute deviation between the finite difference and 3b:Mb procedure does not exceed 0.5 cm^{-1}, and for only one mode (in the cyclic tetramer) did a deviation exceed 1 cm^{-1}. For the Prism, Cage and Book1 isomers, however, the differences grow as large as 10–20 cm^{-1} for a couple of intermolecular modes, while the intramolecular modes deviate by at most 3 cm^{-1}, corresponding to bending motions. All 3b:Mb frequencies differing by more than 2 cm^{-1} from the values reported in Reference 252 are tabulated in the Supplementary Material.[251] We also note that for the $(H_2O)_n$ isomers not examined in that work, CCSD(T) vibrational frequencies agree with previous MP2 results on the nature of each stationary point (C_3 and C_{3h} trimers;[253] C_i, C_4 and C_{4h} tetramers;[161] C_5 and C_{5h} pentamers[254] and the Bag and Boat1 hexamers.[166]

2.3.2.2 Hydrogen Fluoride Clusters

Deviations of harmonic vibrational frequencies for the $(HF)_n$ clusters ($n = 3 - 6$) computed with the 2b:Mb, 3b:Mb and 4b:Mb methods are summarized in Table 2.5. As with the gradient calculations, the harmonic vibrational frequencies of $(HF)_n$ exhibit larger deviations from CCSD(T) compared to $(H_2O)_n$ results. The MaxAD for the 2b:Mb method exceeds 6 cm^{-1} for the tetramer and 10 cm^{-1} for the pentamer. These deviations for $(HF)_6$ isomers approach 8 cm^{-1}. As with gradients, the vibrational frequencies computed with the N-body:Many-body approach can be systematically improved. The MaxADs decrease to around 2 cm^{-1} for the 3b:Mb frequencies and to less than 1 cm^{-1} for the 4b:Mb frequencies. Speedups are not reported for these systems because all of the computations were not performed on the same type of node.

2.4 Conclusions

The N-body:Many-body approach has been applied to geometry optimizations and computation of harmonic vibrational frequencies of $(H_2O)_n$ and $(HF)_n$ clusters. Combining CCSD(T) computations to capture the leading terms in the many-body expansion and MP2 for the higher-order terms is shown to be an efficient and accurate strategy for reproducing

CCSD(T) optimized structures and harmonic vibrational frequencies. For $(H_2O)_n$ clusters ($n = 3-7$), the 2-body:Many-body CCSD(T):MP2 geometry optimizations essentially yield CCSD(T) optimized structures when adopting a convergence criterion of 2.5×10^{-4} E_h a_0^{-1} for the maximum absolute Cartesian force component. For $(H_2O)_7$ optimized structures, the maximum absolute deviations between CCSD(T) and 2-body:Many-body Cartesian gradient components are less than 2.4×10^{-4} E_h ${a_0}^{-1}$ whereas the average absolute deviations are nearly an order of magnitude smaller ($\leq 4.6 \times 10^{-5}$ E_h a_0^{-1}). The 57 independent computations required for the 2-body:Many-body $(H_2O)_7$ gradient (42 pairs, 14 monomers and the cluster) require significantly less time than the single CCSD(T) supermolecular gradient calculation. Each 2-body:Many-body gradient calculation for $(H_2O)_7$ was nearly 40 times faster on a single node than the full CCSD(T) computation on the same computer. The 3-body:Many-body forces are even closer to the canonical CCSD(T) results, with the maximum absolute deviation in Cartesian force components less than 7×10^{-5} E_h ${a_0}^{-1}$. Deviations are somewhat larger for the $(HF)_n$ clusters but can systematically be converged to the CCSD(T) results by including more terms from the many-body expansion with CCSD(T) computations.

Harmonic vibrational frequencies computed with the N-body:Many-body approaches are also in excellent agreement with canonical CCSD(T) supermolecular results for water clusters. The maximum absolute deviations for the 2-body:Many-body method never exceed 6 cm^{-1}. The average absolute deviation is only 0.8 cm^{-1} for all modes of all the H_2O clusters examined here. Deviations for the bending modes and the free OH stretches never exceed 0.5 cm^{-1}, and the intermolecular modes show deviations less than 1 cm^{-1} on average. The largest deviations for the 2-body:Many-body frequencies are seen in the bound OH stretches, where the maximum deviations increase from around 1 cm^{-1} for $(H_2O)_3$ to 6 cm^{-1} for $(H_2O)_6$. The 2-body:Many-body results are obtained on a single node at a fraction of the wall time required for the canonical CCSD(T) calculation, typically more than 100 times faster for the hexamer isomers.

The deviations associated with the 3-body:Many-body CCSD(T):MP2 harmonic vibrational frequencies are significantly smaller than those for the 2-body:Many-body method. The 3-body:Many-body approach exhibits a maximum absolute deviation for the bound OH stretches of only 0.9 cm^{-1} and 1.3 cm^{-1} for the low frequency intermolecular modes. The average absolute deviation for all modes of all water clusters examined here is only 0.1 cm^{-1}. Speedups for the 3-body:Many-body technique approach a factor of 10 for the $(H_2O)_6$ clusters, but this is easily improved with parallelization. Applied to clusters with significant cooperative effects such as $(HF)_n$, N-body:Many-body gradients and frequencies can demonstrate larger deviations relative to CCSD(T), but they can be systematically improved by including more terms from the many-body expansion at the CCSD(T) level.

These encouraging results suggest that the N-body:Many-body approach can be used to generate benchmark-quality geometries and harmonic vibrational frequencies of clusters with more modest computational resources and in less time than is required for supermolecular CCSD(T) computations.

Table 2.1. Maximum absolute deviations from components of CCSD(T) Cartesian gradients (all values in E_h a_0^{-1})

	2b	3b	2b:Mb	3b:Mb
$(H_2O)_3$				
C_1	0.004001	–	0.000048	–
C_3	0.003555	–	0.000042	–
C_{3h}	0.002752	–	0.000032	–
$(H_2O)_4$				
S_4	0.008134	0.000959	0.000077	0.000030
C_i	0.006672	0.000812	0.000073	0.000024
C_4	0.005936	0.000675	0.000067	0.000018
C_{4h}	0.006107	0.000700	0.000059	0.000013
$(H_2O)_5$				
C_1	0.009351	0.001579	0.000106	0.000034
C_5	0.008794	0.001466	0.000097	0.000027
C_{5h}	0.007878	0.001286	0.000086	0.000016
$(H_2O)_6$				
Ring	0.010280	0.001939	0.000111	0.000030
Book1	0.011707	0.001741	0.000186	0.000047
Cage	0.012049	0.001115	0.000162	0.000046
Boat1	0.009595	0.001729	0.000100	0.000030
Bag	0.015519	0.002091	0.000224	0.000061
Prism	0.014650	0.001661	0.000177	0.000052
$(H_2O)_7$				
PR1	0.014149	0.001901	0.000236	0.000067
PR2	0.012684	0.001394	0.000156	0.000045
PR3	0.015040	0.001530	0.000228	0.000063

Table 2.2. Maximum absolute deviations from components of CCSD(T) Cartesian gradients (all values in E_h a_0^{-1})

	2b	3b	4b	2b:Mb	3b:Mb	4b:Mb
$(HF)_3$	0.005863	–	–	0.000064	–	–
$(HF)_4$	0.010970	0.001380	–	0.000174	0.000045	–
$(HF)_5$	0.017378	0.003005	0.000328	0.000256	0.000080	0.000018
Planar	0.018502	0.003629	0.000551	0.000260	0.000079	0.000028
Chair	0.017816	0.003327	0.000479	0.000265	0.000078	0.000041

Table 2.3. Maximum absolute deviations from CCSD(T) harmonic vibrational frequencies (in cm^{-1}) and relative speedups[a] for N-body:Many-body CCSD(T):MP2 computations

	2b:Mb		3b:Mb	
	MaxAD	Speedup	MaxAD	Speedup
$(H_2O)_3$				
C_1	1.3	6	–	–
C_3	1.6	7	–	–
C_{3h}	1.5	6	–	–
$(H_2O)_4$				
S_4	2.5	11	0.7	1
C_i	2.5	9	0.6	1
C_4	2.3	11	0.5	1
C_{4h}	1.9	10	0.3	1
$(H_2O)_5$				
C_1	3.4	66	0.9	4
C_5	3.0	94	0.7	6
C_{5h}	2.7	83	0.5	6
$(H_2O)_6$				
Ring	3.6	85	0.9	4
Book1	4.4	159	0.8	8
Cage	5.8	163	0.7	8
Boat1	3.7	175	1.3	9
Bag	6.0	162	0.8	8
Prism	5.5	145	0.5	7

[a] Speedup is the ratio of CCSD(T) wall time to Nb:Mb total wall time

Table 2.4. Maximum absolute deviations (cm^{-1}) associated with 2- and 3-body:Many-body CCSD(T):MP2/haDZ harmonic vibrational frequencies

	Low		Bends		OH_b Str.		OH_f Str.	
	2b:Mb	3b:Mb	2b:Mb	3b:Mb	2b:Mb	3b:Mb	2b:Mb	3b:Mb
$(H_2O)_3$								
C_1	1.3	–	0.1	–	1.0	–	0.1	–
C_3	1.6	–	0.1	–	0.8	–	0.1	–
C_{3h}	1.5	–	0.1	–	0.6	–	0.1	–
$(H_2O)_4$								
S_4	1.9	0.3	0.2	0.1	2.5	0.7	0.3	0.0
C_i	1.7	0.3	0.1	0.1	2.5	0.6	0.3	0.0
C_4	2.0	0.2	0.3	0.0	2.3	0.5	0.3	0.0
C_{4h}	1.8	0.3	0.3	0.0	1.9	0.3	0.4	0.1
$(H_2O)_5$								
C_1	1.9	0.4	0.3	0.1	3.4	0.9	0.4	0.1
C_5	2.4	0.5	0.3	0.1	3.0	0.7	0.5	0.1
C_{5h}	1.8	0.3	0.3	0.1	2.7	0.5	0.5	0.1
$(H_2O)_6$								
Ring	1.9	0.4	0.5	0.1	3.6	0.9	0.4	0.1
Book1	2.0	0.4	0.3	0.1	4.4	0.8	0.3	0.1
Cage	2.2	0.7	0.4	0.1	5.8	0.3	0.2	0.0
Boat1	1.9	1.3	0.3	0.1	3.7	0.8	0.4	0.1
Bag	1.6	0.8	0.3	0.1	6.0	0.6	0.5	0.2
Prism	2.0	0.5	0.3	0.3	5.5	0.4	0.1	0.1

Acknowledgements

This material is based upon work supported by the National Science Foundation under EPS-0903787 and CHE-0957317. The authors would also like to thank the Mississippi Center for Supercomputing Research for a generous amount of CPU time.

Table 2.5. Maximum absolute deviations from CCSD(T) harmonic vibrational frequencies (cm^{-1}) for N-body:Many-body CCSD(T):MP2 computations for $(HF)_n$ clusters ($n = 3 - 6$)

	2b:Mb	3b:Mb	4b:Mb
$(HF)_3$	2.1	–	–
$(HF)_4$	6.3	1.4	–
$(HF)_5$	10.3	2.1	0.5
Chair	8.0	2.4	0.5
Planar	7.7	2.3	0.7

CHAPTER 3

GETTING DOWN TO THE FUNDAMENTALS OF HYDROGEN BONDING: ANHARMONIC VIBRATIONAL FREQUENCIES OF $(HF)_2$ AND $(H_2O)_2$ FROM *AB INITIO* ELECTRONIC STRUCTURE COMPUTATIONS

This work presents a systematic investigation into the basis set convergence of harmonic vibrational frequencies of $(H_2O)_2$ and $(HF)_2$ computed with second-order Møller-Plesset perturbation theory (MP2) and the coupled-cluster singles and doubles method with perturbative connected triples, CCSD(T), while employing correlation-consistent basis sets as large as aug-cc-pV6Z. The harmonic vibrational frequencies presented here are expected to lie within a few cm^{-1} of the complete basis set (CBS) limit. For these important hydrogen-bonding prototype systems, a basis set of at least quadruple-ζ quality augmented with diffuse functions is required to obtain harmonic vibrational frequencies within 10 cm^{-1} of the CBS limit. In addition, second-order vibrational perturbation theory (VPT2) anharmonic corrections yield CCSD(T) vibrational frequencies in excellent agreement with experimental spectra, differing by no more than a few cm^{-1} for the intramonomer fundamental vibrations. D_0 values predicted by CCSD(T) VPT2 computations with a quadruple-ζ basis set reproduce the experimental values of $(HF)_2$ and $(H_2O)_2$ to within 2 cm^{-1} and 21 cm^{-1}, respectively.

3.1 Introduction

Small model systems have long provided critical insight into the nature of noncovalent interactions. The interactions that stabilize these systems play crucial roles in a variety of biological and environmental phenomena.[3,5,255] $(H_2O)_2$ and $(HF)_2$ have served as prototypes for hydrogen bonding interactions since the first *ab initio* quantum mechanical (QM) calculations on H-bonded dimers in the 1960s.[108,109,111,112,142,143,256–258] The global minimum energy conformations of these dimers are depicted in 3.1, where a mirror plane includes the nuclei associated with the hydrogen bond (X-H$\cdots$X). The hydrogen atoms in the figure are marked to distinguish them as "donor" involved in the hydrogen bond or as a "free" or "acceptor" atom. An accurate description of the water dimer is especially important in the construction of reliable water potentials,[121,233–240,259–261] since the stabilizing intermonomer interactions of larger clusters are dominated by pairwise interactions.[18,112,129,144,150,172,262–266]

The relatively small size of these complexes makes them attractive targets for robust electronic structure methods. The computational demands of accurate coupled-cluster[30] (CC) calculations place significant limits on the number of atoms and the size of the basis set, typically requiring excitations to be limited to singles and doubles (CCSD) and possibly a perturbative treatment of the connected triple excitations as in the "gold standard" CCSD(T) method.[267] However for $(H_2O)_2$, higher orders of electron correlation through full quadruple excitations (CCSDTQ) have been included in benchmark calculations of the

Reproduced with permission from "Getting down to the Fundamentals of Hydrogen Bonding: Anharmonic Vibrational Frequencies of $(HF)_2$ and $(H_2O)_2$ from Ab Initio Electronic Structure Computations," J. Coleman Howard, J.L. Gray, A.J. Hardwick, L.T. Nguyen and Gregory S. Tschumper, *J. Chem. Theory Comput.* *10*(12), 5426–5435. http://dx.doi.org/10.1021/ct500860v Copyright © 2014, American Chemical Society

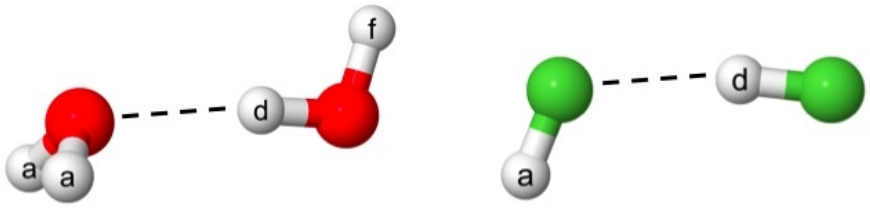


Figure 3.1. Global minimum geometries of $(H_2O)_2$ and $(HF)_2$

dimer's interaction energy and geometry.[141,268] Relativistic and non-Born-Oppenheimer corrections have been examined with CC computations as well.[93]

Several benchmark interaction energy databases have been compiled from high-accuracy QM calculations on noncovalent dimer complexes.[46,68,269–273] By documenting the performance of lower-scaling *ab initio* methods or density functional theory (DFT) approximations across a diverse set of chemical systems, these databases serve as useful guides for selecting appropriate methods for larger systems. This reference data is also of interest to those designing DFT functionals or parameterizing semi-empirical methods. While the collection of benchmark energetics for dimer systems in the literature is substantial, vibrational frequencies of noncovalent complexes computed from high-accuracy *ab initio* QM methods are far less common. Instead of comparing theory to theory, as is necessary for a quantity like interaction energy, computed vibrational frequencies can be compared directly to an experimentally measurable property.

The additional computational effort associated with obtaining optimized molecular geometries and calculating force constants can be a substantial investment for a method such as CCSD(T). Recent experimental success[123,124,274–278] in the formation and spectroscopic characterization of noncovalent clusters provides an excellent opportunity to benchmark *ab initio* methods against experimentally measured properties.

Theoretical studies of the vibrational spectra of hydrogen-bonded clusters often focus on the X-H stretching region, with particular interest in the shift to lower energy (commonly referred to as a "red shift") associated with the donor X-H bond lengthening upon hydrogen bond formation. For water clusters, harmonic vibrational frequency computations indicate that the MP2 method overestimates the shift of the donor O-H stretch to lower energy, relative to CCSD(T) harmonic values and experiment.[252] A proper comparison between calculated and experimental vibrational spectra requires an appropriate anharmonic treatment of the vibrational modes, such as Rayleigh-Schrödinger vibrational second-order perturbation theory (VPT2).[279–282] For the water dimer, Kjaergaard and collaborators[283–289] have

improved on the usual double harmonic approximation to probe the O-H stretching region with a variety of impressive techniques, using reduced dimensionality approaches such as local modes (HCAO) and numerically solving a 1-dimensional vibrational Schrödinger equation for the the donor O-H stretch, in addition to anharmonic treatments of the complete set of water dimer normal modes through perturbative approaches VPT2 and cc-VSCF with high-accuracy QM methods including CCSD(T). Cremer and coworkers have also investigated the $(H_2O)_2$ spectrum in terms of local and normal mode of vibrations with CCSD(T) calculations and VPT2 corrections.[290] VPT2 gives the exact energy levels of a Morse oscillator and thus is well-suited to describe O-H stretching modes. $(H_2O)_2$ stretching frequencies computed with VPT2 in conjunction with the CCSD(T) method and a triple-ζ basis set compare favorably with experimental values, with a maximum deviation less than 30 cm^{-1} in the fundamental modes.[286]

In addition to the water dimer, the HF dimer is a useful hydrogen-bonding model for pursuing near-CBS vibrational frequencies with correlated QM methods. The structure of this dimer is similar to $(H_2O)_2$, possessing C_s point group symmetry where all atoms are contained in the mirror plane. $(HF)_2$ has been the subject of a number of ab initio investigations.[256,258,291–306] Accurate semiempirical and ab initio potentials[307–311] have been developed from QM calculations and $(HF)_2$ experimental spectra.

In this work, we address the issue of basis set convergence of harmonic vibrational frequencies in these two hydrogen-bonded dimers with a systematic approach towards the CBS limit with wavefunction methods MP2, CCSD and CCSD(T). We also make direct comparison to experimental spectra with VPT2 computations.

3.2 Theoretical Methods

Fully optimized geometries were obtained to compute harmonic and anharmonic vibrational frequencies, utilizing analytic first and second geometrical derivatives. Coupled-cluster and MP2 computations were performed with CFOUR[248] and Gaussian 09,[250] re-

spectively. Anharmonic frequencies were computed within the VPT2 method by double-sided numerical differentiation of analytic second derivatives along normal modes.[312] To examine the basis set convergence of the harmonic vibrational frequencies, we employed the correlation-consistent family of basis sets.[57] In particular, we used the cc-pVXZ ($X = \mathrm{D} - 6$, i.e., D,T,Q,5,6) basis sets, as well as the fully augmented aug-cc-pVXZ (aXZ) and "heavy-augmented" versions (ha$X$Z), where diffuse functions were only added to the heavy (i.e., non-hydrogen) atoms. MP2 optimized geometries and harmonic vibrational frequencies were also calculated on a counterpoise-corrected potential energy surface for each basis set to firmly establish CBS estimates and to compare the convergence of the CP-corrected values to those obtained from the standard uncorrected potential energy surfaces. Spherical harmonic d, f, g, h and i functions were used instead of their Cartesian counterparts. All calculations were performed with the "frozen core" approximation (i.e., core electrons were not correlated in post-HF computations). CCSD(T)/ha5Z harmonic frequencies of $(H_2O)_2$ were not calculated analytically. Instead, these frequencies were evaluated by 3-point numerical differentiation of energies computed with Molpro[106] at displaced geometries generated by PSI4.[313] The accuracy of this approach was calibrated using the smaller basis sets, for which CCSD(T) analytic second derivative computations were feasible. Harmonic frequencies computed via this finite difference procedure never differed from the analytic values by more than 1 cm^{-1}.

3.3 Results and Discussion

3.3.1 Geometries

3.1 reports key distances for each monomer and dimer system. The covalent bond lengths are tabulated for the CCSD(T), MP2 and CCSD methods computed with the ha5Z basis set, along with the O$\cdots$O distance for $(H_2O)_2$ and the F$\cdots$F distances for $(HF)_2$. Of particular interest in this study are the changes in the OH and FH bond lengths when donating a hydrogen bond. Changes relative to the monomer are given in parentheses in

3.1, along with the unique FH and OH covalent bond lengths in the dimers. Relative to CCSD(T), MP2 slightly overestimates the distortion of the covalent bonds when the hydrogen bond is formed, whereas the CCSD method does the opposite. The CCSD method overestimates the distance between the the heavy atoms in both dimers relative to CCSD(T) values. The MP2/ha5Z $(HF)_2$ geometry predicts a slightly longer F$\cdots$F separation relative to CCSD(T), while the $(H_2O)_2$ O$\cdots$O distance is slightly shorter at the MP2 level. The CCSD(T)/ha5Z geometries here are in excellent agreement with previous benchmark calculations,[93,94,141,306] with maximum deviations from previous CCSD(T) CBS estimates on the order of 0.005 Å for the heavy atom distances, and the agreement improves by an order of magnitude for the intramonomer distances. Cartesian coordinates are provided in the Supporting Information.[314]

Table 3.1. Covalent bond lengths[a] and key intermonomer distances (in Å) associated with equilibrium monomer and dimer structures optimized with CCSD(T), MP2 and CCSD methods using the ha5Z basis set.

	Parameter	CCSD(T)	MP2	CCSD
HF monomer	R(FH)	0.9173	0.9183	0.9144
H_2O monomer	R(OH)	0.9584	0.9583	0.9555
$(HF)_2$	R(FH_d)	0.9230 (+0.0057)	0.9246 (+0.0063)	0.9195 (+0.0051)
	R(FH_f)	0.9202 (+0.0029)	0.9214 (+0.0031)	0.9172 (+0.0027)
	R(F$\cdots$F)	2.7369	2.7413	2.7509
$(H_2O)_2$	R(OH_d)	0.9647 (+0.0063)	0.9657 (+0.0074)	0.9611 (+0.0056)
	R(OH_f)	0.9576 (−0.0008)	0.9574 (−0.0009)	0.9547 (−0.0009)
	R(OH_a)	0.9591 (+0.0007)	0.9592 (+0.0009)	0.9562 (+0.0007)
	R(O$\cdots$O)	2.9127	2.9061	2.9319

[a] Changes relative to monomer given in parentheses

3.3.2 Basis Set Convergence of MP2 Harmonic Vibrational Frequencies

3.3.2.1 Estimated MP2 CBS Limit

The MP2 CBS estimate of each harmonic vibrational frequency in both the monomers and dimers is listed in 3.2. For each normal mode, the MP2 CBS estimate was calculated

as the mean of the MP2/a6Z and MP2/ha6Z frequencies. For the dimers, these two values were averaged along with their counterpoise-corrected counterparts (CP-MP2/a6Z and CP-MP2/ha6Z) for the CBS limit estimate (4 values altogether). Even though the fully-augmented a6Z basis set has more basis functions, we include the ha6Z frequencies in the CBS estimates because these selectively augmented basis sets tend to provide interaction energies and geometries closer to the CBS limit for hydrogen-bonded clusters.[21,94] For each normal mode, Table 2 also shows the maximum absolute deviation of the MP2/a6Z, MP2/ha6Z, CP-MP2/a6Z and CP-MP2/ha6Z frequencies from the corresponding CBS estimate. For the HF and H_2O monomer modes, the MP2/ha6Z and a6Z harmonic frequencies agree to within 1 cm^{-1}. In the dimers, the MP2/ha6Z, MP2/a6Z and CP-corrected frequencies are, on average, within 2 cm^{-1} of the CBS estimate for $(HF)_2$ and 1 cm^{-1} for $(H_2O)_2$.

3.3.2.2 HF and H_2O Monomer Modes

The basis-set convergence of the MP2 harmonic frequencies towards the CBS limit is illustrated in Figures 2 - 4. For each normal mode, the MP2 aXZ and ha$X$Z deviations from the MP2 CBS limit estimates are plotted across the D$-$6 progression, marked by circles and squares connected by solid lines, respectively. For the dimers, the corresponding CP-corrected deviations are shown as triangles with dotted lines. Under the label for each normal mode, the MP2 CBS values are given in parentheses.

3.2 illustrates the MP2 basis set convergence in the H_2O and HF monomer harmonic vibrational frequencies. The water bend (ν_2) converges quickly, as the MP2/haDZ ν_2 frequency is within 2 cm^{-1} of the MP2 CBS estimate. The HF and OH stretching frequencies converge much more slowly. The MP2/aDZ values underestimate the corresponding CBS frequencies by more than 30 cm^{-1} in each stretching mode. Increasing the basis set to haTZ or aTZ improves the predicted MP2 stretching frequencies to within *ca.* $10 - 20$ cm^{-1}. Agreement to within 5 cm^{-1} of the MP2 CBS limit is achieved with quadruple-ζ basis sets for each mode.

3.3.2.3 $(HF)_2$ and $(H_2O)_2$ Intermonomer Modes

There are 4 intermonomer modes in the HF dimer (ν_5,ν_4,ν_6 and ν_3) and 6 in $(H_2O)_2$ (ν_{12},ν_{11},ν_8, ν_7,ν_6 and ν_{10}). See References 286 and 311 or tables S7 and S13 in the Supporting Information[314] for a description of each of these intermonomer normal modes. As can be seen in Figures 3 and 4, the basis-set convergence of the intermonomer frequencies towards the MP2 CBS limit is very similar in both complexes. For both dimers, the slowest basis-set convergence for the MP2 harmonic frequencies is seen in the 2 highest-frequency intermonomer modes *ca.* $350-650$ cm^{-1} (ν_6 and ν_3 for $(HF)_2$ and ν_6 and ν_{10} for $(H_2O)_2$). In each complex, these 2 modes consist of one in-plane (a') and one out-of-plane (a'') bending motion of the X-H$\cdots$X angle associated with the donor hydrogen involved in the H-bond. Although MP2 frequencies computed with triple-ζ basis sets are within 10 cm^{-1} for the lowest-energy intermonomer modes (< 250 cm^{-1}), basis sets of at least quadruple-ζ quality are required to be within 10 cm^{-1} of the MP2 CBS limit for the more challenging higher-energy intermonomer modes above 350 cm^{-1}. We note that the energetic ordering of the nearby modes ν_{11} (a'') and ν_8 (a') in $(H_2O)_2$ is reversed in the harmonic frequencies computed using the double-ζ basis sets, with the exception of CP-MP2/aDZ.

The CP-corrected and uncorrected values tend to converge from opposite sides of the CBS limit, reminiscent of the convergence of interaction energies.[21] The net effects of the CP procedure, however, are somewhat mixed. In $(H_2O)_2$, for example, the lowest-energy frequencies (ν_{12}, ν_{11} and ν_8) computed with smaller basis sets are usually shifted closer to the MP2 CBS limit when the procedure is applied, particularly for MP2/haDZ frequencies. In contrast, the CP procedure increases the deviations from the CBS limit for the lowest-frequency modes in $(HF)_2$ (ν_5 and ν_4).

The largest differences in MP2 and CP-MP2 intermonomer frequencies are observed for the two highest-frequency intermonomer modes (*ca.* $350-650$ cm^{-1}) in each dimer. Even with the large a6Z basis set, the difference between CP corrected and uncorrected frequencies for $(HF)_2$ grows as large as 8 cm^{-1} for ν_6 and 6 cm^{-1} for ν_3. In $(H_2O)_2$, the largest difference

between MP2/a6Z and CP-MP2/a6Z frequencies also occurs in the out-of-plane bending mode (4 cm^{-1} for ν_{10}).

3.3.2.4 $(HF)_2$ and $(H_2O)_2$ Intramonomer Modes

The basis-set dependence of the intramonomer frequencies in these dimers closely follows what was seen for the isolated monomers in Figure 2 at the MP2 level of theory. The harmonic frequencies of the bending modes in the water dimer are within 10 cm^{-1} of the MP2 CBS estimate with the double-ζ basis sets and within 5 cm^{-1} with the triple- ζ basis sets. The two bending modes in $(H_2O)_2$ are excluded from Figure 4 due to this rapid convergence, but the data can be found in the Supporting Information.[314]

The intramonomer stretching frequencies are underestimated with the smaller basis sets, by more than 60 cm^{-1} for FH donor stretch and 38 cm^{-1} for the symmetric OH acceptor stretch computed with the aDZ basis set. As with the monomers, the quadruple-ζ basis sets are needed to obtain OH and FH stretching frequencies within 5 cm^{-1} of the MP2 CBS limit.

The "red-shifted" donor hydrogen stretches (ν_2 in $(HF)_2$ and ν_3 in $(H_2O)_2$) are affected by the CP procedure much more than the other intramonomer FH and OH stretching modes not directly participating in the hydrogen bond. As shown in 3.3 for $(HF)_2$, the CP procedure hardly affects the MP2 frequency of ν_1, regardless of the basis set. The maximum shift caused by the CP procedure for this free FH stretch in any basis set is less than 4 cm^{-1}. For the donor FH stretch ν_2, on the other hand, the MP2 and CP-MP2 frequencies differ by nearly 30 cm^{-1} with the aDZ and haDZ basis sets. A similar trend is observed for the OH stretches in $(H_2O)_2$. The CP procedure only noticeably affects the OH donor stretch (ν_3), improving the haDZ and aDZ frequencies by roughly 15 cm^{-1} relative to the MP2 CBS limit.

Although the CP procedure tends to move the double and triple-ζ donor stretching frequencies (ν_2 in $(HF)_2$ and ν_3 in $(H_2O)_2$) closer to the MP2 CBS limit, it leads to an appreciable underestimation of the shift to lower energy of these modes relative to the monomer stretching frequencies ($\Delta\nu(FH_d)$ and $\Delta\nu(OH_d)$ in Table 2). The calculated CP-

MP2 double-ζ and triple-ζ values underestimate the CBS shift by more than 20 cm^{-1} and 10 cm^{-1}, respectively. The $\Delta\nu(\mathrm{FH}_d)$ values from the standard MP2 frequencies are within 8 cm^{-1} and 4 cm^{-1} with these basis sets. The same pattern can be seen for the $\Delta\nu(\mathrm{OH}_d)$ in $(\mathrm{H_2O})_2$ as well.

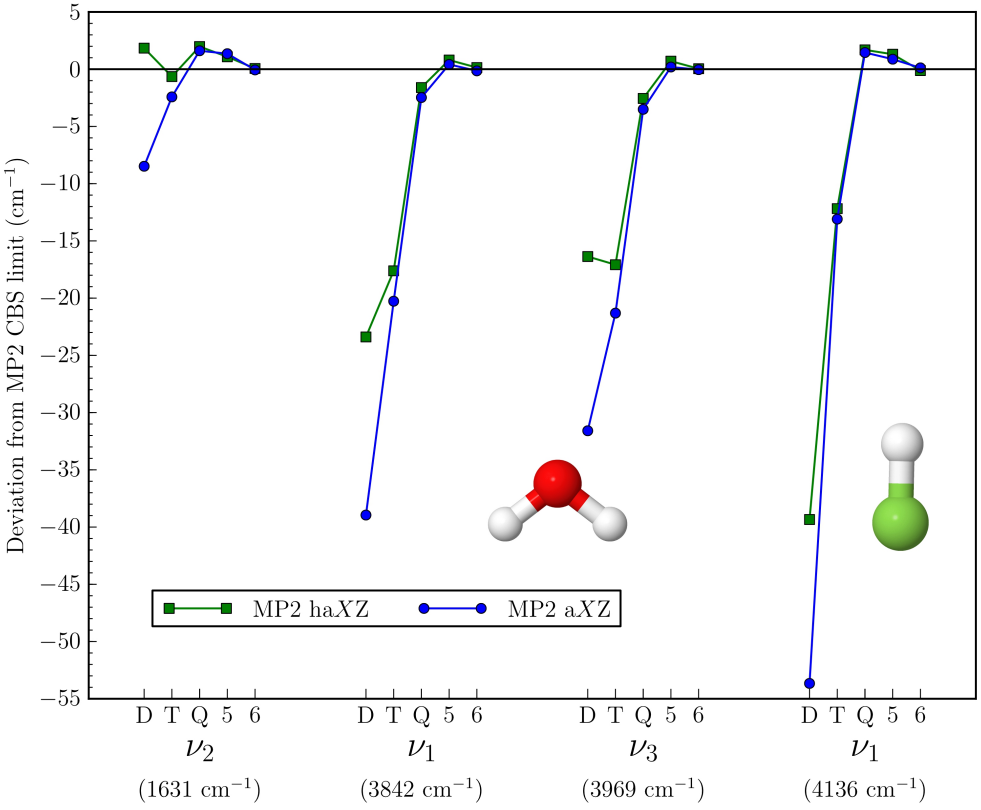


Figure 3.2. Basis set convergence of MP2 harmonic vibrational frequencies of HF (far right) and $\mathrm{H_2O}$ monomers depicted as deviations from the estimated CBS limit (in parentheses)

3.3.3 Comparison of Harmonic MP2 and CCSD(T) Frequencies

Because CCSD(T) harmonic frequency computations with sextuple-ζ basis sets were infeasible, the comparison between MP2 and CCSD(T) frequencies utilizes results obtained with the ha5Z basis set. The MP2/ha5Z harmonic frequencies in the penultimate column of 3.2 are within a few cm^{-1} of the MP2 CBS limit. Even though the effects of the CP procedure will not be identical for the MP2 and CCSD(T) methods, the CCSD(T)/ha5Z harmonic frequencies in the last column of 3.2 are expected to lie close to the CCSD(T) CBS

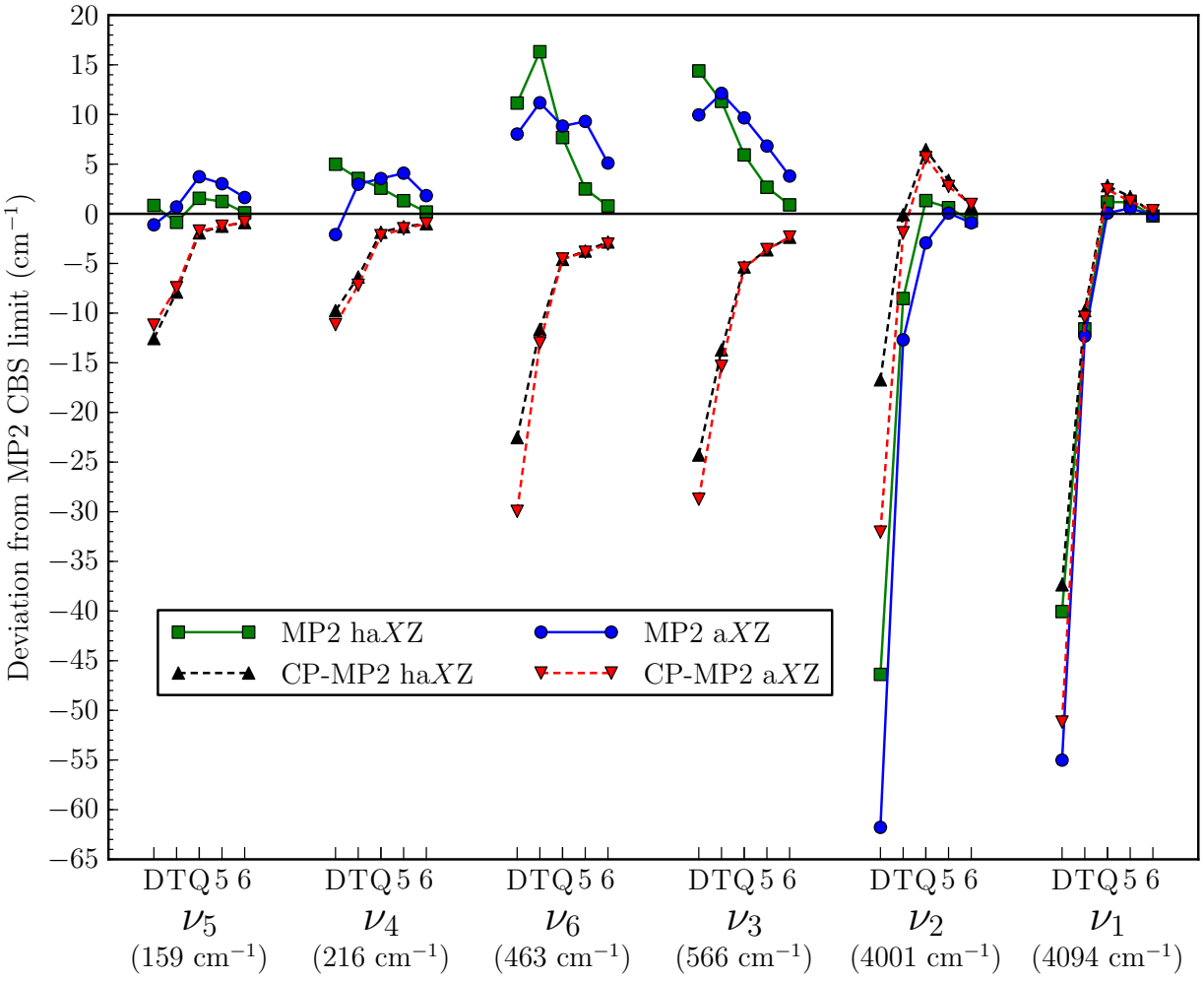


Figure 3.3. Basis set convergence of MP2 harmonic vibrational frequencies of $(HF)_2$ depicted as deviations from the estimated CBS limit (in parentheses)

limit.

For the HF monomer stretch listed first in 3.2, the MP2 value agrees well with the CCSD(T) result, underestimating the harmonic frequency by 6 cm^{-1}. However, the MP2/ha5Z H_2O monomer harmonic frequencies show larger deviations from the corresponding CCSD(T)/ha5Z values. MP2 underestimates the bending mode (ν_2) by 18 cm^{-1} but overestimates both of the stretching frequencies (by as much as 25 cm^{-1} for ν_3).

Comparing the $(HF)_2$ and $(H_2O)_2$ CCSD(T) and MP2 intermonomer modes, the MP2 frequency values are in excellent agreement with CCSD(T). Only ν_{10} in $(H_2O)_2$ differs by 10 or more cm^{-1} between the methods. The average absolute deviation between MP2

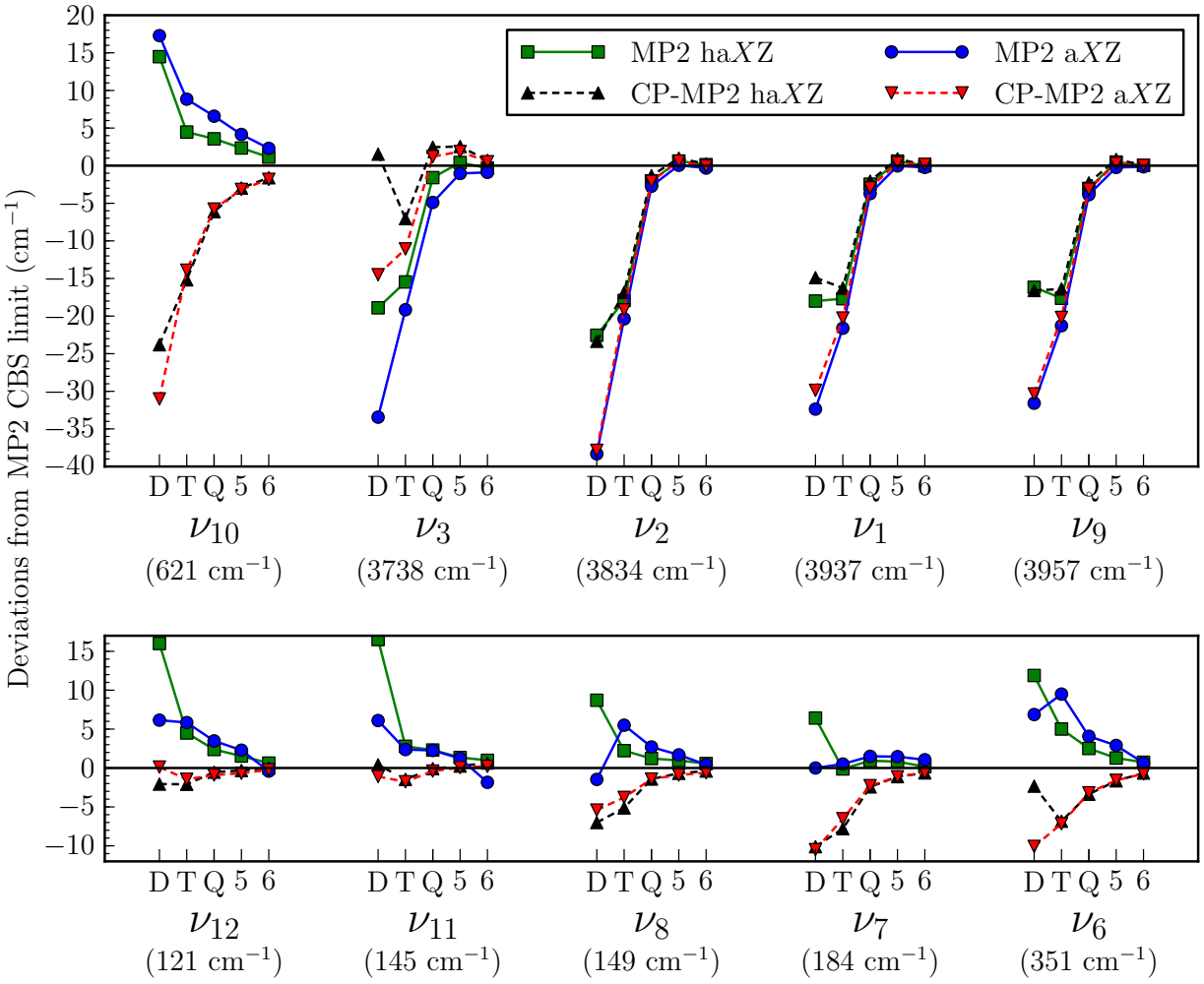


Figure 3.4. Basis set convergence of MP2 harmonic vibrational frequencies of $(H_2O)_2$ depicted as deviations from the estimated CBS limit (in parentheses)

and CCSD(T) intermonomer frequencies is only 3 cm^{-1} with the ha5Z basis set for all 10 intermonomer modes (ν_5, ν_4, ν_6 and ν_3 for $(HF)_2$ and ν_{12}, ν_{11}, ν_8, ν_7, ν_6 and ν_{10} for $(H_2O)_2$).

For the intramonomer bending and stretching modes, much larger discrepancies are observed between MP2/ha5Z and CCSD(T)/ha5Z harmonic frequencies, and they tend to mimic the differences seen for the isolated monomers. MP2 consistently overestimates the free OH stretches relative to CCSD(T). The discrepancy is as small as 5 cm^{-1} for the symmetric acceptor OH stretch (ν_2), and it is large as $20-21$ cm^{-1} for the two highest-frequency OH stretching modes in $(H_2O)_2$ (ν_9 and ν_1). For the HOH bending modes in $(H_2O)_2$, the

MP2/ha5Z harmonic frequencies are always smaller by $17 - 18$ cm^{-1} than the CCSD(T) results, just like the monomer. In $(HF)_2$, the MP2 computations underestimate the free FH stretch by 10 cm^{-1}, while the bound FH stretch is underestimated by 24 cm^{-1} relative to CCSD(T).

The last row in 3.2 for each dimer gives the shift of the donor stretching mode relative to the monomer stretch, denoted $\Delta\nu(FH_d)$ and $\Delta\nu(OH_d)$, computed at the same level of theory (symmetric OH stretch in the case of H_2O). A negative value for $\Delta\nu$ implies a shift to lower energy (a.k.a. a red-shift). The magnitude of the shift is overestimated in both dimer systems at the MP2 level. The overestimation is more severe in the water dimer (-105 cm^{-1} vs -81 cm^{-1} at the MP2 and CCSD(T) levels, respectively). Recall the XH bond lengthening (3.1) associated with hydrogen bond formation (and the red-shifted XH stretching mode) is also larger at the MP2 level for each dimer complex here, as discussed in Section 3.1.

3.4 Anharmonic Vibrational Frequencies and Dissociation Energies

Tables 3-6 summarize the VPT2 frequencies of the HF and H_2O monomers and dimers computed with CCSD(T), MP2 and CCSD methods and the haQZ basis set. The VPT2 results for each electron correlation method are given as deviations from the experimental values, where the sign convention is such that a positive value implies an overestimated frequency relative to the experimental reference. For the reference values, gas-phase values are given when possible.

3.4.1 Monomers

The first row of data in 3.3 compares the experimental HF stretching fundamental to VPT2 results. The CCSD(T) and MP2 values computed with the haQZ basis set are in excellent with the experimental value of 3961 cm^{-1}, each within 2 cm^{-1} of the reference. The CCSD method overestimates the ν_1 position by more than 50 cm^{-1}. For the first overtone

Table 3.2. CBS limit MP2 harmonic vibrational frequencies (in cm^{-1}) of $(HF)_2$ and $(H_2O)_2$, along with MP2 and CCSD(T) values obtained with the ha5Z basis set

	Mode	MP2 CBS est.[a]	Max Abs. dev.[b]	MP2 ha5Z	CCSD(T) ha5Z
HF	ν_1	4136	0	4137	4143
H_2O	ν_2 (a_1)	1631	0	1632	1650
	ν_1 (a_1)	3842	0	3843	3835
	ν_3 (b_2)	3969	0	3970	3945
$(HF)_2$	ν_5 (a')	159	2	160	161
	ν_4 (a')	216	2	217	217
	ν_6 (a'')	463	5	465	461
	ν_3 (a')	566	4	568	566
	ν_2 (a')	4001	1	4002	4026
	ν_1 (a')	4094	0	4095	4105
	$\Delta\nu(FH_d)$	−135	1	−135	−117
$(H_2O)_2$	ν_{12} (a'')	121	1	123	125
	ν_{11} (a'')	145	2	146	142
	ν_8 (a')	149	1	150	148
	ν_7 (a')	184	1	184	185
	ν_6 (a')	351	1	352	351
	ν_{10}(a'')	621	2	624	614
	ν_5 (a')	1632	0	1633	1650
	ν_4 (a')	1651	0	1652	1670
	ν_3 (a')	3738	1	3738	3754
	ν_2 (a')	3834	0	3835	3830
	ν_1 (a')	3937	0	3937	3917
	ν_9 (a'')	3957	0	3957	3936
	$\Delta\nu(OH_d)$	−105	1	−105	−81

[a]Mean of the MP2/ha6Z, CP-MP2/ha6Z, MP2/a6Z and CP-MP2/a6Z values for dimers and mean of the MP2/ha6Z and MP2/a6Z values for monomers

[b]Maximum absolute deviation of the 2 monomer or 4 dimer values used to estimate the MP2 CBS limit from their mean

of the HF stretch in 3.4, the absolute deviation from experiment approximately doubles for each method, as the CCSD(T) and MP2 $2\nu_1$ positions are are within 4 cm^{-1} and 2 cm^{-1} of the experimental spectrum, respectively.

The H_2O VPT2 fundamental frequencies are shown in the first rows of Tables 5. For the 3 normal modes, the CCSD(T)/haQZ values are much closer to experiment than the MP2 predicted frequencies. The largest deviation with the CCSD(T) method is only a 6 cm^{-1} underestimation of the antisymmetric OH stretch ν_3. As with the harmonic vibrational frequencies, the MP2 method predicts OH stretching frequencies too high, compared to CCSD(T) calculations, while the bending mode ν_2 is predicted too low at the MP2/haQZ level. The CCSD/haQZ fundamentals are the farthest from experiment, with the frequencies of the OH stretching modes overestimated by more than 40 cm^{-1} on average.

The first overtones and combination bands of the H_2O vibrational modes are shown in Table 6. For the first bending overtone ($2\nu_2$) and the combination bands involving the bending mode ($\nu_1+\nu_2$ and $\nu_2+\nu_3$), the CCSD(T)/haQZ VPT2 results are again in good agreement with experiment. $2\nu_2$ is within 10 cm^{-1} of experiment, while the two combinations bands only deviate by 2 and 3 cm^{-1}. Because the MP2 method overestimates the fundamental stretching frequencies and underestimates the H_2O bending frequency, rather good agreement is obtained for the combination bands between 5200 cm^{-1} and 5350 cm^{-1} due to a rather favorable cancellation of errors. For example, MP2 underestimates the bending mode ν_2 by -14 cm^{-1} and overestimates the symmetric OH stretch ν_1 by $+12$ cm^{-1}, and the deviation from the experimental value of the combination band $\nu_1+\nu_2$ is -2 cm^{-1}.

The first OH stretching overtones in H_2O calculated with CCSD(T) show much larger deviations relative to experiment. This is due to strong Darling-Dennison resonance,[316] which is unaccounted for in the standard VPT2 implementation used here. An approach such as that undertaken in Reference 316 would shift the CCSD(T) $2\nu_1$ and $2\nu_3$ values closer to the experiment. The OH stretching combination band $\nu_1+\nu_3$ is unaffected by this resonance, and the CCSD(T)/haQZ frequency of 7239 cm^{-1} is in good agreement with the measured

value of 7250 cm^{-1}. The MP2 and CCSD methods overestimate the energy of the $\nu_1+\nu_3$ transition by more than 40 cm^{-1} and 80 cm^{-1}, respectively.

3.4.2 $(HF)_2$

3.3 compares VPT2 fundamental frequencies of $(HF)_2$ computed with CCSD(T), MP2 and CCSD using the haQZ basis set to experimental results.[310,311,317–323] Note, however, that the experimental values for the lowest-energy transitions are actually indirect estimates.[310,311,317,318] The largest deviations from the experimental values occur for the same intermonomer modes that exhibited the slowest convergence to the MP2 CBS limit (ν_6 and ν_3 in Figure 3). MP2 overestimates the ν_6 fundamental by +23 cm^{-1} whereas CCSD and CCSD(T) underestimate ν_3 by −34 cm^{-1} and −22 cm^{-1}, respectively.

In contrast to the monomer, the CCSD(T)/haQZ VPT2 intramonomer frequencies of $(HF)_2$ are in far better agreement with experiment than MP2/haQZ values. The CCSD(T) calculated FH stretching fundamentals are within 1 cm^{-1} of the experimental values, and the $\Delta\nu(FH_d)$ is within 1 cm^{-1} as well. MP2 underestimates both HF stretching fundamental frequencies by −7 cm^{-1} and −17 cm^{-1} for FH_f and FH_d stretches, respectively. In contrast, the CCSD method overestimates these stretching modes by +51 cm^{-1} for ν_1 and +65 cm^{-1} for ν_2.

In the last row of 3.3, the calculated D_0 values are compared to the experimental value of 1062 cm^{-1}.[320] The CCSD(T) D_0 is within 2 cm^{-1} of this reference value. Although this remarkable level of agreement is somewhat fortuitous, it can be understood in light of the recent work of Řezáč et al.[306] They have computed the dissociation energy of the HF dimer with a series of high-level ab initio calculations, including full quadruple excitations via CCSDTQ, relativistic effects and a diagonal Born-Oppenheimer correction. They concluded that the sum of those corrections to D_0 amounts to less than 2 cm^{-1}. That work also computed CCSD(T) VPT2 frequencies with the fully augmented aug-cc-pVQZ basis set, which overall agree with those computed here. The only substantial differences between the

Table 3.3. Deviations between VPT2 fundamental transition energies and experimental vibrational frequencies of HF and $(HF)_2$ and dissociation energy (D_0) (in cm^{-1}) of $(HF)_2$ computed with CCSD(T), MP2 and CCSD methods with the haQZ basis set

		Exp	CCSD(T)	MP2	CCSD
HF	ν_1	3961[a]	+2	+1	+52
$(HF)_2$	ν_5 (a')	125[b]	+7	+6	+4
	ν_4 (a')	161[c]	+11	+11	+3
	ν_6 (a'')	380[d]	+9	+23	+7
	ν_3 (a')	475[e]	−22	−14	−34
	ν_2 (a')	3868[f]	+1	−17	+65
	ν_1 (a')	3931[f]	−1	−7	+51
	$\Delta\nu(FH_d)$	−93[f]	−1	−18	+13
	D_0	1062[g]	−2	−24	−60

[a]Reference 315 (Gas)
[b]Reference 311 (Gas)
[c]Reference 310 (Gas)
[d]Reference 317 (Gas)
[e]Reference 318 (Gas)
[f]Reference 319 (Gas)
[g]Reference 320 (Gas)

two calculations occur in the intermonomer modes, where ν_6 is shifted 20 cm^{-1} higher with the aug-cc-pVQZ basis set. However, the computed frequencies for ν_6 have been shown to be extremely sensitive to small changes in the basis set.[304]

Vibrational overtone and combination band transition energies for the HF dimer are presented in 3.4. For this dimer, the CCSD(T) VPT2 energy levels predict the $2\nu_1$ and $2\nu_2$ transitions to within 8 cm^{-1} of experiment. The first overtone of the donor FH stretch observed at 7550 cm^{-1} is perfectly matched at the CCSD(T)/haQZ level. The combination bands in the 4000 - 4100 cm^{-1} range, corresponding to excitations in one HF stretching mode and one intermonomer mode, are also captured very well by the CCSD(T) method, with a maximum absolute deviation from experiment of only 6 cm^{-1}. The combination band involving ν_3 in the last row reveals larger discrepancies, although only slightly more than the deviation seen for the ν_3 fundamental in 3.3. With the exception of the $2\nu_2$ transition predicted 39 cm^{-1} too low, the MP2 overtones and combination bands are also rather close

Table 3.4. Select VPT2 overtone and combination levels (in cm^{-1}) of HF and $(HF)_2$ computed with MP2, CCSD and CCSD(T) methods with the haQZ basis set

		Exp	CCSD(T)	MP2	CCSD
HF	$2\nu_1$	7751[a]	−4	−2	+98
$(HF)_2$	$2\nu_2$	7550[b]	0	−39	+136
	$2\nu_1$	7683[c]	+8	−2	+117
	$\nu_2+\nu_5$	4001[d]	+5	−12	+65
	$\nu_2+\nu_4$	4049[d]	+6	−8	+63
	$\nu_1+\nu_5$	4058[d]	+4	−2	+53
	$\nu_1+\nu_4$	4099[d]	+3	−2	+48
	$\nu_1+\nu_3$	4418[e]	−27	−24	+13

[a]Reference 315 (Gas)
[b]Reference 322 (Gas)
[c]Reference 323 (Gas)
[d]Reference 321 (Gas)
[e]Reference 319 (Gas)

to the experimental values. The CCSD method consistently overestimates the stretching overtones and combination bands, by more than 100 cm^{-1} in the case of $2\nu_2$ and $2\nu_1$.

3.4.3 $(H_2O)_2$

3.5 compares VPT2 fundamental frequencies of $(H_2O)_2$ computed with CCSD(T), MP2 and CCSD using the haQZ basis set to experimental results.[277,324–328] Note that experimental values for $(H_2O)_2$ vibrational frequencies can vary by several cm^{-1} due to differing experimental conditions. For example, the rotational temperatures have been estimated to be around 5K for gas-phase measurements of the intermonomer fundamentals,[325,326] but the vibrational degrees of freedom "may have different effective temperatures or even non-thermal distributions"[277] depending on the expansion conditions. References 285 and 324 provide thorough reviews of the various experimental vibrational frequencies that can be found in the literature. Also note that several $(H_2O)_2$ vibrational transitions have only been measured in Ne matrix environments, where interactions with the dimer shift $(H_2O)_2$ fundamentals by as much as 20% relative to gas-phase values.[327,329] Lastly, the relatively large tunneling splittings[330] in $(H_2O)_2$ are unaccounted for in the VPT2 calculations.

As with $(HF)_2$, the CCSD(T) fundamental vibrational energies of $(H_2O)_2$ tend to agree quite well with the measured transitions (column 3 of 3.5). For the four lowest-energy vibrations, the CCSD(T) results show a maximum deviation from experiment of -9 cm^{-1} (ν_{12}) and an average absolute deviation around 3 cm^{-1}. For the MP2 and CCSD VPT2 frequencies (last two columns in 3.5), the maximum deviations in these 4 modes are appreciably larger at $+39$ cm^{-1} and -26 cm^{-1}, respectively (ν_{11} and ν_{12}).

Table 3.5. Deviations between VPT2 fundamental transition energies and experimental vibrational frequencies and dissociation energy (D_0) (in cm^{-1}) of $(H_2O)_2$ computed with CCSD(T), MP2 and CCSD methods with the haQZ basis set

		Exp	CCSD(T)	MP2	CCSD
H_2O	ν_2	1595[a]	+3	−14	+18
	ν_1	3657[a]	−2	+12	+46
	ν_3	3756[a]	−6	+27	+38
$(H_2O)_2$	ν_{12}	88[b]	−9	−16	−26
	ν_{11}	108[b]	+3	+39	−6
	ν_8	103[b]	0	0	0
	ν_7	143[c]	0	−1	−5
	ν_6	311[d]	−18	−27	−25
	ν_{10}	523[d]	−28	−18	−50
	ν_5	1599[d]	0	−10	+25
	ν_4	1616[d]	0	−21	+14
	ν_3	3602[e]	+3	−6	+61
	ν_2	3651[e]	−1	+15	+47
	ν_1	3730[e]	−3	+37	+47
	ν_9	3745[e]	+2	+14	+48
	$\Delta\nu(OH_d)$	−55[e]	+6	−18	+15
	D_0	1105[f]	−21	−2	−81

[a]HITRAN 2006 values from Reference 286
[b]Reference 325 (Gas)
[c]Reference 326 (Gas)
[d]Reference 327 (Ne matrix)
[e]Reference 277 (Gas)
[f]Reference 328 (Gas)

The two highest-energy intermonomer vibrations (ν_6 and ν_{10}) are particularly challenging. Although all three methods consistently underestimate these frequencies with respect to experiment, none of them reproduce either of the experimental ν_6 or ν_{10} frequencies

to within 17 cm^{-1}. This results is perhaps not too surprising in light of the slow basis set convergence exhibited by the MP2 harmonic frequencies for these two modes as discussed in Section 3.2.3.

Looking at the intramonomer fundamental frequencies near the bottom of 3.5, the CCSD(T) VPT2 transitions for these six modes show remarkable agreement with the experimental values. All of the OH-stretching fundamental transitions computed with CCSD(T) are within ± 3 cm^{-1} of the gas-phase positions, with $\Delta\nu(OH_d)$ within 6 cm^{-1} of experiment. Deviations from experimental intramonomer vibrational frequencies are significantly larger for the VPT2 results computed with the MP2 and CCSD methods, as with the H_2O monomer. The average absolute deviation in the six intramonomer modes is 17 cm^{-1} with MP2 and 41 cm^{-1} with CCSD. MP2 overestimates the shift associated with the donor stretching mode ($\Delta\nu(OH_d)$) by nearly 20 cm^{-1}, whereas CCSD underestimates the shift by -15 cm^{-1}. These deviations are nearly identical to those seen for the harmonic frequencies in 3.2. It is interesting to note that the order of the two highest-energy MP2 stretching frequencies changes with the anharmonic corrections. This reordering also occurs for VPT2 frequencies computed with the haDZ and haTZ basis sets (see Supporting Information).[314]

D_0 values computed with each correlated method are given in the last row of 3.5. The D_0 values computed with CCSD(T) and MP2 methods compare fairly well with the experimental value[328] of 1105 cm^{-1} (-21 cm^{-1} and -2 cm^{-1}, respectively). With the haQZ basis set, the CCSD(T) electronic dissociation energy (D_e) of 1752 cm^{-1} is slightly larger than the estimated CCSD(T) CBS limit (within the frozen core approximation). With the ha5Z basis set, the CCSD(T) D_e computed here is reduced to 1745 cm^{-1}, in good agreement with the CCSD(T) CBS estimate of 1743 cm^{-1} obtained by Lane[141] utilizing geometries near the CCSD(T) CBS limit. The MP2 D_e values obtained in the present study with the haQZ and ha5Z basis sets are 1747 cm^{-1} and 1740 cm^{-1}, respectively. Despite the much larger errors for MP2/haQZ fundamental frequencies, some favorable error cancellation leads to an MP2/haQZ D_0 that is within 2 cm^{-1} of the experimental value.

Table 3.6. Select VPT2 overtone and combination levels (in cm^{-1}) of $(H_2O)_2$ computed with MP2, CCSD and CCSD(T) methods with the haQZ basis set

		Exp	CCSD(T)	MP2	CCSD
H_2O	$2\nu_2$	3152[a]	+10	−23	+38
	$2\nu_1$	7202[a]	+21	+52	+122
	$2\nu_3$	7445[a]	−43	+27	+46
	$\nu_1+\nu_2$	5235[a]	+2	−2	+66
	$\nu_2+\nu_3$	5331[a]	−3	+12	+56
	$\nu_1+\nu_3$	7250[a]	−11	+43	+85
$(H_2O)_2$	$2\nu_5$	3163[b]	+5	−14	+53
	$2\nu_4$	3194[b]	+3	−37	+31
	$2\nu_3$	7018[b]	+46	+17	+172
	$2\nu_2$	7193[c]	+24	+57	+122
	$2\nu_1$	7362[b]	−49	+34	+56
	$2\nu_9$	7442[b]	−44	−18	+50
	$\nu_3+\nu_4$	5219[d]	−10	−41	+63
	$\nu_2+\nu_5$	5243[b]	−9	−5	+66
	$\nu_9+\nu_5$	5329[d]	−3	+4	+69
	$\nu_1+\nu_4$	5345[d]	−19	+1	+46
	$\nu_1+\nu_3$	7240[c]	+18	+57	+123
	$\nu_9+\nu_2$	7250[c]	−17	+17	+80

[a]HITRAN 2006 values from Reference 286
[b]Reference 327 (Ne matrix)
[c]Reference 331 (Gas)
[d]Reference 332 (Gas)

The bottom part of 3.6 shows $(H_2O)_2$ combination band and overtone transition frequencies for which experimental data[327,331,332] is available for comparison. The CCSD(T) bending overtones ($2\nu_5$ and $2\nu_4$) are within 5 cm^{-1} of the experimental positions.[327] However, the stretching overtones ($2\nu_3$, $2\nu_2$, $2\nu_1$ and $2\nu_9$) show much larger deviations at the CCSD(T) level. As discussed for the H_2O monomer overtones in Section 4.1, a Darling-Dennison resonance is the origin of these large discrepancies.[316]

The VPT2 transition energies of the intramonomer combination bands are shown near the bottom of 3.6. The CCSD(T) transitions involving the bending modes between 5200 cm^{-1} and 5400 cm^{-1} compare favorably with experiment, underestimating the energy of each transition by a maximum of 20 cm^{-1} and less than 10 cm^{-1} on average. CCSD(T)/haQZ

VPT2 results for the stretching combination bands in the last two rows of 3.5 are in better agreement with experiment than the stretching overtones, reproducing the experimental values to within 20 cm^{-1}. The performance of MP2 for these combination bands is inconsistent. Three of the MP2 transitions are predicted to within 5 cm^{-1} of experiment, while the lowest-energy combination band deviates from the experimental value by -41 cm^{-1}, and the combination band arising from excitations in both donor stretching modes ($\nu_1+\nu_3$) deviates by $+60$ cm^{-1}.

3.5 Conclusions

In this work, we have established benchmark values for the harmonic vibrational frequencies of $(H_2O)_2$ and $(HF)_2$ with correlated wavefunction methods and large correlation-consistent basis sets. MP2 harmonic vibrational frequencies have been computed with basis sets as large as aug-cc-pV6Z on both standard and counterpoise-corrected potential energy surfaces. In addition, we have calculated anharmonic vibrational frequencies with these correlated methods and the haQZ basis set by way of VPT2 computations. The main conclusions for this investigation are as follows:

(i) To obtain harmonic vibrational frequencies that are consistently converged to within 10 cm^{-1} of the CBS limit, a basis set of quadruple-ζ quality is needed according to our analysis at the MP2 level of theory with the ha$X$Z and aXZ families of correlation-consistent basis sets. For these dimers, the haQZ basis set gives frequencies exhibiting average absolute deviations from the MP2 CBS limit of less than 3 cm^{-1}. In particular, the harmonic frequencies of the OH and FH stretching modes are greatly improved with these quadruple-ζ basis sets compared to their triple-ζ counterparts. The maximum absolute errors in the stretching modes decreases from $10-20$ cm^{-1} in the haTZ basis set to $1-3$ cm^{-1} with the haQZ basis set. Use of the fully augmented aQZ basis set or application of the counterpoise procedure does not improve the MP2/haQZ harmonic frequencies relative to the CBS values in either of the dimers, despite the additional computational effort.

(ii) Counterpoise corrections affect the MP2 harmonic vibrational frequencies in different ways, depending on the nature of the normal mode. The standard MP2 harmonic calculations tend to overestimate the frequencies of the intermonomer modes, while the CP corrections tend to result in underestimated intermonomer frequencies. This is similar to the convergence patterns seen for binding energies in hydrogen-bonded dimers.[21] The intramonomer modes are hardly affected by the CP correction, with the exception of the donor stretch. The CP procedure for the double- and triple-ζ basis sets examined here moves the donor stretch frequency position to higher energy and closer to the MP2 CBS estimate. Unfortunately, the net result is an underestimated "red shift" that is farther from the estimated CBS limit.

(iii) CCSD(T) anharmonic vibrational frequencies computed with VPT2 and the haQZ basis set show excellent agreement with available experimental values, particularly for the intramonomer fundamental vibrational transitions. The CCSD(T) intramonomer frequencies agree with experiment to within a few cm^{-1} for both the $(HF)_2$ and $(H_2O)_2$ dimers. The low-energy intermonomer fundamental frequencies are predicted well by CCSD(T), typically within 10 cm^{-1}, while the deviations exceed 20 cm^{-1} for the highest-energy intermonomer modes.

(iv) The accuracy of the CCSD(T) predicted anharmonic vibrational transitions decreases slightly for the overtones and combination bands. The combination bands and overtones between 3000 cm^{-1} and 6000 cm^{-1} are typically within 10 cm^{-1} of experiment, while the deviations can exceed 50 cm^{-1} for the first OH stretching overtones located above 7000 cm^{-1}, calculated within the standard VPT2 implementation.[248] All of the $(HF)_2$ overtones and combination bands agree well with experimental spectra, with no computed transitions differing by more than 20 cm^{-1} from experiment and only one transition differing by more than 8 cm^{-1}.

(v) The CCSD(T)/haQZ dissociation energies computed with VPT2 for these dimers agree very well with experiment. The predicted D_0 for $(H_2O)_2$ of 1084 cm^{-1} underestimates

the experimental value by only 21 cm^{-1} and that for $(HF)_2$ of 1060 cm^{-1} is only 2 cm^{-1} below experiment. These results demonstrate the accuracy that can be realized for computing vibrational properties of small clusters, using the CCSD(T) method in conjunction with sufficiently large basis sets as well as an appropriate treatment of anharmonic effects with VPT2.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant nos. EPS-0903787 and CHE-1156713. The authors also thank the Mississippi Center for Supercomputing Research for providing computational resources.

CHAPTER 4

BENCHMARK STRUCTURES AND HARMONIC VIBRATIONAL FREQUENCIES NEAR THE CCSD(T) COMPLETE BASIS SET LIMIT FOR SMALL WATER CLUSTERS: $(H_2O)_{n=2,3,4,5,6}$

A series of $(H_2O)_n$ clusters ranging from the dimer to the hexamer have been characterized with the CCSD(T) and the 2-body:Many-body CCSD(T):MP2 methods near the complete basis set (CBS) limit to generate benchmark-quality optimized structures and harmonic vibrational frequencies for these important systems. Quadruple-ζ correlation-consistent basis sets that augment the O atoms with diffuse functions have been employed in the analytic computation of harmonic vibrational frequencies for the global minima of the dimer, trimer, tetramer and pentamer as well as the Ring, Book, Cage and Prism isomers of the hexamer. Prior calibration [*J. Chem. Phys.*, **139**, 184113 (2013) and *J. Chem. Theor. Comput.*, **10**, 5426 (2014)] suggests that harmonic frequencies computed with this approach will lie within a few cm^{-1} of the canonical CCSD(T) CBS limit. These data are used as reference values to gauge the performance of harmonic frequencies obtained with other *ab initio* methods (e.g., LCCSD(T) and MP2) and water potentials (e.g., TTM3-F and WHBB). This comparison reveals that it is far more challenging to converge harmonic vibrational frequencies for the bound OH stretching modes in these $(H_2O)_n$ clusters to the CCSD(T) CBS limit than the free OH stretches, the n intramonomer HOH bending modes and even the $6n-6$ intermonomer modes. Deviations associated with the bound OH stretch-

Reproduced with permission from "Benchmark Structures and Harmonic Vibrational Frequencies Near the CCSD(T) Complete Basis Set Limit for Small Water Clusters: $(H_2O)_n = 2, 3, 4, 5, 6$," J. Coleman Howard and Gregory S. Tschumper, *J. Chem. Theory Comput.* *11*(5), 2126–2136. http://dx.doi.org/10.1021/acs.jctc.5b00225 Copyright © 2015, American Chemical Society

ing harmonic frequencies increase rapidly with the size of the cluster for all methods and potentials examined, as do the corresponding frequency shifts relative to the monomer OH stretches.

4.1 Introduction

Hydrogen bonding plays critical roles in a variety of biological and environmental phenomena.[3,5,255,333–335] As the noncovalent interaction that binds together H_2O molecules, hydrogen bonding is ultimately responsible for the unique properties of water. Obtaining reliable theoretical descriptions of the simplest water clusters is an essential step to understanding the complex hydrogen bonding dynamics in liquid water.[336] To this end, *ab initio* quantum mechanical (QM) wavefunction methods have provided valuable insight into the structures and energetics of small water clusters.[153]

The water dimer is the smallest water cluster and has long served as a model for hydrogen bonding interactions since the first *ab initio* investigations into $(H_2O)_2$.[108–114,125,126,257] As a small model system, the $(H_2O)_2$ global minimum has been the focus of benchmark electronic structure calculations,[93,141,290] at levels of electron correlation typically too demanding for larger chemical systems. An accurate description of $(H_2O)_2$ is crucial in the development of water potentials, as pairwise interactions are the dominant stabilizing force in larger clusters.[18,112,129,144,150,172,262–266] However, larger clusters offer insight into an important phenomenon missing in the dimer, hydrogen bonding cooperativity.[18]

The global minima for $(H_2O)_n$ $(n = 3 - 5)$ are characterized by cyclic homodromic hydrogen-bonding networks (i.e., each H_2O accepts and donates one hydrogen bond). Energetics of these small, cyclic minima (see 4.1) have been examined at correlated levels in detail.[27–29,149,150,155–163,165,253] The efforts of Xantheas and coworkers[84,129,131,193,337] have been particularly important in establishing benchmark energetics and basis set convergence patterns in these systems. Second-order Møller-Plesset perturbation theory (MP2)[31] within the "frozen-core" approximation is sufficient for calculating high-accuracy energetics for the

lowest energy configurations of these small cyclic clusters, as the effect of correlating the core electrons is of the same small magnitude but opposite sign of including higher-order dynamical electron correlation in the valence space.[152,253] Infrared vibrational spectra of the neutral $(H_2O)_m$ $(m = 3 - 5)$ cyclic minima have been measured in the gas phase[324,338] in helium droplets,[339,340] and also in solid matrix environments.[341,342]

When the hydrogen-bonding topology ceases to follow the homodromic pattern seen for water trimer, tetramer and pentamer minima, MP2 is no longer sufficient for establishing benchmark-quality *ab initio* energies because the higher-order correlation effects can be appreciable and quite different for each type of network.[166] The water hexamer represents an important structural transition for water clusters. Six water molecules form the smallest group for which 3-dimensional hydrogen bonding networks (see 4.2) possess lower electronic energies than configurations exhibiting the 2-dimensional homodromic motif seen in the trimer, tetramer and pentamer.[84,166,172–174,176,177] (Note that "2-dimensional" here refers to a planar or quasi-planar arrangement of the oxygen atoms.)

The first isolated 3-dimensional structure characterized experimentally was the "Cage" form observed by Liu *et al.*[175] CCSD(T) CBS limit energies[166] and full-dimensional quantum simulations[54] of low-lying hexamer isomers suggested that the "Prism" form has a lower electronic energy and should be nearly isoenergetic with the Cage after accounting for zero-point vibrational energy (ZPVE). More recently both the Cage and the Prism have been observed in broadband rotational spectra.[123] The six-membered analog of the smaller homodromic global minima is the higher-energy "Ring" structure in 4.2, for which IR spectra have been recorded in liquid helium droplets[167,339] and in matrix environments.[341–343] Firm assignment of vibrational transitions associated with $(H_2O)_6$ is more challenging not only due to the increased size of the system but also due to the potential presence of multiple isomers.[122,344,345] Nevertheless, exciting progress continues to be made in the vibrational spectroscopy of water clusters.[324] In fact, the first Raman vibrational spectra of isolated water clusters were reported in 2014.[346]

While the literature contains a large volume of high-quality *ab initio* data for water cluster energetics,[153] computed vibrational frequencies of small water clusters at the CCSD(T) level are scarce.[236,252,347–349] Bowman and coworkers have made significant theoretical contributions, designing intermolecular *ab initio* potentials fit to CCSD(T) energies and providing valuable insight into experimental spectra with coupled local mode models which include anharmonicity.[121,122,236,347,350] Another *ab initio*-based water potential, TTM3-F, has been applied to a range of water clusters from the dimer to the liquid phase.[351]

A reliable description of vibrational frequencies that can be systematically improved and compared to experiment first requires a high-quality harmonic potential. A study of $(HF)_2$ and $(H_2O)_2$ found that anharmonic vibrational frequencies computed with CCSD(T) and a selectively augmented quadruple-ζ basis set yields fundamental frequencies within a few cm^{-1} of experiment for intramonomer modes.[349] In this work, we compute benchmark harmonic vibrational frequency values for the cyclic water trimer, tetramer and pentamer global minima, as well as the Ring, Book, Cage and Prism isomers of the hexamer. In addition to CCSD(T) analytic harmonic frequencies,[267] we use an accurate QM:QM approach[348] based on the many-body expansion[112] to obtain CCSD(T)-quality results analytically with a larger correlation-consistent quadruple-ζ basis set. These current benchmark values are compared to the results of several *ab initio* methods and two water potentials.

4.2 Theoretical Methods

Fully optimized geometries and harmonic vibrational frequencies for eight $(H_2O)_m$ ($m = 2 - 6$) minima (see 4.1 and 4.2) were computed at the MP2 level,[31] employing Dunning's correlation-consistent[57] cc-pVXZ (X=D,T,Q) basis set for hydrogen atoms and aug-cc-pVXZ for the "heavy" atoms (i.e., oxygen). This application of the correlation-consistent family of basis sets, hereafter referred to as ha$X$Z, has been shown to provide dissociation energies and geometries closer to the CBS limit for water clusters than their fully augmented counterparts.[21,94] For each $(H_2O)_m$ isomer, optimal geometries and harmonic vibrational fre-

quencies were also computed with the 2b:Mb CCSD(T):MP2 method (*vide infra*) and the TTM3-F potential.[351]

All geometry optimizations and harmonic vibrational frequencies were computed with Gaussian09,[250] utilizing analytic gradients and Hessians for the MP2 and CCSD(T):MP2 methods. For the $(H_2O)_5$ and $(H_2O)_6$ clusters, MP2/haQZ vibrational frequencies were computed by finite differences of gradients evaluated with MPQC[249] at displacements generated with PSI4.[313] The CCSD(T) analytic gradients and Hessians needed for the CCSD(T):MP2 calculations were provided by CFOUR.[248] Structures and harmonic frequencies for the TTM3-F potential[351] were computed via finite differences generated and evaluated in Gaussian09 using analytic gradients computed with a freely available copy of the potential.[352] All computations were performed in the "frozen core" approximation (i.e., the 1s-like core electrons of O were not correlated in post-HF calculations). For optimized geometries, the maximum absolute Cartesian force component never exceeds 2.5×10^{-4} E_h a_0^{-1}.

To extend the computationally demanding CCSD(T) treatment of electron correlation to larger basis sets, we have applied a QM:QM scheme, based on the traditional many-body expansion for the interaction energy of weakly-bound clusters,[112] which we refer to as N-body:Many-body QM:QM.[225–231,348] Here, the many-body expansion is carried out from 1-body (monomer) terms through N-body terms using a high-level QM method, and any higher-order cooperativity in the cluster ($> N$-body) is captured with a less-demanding QM method. While a reasonable estimate for the energy of a neutral cluster composed of m fragments ($E = E_1 + E_2 + \cdots + E_m$) can sometimes be obtained by truncating the many-body expansion after the inclusion of 2- or 3-body terms (e.g., $E \approx E_1 + E_2 + E_3$),[18,112,228] a much more accurate approximation can be realized by including the trailing terms with a lower-level QM method.[217,224,228,232,348]

This N-body:Many-body (Nb:Mb) approach applied to a homogeneous or heterogeneous cluster composed of m fragments can be viewed as correcting a many-body expansion truncated at order N by computing all higher-order ($N+1$ through m) interactions with a

lower level method, as in Equation 4.1.

$$E^{\mathrm{Hi:Lo}}_{N\mathrm{b:Mb}} = \sum_{i=1}^{N} E_i^{\mathrm{Hi}} + \sum_{j=N+1}^{m} E_j^{\mathrm{Lo}} \tag{4.1}$$

An alternative interpretation of the procedure is as an electron correlation correction, as in Equation 2, where the cluster's total supermolecular energy at a lower level of theory ($\mathrm{E^{Lo}}$) is corrected term-by-term with a more complete treatment of electron correlation in a many-body expansion truncated at order N.

$$E^{\mathrm{Hi:Lo}}_{N\mathrm{b:Mb}} = E^{\mathrm{Lo}} + \sum_{i=1}^{N} (E_i^{\mathrm{Hi}} - E_i^{\mathrm{Lo}}) \tag{4.2}$$

The linear expressions in Equations 4.1 and 4.2 are easily differentiated to obtain analogous expressions for the gradient and Hessian. Explicit expressions for the geometrical derivatives have already been presented elsewhere.[229,231,348]

This study uses CCSD(T) as the "high-level" method to compute up through the 2-body (2b) terms and MP2 as the "low-level" method to capture all higher-order terms, which defines the 2b:Mb CCSD(T):MP2 method. The 2b:Mb CCSD(T):MP2 technique was applied to the computation of optimized geometries and harmonic vibrational frequencies of $(H_2O)_n$ clusters ($n = 3 - 6$).[348] With the haDZ basis set, 2b:Mb CCSD(T):MP2 optimized structures and vibrational frequencies were virtually identical to canonical CCSD(T) results, with a maximum deviation in the harmonic frequencies of less than 6 $\mathrm{cm^{-1}}$ and an average absolute deviation less than 1 $\mathrm{cm^{-1}}$ for all normal modes in a set of 16 different water clusters.[348]

In this work, we extend the 2b:Mb treatment to the larger haTZ and haQZ basis sets. For all minima, 2b:Mb CCSD(T):MP2/haTZ and haQZ optimized geometries and harmonic vibrational frequencies were computed. Since all of the QM:QM computations were performed within the 2-body:Many-body approximation, CCSD(T):MP2 hereafter implies 2b:Mb CCSD(T):MP2.

4.3 Results and discussion

4.3.1 Water cluster geometries

Since the vibrational frequencies of water clusters are sensitive to geometrical changes (particularly for the OH stretching modes),[252] we first briefly discuss the the basis set convergence of the CCSD(T):MP2 geometries of in terms of the covalent bond lengths (R(OH)) and the hydrogen bond distances (R(O$\cdots$H)).

The R(OH) haDZ bond lengths are overestimated by 0.006 to 0.007 Å relative to corresponding haQZ values for the dimer and trimer. In the larger clusters, the maximum R(OH) deviations are very similar, but the deviations cover a much wider range for the clusters with more hydrogen bonds (Prism and Cage). In most cases, the R(OH) haTZ deviations relative to haQZ are roughly half the magnitude of the corresponding haDZ deviations.

The CCSD(T):MP2/haDZ optimized R(O$\cdots$H) distances are too long relative to the haQZ values in every case, with the H-bond distances in the dimer and cyclic minima overestimated by 0.01 Å to 0.02 Å. The haDZ deviations are larger in the 3-dimensional hexamers. Maximum deviations in the Prism, Cage and Book R(O$\cdots$H) are 0.07 Å, 0.04 Å and 0.03 Å. The larger haTZ basis set offers significant improvement in the intermonomer separations, as the largest absolute deviation from the corresponding haQZ reference is only 0.004 Å (Cage). The basis set convergence of the R(OH) and R(O$\cdots$H) parameters can be seen in more detail in Figure S1 of the Supporting Information.[353]

4.3.2 Harmonic Vibrational Frequencies

When a particular $(H_2O)_n$ cluster coalesces from n monomers, a total of $6n-6$ new low-energy vibrational modes are formed (*ca.* 10–1000 cm^{-1}) that are predominantly of an intermonomer nature. That leaves $3n$ higher-energy intramonomer vibrational modes: n intramonomer bending vibrations (*ca.* 1600–1800 cm^{-1}) and $2n$ OH stretching vibrations (*ca.* 3200–3900 cm^{-1}). The changes in the monomer OH stretching frequencies induced by the formation of hydrogen bonds provide valuable spectral signatures for these water

clusters. When these small H_2O clusters form, the stretching frequencies of the OH groups participating in the hydrogen bonds (i.e., the "bound OH" stretches or "H-bonded OH" stretches) are shifted to substantially lower energies (on the order of 10^2 cm^{-1}) relative to the OH stretching frequencies of an isolated water monomer ($\nu_1 = 3657$ cm^{-1} and $\nu_3 = 3756$ cm^{-1}).[354] In contrast, the perturbations are much smaller (on the order of a few cm^{-1}) if the OH group is not participating in a hydrogen bond (i.e., the "free OH" stretches). The large shift to lower energies of the bound OH stretches is commonly referred to as a "red shift."

4.3.2.1 Basis Set Convergence

The accuracy of the CCSD(T):MP2 harmonic frequencies for these clusters was previously demonstrated for the haDZ basis set,[348] where the largest deviations from canonical CCSD(T) typically occurred for the lowest-energy OH stretching mode but never exceeded a few cm^{-1}. In the present study, it was feasible to compute canonical CCSD(T)/haTZ geometries and harmonic frequencies for the $(H_2O)_3$ and $(H_2O)_4$ clusters, The excellent agreement between the canonical CCSD(T) and CCSD(T):MP2 vibrational frequencies clearly extends to the larger haTZ basis set for which the frequencies never differ by more than 1 cm^{-1} for the trimer and 4 cm^{-1} for the tetramer. The average absolute deviation between CCSD(T):MP2/haTZ harmonic frequencies and canonical CCSD(T) results for all of the normal modes in these two clusters is only 0.9 cm^{-1}. (All CCSD(T) and CCSD(T):MP2 unscaled harmonic vibrational frequencies computed with the haTZ basis set are provided in the Supporting Information).[353]

The basis-set convergence of the CCSD(T):MP2 harmonic frequencies is consistent with another study[349] that examined the basis set convergence of the MP2 and CCSD(T) harmonic vibrational frequencies for $(H_2O)_2$ and $(HF)_2$. That work demonstrated that basis sets of at least quadruple-ζ quality (with or without counterpoise corrections) are required to obtain MP2 frequencies that are consistently within 10 cm^{-1} of the estimated CBS limit values obtained from basis sets as large as aug-cc-pV6Z.

The absolute deviations of the CCSD(T):MP2/haDZ and haTZ frequencies relative

to haQZ are plotted in 4.3 (all of which are provided in the Supporting Information).[353] For the dimer, the CCSD(T):MP2/frequencies are actually pure CCSD(T) frequencies computed in Reference 349. Tabulated below the figure are the number of OH stretches (□ symbols), bending modes (× symbols) and intermonomer modes (◯ symbols). The deviations of the harmonic OH stretching frequencies from haQZ values exceed 25 cm^{-1} with the haDZ basis set and approach nearly 20 cm^{-1} with the haTZ basis set. None of the OH stretching frequencies computed with the haTZ basis set lies within 10 cm^{-1} of the haQZ value. The haTZ basis set does offer significant improvement over haDZ for intermonomer frequencies. The maximum deviations in these modes with the haTZ basis set are typically less than half the magnitude of the corresponding maximum deviations associated with the haDZ basis set.

4.3.2.2 Performance of Other Electronic Structure Methods

In 4.4, harmonic vibrational frequencies of $(H_2O)_n$ ($n = 2-5$) clusters computed with MP2, LCCSD(T) and the CCSD(T) methods using double-, triple- and quadruple-ζ basis sets are compared to the benchmark CCSD(T):MP2/haQZ frequencies. The deviations from the reference values are plotted along the x-axis as vertical bars, with each bar corresponding to a single harmonic frequency. Each row of 4.4 presents the results for a particular method and basis set (labeled on the far right) within a particular $(H_2O)_n$ cluster (labeled on the far left). Within each of these rows, the results are further divided into the deviations for intermonomer modes (blue bars on the top of each row), bending modes (green bars in the middle of each row) and OH stretching modes (red bars at the bottom of each row).

For a given cluster in 4.4, the first 3 rows give the deviations for MP2/haDZ, MP2/haTZ and MP2/haQZ values, respectively. The largest discrepancies between MP2 and the benchmark CCSD(T):MP2/haQZ frequencies occur for OH stretching frequencies. For example, the very first two rows of red bars in 4.4 show that MP2/haDZ and MP2/haTZ computations underestimate the OH stretching frequencies by roughly 30 cm^{-1} for the dimer. With the haQZ basis set (3rd row of red bars), the MP2 stretching frequencies are slightly closer

to the reference values, but the largest deviations in this case are due to 2 OH stretches overestimated by ca. 20 cm^{-1}. For the intermonomer modes (blue bars at the top of each row), the MP2 deviations are virtually indistinguishable for the haTZ and haQZ basis sets. The average absolute deviation in the intermonomer modes of $(H_2O)_2$ is less than 5 cm^{-1} for MP2/haTZ and MP2/haQZ frequencies. The MP2 bending frequencies (green bars in the middle of the each row) are consistently underestimated relative to CCSD(T) by 16 - 20 cm^{-1} in the dimer .

The MP2 results for $(H_2O)_3$, $(H_2O)_4$ and $(H_2O)_5$ indicate that the deviations from CCSD(T) in the OH stretching frequencies (red bars) increase with cluster size. In particular, the bound OH stretching modes are responsible for the largest discrepancies in clusters larger than the dimer. For these $(H_2O)_n$ clusters, where $n = 3, 4, 5$, previous comparison of CCSD(T) and MP2 harmonic frequencies with the aug-cc-pVDZ basis set has shown that the MP2 method overestimates the red shift of the hydrogen-bonded bound stretching modes.[252] In 4.4, the n leftmost red bars on each MP2 row correspond to the deviations in bound OH stretching frequencies for a given $(H_2O)_n$ cluster ($n > 2$). (See Supporting Information[353] for a complete list of harmonic frequencies.) In the trimer, tetramer and pentamer, the most red-shifted OH stretching frequency is underestimated at the MP2/haQZ level by 33 cm^{-1}, 55 cm^{-1} and 67 cm^{-1}, respectively.

For the intermonomer modes (blue bars), the MP2 deviations increase more slowly with cluster size compared to OH stretching modes (red bars). From the dimer to the pentamer, the MP2/haQZ average absolute deviation only increases from 4 cm^{-1} to 9 cm^{-1} (and from 5 cm^{-1} to 11 cm^{-1} with haTZ). In stark contrast, however, intermonomer frequencies computed with the haDZ basis set are overestimated relative to the CCSD(T):MP2/haQZ benchmark values by as much as 40 cm^{-1} in $(H_2O)_5$. The accuracy of MP2 water bending frequencies remains consistent with increasing cluster size. Average and maximum absolute deviations in these modes are between 16 cm^{-1} and 20 cm^{-1} for $(H_2O)_3$, $(H_2O)_4$ and $(H_2O)_5$ at the MP2/haQZ level.

4.5 compares MP2 harmonic frequencies for the hexamer structures to benchmark CCSD(T):MP2/haQZ values and is organized in the same manner as 4.4. The first four rows of 4.5 give the frequency deviations for the Ring hexamer, a cyclic isomer analogous to the global minimum configurations of $(H_2O)_3$, $(H_2O)_4$ and $(H_2O)_5$. The deviations of the MP2 frequencies of the Ring are similar to those seen in $(H_2O)_5$. With each of the 3 basis sets employed in the MP2 calculations, the average absolute deviation across all harmonic frequencies (17 – 20 cm^{-1}) is only 1 cm^{-1} larger in the $(H_2O)_6$ Ring than in the pentamer. The MP2 deviations for each subset of normal modes (intermonomer with blue bars, bend with green bars and OH stretches with red bars) are also very similar between the pentamer and Ring structures, with the largest deviations corresponding to underestimated bound stretching frequencies (far left red bars in each row).

The Book, Cage and Prism hexamers are not homodromic cyclic structures, instead having 7, 8 and 9 hydrogen bonds (and bound stretching modes), respectively. Like the cyclic minima, the largest differences between MP2 and CCSD(T) frequencies also occur in the most red-shifted stretching modes. The most red-shifted OH stretching frequencies computed for these 3 hexamer isomers at the MP2/haQZ level fall 79–88 cm^{-1} lower than the current benchmark reference values. This discrepancy actually exceeds 100 cm^{-1} with the haTZ basis set for the Cage isomer. For the bending modes and intermonomer modes, the deviations between CCSD(T):MP2 and MP2 do not significantly increase for the hexamer isomers relative to the cyclic pentamer. With the haQZ basis set, the average absolute deviations for the bending frequencies are 16–17 cm^{-1}, and the average absolute deviations for the intermonomer modes range from 7–9 cm^{-1}.

The CCSD(T)/aDZ bound stretching deviations in 4.5 are interesting because the basis set incompleteness error in these modes is apparently less than the intrinsic error of the MP2 method. For the most red-shifted OH mode in each hexamer, the CCSD(T)/aDZ values are roughly 30–40 cm^{-1} closer to the benchmark reference frequencies than the MP2/haQZ values. In addition, the intermonomer modes calculated at the CCSD(T)/aDZ level are

closer to the reference values, with smaller average and maximum absolute deviation values for every hexamer in 4.5. This is not the case for the free OH stretching modes, as the CCSD(T)/aDZ harmonic frequencies are underestimated by more than 30 cm^{-1} in each $(H_2O)_6$ isomer.

Harmonic vibrational frequencies have also been reported[236] for $(H_2O)_2$,$(H_2O)_3$, $(H_2O)_4$ and the Prism isomer of $(H_2O)_6$ that were obtained numerically with a local treatment of electronic correlation at the CCSD(T) level of theory, denoted LCCSD(T). The T0 approximation[355] was employed for the triple excitations in those computations. The deviations between these LCCSD(T) harmonic frequencies and CCSD(T):MP2/haQZ values are also shown in 4.4 and 4.5 below the MP2 results. The LCCSD(T) frequencies from Reference 236 were computed using fully-augmented aXZ basis sets. As such, the deviations of available canonical CCSD(T) frequencies computed with aXZ basis sets (from References 236 and 252) are also shown in the row(s) following the LCCSD(T) deviations. Together they demonstrate that the largest deviations associated with the LCCSD(T) frequencies are not due to the small differences in the basis set (*vide infra*).

For the dimer, the OH stretching frequencies computed with the LCCSD(T) method compare favorably with the canonical CCSD(T) results obtained with the same basis set. The maximum LCCSD(T)/aTZ deviation from the reference values is no larger than that for CCSD(T)/aTZ. In fact, the average absolute deviation is actually of smaller magnitude for the former than the latter. The LCCSD(T)/aQZ OH stretches for $(H_2O)_2$ are, on average, within 5 cm^{-1} of the reference benchmark values. However, the largest deviation associated with the LCCSD(T)/aQZ OH stretching modes in the dimer is due to the important bound stretching mode that is too large by more than 13 cm^{-1}. The free stretches at this level of theory are all within 2 cm^{-1} of the CCSD(T)/haQZ values.

The LCCSD(T) frequencies of the trimer and tetramer show much larger deviations relative to the benchmark CCSD(T):MP2/haQZ values. The maximum absolute deviations for LCCSD(T)/aTZ and aQZ frequencies are more than twice what is seen for the same

basis sets in $(H_2O)_2$. The LCCSD(T)/aTZ frequencies reported for $(H_2O)_3$ and $(H_2O)_4$ underestimate several intermonomer modes by more than 50 cm^{-1} (blue bars on the far left). The aQZ basis set offers some improvement for these modes in the trimer, but the largest deviation still approaches 50 cm^{-1}. The LCCSD(T) method actually reproduces the canonical CCSD(T) bending frequencies very well when computed with the same basis set, always within 5 cm^{-1} for the clusters in 4.4.

The only LCCSD(T) values available for these hexamer isomers are the haTZ harmonic frequencies of the Prism.[236] This data helps show how the the LCCSD(T) deviations corresponding to OH stretching modes (red bars) grow as the cluster size increases. The largest deviations from the benchmark CCSD(T):MP2/haQZ stretching frequencies are due to bound OH stretches (red bars to the far right of the LCCSD(T) rows in 4.4 and 4.5). The maximum overestimation of these OH stretching frequencies increases from 33 cm^{-1} in the trimer to 61 cm^{-1} in the tetramer with the aTZ basis set and to more than 100 cm^{-1} in the Prism with the haTZ basis set. The LCCSD(T)/haTZ deviations in the other modes (green and blue bars) of the Prism, however, are comparable to those seen for the smaller water clusters.

For $(H_2O)_n$ ($n = 2-5$) clusters, intramonomer frequencies computed with a modified version[356] of the Gaussian-3 composite method[357] have also been reported in the literature. The harmonic frequencies computed with this model chemistry, hereafter denoted G3*, reproduce the CCSD(T):MP2/haQZ bending frequencies (green bars) to within *ca.* 10 cm^{-1} regardless of cluster size. The G3* OH stretching frequencies (red bars), on the other hand, are consistently larger than the reference values, with the error increasing to nearly 100 cm^{-1} for the pentamer. No blue bars are plotted for the G3* method because intermonomer frequencies were not reported.[356]

4.3.2.3 Comparison of Hydrogen Bond Induced Frequency Shifts

It is important to keep in mind that the deviations reported for the OH stretching frequencies of these water clusters in Figures 4 and 5 (red bars) do not necessarily provide

insight into the reliability of red shifts computed with the same method. For example, error cancellation between the computed frequencies for the monomer and cluster could produce an accurate red shift even when a method experiences large deviations for the absolute OH stretching frequencies of a cluster.

The magnitude of the largest red shift has been calculated for each cluster by subtracting the lowest-energy OH stretching frequency from ν_3 of the H_2O monomer computed at the same level of theory. Note that application of the CCSD(T):MP2 scheme to a monomer is simply a canonical CCSD(T) computation. As such, the benchmark CCSD(T):MP2/haQZ red shifts are calculated relative to CCSD(T)/haQZ monomer frequencies. At this level of theory, the magnitude of the largest red shift for each cluster ranges from 190 cm^{-1} for the dimer to 640 cm^{-1} for the Prism hexamer. These benchmark values are listed at the bottom of Figure 6, which shows the deviations of other methods from the reference maximum red shifts. The large positive deviations associated with the MP2 method indicate that it consistently overestimates the magnitude of the red shifts in these clusters by roughly 40–120 cm^{-1}. This trend is consistent with the MP2 data presented in Figures 4 and 5 where the bound OH stretching frequencies were appreciably underestimated at the MP2 level of theory (leftmost red bars in the MP2 rows).

In contrast, one sees in Figure 6 that the LCCSD(T) frequencies from Reference 236 underestimate the benchmark red shifts. With triple-ζ quality basis sets, the LCCSD(T) deviation associated with the largest red shift exceeds 80 cm^{-1} for the tetramer and 130 cm^{-1} for the Prism. Again, this trend for the LCCSD(T) data in in Figure 6 is consistent with the sizeable overestimation of the OH stretching frequencies seen in Figures 4 and 5 (rightmost red bars in the LCCSD(T) rows). The Supporting Information[353] provides the data needed to calculate every OH frequency shift for each cluster at each level of theory.

Regarding the basis set effects, it is interesting to note that the MP2/haTZ deviations in the bound stretching modes are always larger than the deviations in haDZ or haQZ for clusters larger than the dimer in 4.4 and 4.5. Examining the shifts in 4.6, the haDZ basis

set underestimates the red shifts in these clusters, compared with triple-ζ and quadruple-ζ basis sets in each method. Thus the use of the haDZ basis set slightly compensates for the tendency of MP2 to overestimate the red shifts, producing somewhat smaller deviations than the MP2/haTZ and MP2/haQZ maximum red shifts.

4.3.2.4 Performance of Water Potentials

The harmonic frequencies computed with two potentials designed by fitting to *ab initio* data were also included for comparison in 4.7. The parameterization of the TTM3-F potential[351] is, in part, based on MP2/aDZ frequencies. The WHBB potential[121,347] employs an accurate semi-empirical 1-body potential[358] along with fits to CCSD(T)/aTZ energies and MP2/aTZ energies for the 2-body and 3-body potentials, respectively. In the WHBB potential, TTM3-F is used for higher-order (beyond 3-body) interactions and long-range interactions. Note that the construction of the WHBB potential is in some ways similar to the 2-body:Many-body CCSD(T):MP2 technique being used to generate our current benchmark values. The WHBB harmonic frequencies used in this comparison were taken from References 236, 347 and 350.

The WHBB deviations from the benchmark reference harmonic frequencies for $(H_2O)_2$ are quite small. The maximum absolute deviation is less than 7 cm^{-1}, corresponding to the bound OH stretching mode. The intermonomer frequencies and bending frequencies are within 5 cm^{-1} and 3 cm^{-1}, respectively, of the reference data.

In 4.7, the results include two WHBB variants that are denoted 3b5 and 3b6, which differ in the order of the polynomial expression for the 3-body interaction. Both forms of the WHBB potential give overall good agreement with the CCSD(T):MP2/haQZ frequencies. The largest change between the 3b5 and 3b6 variants is the shift in bound OH stretching frequencies to lower energy, with the 3b6 form decreasing the average absolute deviation from 30 cm^{-1} to 10 cm^{-1}.

For larger clusters, the WHBB-3b5 harmonic frequencies overestimate the bound OH stretches. The largest deviation from CCSD(T):MP2 bound stretches occurs in the hexamer

Ring with one mode overestimated by more than 120 cm^{-1}. This contrasts with the MP2 frequencies for which the largest deviations were associated with the non-cyclic hexamers. It should be noted that among these hexamer isomers, the cooperative binding effects (i.e., beyond 2-body interactions) are largest in the cyclic Ring.[18] Given the accuracy of this potential for the dimer, the treatment of cooperative effects in WHBB is the most likely source of these large discrepancies.

For the bending modes, the WHBB potential shows excellent agreement with the benchmark values, with a maximum absolute deviation less than 12 cm^{-1} for the Ring isomer. The WHBB intermonomer frequencies are only available for comparison in $(H_2O)_3$, and these are similar to CCSD(T)/aTZ deviations. The TTM3-F vibrational frequencies show much larger deviations. This is not surprising since the potential was parameterized based on MP2/aDZ values. As with the MP2 values, the bound OH stretching frequencies are shifted too low. The largest deviation actually occurs in the $(H_2O)_3$ with the lowest bound stretch predicted over 240 cm^{-1} too low. The TTM3-F deviations for the larger clusters are all less than 200 cm^{-1}.

4.4 Conclusions

The 2b:Mb CCSD(T):MP2 method has been applied to water clusters ranging from $(H_2O)_3$ to $(H_2O)_6$ with basis sets as large as haQZ. Based on results for $(H_2O)_2$ and $(HF)_2$[349] and prior calibration,[348] we expect these harmonic vibrational frequencies to lie close to the CCSD(T) CBS limit. These benchmark-quality harmonic vibrational frequencies served as a reference point to gauge the performance of other harmonic frequencies reported in the literature from other high-level *ab initio* computations and various water potentials.

MP2 harmonic vibrational frequencies of these water clusters depart significantly from the current benchmark values, particularly for the red-shifted hydrogen-bonded OH stretching modes, where MP2/haQZ computations underestimate the frequencies by more than 80 cm^{-1} and the corresponding red shifts by more than 100 cm^{-1} in the 3-dimensional

water hexamer structures. CCSD(T) and CCSD(T):MP2 harmonic frequencies computed with modest double-ζ basis sets are in better agreement with benchmark reference values than MP2 frequencies computed with quadruple-ζ basis sets. Compared to the bound OH stretches, MP2 intermonomer, bending and free OH stretching harmonic frequencies computed with the haQZ basis set are in much better agreement with the benchmark values, never differing by more than 25 cm^{-1}, 20 cm^{-1} and 21 cm^{-1} respectively.

For $n \geq 3$, LCCSD(T) harmonic frequencies from the literature deviate substantially from the benchmark values in both intermonomer modes and bound OH stretches due to underestimated red shifts. We plan to investigate how the definition of orbital domains and the degree of electron correlation can be improved to obtain harmonic frequencies in better agreement with canonical CCSD(T) values for these clusters.

The values calculated in this work should guide development and refinement of water potentials. In particular, comparison between CCSD(T)-quality harmonic vibrational frequencies and those from two semi-empirical potentials (TTM3-F and WHBB) suggests that the treatment of cooperative effects in water potentials warrants further investigation.

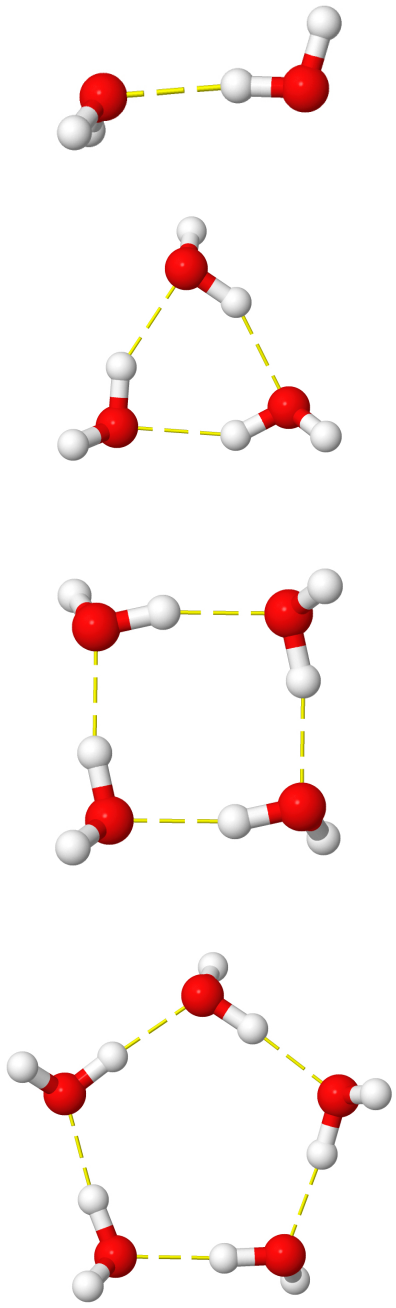
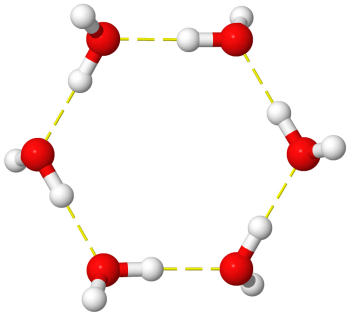
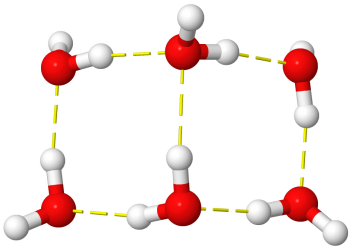


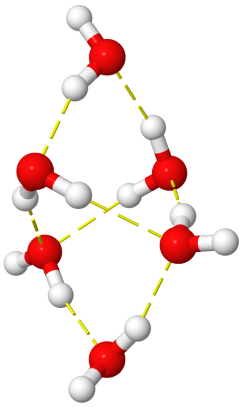
Figure 4.1. $(H_2O)_n$ minima $(n = 2 - 5)$ examined in this study



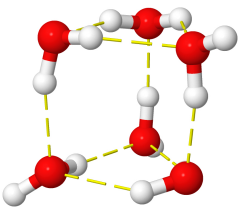
(a) Ring [S_6]



(b) Book [C_1]



(c) Cage [C_1]



(d) Prism [C_1]

Figure 4.2. $(H_2O)_6$ minima examined in this study with point group symmetries in square brackets

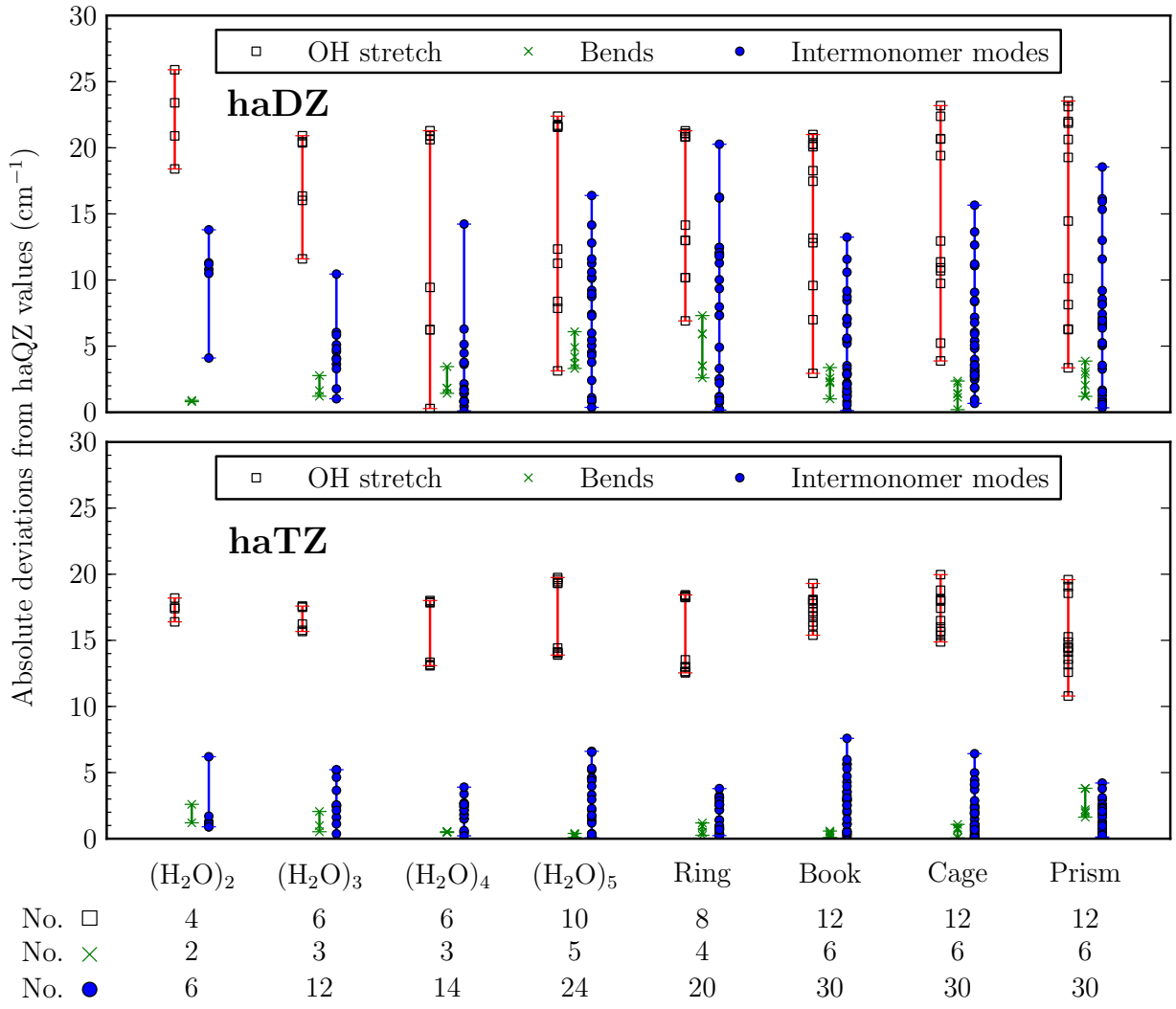


Figure 4.3. Deviations between 2b:Mb CCSD(T):MP2/haQZ harmonic vibrational frequencies and corresponding haDZ values (top) and haTZ values (bottom). The connecting lines only serve as visual aids. (The number of OH stretches, bends and intermonomer modes is indicated below each isomer.)

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant nos. EPS-0903787, CHE-0957317 and CHE-1338056. The authors would also like to thank the Mississippi Center for Supercomputing Research for providing computational resources.

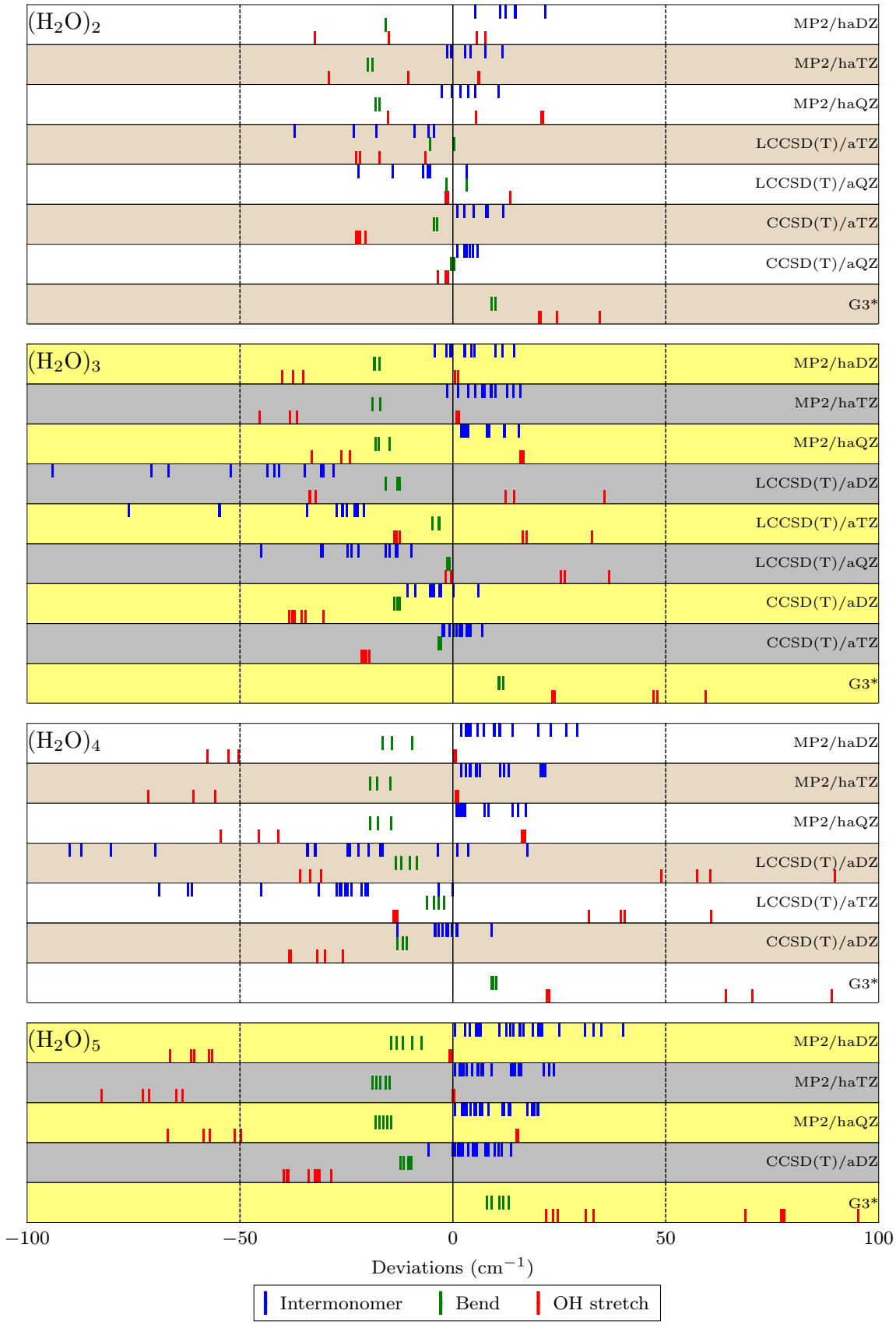


Figure 4.4. Deviations from the 2b:Mb CCSD(T):MP2/haQZ harmonic vibrational frequencies of $(H_2O)_n$ isomers ($n = 2 - 5$) for various levels of theory

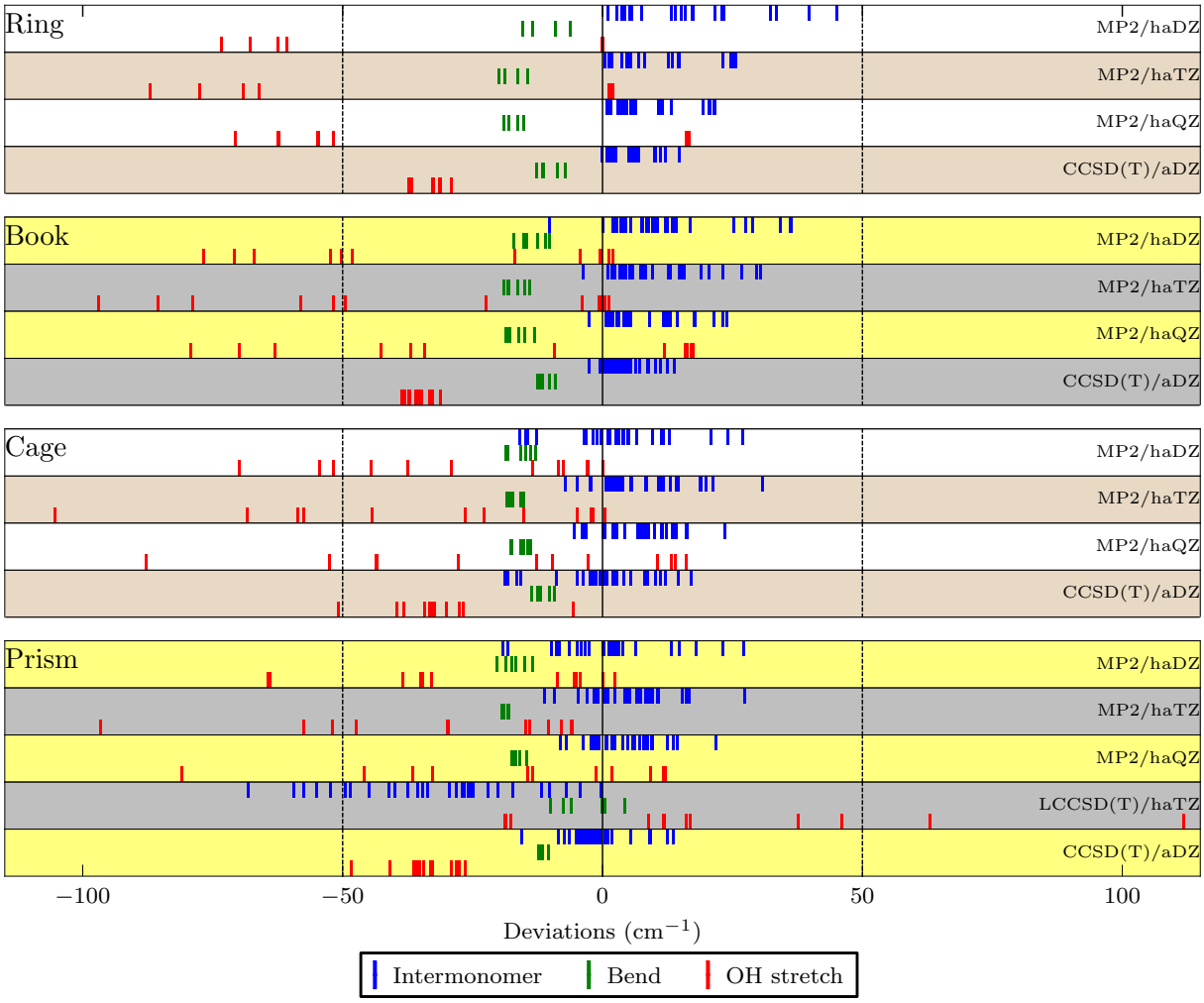


Figure 4.5. Deviations from the 2b:Mb CCSD(T):MP2/haQZ harmonic vibrational frequencies of $(H_2O)_6$ isomers for various levels of theory

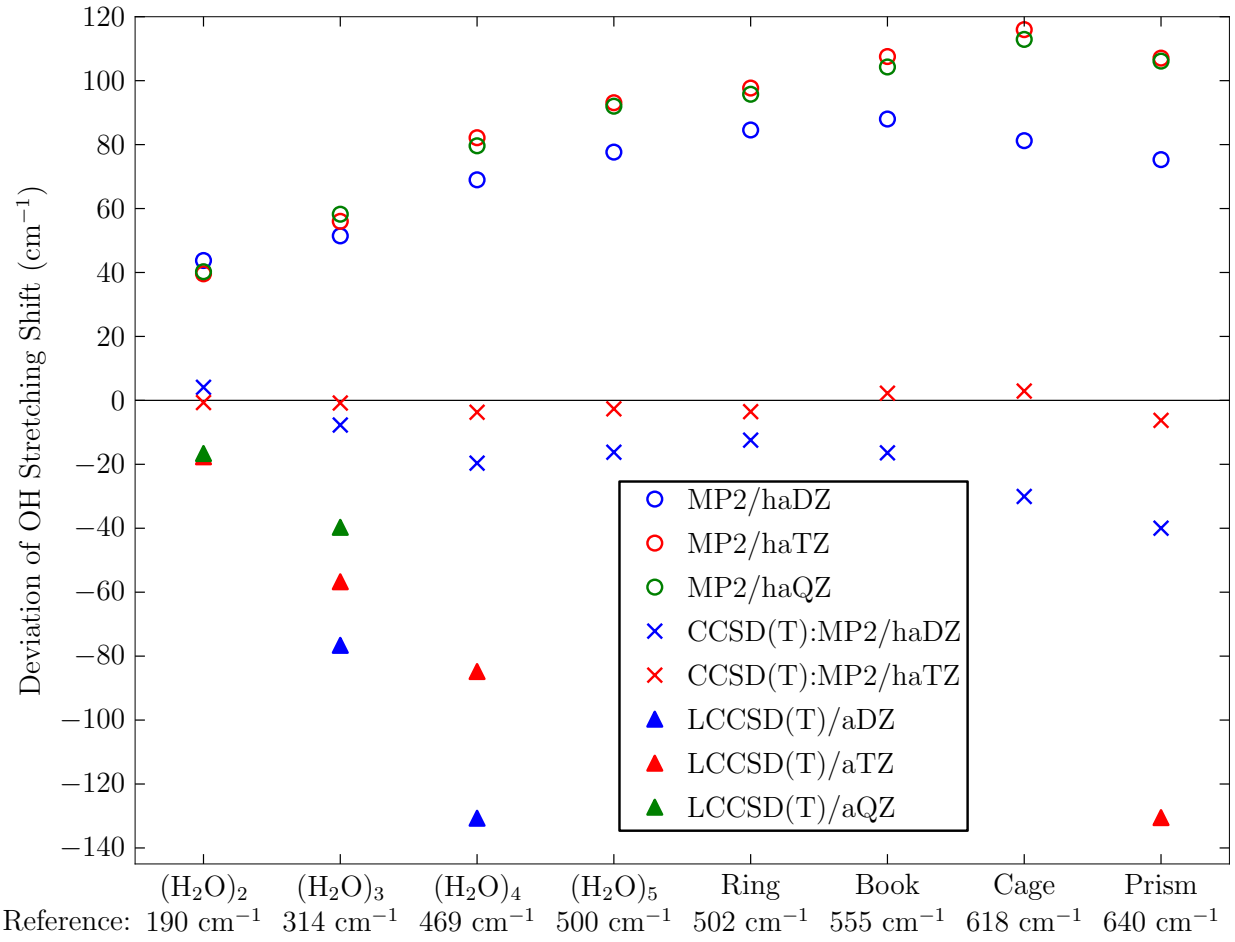


Figure 4.6. Deviation from the largest 2b:Mb CCSD(T):MP2/haQZ red shift (relative to H_2O ν_3) for various levels of theory

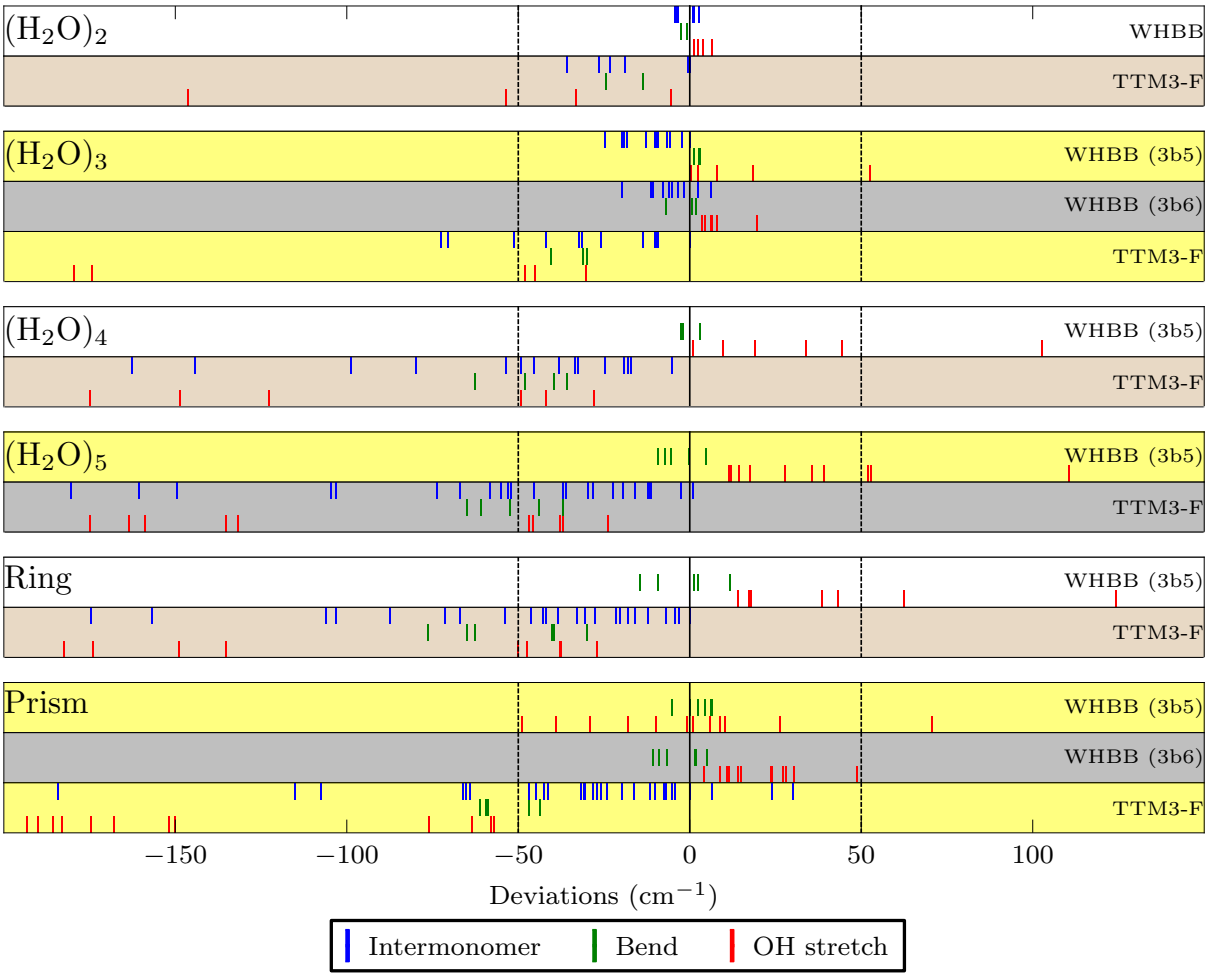


Figure 4.7. Deviations from the 2b:Mb CCSD(T):MP2/haQZ harmonic vibrational frequencies of select $(H_2O)_n$ isomers ($n = 2 - 6$) for WHBB and TTM3-F potentials

CHAPTER 5

ASSESSING THE ACCURACY OF SOME POPULAR DFT METHODS FOR COMPUTING HARMONIC VIBRATIONAL FREQUENCIES OF WATER CLUSTERS

A wide range of density functional theory (DFT) methods (37 altogether), including pure, hybrid, range-separated hybrid, double-hybrid and dispersion-corrected functionals, have been employed to compute the harmonic vibrational frequencies of eight small water clusters ranging in size from the dimer to four different isomers of the hexamer. These computed harmonic frequencies have been carefully compared to recently published benchmark values that are expected to be very close to the CCSD(T) complete basis set (CBS) limit. Of the DFT methods examined here, ωB97 and ωB97X are the most consistently accurate, deviating from the reference values by less than 20 cm^{-1} on average and never more than 60 cm^{-1}. The performance of double-hybrid methods including B2PLYP and mPW2-PLYP is only slightly better than more economical approaches, such as the M06-L pure functional and the M06-2X hybrid functional. Additionally, dispersion corrections offer very little improvement in computed frequencies.

5.1 Introduction

The vibrational frequencies of a water molecule are particularly sensitive to the surrounding environment. Consequently, vibrational spectroscopy represents a powerful tool for investigating both molecular structure and dynamics of aqueous systems from the gas to

Reprinted with permission from "Assessing the Accuracy of Some Popular DFT Methods for Computing Harmonic Vibrational Frequencies of Water Clusters" J. Coleman Howard, Jordan D. Enyard and Gregory S. Tschumper, *J. Chem. Phys. 143* (21), 214103. http://dx.doi.org/10.1063/1.4936654 Copyright © 2015, American Institute of Physics

the condensed phases.[336,359–366] However, a unique assignment of the spectra measured for water under different conditions and in different environments is often nontrivial due to the dynamical nature of the underlying hydrogen-bonding network in which both the number and the strength of the hydrogen bonds continually fluctuate.[367] The difficulty in making an unambiguous assignment of the observed spectroscopic features has led to numerous controversies, including those about the nature of association bands,[368–370] and the relationship between structural order and spectral intensity at the air/water interface.[365,371–375]

Since the advent of computer simulations, a myriad of molecular models for water based on either force fields (including different degrees of empiricism) and *ab initio* representations have been proposed. Although a quantitative assessment of the accuracy of several common PESs for water has been reported based on the many-body expansion of the interaction energy,[240] a quantitative assessment of the associated spectroscopic accuracy remains elusive. The scaling of accurate wavefunction methods necessarily limits their application for computing benchmark-quality properties of water clusters, useful for calibrating more economical approaches including water PESs and density functional theory (DFT) methods. DFT calibration studies on water clusters[52,179,376–383] often focus on interaction or binding energies since vibrational frequencies computed with the "gold standard" CCSD(T) method are rare in the literature.[236,252,347–349] Recently, however, accurate *ab initio* reference harmonic frequencies for small water clusters have been obtained through 2-body:Many-body CCSD(T):MP2 calculations[225,226,228–231,348] with the aug-cc-pVQZ basis set for O and cc-pVQZ for H,[384] which are expected to lie within a few cm^{-1} of the CCSD(T) complete basis set (CBS) limit. In this work, we compute harmonic vibrational frequencies of small water clusters with a large number of DFT methods for comparison with the benchmark 2-body:Many-body CCSD(T):MP2 values.

The water clusters examined in this study are shown in Figure 1. The global minima of $(H_2O)_n$ $(n = 3 - 5)$ consist of cyclic structures with each water monomer donating and receiving one hydrogen bond.[153] The "free" hydrogens not participating in hydrogen bonds

alternately lie above and below the quasi-plane composed of oxygen atoms. In terms of electronic energy, Figures 1(f)-1(h) show the 3 most stable hexamer isomers in decreasing electronic energy, with the Cage and Prism essentially isoenergetic.[166] In helium and neon expansions, the Prism:Cage:Book ratio was observed to be 1:1:0.25,[123] while only the Cage has been observed in argon expansions.[123,175] In addition, at temperatures over 150 K, the Book isomer is predicted to become the most prevalent hexamer isomer,[54] while cyclic structures with six hydrogen bonds have lower Gibbs free energies above 200 K.[28] Since the isomer compositions are so dependent on experimental conditions, an accurate theoretical treatment is crucial to distinguishing among them.

5.2 Computational Methods

For eight isomers of $(H_2O)_n$ clusters ($n = 2 - 6$), full geometry optimizations and harmonic frequency computations were performed with the analytic gradients and Hessians available in the Gaussian09 software package.[250] For each optimized minimum, the maximum Cartesian force component was less than 1×10^{-5} E_h ${a_0}^{-1}$. For the numerical integrations, the computations implemented a pruned grid of 99 radial shells and 590 angular points per shell. In one case (the M11 $(H_2O)_5$ harmonic frequency computation), a larger grid of 150 radial shells and 974 points per shell was needed to remove a spurious imaginary frequency. Dunning's correlation-consistent[57] basis sets were used with diffuse functions added to only oxygen atoms (aug-cc-pVXZ for O and cc-pVXZ for H). These basis sets are hereafter referred to as ha$X$Z basis sets and, in this study, include haDZ and haTZ, with 33 and 74 basis functions for a single water molecule, respectively. A variety of DFT methods were employed, including M06-L,[385] M11-L,[386] N12,[387] MN12-L,[388] PBE,[389] BLYP,[390–392] SOGGA11,[393] B97D,[394] B3LYP,[391,392,395] X3LYP,[396] M06-2X,[397] PBE0,[398] SOGGA11-X,[399] ωB97,[400] ωB97XD,[401] CAM-B3LYP,[402] LC-ωPBE,[403–405] MN12-SX,[406] M11,[407] HSEH1PBE,[408–413] HISSbPBE,[414] N12-SX,[406] B2PLYP[415] and mPW2-PLYP.[416] In addition, the same methods were used in conjunction with a dispersion correction using Grimme's original D3 damping function[417]

for those functionals with D3 parameters as defined in Gaussian09.[250]

5.3 Results and discussion

For each $(H_2O)_n$ isomer, the DFT harmonic vibrational frequencies computed with the haTZ basis sets were compared to the corresponding benchmark values from Reference 384. All DFT frequencies, including those obtained with the smaller haDZ basis set, are reported in the Supplementary Material.[418] The average and maximum absolute deviations (AvgADs and MaxADs, respectively) were determined by directly comparing the DFT haTZ and benchmark frequencies when ordered by energy because the character of the normal modes, particularly the low-energy intermolecular vibrations, can vary appreciably across the range of DFT methods utilized here. As such, it is possible the deviations could increase slightly in some cases if the order of the DFT vibrational modes is different than that of the reference computations. For the important redshifted "bound" OH stretching frequencies which are useful in differentiating water cluster isomers, we have verified that in ordering those vibrational modes by energy, the normal mode character of each bound stretching vibration remains consistent across all methods used here.

The AvgADs and MaxADs are collected in Table 1 and Figure 2 along with the corresponding MP2/haTZ deviations from Reference 384 The functionals are grouped into 4 classes according to their functional type. From top to bottom in Table 1 and left to right in Figure 2, these groups are the pure functionals, hybrid functionals, range-separated hybrids and the double-hybrid functionals. For easy reference, the smallest AvgAD and MaxAD values in Table 1 associated with each functional class for each isomer are in bold font.

5.3.1 Pure Functionals

The first 4 rows in Table 1 give the AvgAD values for the pure functionals. These are followed by variants which include long-range corrections (LC),[419] Grimme's D3 dispersion corrections (D3)[417] or both. For these clusters, M06-L is clearly the most accurate pure functional considered here. The AvgAD values all fall between 18 and 22 cm^{-1}, whether or

not dispersion corrections are included. Notice that the dispersion correction has essentially no effect on the M06-L AvgAD values (≤ 1 cm^{-1}). The MaxAD values in Table 1 show that the M06-L and M06-L-D3 deviations reach a maximum of 110 cm^{-1} for the $(H_2O)_6$ Book isomer, which still represents the best performance of any of the pure functionals, as seen by the first group of bars in the right panel of Figure 2. The M06-L functional's accuracy for computing binding energies of water clusters has been noted previously.[380] In terms of the different types of vibrational modes, the M06-L functional has the lowest AvgAD values of the pure functionals in each case of intermolecular modes (11 cm^{-1}), bending modes (4 cm^{-1}) and OH stretching modes (46 cm^{-1}). An analysis of the AvgAD and MaxAD values for different types of modes (intermolecular, bending and OH stretching) can be found in the Supplementary Material.[418]

The M11-L functional is the second-most accurate of these pure functionals, with AvgAD values between 34 cm^{-1} and 37 cm^{-1}. The rest of the pure functionals in Table 1 exhibit larger deviations. For these other methods, the AvgAD values are never smaller than 43 cm^{-1} for clusters larger than the dimer. The largest deviations are seen for the PBE and BLYP functionals, where the AvgAD exceeds 70 cm^{-1} even for the smallest cluster. As seen in Table 1, the AvgAD for PBE harmonic frequencies reaches 100 cm^{-1} for the tetramer and larger isomers, while maximum deviations exceed 500 cm^{-1} for the Book and Prism isomers (Figure 2).

We note that there are three additional DFT methods tested with which we were unable to locate all of the minima. These include SOGGA11, B97D3 and LC-B97D3. In these cases, stationary points corresponding to some of the cyclic structures in Figure 1 were saddle points with 1 or more imaginary frequencies that remained even when more dense numerical integration grids were employed. The data for these functionals can be found in the Supplementary Material[418] even though they've been omitted from Table 1 and Figure 2.

5.3.2 Hybrid Functionals

The AvgAD and MaxAD values for the hybrid functionals investigated here are included in the second section of Table 1, as well as the second section of Figure 2. These include B3YLP, X3LYP, M06-2X and PBE0. In the case of the hybrid functionals, a Minnesota functional once again is the most accurate, with M06-2X yielding AvgAD values between 13 and 22 cm^{-1} for each isomer, regardless of the dispersion treatment. For the Book, Cage and Prism isomers, the M06-2X MaxADs (142 cm^{-1} – 162 cm^{-1}) are actually much larger than the corresponding M06-L values. As with the M06-L harmonic frequencies, the M06-2X AvgAD values are hardly affected by the dispersion treatment, as the largest difference in M06-2X and M06-2X-D3 values is less than 1 cm^{-1}. The M06-2X functional is particularly effective for computing the intermolecular frequencies, as the M06-2X and M06-2X-D3 AvgAD values for these modes (6 cm^{-1}) is the lowest of all methods tested here. A breakdown of the intermolecular, bending and stretching frequencies can be seen in Table S1 of the Supplementary Material.

The B3LYP and X3LYP functionals produce harmonic frequencies near the accuracy of M06-2X in $(H_2O)_2$, but the AvgAD values quickly increase with cluster size, reaching 36 cm^{-1} and 45 cm^{-1}, respectively, for the $(H_2O)_5$ global minimum. The AvgAD values for B3LYP and X3LYP are always within 2 cm^{-1} of each other, except in the case of $(H_2O)_5$. The PBE0 functional gives the best agreement with the reference values for $(H_2O)_2$ with an AvgAD of 18 cm^{-1} but becomes less accurate with cluster size. For each cluster larger than $(H_2O)_3$, the AvgAD exceeds 40 cm^{-1} for the PBE0 harmonic frequencies. For the B3LYP and PBE0 functionals, the D3 correction consistently increases the AvgAD values slightly for each isomer. The SOGGA11-X hybrid functional is an interesting case. This is the least accurate hybrid functional for the harmonic frequencies of $(H_2O)_2$ with an AvgAD just below 30 cm^{-1}. The SOGGA11-X AvgAD values decrease for the larger clusters, becoming the most accurate functional for $(H_2O)_5$ and the Ring and Book $(H_2O)_6$ isomers. The SOGGA11-X MaxAD values in Table 1 are remarkably consistent across the 8 isomers, always from 66

cm^{-1} to 68 cm^{-1}. For the Book, Cage and Prism isomers, the MaxAD values of all other hybrid functionals are more than double the corresponding SOGGA11-X values, as can be seen in Figure 2.

5.3.3 Range-separated Hybrids

The third block of Table 1 gives the results for the range-separated hybrid functionals. The first 3 range-separated hybrid functionals in Table 1 are the ωB97 functionals of Head-Gordon and coworkers. ωB97X differs from ωB97 by including a small amount of short-range Hartree-Fock exchange, and ωB97XD includes Grimme's D2 dispersion correction.[394] These three functionals demonstrate consistent accuracy for these small water clusters. The ωB97 and ωB97X AvgADs range from 13 cm^{-1} to 20 cm^{-1}. The inclusion of dispersion corrections typically increases the deviations for these water clusters. However, the ωB97XD method tends to yield more reliable binding energies[400,401,420] and geometries[421] for a more diverse set of hydrogen-bonded complexes and noncovalent interactions. In the more compact Prism and Cage $(H_2O)_6$ structures, the ωB97X AvgAD values of 13 cm^{-1} and 15 cm^{-1}, respectively, are the lowest of any of the DFT methods considered in this study, excluding double-hybrid methods. In addition, the AvgAD values of the OH stretching modes for these 3 DFT methods range from 30 cm^{-1} to 37 cm^{-1}, while that value is never smaller than 40 cm^{-1} for other DFT methods considered in this work, except for mPW2-PLYP (see Table S1 in the Supplementary Material). Table 1 shows that the smallest MaxAD values for every cluster larger than $(H_2O)_2$ are given by either ωB97 or ωB97X. On this metric, these functionals even outperform MP2 relative to the benchmark harmonic frequencies.

Also worth noting in this group of functionals is M11, for which the harmonic frequencies produce AvgAD values between 19 and 21 cm^{-1}, comparable to the accuracy of M06-L and M06-2X, while maximum deviations are below 150 cm^{-1} for the $(H_2O)_6$ isomers. The performance of the other range-separated hybrids is uneven. For example, CAM-B3LYP has the smallest AvgAD value for $(H_2O)_2$ apart from the double-hybrid methods. However,

the AvgAD increases more than twofold for the water tetramer. As indicated by Figure 2, the CAM-B3LYP MaxAD value across all 8 isomers is the second-largest among the range-separated hybrids.

5.3.4 Double-hybrid Functionals

The double-hybrid functionals B2PLYP and mPW2-PLYP compose the last class of functionals. The B2PLYP harmonic frequencies show consistent accuracy with no AvgAD values above 23 cm^{-1}, but this performance is no better than many of the more computationally efficient methods of the other functional categories. The B2PLYPD method demonstrates good accuracy for the water dimer. Of the other classes of functionals, only CAM-B3LYP has a smaller AvgAD for $(H_2O)_2$ than the B2PLYP value of 17 cm^{-1}. The accuracy of B2PLYPD decreases with cluster size as the AvgAD values exceed 30 cm^{-1} for each cluster larger than $(H_2O)_3$ except for the Prism. With the D3 correction, the harmonic frequencies are improved relative to B2PLYPD, and the AvgAD values do not exceed 26 cm^{-1}. However, this is only an improvement over B2PLYP for one isomer, the water tetramer mimimum. Both B2PLYPD and B2PLYP-D3 produce harmonic frequencies with MaxAD values larger than B2PLYP for every isomer.

The double-hybrid mPW2-PLYP functional produces harmonic frequencies in better agreement with the reference values by a few cm^{-1} for each isomer, never more than 20 cm^{-1}. These AvgAD values are competitive with the MP2/haTZ values listed at the bottom of Table 1. When considering all of the vibrational frequencies of all the isomers in this study, the mPW2-PLYP AvgAD value is actually smaller than that of MP2 (last section in Figure 2). The maximum deviations for mPW2-PLYP are smaller than those from MP2 computations for every isomer. The addition of a D2 dispersion correction in the mPW2-PLYPD method only increases the AvgAD and MaxAD values, but to a lesser extent than observed for B2PLYP. When looking at the different types of vibrational modes, the most accurate double-hybrid depends on the nature of the mode. For intermolecular frequencies, B2PLYP

is slightly more accurate than mPW2-PLYP with an AvgAD of 10 cm^{-1} across all isomers in this study. However, the bending and stretching vibrational frequencies are computed more accurately with the mPW2-PLYP functional (see Table S1 in the Supplementary Material).

5.4 Conclusions

37 DFT methods were employed to compute harmonic vibrational frequencies of $(H_2O)_n$ clusters ranging from 2 to 6 water molecules. A variety of functionals including pure functionals, hybrids, range-separated hybrids and double-hybrid functionals along with variants including empirical dispersion corrections were tested for their accuracy in comparison with benchmark harmonic vibrational frequencies expected to lie close to the CCSD(T) CBS limit.

Among the pure functionals, the M06-L functional is the most accurate, yielding harmonic vibrational frequencies within 20 cm^{-1} of the reference values on average. M06-2X is the most consistently accurate of the hybrid functionals investigated here, typically on par with the pure M06-L functional. Among the hybrids, SOGGA11-X demonstrates the worst agreement with the benchmark values for $(H_2O)_2$ but yields frequencies with maximum absolute deviations always from 66 cm^{-1} to 68 cm^{-1}, less than half of what is seen for the other hybrid functionals in the Book, Cage and Prism $(H_2O)_6$ isomers.

Two range-separated hybrid functionals perform the most consistently of all methods in this study, ωB97 and ωB97X, with AvgAD values never exceeding 20 cm^{-1} for any of the water cluster isomers. For $(H_2O)_3$ and larger clusters, the smallest MaxAD value of any method in this study is always that of ωB97 or ωB97X (< 60 cm^{-1}). In comparison, the MP2 harmonic frequencies exhibit MaxAD values exceeding 80 cm^{-1} for every cluster larger than $(H_2O)_4$. The ωB97 and ωB97X deviations are also much smaller than those for some flexible water potentials parameterized with *ab initio* data.[384] However, a paper appeared during the preparation of this manuscript that demonstrated the outstanding performance of the MB-pol potential, which deviates from the reference values by only 11 cm^{-1} on average

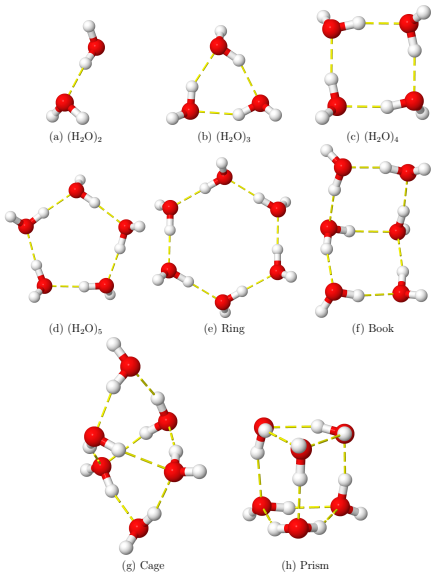


Figure 5.1. $(H_2O)_n$ minima ($n = 2 - 6$) examined in this study

(AvgAD) and never by more than 50 cm^{-1} (MaxAD).[422]

Of the two double-hybrid DFT methods investigated, B2PLYP and mPW2-PLYP, the mPW2-PLYP frequencies are in better agreement with the CCSD(T)-quality reference data, with accuracy comparable to that of MP2 computations. However, this accuracy is not significantly better than many of the more economical functionals examined here. The empirical dispersion corrections applied to the functionals in this study generally have a very small effect on AvgAD values and, in most cases, only slightly increase the deviations from the reference harmonic frequencies.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant numbers EPS-0903787 and CHE-1338056. The authors would also like to thank the Mississippi Center for Supercomputing Research for providing computational resources. The authors also thank Prof. Francesco Paesani and Greg Medders from the University of California, San Diego for helpful comments and suggestions.

Table 5.1. Average absolute deviations (in cm^{-1}) of DFT harmonic frequencies computed with the haTZ basis set from benchmark harmonic frequencies of $(H_2O)_n$ clusters ($n = 2 - 6$). Horizontal lines separate the methods from top to bottom into classes of pure functionals, hybrid functionals, range-separated hybrids, double-hybrid functionals and MP2 values. The smallest values within each class for each isomer appear in bold.

Functional	$(H_2O)_2$	$(H_2O)_3$	$(H_2O)_4$	$(H_2O)_5$	Ring	Book	Cage	Prism
M06-L	**18**	22	20	20	**18**	19	20	18
M11-L	36	37	35	34	36	33	34	34
N12	32	57	66	69	69	65	58	59
MN12-L	48	43	50	48	48	49	49	50
BLYP	75	80	83	84	81	79	74	80
PBE	71	96	105	106	103	102	95	94
B97D	53	67	79	81	76	75	67	60
M06-L-D3	**18**	**21**	**19**	**19**	19	**18**	**19**	**17**
LC-B97D	45	54	57	58	59	59	61	63
PBE-D3	72	96	107	109	104	106	99	90
BLYP-D3	76	81	85	84	80	83	81	74
B3LYP	22	30	34	36	35	33	28	26
X3LYP	21	29	34	45	36	34	29	26
M06-2X	19	**13**	**19**	21	22	20	**18**	**19**
PBE0	**18**	32	41	44	44	40	33	27
SOGGA11-X	29	24	20	**19**	**18**	**19**	19	20
B3LYP-D3	24	31	35	37	36	36	32	28
M06-2X-D3	19	**13**	**19**	20	22	20	**18**	**19**
PBE0-D3	20	33	42	45	44	43	36	30
ωB97	18	20	**17**	**16**	**16**	**16**	**15**	14
ωB97X	19	**19**	18	19	19	17	**15**	**13**
ωB97XD	22	20	25	25	24	24	21	17
CAM-B3LYP	**16**	29	35	37	38	35	31	27
LC-ωPBE	20	21	27	28	29	25	20	19
MN12-SX	35	33	38	36	35	36	37	38
M11	20	**19**	21	21	19	20	21	20
HSEH1PBE	20	37	43	45	44	43	37	31
HISSbPBE	45	39	34	35	36	34	32	32
N12-SX	35	29	40	43	44	38	32	29
B2PLYP	13	18	21	23	23	22	18	16
mPW2-PLYP	**10**	**14**	**18**	**20**	**20**	**18**	**14**	**11**
B2PLYPD	17	27	31	32	31	31	28	24
B2PLYP-D3	14	20	19	26	25	25	22	19
mPW2-PLYPD	12	20	25	26	26	25	21	17
MP2	10	13	17	18	18	18	16	14

Table 5.2. Maximum absolute deviations (in cm^{-1}) of DFT harmonic frequencies computed with the haTZ basis set from benchmark harmonic frequencies of $(H_2O)_n$ clusters ($n = 2 - 6$). Horizontal lines separate the methods from top to bottom into classes of pure functionals, hybrid functionals, range-separated hybrids, double-hybrid functionals and MP2 values. The smallest values within each class for each isomer appear in bold.

Functional	$(H_2O)_2$	$(H_2O)_3$	$(H_2O)_4$	$(H_2O)_5$	Ring	Book	Cage	Prism
M06-L	55	90	104	**105**	**104**	**110**	103	87
M11-L	74	76	126	134	141	146	152	172
N12	57	136	235	257	263	283	287	318
MN12-L	92	105	140	153	151	150	178	167
BLYP	224	268	327	347	351	360	340	401
PBE	218	318	428	462	469	501	487	516
B97D	177	214	320	343	340	359	301	312
M06-L-D3	**54**	**89**	**102**	**105**	**104**	**110**	**102**	**86**
LC-B97D	104	143	214	249	263	263	191	225
PBE-D3	218	320	436	474	477	522	516	521
BLYP-D3	221	273	342	355	355	390	406	395
B3LYP	72	95	138	155	162	160	145	143
X3LYP	67	93	138	176	165	164	151	148
M06-2X	36	**35**	64	91	102	142	162	158
PBE0	30	80	148	173	181	191	181	180
SOGGA11-X	68	65	67	**67**	**68**	**68**	**67**	**68**
B3LYP-D3	71	102	144	161	165	177	178	174
M06-2X-D3	36	**35**	**62**	90	103	141	160	155
PBE0-D3	**29**	83	151	178	184	202	200	201
ωB97	**28**	**34**	46	48	51	57	56	43
ωB97X	39	38	**40**	**43**	**48**	**53**	**50**	**40**
ωB97XD	47	54	50	49	54	73	68	53
CAM-B3LYP	41	81	135	156	166	169	165	161
LC-ωPBE	50	45	68	85	94	95	88	87
MN12-SX	74	94	117	119	116	118	122	128
M11	57	52	67	79	86	109	144	133
HSEH1PBE	42	98	159	181	188	199	193	192
HISSbPBE	125	124	127	127	127	127	126	125
N12-SX	82	80	81	107	118	120	113	109
B2PLYP	44	60	89	103	108	110	104	99
mPW2-PLYP	**19**	**34**	**64**	**78**	**84**	**84**	**79**	**72**
B2PLYPD	56	77	124	139	141	152	139	136
B2PLYP-D3	46	66	102	115	119	127	124	118
mPW2-PLYPD	27	62	88	103	108	114	104	99
MP2	29	45	72	83	87	97	105	98

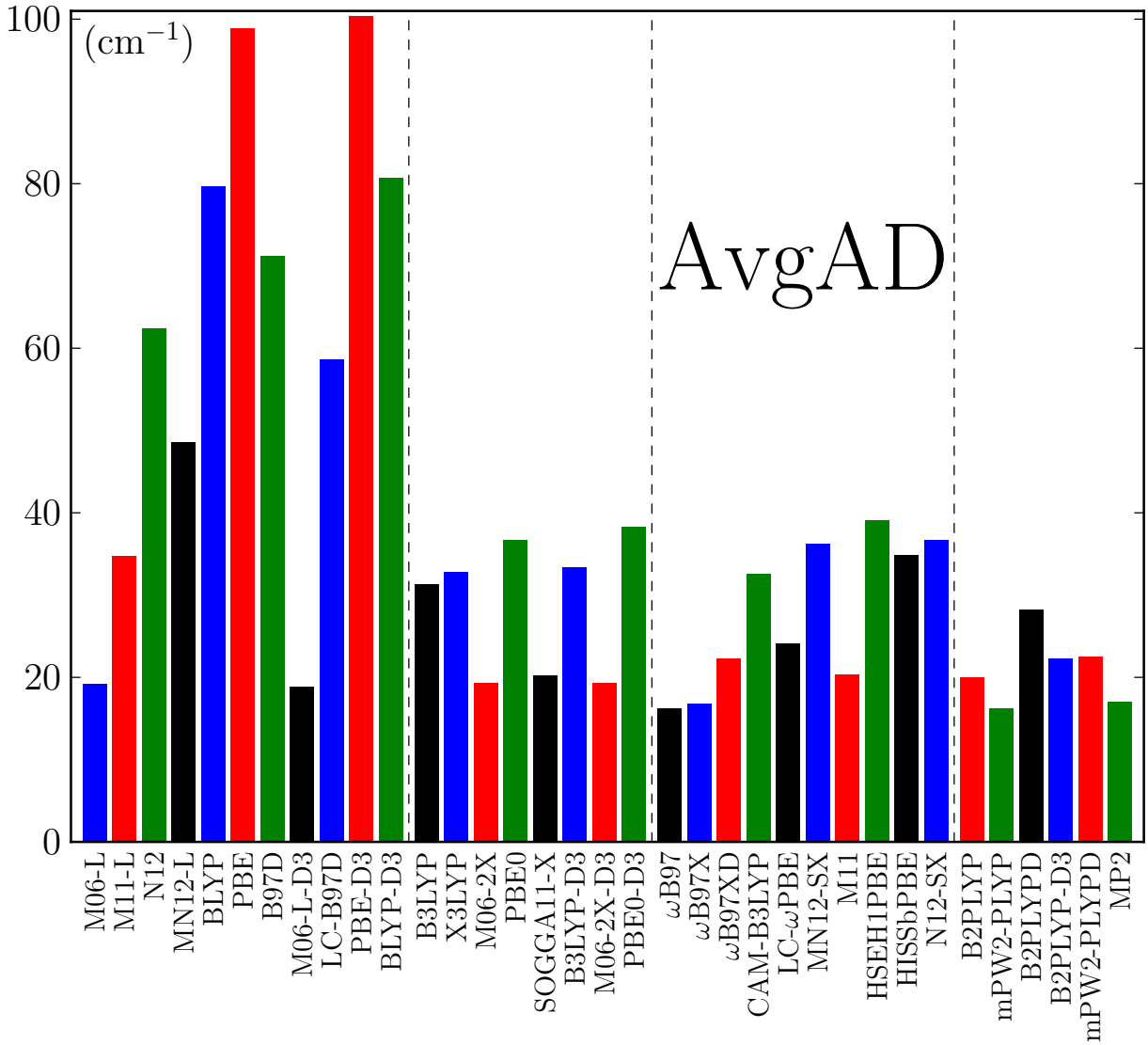


Figure 5.2. Average absolute deviations (in cm^{-1}) of DFT harmonic frequencies computed with the haTZ basis set from benchmark harmonic frequencies of all 8 $(H_2O)_n$ $(n = 2 - 6)$ isomers in this study with vertical lines separating pure functionals, hybrid functionals, range-separated hybrids and double-hybrid functionals

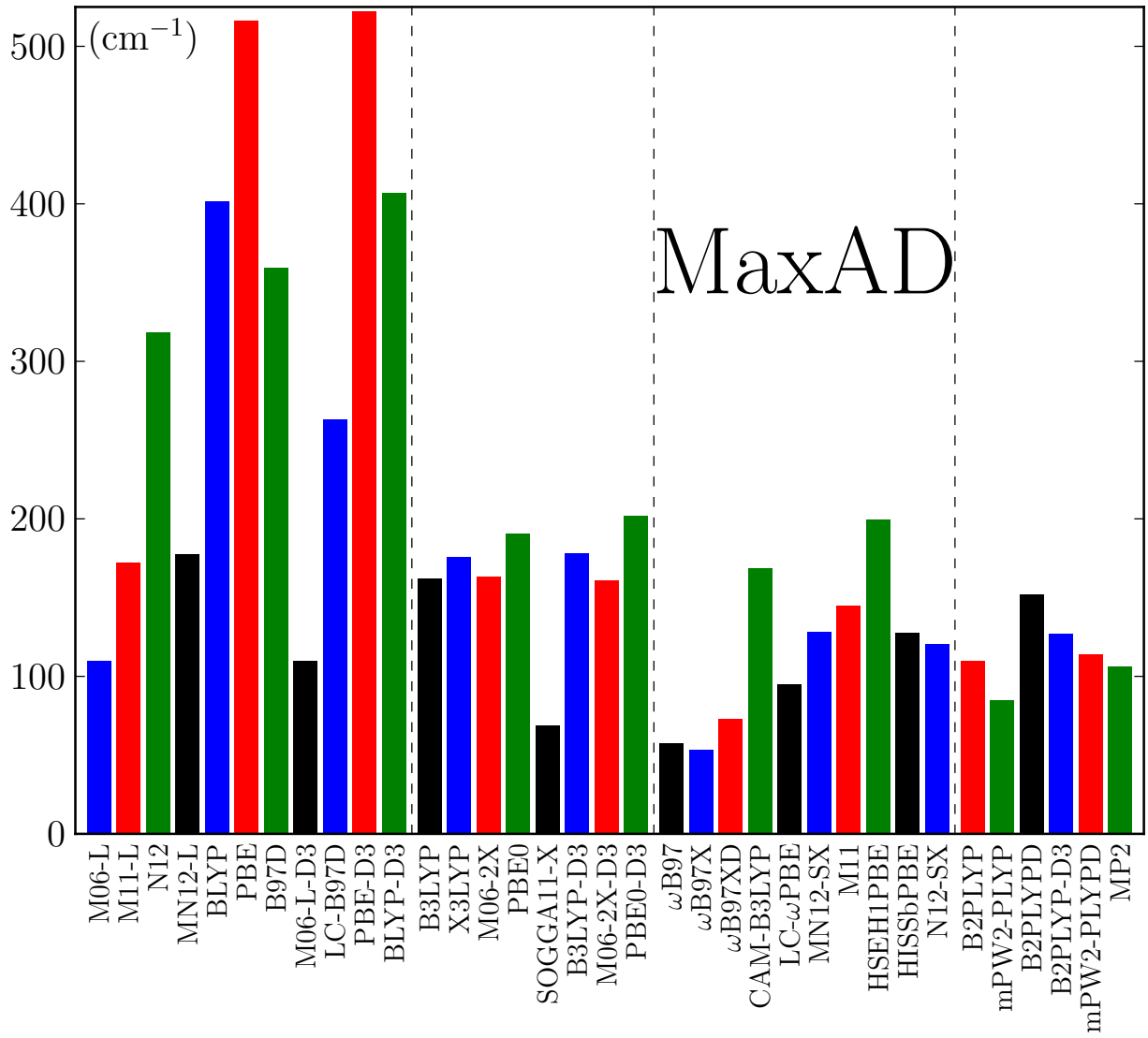


Figure 5.3. Maximum absolute deviations (in cm^{-1}) of DFT harmonic frequencies computed with the haTZ basis set from benchmark harmonic frequencies of all 8 $(H_2O)_n$ $(n = 2 - 6)$ isomers in this study with vertical lines separating pure functionals, hybrid functionals, range-separated hybrids and double-hybrid functionals

CHAPTER 6

CONCLUSIONS

The work presented here has focused on the computation of high-accuracy molecular properties with *ab initio* wavefunction methods, particularly vibrational frequencies and dissociation energies of small hydrogen-bonded clusters. It was demonstrated that CCSD(T)-quality geometries and harmonic frequencies could be computed efficiently through a many-body approach via the N-body:Many-body QM:QM technique. For $(HF)_n$ and $(H_2O)_n$ $(n = 3 - 6)$ clusters, 2-body:Many-body CCSD(T):MP2 optimized geometries were essentially identical to those from canonical CCSD(T). When applied to harmonic vibrational frequencies, 2-body:Many-body CCSD(T):MP2 values differed from CCSD(T) results by an average of only 0.8 cm^{-1} for all water clusters examined. In addition, the 2-body:Many-body frequencies were obtained on a single node over 100 times faster than CCSD(T) computations for $(H_2O)_6$ isomers. The independent nature of the many-body calculations could further increase this speedup to a thousandfold.

The research here has also used the $(H_2O)_2$ and $(HF)_2$ dimers to investigate the basis set convergence of harmonic vibrational frequencies with basis sets as large as aug-cc-pV6Z. This project served the purposes of both establishing benchmark values for these systems and also determining the most economical basis sets for computing harmonic frequency values that lie near the complete basis set (CBS) limit. It was determined that a basis set of at least quadruple-ζ quality was needed to yield values within *ca.* 10 cm^{-1} of the CBS limit. When the haQZ basis set was employed in conjunction with CCSD(T) and VPT2 to compute anharmonic vibrational frequencies, it was demonstrated that this approach could predict fundamental frequencies to within a few cm^{-1} of experimental values for both $(HF)_2$

and $(H_2O)_2$. In addition, dissociation energies calculated within this approach also were in excellent agreement with experiment, as the errors relative to experiment were only 21 cm^{-1} and 2 cm^{-1} for $(H_2O)_2$ and $(HF)_2$, respectively.

Having established the reliability of the 2-body:Many-body CCSD(T):MP2 approach as well as the basis set convergence trends in harmonic frequency computations, the conclusions from the previous studies were combined to establish an effective strategy for computing benchmark harmonic frequency values for a series water clusters up to the size of the water hexamer. 2-body:Many-body CCSD(T):MP2/haQZ values were computed for 8 water cluster isomers and used to calibrate the accuracy of other *ab initio* methods (LCCSD(T) and MP2), *ab initio*-based water potentials and, lastly, a variety of DFT methods. When comparing other *ab initio* methods to the benchmark values, an interesting result was that CCSD(T) computations with small basis sets were actually in better agreement with the benchmark data when compared to MP2/haQZ computations. LCCSD(T) harmonic frequencies showed much larger deviations, particularly for bound OH stretches and intermonomer modes. Neither of these methods yielded satisfactory red shifts for the bound stretches as MP2 and LCCSD(T) overestimated and underestimated their magnitudes, respectively. For the two water potentials tested (TTM3-F and WHBB), the WHBB values were in much better agreement with the reference data. However, the performance of that potential deteriorated with increasing cluster size, likely reflecting an inadequate description of many-body effects. For the DFT methods investigated, the ωB97 and ωB9X functionals were the most consistently accurate across all water cluster isomers, comparable to the accuracy of MP2. Also encouraging was the performance of the M06-L functional, which, as a pure functional, offers a very attractive accuracy-to-cost ratio.

What is perhaps most important in the work presented here is that it establishes a clear strategy for efficiently computing reliable properties of clusters that will be used to predict the results of experiments that have not yet been performed and also to aid in the interpretation of experimental data where robust theory is needed. One example is in the vi-

brational spectra of water clusters, where assignment of spectral features can be challenging. With the N-body:Many-body QM:QM approach to compute harmonic vibrational frequencies and an understanding of what is required to compute accurate anharmonic vibrational frequencies with vibrational perturbation theory, all of the tools needed are present to extend *ab initio* wavefunction methods to larger chemical systems, with a level of accuracy previously unattainable.

BIBLIOGRAPHY

[1] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, Chichester, 2000)

[2] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (McGraw-Hill, New York, 1989)

[3] G. A. Jeffrey and W. Saenger, *Hydrogen Bonding in Biological Structures* (Springer-Verlag, Berlin, 1994)

[4] G. R. Desiraju, Angew. Chem. Int. Ed. **50**, 52 (2011)

[5] G. C. Pimentel and A. L. McClellan, *The Hydrogen Bond* (W.H. Freeman, San Francisco, 1960)

[6] P. A. Kollman and L. C. Allen, Chem. Rev. **72**, 283 (1972)

[7] B. Jeziorski, R. Moszynski, and K. Szalewicz, Chem. Rev. **94**, 1887 (1994)

[8] B. Jeziorski and K. Szalewicz, in *Encyclopedia of Computational Chemistry*, edited by P. v. R. Schleyer (John Wiley and Sons, Chichester, England, 1998) pp. 1376–1398

[9] B. Jeziorski and K. Szalewicz, in *Handbook of Molecular Physics and Quantum Chemistry*, Vol. 3, edited by S. Wilson (John Wiley and Sons, Chichester, England, 2003) pp. 232–279

[10] K. Szalewicz, K. Patkowski, and B. Jeziorski, in *Intermolecular Forces and Clusters II*, Structure and Bonding, Vol. 116, edited by D. J. Wales (Springer, Germany, 2005) pp. 44–117

[11] K. Szalewicz, WIREs Comput. Mol. Sci. **2**, 254 (2012)

[12] R. Moszynski, P. E. S. Wormer, B. Jeziorski, and A. van der Avoird, J. Chem. Phys. **103**, 8058 (1995)

[13] R. Moszynski, P. E. S. Wormer, T. G. A. Heijmen, and A. van der Avoird, J. Chem. Phys. **108**, 579 (1998)

[14] V. F. Lotrich and K. Szalewicz, J. Chem. Phys. **106**, 9668 (1997)

[15] R. Moszynski, P. E. S. Wormer, B. Jeziorski, and A. van der Avoird, J. Chem. Phys. **107**, 672 (1997)

[16] R. Moszynski, P. E. S. Wormer, B. Jeziorski, and A. van der Avoird, J. Chem. Phys. **112**, 3159 (2000)

[17] R. Podeszwa and K. Szalewicz, J. Chem. Phys. **126**, 194101 (2007)

[18] S. Xantheas, Chem. Phys. **258**, 225 (2000)

[19] J. A. Pople, in *Energy, Structure, and Reactivity*, edited by D. W. Smith and W. B. McRae (Wiley, New York, 1973) pp. 51–61

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

[21] G. S. Tschumper, in *Reviews in Computational Chemistry*, Vol. 26, edited by K. B. Lipkowitz and T. R. Cundari (Wiley-VCH, Inc., Hoboken, NJ, 2009) pp. 39–90

[22] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (John Wiley and Sons, Inc., New York, 1954)

[23] A. Lenz and L. Ojamäe, J. Chem. Phys. **131**, 134302 (2009)

[24] R. S. Grev, C. L. Janssen, and H. F. Schaefer, J. Chem. Phys. **95**, 5128 (1991)

[25] A. P. Scott and L. Radom, J. Phys. Chem. **100**, 16502 (1996)

[26] M. E. Dunn, T. M. Evans, K. N. Kirschner, and G. C. Shields, J. Phys. Chem. A **110**, 303 (2006)

[27] B. Temelso and G. C. Shields, J. Chem. Theory Comput. **7**, 2804 (2011)

[28] B. Temelso, K. Archer, and G. Shields, J. Phys. Chem. A **115**, 12034 (2011)

[29] R. Shields, B. Temelso, K. Archer, T. Morrell, and G. Shields, J. Phys. Chem. A **114**, 11725 (2010)

[30] R. Bartlett, WIREs Comput. Mol. Sci. **2**, 126 (2012)

[31] D. Cremer, WIREs Comput. Mol. Sci. **1**, 509 (2011)

[32] J. Paldus, J. Čížek, and B. Jeziorski, J. Chem. Phys. **90**, 4356 (1989)

[33] R. J. Bartlett, J. Phys. Chem. **93**, 1697 (1989)

[34] J. Paldus, Mol. Phys. **108**, 2941 (2010)

[35] P. Jurečka, J. Šponer, J. Černý, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006)

[36] P. Hobza, H. Selzle, and E. Schlag, J. Phys. Chem. **100**, 18790 (1996)

[37] J. Šponer and P. Hobza, Chem. Phys. Lett. **267**, 263 (1997)

[38] P. Hobza and J. Šponer, Chem. Rev. **99**, 3247 (1999)

[39] M. Sinnokrot, E. Valeev, and C. Sherrill, J. Am. Chem. Soc. **124**, 10887 (2002)

[40] B. W. Hopkins and G. S. Tschumper, J. Phys. Chem. A **108**, 2941 (2004)

[41] E. J. Carrell, C. M. Thorne, and G. S. Tschumper, J. Chem. Phys. **136**, 014103 (2012)

[42] K. E. Riley and P. Hobza, WIREs Comput. Mol. Sci. **1**, 3 (2011)

[43] E. G. Hohenstein and C. D. Sherrill, WIREs Comput. Mol. Sci. **2**, 304 (2012)

[44] C. D. Sherrill, in *Reviews in Computational Chemistry*, Vol. 26, edited by K. B. Lipkowitz and T. R. Cundari (Wiley-VCH, Inc., Hoboken, NJ, 2009) pp. 1–38

[45] K. E. Riley, M. Pitonák, P. Jurecka, and P. Hobza, Chem. Rev. **110**, 5023 (2010)

[46] P. Hobza, Acc. Chem. Res. **45**, 663 (2012)

[47] T. Janowski and P. Pulay, Chem. Phys. Lett. **447**, 27 (2007)

[48] J. Olsen, O. Christiansen, H. Koch, and P. Jørgensen, J. Chem. Phys. **105**, 5082 (1996)

[49] A. Halkier, H. Larsen, J. Olsen, and P. Jørgensen, J. Chem. Phys. **110**, 7127 (1999)

[50] M. L. Leininger, W. D. Allen, H. F. Schaefer, and C. D. Sherrill, J. Chem. Phys. **112**, 9213 (2000)

[51] I. Gurtubay and R. Needs, J. Chem. Phys. **127**, 124306 (2007)

[52] B. Santra, A. Michaelides, M. Fuchs, A. Tkatchenko, and C. Filippi, J. Chem. Phys. **129**, 194111 (2008)

[53] B. Santra, J. Klimeš, D. Alfè, A. Tkatchenko, B. Slater, A. Michaelides, R. Car, and M. Scheffler, Phys. Rev. Lett. **107**, 185701 (2011)

[54] Y. Wang, V. Babin, J. Bowman, and F. Paesani, J. Am. Chem. Soc. **134**, 11116 (2012)

[55] M. Gillan, F. Manby, M. Towler, and D. Alfè, J. Chem. Phys. **136**, 244105 (2012)

[56] F. Jensen, WIREs Comput. Mol. Sci., Early View(2012)

[57] T. Dunning Jr., J. Chem. Phys. **90**, 1007 (1989)

[58] R. Kendall, T. Dunning Jr., and R. Harrison, J. Chem. Phys. **96**, 6796 (1992)

[59] D. Woon and T.H. Dunning Jr., J. Chem. Phys. **98**, 1358 (1993)

[60] D. Woon and T.H. Dunning, J. Chem. Phys. **100**, 2975 (1994)

[61] K. Peterson, D. Woon, and T. Dunning Jr., J. Chem. Phys. **100**, 7410 (1994)

[62] D. Woon and T. Dunning Jr., J. Chem. Phys. **103**, 4572 (1995)

[63] A. Wilson, T. van Mourik, and T. Dunning Jr., J. Mol. Struct. (THEOCHEM) **388**, 339 (1996)

[64] K. Peterson and T. Dunning, J. Chem. Phys. **117**, 10548 (2002)

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

[66] K. Müller-Dethlefs and P. Hobza, Chem. Rev. **100**, 143 (2000)

[67] P. Hobza, R. Zahradník, and M. Müller-Dethlefs, Collection of Czechoslovak Chemical Communications **71**, 443 (2006)

[68] J. Řezáč, K. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 2427 (2011)

[69] E. Papajak and D. G. Truhlar, J. Chem. Theory Comput. **7**, 10 (2011)

[70] D. Feller, J. Chem. Phys. **98**, 7059 (1993)

[71] S. Xantheas and T. Dunning, Jr., J. Phys. Chem. **97**, 18 (1993)

[72] J. Martin, Chem. Phys. Lett. **259**, 669 (1996)

[73] A. Halkier, H. Koch, P. Jorgensen, O. Christiansen, I. Nielsen, and T. Helgaker, Theor Chem Acc **97**, 150 (1997)

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

[75] J. Martin and P. Taylor, J. Chem. Phys. **106**, 8620 (1997)

[76] W. Klopper and H. Lüthi, Mol. Phys. **96**, 559 (1999)

[77] A. Tajti, P. Szalay, A. Császár, M. Kállay, J. Gauss, E. Valeev, B. Flowers, J. Vázquez, and J. Stanton, J. Chem. Phys. **121**, 11599 (2004)

[78] D. Feller, K. Peterson, and J. Hill, J. Chem. Phys. **135**, 044102 (2011)

[79] C. Schwartz, Phys. Rev. **126**, 1015 (1962)

[80] D. Carroll, H. Silverstone, and R. Metzger, J. Chem. Phys. **71**, 4142 (1979)

[81] D. Truhlar, Chem. Phys. Lett. **294**, 45 (1998)

[82] D. W. Schwenke, J. Chem. Phys. **122**, 014107 (2005)

[83] S. Xantheas, J. Chem. Phys. **104**, 8821 (1996)

[84] S. Xantheas and C. Burnham, J. Chem. Phys. **116**, 1493 (2002)

[85] S. Xantheas and E. Aprà, J. Chem. Phys. **120**, 823 (2004)

[86] N. R. Kestner, J. Chem. Phys. **48**, 252 (1968)

[87] H. B. Jansen and P. Ros, Chem. Phys. Lett. **3**, 140 (1969)

[88] S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970)

[89] B. Liu and A. D. McLean, J. Chem. Phys. **59**, 4557 (1973)

[90] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, Chem. Rev. **94**, 1873 (1994)

[91] N. R. Kestner and J. E. Combariza, in *Reviews in Computational Chemistry*, Vol. 13, edited by K. B. Lipkowitz and D. B. Boyd (Wiley-VCH Publishers, New York, 1999) Chap. 2, pp. 99–132

[92] A. Halkier, W. Klopper, T. Helgaker, P. Jorgensen, and P. Taylor, J. Chem. Phys. **111**, 9157 (1999)

[93] G. S. Tschumper, M. L. Leininger, B. C. Hoffman, E. F. Valeev, H. F. Schaefer, and M. Quack, J. Chem. Phys. **116**, 690 (2002)

[94] J. R. Lane and H. G. Kjaergaard, J. Chem. Phys. **131**, 034307 (2009)

[95] E. Valeev and J. Zhang, J. Chem. Theory Comput. **8**, 3175 (2012)

[96] E. Hyleraas, Z. Phys. **54**, 347 (1929)

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

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

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

[100] C. Hättig, W. Klopper, A. Köhn, and D. Tew, Chem. Rev. **112**, 4 (2012)

[101] E. Valeev and T. Crawford, J. Chem. Phys. **128**, 244113 (2008)

[102] M. Marshall, L. Burns, and C. Sherrill, J. Chem. Phys. **135**, 194102 (2011)

[103] M. S. Marshall and C. D. Sherrill, J. Chem. Theory Comput. **7**, 3978 (2011)

[104] T. Adler, G. Knizia, and H.-J. Werner, J. Chem. Phys. **127**, 221106 (2007)

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

[106] H.-J. Werner, P. J. Knowles, F. R. Manby, M. Schütz, P. Celani, G. Knizia, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and A. Wolf, "Molpro, version 2010.1, a package of ab initio programs," (2010), see http://www.molpro.net

[107] O. Marchetti and H.-J. Werner, J. Phys. Chem. A **130**, 11580 (2009)

[108] K. Morokuma and L. Pedersen, J. Chem. Phys. **48**, 3275 (1968)

[109] P. A. Kollman and L. C. Allen, J. Chem. Phys. **51**, 3286 (1969)

[110] G. H. F. Diercksen, Chem. Phys. Lett. **4**, 373 (1969)

[111] K. Morokuma and J. R. Winick, J. Chem. Phys. **52**, 1301 (1970)

[112] D. Hankins, J. W. Moskowitz, and F. H. Stillinger, J. Chem. Phys. **53**, 4544 (1970)

[113] G. H. F. Diercksen, Theor. Chim. Acta **21**, 335 (1971)

[114] J. Del Bene, J. Chem. Phys. **55**, 4633 (1971)

[115] G. Fanourgakis, E. Aprà, and S. Xantheas, J. Chem. Phys. **121**, 2655 (2004)

[116] E. Aprà, A. Rendell, R. Harrison, V. Tipparaju, W. deJong, and S. Xantheas, in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis* (ACM, New York, 2009) p. 66

[117] A. East and W. Allen, J. Chem. Phys. **99**, 4638 (1993)

[118] A. Császár, W. Alen, and H. Schaefer, J. Chem. Phys. **108**, 9751 (1998)

[119] K. Diri, E. Myshakin, and K. Jordan, J. Phys. Chem. A **109**, 4005 (2005)

[120] M. Dunn, T. Evans, K. Kirschner, and G. Shields, J. Phys. Chem. A **110**, 303 (2006)

[121] Y. Wang and J. Bowman, J. Chem. Phys. **134**, 154510 (2011)

[122] Y. Wang and J. Bowman, J. Phys. Chem. Lett. **4**, 1104 (2013)

[123] C. Pérez, M. Muckle, D. Zaleski, N. Seifert, B. Temelso, G. Shields, Z. Kisiel, and B. Pate, Science **336**, 897 (2012)

[124] C. Pérez, S. Lobsiger, N. Seifert, D. Zaleski, B. Temelso, and G. Shields, Chem. Phys. Lett. **571**, 1 (2013)

[125] H. Popkie, H. Kistenmacher, and E. Clementi, J. Chem. Phys. **59**, 1325 (1973)

[126] O. Matsuoka, E. Clementi, and M. Yoshimine, J. Chem. Phys. **64**, 1351 (1976)

[127] J. Odutola and T. Dyke, J. Chem. Phys. **72**, 5062 (1980)

[128] B. Smith, D. Swanton, J. Pople, and H. Schaefer, J. Chem. Phys. **92**, 1240 (1990)

[129] S. Xantheas, J. Chem. Phys. **100**, 7523 (1994)

[130] P. Hobza, O. Bludský, and S. Suhai, Phys. Chem. Chem. Phys. **1**, 3073 (1999)

[131] S. Xantheas and T. Dunning, Jr., J. Chem. Phys. **99**, 8774 (1993)

[132] S. Simon and M. Duran, J. Chem. Phys. **105**, 11024 (1996)

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

[134] S. K. Min, E. C. Lee, H. M. Lee, D. Y. Kim, D. Kim, and K. S. Kim, J. Comput. Chem. **29**, 1208 (2008)

[135] W. Klopper, J. van Duijneveldt-van de Rijdt, and F. van Duijneveldt, Phys. Chem. Chem. Phys. **2**, 2227 (2000)

[136] W. Klopper, M. Schütz, H. Lüthi, and S. Leutwyler, J. Chem. Phys. **103**, 1085 (1995)

[137] B. Wladkowski, W. Allen, and J. Brauman, J. Phys. Chem. **98**, 13532 (1994)

[138] S. Klippenstein, A. East, and W. Allen, J. Chem. Phys. **118**, 118 (1996)

[139] N. Allinger, J. Fermann, W. Allen, and H. Schaefer, J. Chem. Phys. **106**, 5143 (1997)

[140] G. Tarczay, A. Császár, W. Klopper, V. Szalay, and W. Allen, J. Chem. Phys. **110**, 11971 (1999)

[141] J. Lane, J. Chem. Theory Comput. **9**, 316 (2013)

[142] J. Del Bene and J. Pople, J. Chem. Phys. **58**, 3605 (1973)

[143] H. Kistenmacher, G. Lie, H. Popkie, and E. Clementi, J. Chem. Phys. **61**, 546 (1974)

[144] E. Clementi, W. Kołos, G. Lie, and G. Ranghino, Int. J. Quantum Chem. **17**, 377 (1980)

[145] S. Tomoda and K. Kimura, Chem. Phys. Lett. **102**, 560 (1983)

[146] E. Honegger and S. Leutwyler, J. Chem. Phys. **88**, 2582 (1988)

[147] G. Chałasiński, M. Szczśniak, P. Cieplak, and S. Scheiner, J. Chem. Phys. **94**, 2873 (1991)

[148] J. Owicki, L. Shipman, and H. Scherage, J. Phys. Chem. **79**, 1794 (1975)

[149] M. Schütz, T. Bürgi, S. Leutwyler, and H. Bürgi, J. Chem. Phys. **99**, 5228 (1993)

[150] O. Mó, M. Yáñez, and J. Elguero, J. Chem. Phys. **97**, 6628 (1992)

[151] S. Xantheas and T. Dunning, J. Chem. Phys. **98**, 8037 (1993)

[152] I. Nielsen, E. Seidl, and C. Janssen, J. Chem. Phys. **110**, 9435 (1999)

[153] J. C. Howard and G. S. Tschumper, WIREs Comput. Mol. Sci. **4**, 199 (2014)

[154] J. A. Anderson, K. Crager, L. Fedoroff, and G. S. Tschumper, J. Chem. Phys. **121**, 11023 (2004)

[155] N. Pugliano and R. Saykally, Science **257**, 1937 (1992)

[156] K. Liu, J. Loeser, M. Elrod, B. Host, J. Rzepiela, N. Pugliano, and R. Saykally, J. Am. Chem. Soc. **116**, 3507 (1994)

[157] D. Wales, J. Am. Chem. Soc. **115**, 11180 (1993)

[158] J. Fowler and H. Schaefer, J. Am. Chem. Soc. **117**, 446 (1995)

[159] D. Wales and T. Walsh, J. Chem. Phys. **105**, 6957 (1996)

[160] T. Taketsugu and D. Wales, Mol. Phys. **100**, 2793 (2002)

[161] M. Schütz, W. Klopper, and H.-P. Lüthi, J. Chem. Phys. **103**, 6114 (1995)

[162] D. Wales and T. Walsh, J. Chem. Phys. **106**, 7193 (1997)

[163] J. Pérez, C. Hadad, and A. Restrepo, Int. J. Quantum Chem. **108**, 1653 (2008)

[164] M. Day, K. Kirschner, and G. Shields, J. Phys. Chem. A **109**, 6773 (2005)

[165] F. Ramírez, C. Hadad, D. Guerra, J. David, and A. Restrepo, Chem. Phys. Lett. **507**, 229 (2011)

[166] D. M. Bates and G. S. Tschumper, J. Phys. Chem. A **113**, 3555 (2009)

[167] K. Nauta and R. Miller, Science **287**, 293 (2000)

[168] K. Kim, M. Dupuis, G. Lie, and E. Clementi, Chem. Phys. Lett. **131**, 451 (1986)

[169] B. Mhin, H. Kim, H. Kim, C. Yoon, and K. Kim, Chem. Phys. Lett. **176**, 41 (1991)

[170] K. Franken, M. Jalaie, and C. Dykstra, Chem. Phys. Lett. **198**, 59 (1992)

[171] A. Vegiri and S. Farantos, J. Chem. Phys. **98**, 4059 (1993)

[172] B. Mhin, J. Kim, S. Lee, J. Lee, and K. Kim, J. Chem. Phys. **100**, 4484 (1994)

[173] K. Kim, K. Joradan, and T. Zwier, J. Am. Chem. Soc. **116**, 11568 (1994)

[174] C. Tsai and K. Jordan, Chem. Phys. Lett. **213**, 181 (1993)

[175] K. Liu, M. Brown, C. Carter, R. Saykally, J. Gregory, and D. Clary, Nature **381**, 501 (1996)

[176] J. Kim and K. Kim, J. Chem. Phys. **109**, 5886 (1998)

[177] R. Olson, J. Bentz, R. Kendall, M. Schmidt, and M. Gordon, J. Chem. Theory Comput. **3**, 1312 (2007)

[178] P. Day, R. Pachter, M. Gordon, and G. Merrill, J. Chem. Phys. **112**, 2063 (2000)

[179] E. Dahlke, R. Olson, H. Leverentz, and D. Truhlar, J. Phys. Chem. A **112**, 3976 (2008)

[180] C. Tsai and K. Jordan, J. Chem. Phys. **99**, 6957 (1993)

[181] K. Kirschner and G. Shields, Int. J. Quantum Chem. **52**, 349 (1994)

[182] D. Wales and H. Scheraga, Science **285**, 1368 (1999)

[183] J. Kim, D. Majumdar, H. Lee, and K. Kim, J. Chem. Phys. **110**, 9128 (1999)

[184] J. Kazimirski and V. Buch, J. Phys. Chem. A **107**, 9762 (2003)

[185] E. Silva, H. Duarte, and J. Belchior, Chem. Phys. **323**, 553 (2006)

[186] M. Kirov, G. Fanourgakis, and S. Xantheas, Chem. Phys. Lett. **461**, 180 (2008)

[187] H. Takeuchi, J. Chem. Inf. Model. **48**, 2226 (2008)

[188] B. Hartke, WIREs Comput Mol Sci **1**, 879 (2011)

[189] L. Darré, M. R. Machado, and S. Pantano, WIREs Comput Mol Sci **2**, 921 (2012)

[190] H. Lee, S. Suh, J. Lee, P. Tarakeshwar, and K. Kim, J. Chem. Phys. **112**, 9759 (2000)

[191] S. Bulusu, S. Yoo, E. Aprà, S. Xantheas, and X. Zeng, J. Phys. Chem. A **110**, 11781 (2006)

[192] G. Fanourgakis, E. Aprà, W. de Jong, and S. Xantheas, J. Chem. Phys. **122**, 134304 (2005)

[193] S. Yoo and S. Xantheas, in *Handbook of Computational Chemistry*, Vol. 2, edited by J. Leszczynski (Springer, New York, 2012) pp. 761–792

[194] B. Hartke, Phys. Chem. Chem. Phys. **5**, 275 (2003)

[195] A. Lagutchenkov, G. Fanourgakis, and S. Xantheas, J. Chem. Phys. **122**, 194310 (2005)

[196] S. Yoo, E. Aprá, X. C. Zeng, and S. S. Xantheas, J. Phys. Chem. Lett. **1**, 3122 (2010)

[197] See www.top500.org

[198] H. Fujimoto, N. Koga, and K. Fukui, J. Am. Chem. Soc. **103**, 7452 (1981)

[199] D. Fedorov, T. Ishida, and K. Kitaura, J. Phys. Chem. A **109**, 2638 (2005)

[200] D. Fedorov and K. Kitaura, J. Phys. Chem. A **111**, 6904 (2007)

[201] D. Fedorov, T. Ishida, M. Uebayasi, and K. Kitaura, J. Phys. Chem. A **111**, 2722 (2007)

[202] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999)

[203] D. Fedorov, L. Slipchenko, and K. Kitaura, J. Phys. Chem. A **114**, 8742 (2010)

[204] S. Pruitt, D. Fedorov, K. Kitaura, and M. Gordon, J. Chem. Theory Comput. **6**, 1 (2010)

[205] S. Gadre, R. Shirsat, and A. Limaye, J. Phys. Chem. **98**, 9165 (1994)

[206] S. Gadre and V. Ganesh, J. Chem. Theory Comput. **5**, 835 (2006)

[207] V. Ganesh, R. Dongare, P. Balanarayan, and S. Gadre, J. Chem. Phys. **2006**, 104109 (2006)

[208] A. Rahalkar, V. Ganesh, and S. Gadre, J. Chem. Phys. **129**, 234101 (2008)

[209] A. Rahalkar, M. Katouda, S. Gadre, and S. Nagase, J. Chem. Phys. **129**, 234101 (2008)

[210] S. Yeole and S. Gadre, J. Chem. Phys. **132**, 094102 (2010)

[211] M. Gordon, D. Fedorov, S. Pruitt, and L. Slipchenko, Chem. Rev. **112**, 632 (2012)

[212] S. Humbel, S. Sieber, and K. Morokuma, J. Chem. Phys. **105**, 1959 (1996)

[213] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, J. Phys. Chem. **100**, 19357 (1996)

[214] P. Karadakov and K. Morokuma, Chem. Phys. Lett. **317**, 589 (2000)

[215] T. Vreven and K. Morokuma, J. Comput. Chem. **21**, 1419 (2000)

[216] T. Vreven, B. Mennucci, C. da Silva, K. Morokuma, and J. Tomasi, J. Chem. Phys. **115**, 62 (2001)

[217] W. Klopper, M. Quack, and M. Suhm, Mol. Phys. **94**, 105 (1998)

[218] W. Guo, A. Wu, and X. Xu, Chem. Phys. Lett. **498**, 203 (2010)

[219] N. Mayhall and K. Raghavachari, J. Chem. Theory Comput. **7**, 1336 (2011)

[220] N. Mayhall and K. Raghavachari, J. Chem. Theory Comput. **8**, 2669 (2012)

[221] G. Beran, J. Chem. Phys. **130**, 164115 (2009)

[222] G. Beran and K. Nanda, J. Phys. Chem. Lett. **1**, 3480 (2010)

[223] J. Řezáč and D. Salahub, J. Chem. Theory Comput. **6**, 91 (2010)

[224] U. Góra, R. Podeszwa, W. Cencek, and K. Szalewicz, J. Chem. Phys. **135**, 224102 (2011)

[225] B. W. Hopkins and G. S. Tschumper, J. Comput. Chem. **24**, 1563 (2003)

[226] B. W. Hopkins and G. S. Tschumper, Mol. Phys. **103**, 309 (2005)

[227] B. W. Hopkins and G. S. Tschumper, Chem. Phys. Lett. **407**, 362 (2005)

[228] G. S. Tschumper, Chem. Phys. Lett. **427**, 185 (2006)

[229] A. M. ElSohly, C. L. Shaw, M. E. Guice, B. D. Smith, and G. S. Tschumper, Mol. Phys. **105**, 2777 (2007)

[230] D. M. Bates, J. R. Smith, T. Janowski, and G. S. Tschumper, J. Chem. Phys. **135**, 044123 (2011)

[231] D. M. Bates, J. R. Smith, and G. S. Tschumper, J. Chem. Theory Comput. **7**, 2753 (2011)

[232] R. Richard and J. Herbert, J. Chem. Phys. **137**, 064113 (2012)

[233] R. Bukowski, K. Szalewicz, G. Groenenboom, and A. van der Avoird, Science **315**, 1249 (2007)

[234] X. Huang, B. Braams, J. Bowman, R. Kelly, J. Tennyson, G. Groenenboom, and A. van der Avoird, J. Chem. Phys. **128**, 034312 (2008)

[235] A. Shank, Y. Wang, A. Kaledin, B. Braams, and J. Bowman, J. Chem. Phys. **130**, 144314 (2009)

[236] Y. Wang, B. Shepler, B. Braams, and J. Bowman, J. Chem. Phys. **131**, 054511 (2009)

[237] Y. Wang, X. Huang, B. Shepler, B. Braams, and J. Bowman, J. Chem. Phys. **134**, 094509 (2011)

[238] V. Babin, G. Medders, and F. Paesani, Chem. Phys. Lett. **3**, 3765 (2012)

[239] C. Leforestier, K. Szalewicz, and A. van der Avoird, J. Chem. Phys. **137**, 14305 (2012)

[240] G. Medders, V. Babin, and F. Paesani, J. Chem. Theory Comput. **9**, 1103 (2013)

[241] L. de la Peña and P. Kusalik, J. Chem. Phys. **121**, 5992 (2004)

[242] H. Stoll, Phys. Rev. B **46**, 6700 (1992)

[243] H. Stoll, B. Paulus, and P. Fulde, J. Chem. Phys. **123**, 144108 (2005)

[244] O. Sode and S. Hirata, J. Chem. Phys. **137**, 174104 (2012)

[245] S. Yeole, N. Sahu, and S. Gadre, J. Phys. Chem. A **117**, 8591 (2013)

[246] G. J. O. Beran and S. Hirata, Phys. Chem. Chem. Phys. **14**, 7559 (2012)

[247] T. Vreven, K. Morokuma, Ö. Farkas, H. Schlegel, and M. Frisch, J. Comput. Chem. **24**, 760 (2003)

[248] J. Stanton, J. Gauss, M. Harding, and P. Szalay, with contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, L. Cheng, O. Christiansen, M. Heckert, O. Heun, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, K. Klein, W.J. Lauderdale, D.A. Matthews, T. Metzroth, L.A. Mück, D.P. O'Neill, D.R. Price, E. Prochnow, K. Ruud, F. Schiffmann, W. Schwalbach, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, and J.D. Watts, the integral packages MOLECULE (J. Almlöf and P.R. Taylor) and PROPS (P.R. Taylor) and 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

[249] C. L. Janssen, I. B. Nielsen, M. L. Leininger, E. F. Valeev, and E. T. Seidl, "The massively parallel quantum chemistry program (mpqc) version 2.3.1," Sandia National Laboratories, Livermore, CA, USA, http://www.mpqc.org (2004)

[250] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, "Gaussian 09 Revision D.01," Gaussian Inc. Wallingford CT (2009)

[251] See supplementary material at http://dx.doi.org/10.1063/1.4829463 for additional tables summarizing deviations relative to CCSD(T) computations.

[252] E. Miliordos, E. Aprà, and S. S. Xantheas, J. Chem. Phys. **139**, 114302 (2013)

[253] J. A. Anderson, K. Crager, L. Fedoroff, and G. S. Tschumper, J. Chem. Phys. **121**, 11023 (2004)

[254] S. Graf, W. Mohr, and S. Leutwyler, J. Chem. Phys. **110**, 7893 (1999)

[255] A. Stone, *The Theory of Intermolecular Forces* (Clarendon, Oxford, 1997)

[256] P. A. Kollman and L. C. Allen, J. Chem. Phys. **52**, 5085 (1969)

[257] J. Del Bene and J. Pople, J. Chem. Phys. **52**, 4858 (1970)

[258] J. Dill, L. Allen, W. Topp, and J. Pople, J. Am. Chem. Soc. **97**, 7220 (1975)

[259] C. J. Burnham and S. S. Xantheas, J. Chem. Phys. **116**, 1500 (2002)

[260] W. Cencek, K. Szalewicz, Leforestier, C. Leforestier, R. van Harrevelt, and A. van der Avoird, Phys. Chem. Chem. Phys. **10**, 4716 (2008)

[261] V. Babin, G. R. Medders, and F. Paesani, J. Chem. Theory Comput. **10**, 1599 (2014)

[262] J. Koeler, W. Saenger, and B. Lesyng, J. Comput. Chem. **8**, 1090 (1987)

[263] K. Hermansson, J. Chem. Phys. **89**, 2149 (1988)

[264] M. Hodges, A. Stone, and S. Xantheas, J. Phys. Chem. A **101**, 9163 (1997)

[265] A. Karpfen, in *Molecular Interactions: From van der Waals to Strongly Bound Complexes*, edited by S. Scheiner (John Wiley and Sons, New York, 1997) Chap. 8, pp. 265–296

[266] M. Masella, N. Gresh, and J.-P. Flament, J. Chem. Soc., Faraday Trans. **94**, 2745 (1998)

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

[268] J. Řezáč and P. Hobza, J. Chem. Theory Comput. **9**, 364 (2012)

[269] P. Jurečka, J. Šponer, J. Černý, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006)

[270] R. Podeszwa, K. Patkowski, and K. Szalewicz, Phys. Chem. Chem. Phys. **12**, 5974 (2010)

[271] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, and C. D. Sherrill, J. Chem. Phys. **132**, 144104 (2010)

[272] K. S. Thanthiriwatte, E. G. Hohenstein, L. A. Burns, and C. D. Sherrill, J. Chem. Theory Comput. **7**, 88 (2011)

[273] M. S. Marshall, L. A. Burns, and C. D. Sherrill, J. Chem. Phys. **135**, 194102 (2011)

[274] J. Thiévin, Y. Cadudal, R. Georges, and A. Vigasin, J. Mol. Spec. **240**, 141 (2006)

[275] M. A. Suhm and F. Kollipost, Phys. Chem. Chem. Phys. **15**, 10702 (2013)

[276] J. Zischang and M. A. Suhm, J. Chem. Phys. **139**, 024201 (2013)

[277] K. E. Otto, Z. Xue, P. Zielke, and M. A. Suhm, Phys. Chem. Chem. Phys. **16**, 9849 (2014)

[278] U. Buck, C. C. Pradzynski, T. Zeuch, J. M. Dieterich, and B. Hartke, Phys. Chem. Chem. Phys. **16**, 6859 (2014)

[279] I. Mills, "Molecular spectroscopy: modern research," (Academic Press, New York, 1972) pp. 115–140

[280] V. Barone, J. Chem. Phys. **122**, 014108 (2005)

[281] J. Vázquez and J. F. Stanton, Mol. Phys. **104**, 377 (2006)

[282] V. Barone, J. Bloino, C. A. Guido, and F. Lipparini, Chem. Phys. Lett. **496**, 157 (2005)

[283] D. P. Schofield and H. G. Kjaergaard, Phys. Chem. Chem. Phys. **5**, 3100 (2003)

[284] D. P. Schofield, J. R. Lane, and H. G. Kjaergaard, J. Phys. Chem. A **111**, 567 (2007)

[285] T. Salmi, V. Hänninen, A. L. Garden, H. G. Kjaergaard, J. Tennyson, and L. Halonen, J. Phys. Chem. A **112**, 6305 (2008)

[286] H. G. Kjaergaard, A. L. Garden, G. M. Chaban, R. B. Gerber, D. A. Matthews, and J. F. Stanton, J. Phys. Chem. A **112**, 4324 (2008)

[287] A. L. Garden, L. Halonen, and H. G. Kjaergaard, J. Phys. Chem. A **112**, 7439 (2008)

[288] J. R. Lane and H. G. Kjaergaard, J. Chem. Phys. **132**, 174304 (2010)

[289] K. Mackeprang, H. G. Kjaergaard, T. Salmi, V. Hänninen, and L. Halonen, J. Chem. Phys. **140**, 184309 (2014)

[290] R. Kalescky, W. Zou, E. Kraka, and D. Cremer, Chem. Phys. Lett. **554**, 243 (2012)

[291] H. Lischka, J. Am. Chem. Soc. **96**, 4761 (1974)

[292] D. R. Yarkony, J. Chem. Phys. **60**, 855 (1974)

[293] H. Lischka, Chem. Phys. Lett. **66**, 108 (1979)

[294] A. Karpfen, A. Beyer, and P. Schuster, Int. J. Quantum Chem. **19**, 1113 (1981)

[295] J. F. Gaw, Y. Yamaguchi, M. A. Vincent, and H. F. Schaefer III, J. Am. Chem. Soc. **106**, 3133 (1984)

[296] M. J. Frisch, J. E. Del Bene, J. S. Binkley, and H. F. Schaefer, J. Chem. Phys. **84**, 2289 (1986)

[297] Z. Latajka and S. Scheiner, Chem. Phys. **122**, 413 (1988)

[298] D. M. Bishop, J. Pipin, and B. Kirtman, J. Chem. Phys. **102**, 6778 (1995)

[299] C. L. Collins, K. Morihashi, Y. Yamaguchi, and H. F. Schaefer, J. Chem. Phys. **103**, 6051 (1995)

[300] K. A. Peterson and T. H. Dunning, J. Chem. Phys. **102**, 2032 (1995)

[301] G. S. Tschumper, Y. Yamaguchi, and H. F. Schaefer, J. Chem. Phys. **106**, 9627 (1997)

[302] P. Hobza and Z. Havlas, Collect. Czechoslov. Chem. Commun. **63**, 1343 (1998)

[303] A. Halkier, W. Klopper, T. Helgaker, P. Jørgensen, and P. R. Taylor, J. Chem. Phys. **111**, 9157 (1999)

[304] G. S. Tschumper, M. D. Kelty, and H. F. Schaefer, Mol. Phys. **96**, 493 (1999)

[305] J. Friedrich, E. Perlt, M. Roatsch, C. Spickerman, and B. Kirchner, J. Chem. Theory Comput. **7**, 843 (2011)

[306] J. Řezáč and P. Hobza, J. Chem. Theory Comput. **10**, 3066 (2014)

[307] M. Kofranek, H. Lischka, and A. Karpfen, Chem. Phys. **121**, 136 (1988)

[308] P. Bunker, M. Kofranek, H. Lischka, and A. Karpfen, J. Chem. Phys. **89**, 3002 (1988)

[309] P. Bunker, P. Jensen, A. Karpfen, M. Kofranek, and H. Lischka, J. Chem. Phys. **92**, 7432 (1990)

[310] M. Quack and M. A. Suhm, Chem. Phys. Lett. **171**, 517 (1990)

[311] M. Quack and M. A. Suhm, J. Chem. Phys. **95**, 28 (1991)

[312] W. Schneider and W. Thiel, Chem. Phys. Lett. **157**, 367 (1989)

[313] J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, N. J. Russ, M. L. Leininger, C. L. Janssen, E. T. Seidl, W. D. Allen, H. F. Schaefer, R. A. King, E. F. Valeev, C. D. Sherrill, and T. D. Crawford, Wiley Interdisciplinary Reviews: Computational Molecular Science **2**, 556 (2012)

[314] See Supporting Information at http://dx.doi.org/10.1021/ct500860v for optimized molecular geometries, harmonic vibrational frequencies, VPT2 fundamental frequencies and zero-point vibrational energies.

[315] K. von Puttkamer and M. Quack, Chem. Phys. **139**, 31 (1989)

[316] D. A. Matthews, J. Vázquez, and J. F. Stanton, Mol. Phys. **105**, 2659 (2007)

[317] M. Quack and M. A. Suhm, Theor. Chim. Acta **93**, 61 (1996)

[318] D. T. Anderson, S. Davis, and D. J. Nesbitt, J. Chem. Phys. **104**, 6225 (1996)

[319] A. S. Pine, W. J. Lafferty, and B. J. Howard, J. Chem. Phys. **81**, 2939 (1984)

[320] R. Miller, Acc. Chem. Res. **23**, 10 (1990)

[321] D. T. Anderson, S. Davis, and D. J. Nesbitt, J. Chem. Phys. **105**, 4488 (1996)

[322] M. Hippler, L. Oeltjen, and M. Quack, J. Phys. Chem. A **111**, 12659 (2007)

[323] M. A. Suhm, J. T. Farrell Jr., A. McIlroy, and D. J. Nesbitt, J. Chem. Phys. **97**, 5341 (1992)

[324] F. Huisken, M. Kaloudis, and A. Kulcke, J. Chem. Phys. **104**, 17 (1996)

[325] L. Braly, K. Liu, M. Brown, F. Keutsch, R. Fellers, and R. Saykally, J. Chem. Phys. **112**, 10314 (2000)

[326] F. N. Keutsch, L. B. Braly, M. G. Brown, H. A. Harker, P. B. Petersen, C. Leforestier, and R. J. Saykally, J. Chem. Phys. **119**, 8927 (2003)

[327] Y. Bouteiller and J. Perchard, Chem. Phys. **305**, 1 (2004)

[328] B. E. Rocher-Casterline, L. C. Ch'ng, A. K. Mollner, and H. Reisler, J. Chem. Phys. **134**, 211101 (2011)

[329] F. Huisken, M. Kaloudis, and A. Vigasin, Chem. Phys. Lett. **269**, 235 (1997)

[330] G. Groenenboom, P. Wormer, A. van der Avoird, E. Mas, R. Bukowski, and K. Szalewicz, J. Chem. Phys. **113**, 6702 (2000)

[331] S. A. Nizkorodov, M. Ziemkiewicz, and D. J. Nesbitt, J. Chem. Phys. **122**, 194316 (2005)

[332] I. V. Ptashnik, K. M. Smith, K. P. Shine, and D. A. Newnham, Quarterly Journal of the Royal Meteorological Society **130**, 2391 (2004)

[333] S. Scheiner, *Hydrogen Bonding: A Theoretical Perspective* (Oxford University Press, New York, 1997)

[334] E. Grabowski, S.J., *Hydrogen Bonding - New Insights, in Challenges and Advances in Computational Chemistry and Physics* (Springer, New York, 2006)

[335] G. Gilli and P. Gilli, *The Nature of the Hydrogen Bond- IUCr Monographs on Crystallography - 23* (Oxford University Press New York, 2009)

[336] F. N. Keutsch and R. J. Saykally, Proceedings of the National Academy of Sciences **98**, 10533 (2001)

[337] S. S. Xantheas and T. H. Dunning Jr., J. Chem. Phys. **98**, 8037 (1993)

[338] J. Paul, C. Collier, R. Saykally, J. Scherer, and A. O'keefe, J. Phys. Chem. A **101**, 5211 (1997)

[339] C. Burnham, S. Xantheas, M. Miller, B. Applegate, and R. Miller, J. Chem. Phys. **117**, 1109 (2002)

[340] M. Slipchenko, K. Kuyanov, B. Sartakov, and A. Vilesov, J. Chem. Phys. **124**, 241101 (2006)

[341] S. Hirabayashi and K. M. Yamada, J. Mol. Struct **795**, 78 (2006)

[342] S. Hirabayashi and K. M. Yamada, Chem. Phys. Lett. **435**, 74 (2007)

[343] M. E. Fajardo and S. Tam, J. Chem. Phys. **115**, 6807 (2001)

[344] E. Diken, W. Robertson, and M. Johnson, J. Phys. Chem. A **108**, 64 (2004)

[345] C. Tainter and J. Skinner, J. Am. Chem. Soc. **137**, 104304 (2012)

[346] K. E. Otto, Z. Xue, P. Zielke, and M. A. Suhm, Phys. Chem. Chem. Phys. **16**, 9849 (2014)

[347] Y. Wang, X. Huang, B. C. Shepler, B. J. Braams, and J. M. Bowman, J. Chem. Phys. **134**, 94509 (2011)

[348] J. C. Howard and G. S. Tschumper, J. Chem. Phys. **139**, 184113 (2013)

[349] J. C. Howard, J. L. Gray, A. J. Hardwick, L. T. Nguyen, and G. S. Tschumper, J. Chem. Theory Comput. **10**, 5426 (2014)

[350] Y. Wang and J. M. Bowman, J. Chem. Phys. **136**, 144113 (2012)

[351] G. S. Fanourgakis and S. S. Xantheas, J. Chem. Phys. **128**, 074506 (2008)

[352] TTM3-F program download link at http://www.pnl.gov/science/ttm3f.asp (accessed January 28, 2015)

[353] See Supporting Information at http://dx.doi.org/10.1021/acs.jctc.5b00225 for optimized geometries and harmonic vibrational frequencies

[354] G. Herzberg, *Molecular Spectra and Molecular Structure Vol III - Electronic Spectra and Electronic Structure of Polyatomic Molecules* (Krieger, Florida, 1991) p. 585

[355] M. Schütz and H.-J. Werner, Chem. Phys. Lett. **318**, 370 (2000)

[356] Y. Watanabe, S. Maeda, and K. Ohno, J. Chem. Phys. **129**, 074315 (2008)

[357] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, J. Chem. Phys. **109**, 7764 (1998)

[358] H. Partridge and D. W. Schwenke, J. Chem. Phys. **106**, 4618 (1996)

[359] M. D. Fayer and N. E. Levinger, Annu. Rev. Anal. Chem. **3**, 89 (2010)

[360] A. M. Jubb, W. Hua, and H. C. Allen, Annu. Rev. Phys. Chem. **63**, 107 (2012)

[361] S. Bourrelly, B. Moulin, A. Rivera, G. Maurin, S. Devautour-Vinot, C. Serre, T. Devic, P. Horcajada, A. Vimont, G. Clet, M. Daturi, J.-C. Lavalley, S. Loera-Serna, R. Denoyel, P. L. Llewellyn, and G. Férey, J. Am. Chem. Soc. **132**, 9488 (2010)

[362] S. Horvath, A. B. McCoy, B. M. Elliott, G. H. Weddle, J. R. Roscioli, and M. A. Johnson, J. Phys. Chem. A **114**, 1556 (2010)

[363] F. Perakis, J. A. Borek, and P. Hamm, J. Chem. Phys. **139**, 014501 (2013)

[364] G. R. Medders and F. Paesani, J. Phys. Chem. Lett. **5**, 2897 (2014)

[365] S. Nihonyanagi, T. Ishiyama, T. Lee, S. Yamaguchi, M. Bonn, A. Morita, and T. Tahara, J. Am. Chem. Soc. **133**, 16875 (2011)

[366] L. Piatkowski, Z. Zhang, E. H. G. Backus, H. J. Bakker, and M. Bonn, Nature Comm. **5**, 4083 (2014)

[367] Y. Maréchal, *The Hydrogen Bond and the Water Molecule: The Physics and Chemistry of Water, Aqueous and Bio-Media* (Elsevier, Amsterdam, 2006)

[368] A. Millo, Y. Raichlin, and A. Katzir, Appl. Spectrosc. **59**, 460 (2005)

[369] W. J. Smit and H. J. Bakker, J. Chem. Phys. **139**, 204504 (2013)

[370] A. B. McCoy, J. Phys. Chem. B **118**, 8286 (2014)

[371] Y. R. Shen and V. Ostroverkhov, Chem. Rev. **106**, 1140 (2006)

[372] J. L. Skinner, P. A. Pieniazek, and S. M. Gruenbaum, Acc. Chem. Res. **45**, 93 (2012)

[373] I. V. Stiopkin, C. Weeraman, P. A. Pieniazek, F. Y. Shalhout, J. L. Skinner, and A. V. Benderskii, Nature **474**, 192 (2011)

[374] M. Vinaykin and A. V. Benderskii, J. Phys. Chem. Lett. **3**, 3348 (2012)

[375] Y. Nagata, C.-S. Hsieh, T. Hasegawa, J. Voll, E. H. G. Backus, and M. Bonn, J. Phys. Chem. Lett. **4**, 1872 (2013)

[376] S. S. Xantheas, J. Chem. Phys. **102**, 4505 (1995)

[377] E. E. Dahlke and D. G. Truhlar, J. Phys. Chem. B. Lett. **109**, 15677 (2005)

[378] B. Santra, A. Michaelides, and M. Scheffler, J. Chem. Phys. **127**, 184104 (2007)

[379] G. Shields and K. Kirschner, Synthesis and Reactivity in Inorganic, Metal-Organic, and Nano-Metal Chemistry **38**, 32 (2008)

[380] V. S. Bryantsev, M. S. Diallo, A. C. van Duin, and W. A. Goddard, III, J. Chem. Theory Comput. **5**, 1016 (2009)

[381] P. L. Silvestrelli, Chem. Phys. Lett. **475**, 285 (2009)

[382] F. Li, L. Wang, J. Zhao, J. R. Xie, K. E. Riley, and Z. Chen, Theor. Chem. Acc. **130**, 341 (2011)

[383] J. A. Plumley and J. Dannenberg, J. Comput. Chem. **32**, 1519 (2011)

[384] J. C. Howard and G. S. Tschumper, J. Chem. Theory Comput. **11**, 2126 (2015)

[385] Y. Zhao and D. G. Truhlar, J. Chem. Phys. **125**, 194101 (2006)

[386] R. Peverati and D. G. Truhlar, J. Phys. Chem. Lett. **3**, 117 (2012)

[387] R. Peverati and D. G. Truhlar, J. Chem. Theory Comput. **8**, 2310 (2012)

[388] R. Peverati and D. G. Truhlar, Phys. Chem. Chem. Phys. **14**, 13171 (2012)

[389] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996)

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

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

[392] B. Miehlich, A. Savin, H. Stoll, and H. Preuss, Chem. Phys. Lett. **157**, 200 (1989)

[393] R. Peverati, Y. Zhao, and D. G. Truhlar, J. Phys. Chem. Lett. **2**, 1991 (2011)

[394] S. Grimme, J. Comput. Chem. **27**, 1787 (2006)

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

[396] X. Xu and W. A. Goddard III, Proc. Natl. Acad. Sci. **101**, 2673 (2004)

[397] Y. Zhao and D. G. Truhlar, Theor. Chem. Acc. **120**, 215 (2008)

[398] C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999)

[399] R. Peverati and D. G. Truhlar, J. Chem. Phys. **135**, 191102 (2011)

[400] J. Chai and M. Head-Gordon, J. Chem. Phys. **128**, 084106 (2008)

[401] J. Chai and M. Head-Gordon, Phys. Chem. Chem. Phys. **10**, 6615 (2008)

[402] T. Yanai, D. P. Tew, and N. C. Handy, Chem. Phys. Lett. **393**, 51 (2004)

[403] O. A. Vydrov and G. E. Scuseria, J. Chem. Phys. **125**, 234109 (2006)

[404] O. A. Vydrov, J. Heyd, A. V. Krukau, and G. E. Scuseria, J. Chem. Phys. **125**, 074106 (2006)

[405] O. A. Vydrov, G. E. Scuseria, and J. P. Perdew, J. Chem. Phys. **126**, 154109 (2007)

[406] R. Peverati and D. G. Truhlar, Phys. Chem. Chem. Phys. **14**, 16187 (2012)

[407] R. Peverati and D. G. Truhlar, J. Phys. Chem. Lett. **2**, 2810 (2011)

[408] J. Heyd and G. E. Scuseria, J. Chem. Phys. **121**, 1187 (2004)

[409] J. Heyd and G. E. Scuseria, J. Chem. Phys. **120**, 7274 (2004)

[410] J. Heyd, J. E. Peralta, G. E. Scuseria, and R. L. Martin, J. Chem. Phys. **123**, 174101 (2005)

[411] A. F. Izmaylov, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **125**, 104103 (2006)

[412] A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, and G. E. Scuseria, J. Chem. Phys. **125**, 224106 (2006)

[413] T. M. Henderson, A. F. Izmaylov, G. Scalmani, and G. E. Scuseria, J. Chem. Phys. **131**, 044108 (2009)

[414] T. M. Henderson, A. F. Izmaylov, G. E. Scuseria, and A. Savin, J. Chem. Theory Comput. **4**, 1254 (2008)

[415] S. Grimme, J. Chem. Phys. **124**, 034108 (2006)

[416] T. Schwabe and S. Grimme, Phys. Chem. Chem. Phys. **8**, 4398 (2006)

[417] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010)

[418] See supplementary material at http://dx.doi.org/10.1063/1.4936654 for all optimized geometries and harmonic vibrational frequencies.

[419] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, J. Chem. Phys. **115**, 3540 (2001)

[420] K. E. Riley, M. Pitoňák, P. Jurečka, and P. Hobza, Chem. Rev. **110**, 5023 (2012)

[421] J. Witte, M. Goldey, J. B. Neaton, and M. Head-Gordon, J. Chem. Theory Comput. **11**, 1481 (2015)

[422] G. R. Medders, A. W. Götz, M. A. Morales, P. Bajaj, and F. Paesani, J. Chem. Phys. **143**, 104102 (2015)

Vita

J. Coleman Howard

Department of Chemistry and Biochemistry
University of Mississippi, University, MS 38677-1848 USA

Phone: 601 507 2661
Electronic Mail: jchoward@go.olemiss.edu

I. Educational Background

Doctor of Philosophy in Chemistry (Advisor: Dr. Gregory S. Tschumper)	University of Mississippi	*2015
Bachelor of Science in Chemistry (Sally McDonnell Barksdale Honors College)	University of Mississippi	2011

II. Professional Experience and Training

Telluride School on Theoretical Chemistry	TSRC	2013
Graduate Research Assistant	University of Mississippi	2011–2015
Graduate Teaching Assistant	University of Mississippi	2010–2011
Undergraduate Research Assistant	University of Mississippi	2008–2010

III. Honors and Awards

Dissertation Fellowship	University of Mississippi	2015
ACS Graduate Research Scholar	University of Mississippi	2015
ACS Graduate Research Award	University of Mississippi	2013
Outstanding Senior Chemistry Graduate	University of Mississippi	2011
Cum Laude Graduate	University of Mississippi	2011
Phi Kappa Phi Honor Society	University of Mississippi	2011
Sigma Alpha Lambda Honors Organization	University of Mississippi	2007
National Society of Collegiate Scholars	University of Mississippi	2006
Alpha Lambda Delta Honor Society	University of Mississippi	2006
W.R. Newman Scholar	University of Mississippi	2006
National Merit Scholar	University of Mississippi	2006
Chancellor's Leadership Class	University of Mississippi	2006

IV. Publications

1. J.C. Howard, N.I. Hammer and G.S. Tschumper, *ChemPhysChem*, **17**, 3262–3272 (2011). "Structures, Energetics and Vibrational Frequency Shifts of Hydrated Pyrimidine." http://dx.doi.org/10.1002/cphc.201100457

2. A.M. Wright, A.A. Howard, J.C. Howard, G.S. Tschumper and N.I. Hammer, *J. Phys. Chem. A*, **117**, 5435–5446 (2013). "Charge Transfer and Blue Shifting Vibrational Frequencies in a Hydrogen Bond Acceptor." http://dx.doi.org/10.1021/jp401642b

3. J.C. Howard and G.S. Tschumper, *J. Chem. Phys.*, **139**, 184113 (2013). "N-body:Many-body QM:QM Vibrational Frequencies: Application to Small Hydrogen-Bonded Clusters." http://dx.doi.org/10.1063/1.4829463

4. J.C. Howard and G.S. Tschumper, *WIREs Comput. Mol. Sci.*, **4**, 199–224 (2014). "Wavefunction methods for the accurate characterization of Water Clusters." http://dx.doi.org/10.1002/wcms.1168

5. J.C. Howard, J.L. Gray, A.J. Hardwick, L.T. Nguyen and G.S. Tschumper, *J. Chem.Theory Comput.*, **10**, 5426–5435 (2014). "Getting Down to the Fundamentals of Hydrogen Bonding: Anharmonic Vibrational Frequencies of $(HF)_2$ and $(H_2O)_2$ from Ab Initio Electronic Structure Computations." http://dx.doi.org/10.1021/ct500860v

6. J.C. Howard and G.S. Tschumper, *J. Chem.Theory Comput.*, **11**, 2126–2136 (2015). "Benchmark Structures and Harmonic Vibrational Frequencies Near the CCSD(T) Complete Basis Set Limit for Small Water Clusters: $(H_2O)_n = 2, 3, 4, 5, 6$." http://dx.doi.org/10.1021/acs.jctc.5b00225

7. J.C. Howard, J.D. Enyard and G.S. Tschumper, *J. Chem. Phys.*, **143**, 214103 (2015). "Assessing the Accuracy of Some Popular DFT Methods for Computing Harmonic Vibrational Frequencies of Water Clusters."

V. Professional Presentations

1. "Structures, Energetics and Vibrational Frequency Shifts of Hydrated Pyrimidine." Poster presentation at annual EPSCoR meeting in Jackson, MS. 2011.

2. "Analytic 1st and 2nd Derivatives for a 2-body:Many-body Fragmentation Method: Optimized Structures and Vibrational Frequencies." Poster presentation at Current Trends in Computational Chemistry Conference in Jackson, MS. 2011.

3. "Analytic 1st and 2nd Derivatives for a 2-body:Many-body QM:QM Method: CCSD(T):MP2 Optimized Structures and Vibrational Frequencies." Poster presentation at the Southeastern Theoretical Chemistry Association Annual Meeting in Athens, GA. 2012.

4. "Analytic 1st and 2nd Derivatives for a 2-body:Many-body QM:QM Method: CCSD(T):MP2 Optimized Structures and Vibrational Frequencies." Poster presentation at the annual EPSCoR meeting in Starkville, MS. 2012.

5. "Estimating CCSD(T) Optimized Structures and Vibrational Frequencies with the N-body:Many-body Approach." Poster presentation at the Sanibel Symposium in St. Simons Island, GA. 2013.

6. "Efficiently and Accurately Reproducing CCSD(T) Optimized Structures and Vibrational Frequencies of Water Clusters with the N-body:Many-body QM:QM Method." Poster presentation at the Southeastern Theoretical Chemistry Association Annual Meeting in Auburn, AL. 2013.

7. "Efficiently and Accurately Reproducing CCSD(T) Optimized Structures and Vibrational Frequencies of Water Clusters with the N-body:Many-body QM:QM Method." Speaker presentation at ACS National Meeting in New Orleans, LA, 2013.

8. "Efficiently and Accurately Reproducing CCSD(T) Optimized Structures and Vibrational Frequencies of Water Clusters with the N-body:Many-body QM:QM Method." Poster presentation at EPSCoR state meeting in Hattiesburg, MS. 2013.

9. "Convergence of Computed Vibrational Properties of Small Molecules and Hydrogen-Bonded Clusters from Correlated Wavefunction Methods." Poster presentation at 25th Austin Symposium on Molecular Structure and Dynamics at Dallas in Dallas, TX. 2014.

10. "Convergence of Computed Vibrational Properties of Small Molecules and Hydrogen-Bonded Clusters from Correlated Wavefunction Methods." Poster presentation at EPSCoR state meeting in Starkville, MS. 2014.

11. "Convergence of Computed Vibrational Properties of Small Molecules and Hydrogen-Bonded Clusters from Correlated Wavefunction Methods." Speaker presentation at Southeastern Theoretical Chemistry Association Annual Meeting in Atlanta, GA. 2014.