UNIVERSITY OF CALIFORNIA, SAN DIEGO

Development of "first principles" methods for modeling vibrational spectra in condensed phases

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Chemistry

by

Gregory R. Medders

Committee in charge:

Professor Francesco Paesani, Chair
Professor Seth Cohen
Professor John Crowell
Professor Michael Gilson
Professor John Weare

2015

The dissertation of Gregory R. Medders is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I first and foremost would like to thank my doctoral research advisor, co-author, and mentor, Professor Francesco Paesani. Knowing that I had little prior knowledge of theoretical and computational chemistry, you took a chance on me by inviting me to join your group. What I have learned in the last five years I owe to the opportunities and support that you have provided. I'm truly grateful for the relationship we've developed and that you pushed me to be the best that I could be.

Some work presented in this dissertation (specifically, Chapters 1–3, corresponding to papers 1–3 in the list below) was performed in collaboration with Dr. Volodymyr Babin, who was a postdoc in the Paesani research group. Volodymyr, I owe much of my understanding of computer programming to your expertise and guidance, and I am thankful both for the role that you played in my graduate education and your friendship during our time together at UC San Diego.

I would like to acknowledge the organizations that published our work (specifically, the American Chemical Society for papers 1–5 and the American Institute of Physics for paper 6 in the list below) and gave permission to reproduce that published material in this dissertation.

1. Medders, G.R.; Babin, V.B.; Paesani, F. "A critical assessment of two-body and three-body interactions in water", *J. Chem. Theory Comput.* **2013** 9, 1103–1114.

2. Babin, V.; Medders, G.R.; Paesani, F. "Development of a first principles water potential with flexible monomers. II: Trimer potential energy surface, third virial coefficient, and small clusters", *J. Chem. Theory Comput.* **2014** 10, 1599–1607.

3. Medders, G.R.; Babin, V.; Paesani, F. "Development of a first principles water potential with flexible monomers. III: Liquid phase properties", *J. Chem. Theory Comput.* **2014** 10, 2906–2910.

4. Medders, G.R.; Paesani, F. “Many-body convergence of the electrostatic properties of water”, *J. Chem. Theory Comput.* **2013** 9, 4844–4852.

5. Medders, G.R.; Paesani, F. “Infrared and Raman spectroscopy of liquid water through first principles many-body molecular dynamics”, *J. Chem. Theory Comput.* **2015** 11, 1145–1154.

6. Medders, G.R.; Paesani, F. “On the interplay of the potential energy and dipole moment surfaces in controlling the infrared activity of liquid water”, *J. Chem. Phys.* **2015** 142, 212411.

Thanks to my group mates who, over the years, provided helpful discussions and fun times. Particular thanks to Jordi and Volodymyr–those 3pm coffee breaks were crucial.

I would like to thank Greg Schenter and Chris Mundy for hosting me (twice!) at Pacific Northwest National Lab and Miguel Morales-Silva for hosting me at Lawrence Livermore National Lab. I learned a ton during those weeks and months, and my graduate education was deeply enriched through the unique perspectives and expertise that you each brought.

To my undergraduate research advisor, Professor Gordon Purser, I can’t thank you enough for both exposing me to computational chemistry and for giving years of friendship and honesty. It is no understatement to say that I wouldn’t be where I am today had I not met you.

On a personal note, I want to thank my partner, Cuong Nguyen. I met you while studying at a coffee shop just three weeks into my first quarter at UC San Diego, and you’ve been by my side ever since. This work would not have been possible without the happiness you (and Merlin the cat) have brought me over the last five years.

To Jeff, my little brother, thanks for the conversations we’ve had over the years.

And finally, to my parents. You taught me to work hard for what I want, and your endless love and support has made me who I am today.

VITA

2010	B. S. in Chemistry, University of Tulsa
2012	M. S. in Chemistry, University of California, San Diego
2015	Ph. D. in Chemistry, University of California, San Diego

PUBLICATIONS

Babin, V.B.; Medders, G.R.; Paesani, F. "Toward a universal water model: First principles simulations from the dimer to the liquid phase", *J. Phys. Chem. Lett.* **2012** 3, 3765–3769.

Medders, G.R.; Babin, V.B.; Paesani, F. "A critical assessment of two-body and three-body interactions in water", *J. Chem. Theory Comput.* **2013** 9, 1103–1114.

Medders, G.R.; Paesani, F. "Many-body convergence of the electrostatic properties of water", *J. Chem. Theory Comput.* **2013** 9, 4844–4852.

Babin, V.; Medders, G.R.; Paesani, F. "Development of a first principles water potential with flexible monomers. II: Trimer potential energy surface, third virial coefficient, and small clusters", *J. Chem. Theory Comput.* **2014** 10, 1599–1607.

Medders, G.R.; Babin, V.; Paesani, F. "Development of a first principles water potential with flexible monomers. III: Liquid phase properties", *J. Chem. Theory Comput.* **2014** 10, 2906–2910.

Medders, G.R.; Paesani, F. "Water dynamics in metal-organic frameworks: Effects of heterogeneous confinement predicted by computational spectroscopy", *J. Phys. Chem. Lett.* **2014** 5, 2897–2902.

Medders, G.R.; Paesani, F. "Infrared and Raman spectroscopy of liquid water through first principles many-body molecular dynamics", *J. Chem. Theory Comput.* **2015** 11, 1145–1154.

Medders, G.R.; Paesani, F. "On the interplay of the potential energy and dipole moment surfaces in controlling the infrared activity of liquid water", *J. Chem. Phys.* **2015** 142, 212411.

ABSTRACT OF THE DISSERTATION

Development of "first principles" methods for modeling vibrational spectra in condensed phases

by

Gregory R. Medders

Doctor of Philosophy in Chemistry

University of California, San Diego, 2015

Professor Francesco Paesani, Chair

Due to the sensitivity of molecular vibrational frequencies and intensities on the surrounding environment, vibrational spectroscopies in principle enable the study of solvation structure and dynamics. Connecting the observed spectral features to a molecular-level picture is, however, often non-trivial. While computer simulations of molecular dynamics represent a potentially powerful tool for developing this molecular-level understanding, the accurate simulation of vibrational spectroscopies in condensed phases poses significant challenges due to the sensitivity of the spectra on both the underlying molecular interactions and the difficulty of obtaining a (statistically meaningful)

treatment of the quantum dynamics.

In this work, we begin by assessing the ability of different molecular models to reproduce thousands of reference two- and three-body interaction energies calculated at the current "gold standard" level of electronic structure theory, CCSD(T). As described in Chapter 2, these results led us to develop a potential energy surface, named MB-pol, that was fitted exclusively to large datasets of CCSD(T) many-body interaction energies. Crucially, MB-pol was designed to be computationally tractable for condensed phase simulations without sacrificing accuracy. MB-pol reproduces experimental measurements of small cluster properties, as well as thermodynamic and dynamical properties of bulk water at ambient conditions, without containing any empirically derived parameters (Chapter 3).

However, unlike the electronic structure calculations to which it is fitted, the MB-pol PES contains no explicit knowledge of the electron distribution, which is required for the calculation of vibrational spectra. To this end, in Chapter 4 we demonstrate that the many-body expansions of the dipole and polarizability also converge for water. Based on this finding, in Chapter 5 we introduce many-body models for the dipole moment and polarizability of water, allowing us to rigorously model IR and Raman spectra from "first principles," through the respective (approximate) quantum time correlation functions. In Chapter 6, we disentangle the contributions of the potential energy and dipole moment surfaces to the IR activity of liquid water. Finally, we conclude in Chapter 7 by reflecting on possible future applications, including the application of the MB-MD approach to the calculation of nonlinear vibrational spectra.

Chapter 1

Introduction

1.1 Connecting theory and experiment through vibrational spectroscopy

From biological function to the Earth's climate, water is a key mediator of many natural phenomena.[1] To a large extent, this is related to the ability of the water molecules to form flexible and dynamic networks of hydrogen bonds. For example, hydration processes facilitated by the rearrangement of the water hydrogen-bonding network play a central role in the organization of ions at environmentally-relevant interfaces,[2,3] in monolayer assembly,[4,5] in biological processes,[6–9] and in the stability and viability of novel materials for energy applications.[10,11]

Since the vibrational frequencies of a water molecule are particularly sensitive to the surrounding environment, vibrational spectroscopy has become a powerful tool for characterizing both molecular structure and dynamics of aqueous systems from the gas to the condensed phases.[5,11–18] However, a unique assignment of the spectra measured for water under different conditions and in different environments is nontrivial due to the fluctuating behavior of the underlying hydrogen-bonding network which gives rise

to several (homogeneously and/or inhomogeneously broadened) features that are often difficult to interpret. This has led to numerous controversies, including those about the nature of association bands,[19–21] the role of Fermi resonances, and the relationship between structural order and spectroscopic features at the air/water interface.[17,22–26] The theoretical interpretation of the water vibrational spectra has, in part, been complicated by the existence of numerous molecular models, which can provide conflicting explanations for the same spectroscopic features. In an attempt to overcome current theoretical limitations, we present here a unified many-body molecular dynamics (MB-MD) simulation approach developed exclusively from "first principles" to calculate vibrational spectra of water at a quantum-mechanical level.

To introduce the MB-MD framework, it is first useful to briefly describe how vibrational spectra can be obtained from molecular simulations. For a vibration to be "infrared active", i.e., for it to adsorb infrared radiation, the charge distribution must change during the vibration.[27,28] Within the dipole approximation, this leads to the familiar selection rule for the fundamental infrared adsorption intensities,

$$I_{\mathrm{IR}}(\omega) \sim |\langle 1|\boldsymbol{\mu}|0\rangle|^2 \tag{1.1}$$

where $\boldsymbol{\mu}$ is the dipole moment and the angular brackets represent integrals over the 0 and 1 vibrational eigenstates. While the dipole moment provides a measure of the anisotropy of a charge distribution, a molecule's polarizability indicates the ability of the molecule to deform its charge distribution in the presence of external electric fields. The geometry dependence of the polarizability is probed in Raman scattering experiments, where the selection rule within the dipole approximation is given by

$$I_{\mathrm{Raman}}(\omega) \sim |\langle 1|\boldsymbol{\alpha}|0\rangle|^2 \tag{1.2}$$

and $\boldsymbol{\alpha}$ is the rank two polarizability tensor. Specifically, in the previous equation, α_{ab} is the "dipole-dipole" polarizability, whose components indicate that an applied external dipole field in direction a generates an induced dipole moment in the b direction. From this point forward, $\boldsymbol{\alpha}$ will be referred to as simply the polarizability.

For condensed phase systems, where it is difficult to define the vibrational eigenstates, infrared and Raman spectra can be obtained within linear response theory through a more convenient time-dependent representation involving Fourier transforms of the corresponding quantum time-correlation functions:

$$I_{\mathrm{IR}}(\omega) \sim \int_{-\infty}^{\infty} dt e^{-i\omega t} \langle \boldsymbol{\mu}(0)\boldsymbol{\mu}(t)\rangle \tag{1.3}$$

and

$$I_{\mathrm{Raman}}(\omega) \sim \int_{-\infty}^{\infty} dt e^{-i\omega t} \langle \boldsymbol{\alpha}(0)\boldsymbol{\alpha}(t)\rangle, \tag{1.4}$$

where the angular brackets imply an ensemble average. Through the time-dependent formulation presented in Eqs. 1.3 and 1.4, it is thus possible to identify three different elements that are necessary for a rigorous calculation of vibrational spectra: 1) an accurate representation of the underlying (Born-Oppenheimer) potential energy surface entering the Hamiltonian that governs the dynamics of the system, 2) accurate representations of the multidimensional dipole moment and polarizability surfaces as a function of the system's degrees of freedom, and 3) a simulation approach capable of correctly describing the quantum dynamics of the system of interest.

In this dissertation, the MB-MD approach for generating accurate, "first principles" representations of the full-dimensional potential energy, dipole moment, and polarizability surface of water is presented. Because these models rely on the Born-Oppenheimer separation of the nuclear and electronic quantum mechanical degrees of freedom (i.e., treating the electronic degrees of freedom as a function of the nuclear

coordinates) and due to the small mass of the hydrogen atoms, inclusion of nuclear quantum effects is critical to accurately describing both thermodynamic and dynamical properties of liquid water. Collectively, and without any empirically derived parameters, these models have so-far accurately predicted (experimentally measured) properties of water ranging from the isomeric equilibria of small clusters to bulk thermodynamic (e.g., the structure and density of water at ambient pressure and temperature) and dynamical properties (e.g., hydrogen bond lifetime, diffusion constants, and infrared and Raman vibrational spectra) at ambient conditions. Furthermore, while this methodology has been developed and applied to water, the techniques described here are expected to be applicable to the wide range of molecular systems.

1.2 The many-body expansion of interaction energies

Connecting small clusters of water and the condensed phases of water through a single molecular model has been a long sought-after goal. The challenges involved in the pursuit of this goal are numerous. For example, at the cluster level, the Born-Oppenheimer energies of topologically distinct isomers of the water hexamer differ by less than 1 kcal/mol,[29–31] indicating that highly-correlated electronic structure calculations are required to quantitatively determine the relative energetic stability of these isomers. In this regard, a faithful description of molecular flexibility appears to be particularly important.[31,32] Furthermore, it has also been shown that nuclear quantum-mechanical effects can impact the structural, thermodynamic, and dynamical properties of both clusters and bulk phases of water.[32–36] The explicit inclusion of nuclear quantum effects in simulations exacerbates the computational expense of a model, providing additional strain on the ability to obtain statistically meaningful results.

Rather than explicitly considering electronic structure of a full N-molecule system,

the total binding energy of the system can be decomposed through an expansion in terms of interacting subunits. In the case of water, the simplest subunit is the water molecule. The binding energy of N interacting molecules is given as E_{N-mer}.[37]

$$E_{N-mer}(1,\ldots,N) = \sum_{i}^{N} V^{1B}(i) + \sum_{i<j}^{N} V^{2B}(i,j) + \sum_{i<j<k}^{N} V^{3B}(i,j,k) + \cdots + V^{NB}(1,\ldots,N) \tag{1.5}$$

Within the above expansion, the total energy of the N-molecule system is represented in terms of its many-body contributions, where, $V^{1B}(i) = E_{mon}(i)$ is the one-body (1B) potential, which describes the energy required to deform individual molecules from their equilibrium geometry. This 1B term depends on the 3M coordinates of molecule, where M is the number of atoms in the molecule, and the molecular coordinates are referred to by a single index, i. In common force-fields of small molecules, the 1B interactions include all bonded terms (i.e., stretches, bends, and torsions). For systems that are easily reduced to distinct molecules, such as water, a 1B configuration is typically referred to as a "monomer", and groups of 2, 3, ... , N interacting monomers are then termed "dimers", "trimers", ..., "N-mers". From this point forward the monomer, dimer, and trimer notation will be dropped and the size of a cluster will be specified by the number of indices.

In Eq. 1.5, higher-order interactions are defined recursively through the lower-order terms. For instance, the two-body (2B) interaction is expressed as

$$V^{2B}(1,2) = E(1,2) - \sum_{i=1}^{2} V^{1B}(i) = E(1,2) - \sum_{i=1}^{2} E(i) \tag{1.6}$$

where $E(1,2)$ is the dimer energy. Similarly, the three-body (3B) interaction is

$$V^{3B}(1,2,3) = E(1,2,3) - \sum_{i<j}^{3} V^{2B}(i,j) - \sum_{i=1}^{3} V^{1B}(i) \tag{1.7}$$
$$= E(1,2,3) - \sum_{i<j}^{3} E(i,j) + \sum_{i=1}^{3} E(i)$$

with $E(1,2,3)$ being the trimer energy. The vast majority of force fields are pairwise additive, meaning that three-body and higher interactions are neglected.

However, provided that it converges quickly, the many-body expansion represents a powerful approach to studying condensed phases, as it allows for the energy of an N-molecule system to be expressed as a sum of lower-order interactions that can, in principle, be calculated with high accuracy. Recently, a detailed study of the convergence of the many-body expansion for water based on the analysis of small clusters was performed using coupled cluster theory with single, double, and perturbative triple excitations [CCSD(T)] and large basis-sets.[31] Consistent with previous calculations,[38–45] it was found that, although two-body interactions dominate the expansion, the three-body term can contribute up to 30% of the total energy of the water hexamer. An estimate of the relative magnitudes of the many-body terms in liquid was obtained through an RI-MP2 analysis of the 21-mer, for which two-body interactions were found to contribute 75-80% of the total interaction energy and three-body interactions comprised 15-20%.[45] For both the water hexamer and the 21-mer, higher-order terms contribute less than 5% of the total interaction energy. It should be noted that, while quickly converging for water, the many-body expansion has been shown to converge slowly and with marked oscillatory behavior for other systems.[46]

1.3 Analysis of two- and three-body water interactions

As perhaps the most studied molecular system, numerous strategies have been proposed and applied (with varying success) to the modeling of water interactions and properties. Here, these modeling strategies are reviewed and assessed for their ability to reproduce benchmark calculations of the many-body interactions of water.

1.3.1 Approaches to modeling water interactions

The many-body expansion provides the underlying framework for common classical force fields. In most cases, including the widely-used TIP4P and SPC families,[47,48] pairwise additivity is assumed, with three- and higher-body interactions being "encoded" into the effective two-body contributions. In addition, the majority of these models treat the water molecules as rigid monomers (i.e., the 1B interactions are set to zero), with only few quantum water models, notably as q-TIP4P/f and qSPC/Fw,[49,50] allowing for molecular flexibility. Nevertheless, pairwise force fields have been surprisingly successful at reproducing, at least qualitatively, the properties of water in homogeneous environments.[51] However, such force fields are expected to be inherently limited in their ability to model the microscopic behavior of aqueous interfaces, water confined at the nanoscale, and clusters, whose properties are sensitive to the detailed interplay of 1B, 2B, 3B, and higher-body interactions.[38–45]

Recent work has focused on improving empirical models through inclusion of three-body interactions, leading to the development of the E3B model.[52,53] Although the inclusion of explicit 3B interactions greatly improves the accuracy of the E3B model relative to pairwise force fields, the use of rigid water monomers and empirical parameterization necessarily misses some of the fundamental properties of the many-body expansion. For example, recent E3B simulations of the isomeric equilibria of

the water hexamer have led to predictions that are markedly different from *ab initio* calculations. Specifically, the prism structure, which corresponds to the energetically lowest-lying isomer at the MP2 and CCSD(T) levels of theory,[30,31] is unstable in the E3B calculations.[54]

Since non-pairwise additive intermolecular interactions arise primarily from electronic polarization at long distances, several methods have been proposed to incorporate this effect into the framework of classical force fields.[55] One common approach is the Applequist polarizable point dipole model,[56] which was elaborated upon by Thole to address the so-called polarization catastrophe.[57] Thole-type polarizable force fields for water include TTM3-F,[58] TTM4-F,[59] and AMOEBA[60] models.

Among methods that attempt to solve directly the many-body problem from "first principles", semiempirical models represent an possible alternative due to their computational efficiency. Semiempirical models such as PM3[61] and PM3-MAIS[62] were derived within the MNDO scheme and differ primarily in the form of the core-core repulsion, as well as in the precise values of their adjustable parameters. These models were parametrized either by fitting experimental data for a wide variety of systems (PM3) or electronic structure reference data in the case of PM3-MAIS. Due to the use of a minimal basis and the explicit neglect of correlation, semiempirical methods are particularly limited in their ability to describe non-bonded interactions. This deficiency has been addressed by the SCP-NDDO model, which augments traditional semiempirical methods with classical polarization.[63] SCP-NDDO has shown success in modeling water clusters and has recently been extended to simulations of bulk properties.[64]

Different DFT methods have also been extensively applied to the study of condensed phases, primarily through the use of GGA functionals such as BLYP[65,66] and PBE.[67] However, common density functionals are, by construction, limited in their ability to describe weakly interacting van der Waals complexes. One attempt to address this

problem involves the addition of a dispersion correction to the energy through the C_6/R^6 term, where the C_6 parameters are atom and basis-set specific.[68,69] These "DFT-D" models have successfully described systems such as the solvation of iodide in water,[70] but are limited by the need to develop parameters for each functional/basis-set.[71] Furthermore, because the correction is pairwise additive, it neglects higher-order dispersion contributions. Recent work to address this limitation has been reported.[72]

A promising alternative to the pairwise DFT-D correction is represented by the non-local van der Waals (nl-vdW) functionals.[73–75] These nl-vdW functionals utilize the electron density to define a non-local correlation contribution to the exchange-correlation functional, leading to a consistent description of both short-range and long-range interactions. Since no atomic or basis-set dependent parameters are required to describe the dispersion interaction due to the explicit dependence of the non-local correlation on the electron density, nl-vdW functionals, in principle, require minimal parameterization and are system-independent. In practice, great care must be taken to avoid double counting of correlation effects in the combination of semi-local and non-local terms. Van der Waals density functionals have recently been applied to the study of liquid water[76] and ice.[77]

One final class of models is represented by the "first principles"-based PESs. These models are built upon a rigorous treatment of the many-body expansion of interactions and are characterized by having a functional form that is sufficiently flexible to accurately describe high-quality electronic structure reference data. Examples of such models are DPP2,[41] CC-pol,[78,79] and WHBB.[80] DPP2 and CC-pol are restricted to the rigid, vibrationally averaged monomer geometries, while WHBB uses permutationally invariant polynomials to represent the flexible monomer 2B and 3B PESs. Such "first principles"-based PESs are often quite computationally demanding and are most commonly used in calculations for gas phase systems,[35] although bulk properties have been obtained from classical simulations with CC-pol[78,81] and DPP2.[41] Recently, a

flexible version of 1B and 2B terms in CC-pol has been developed, CC-pol-8s*f*, and the effects of flexibility on the dimer vibrational-rotation-tunneling (VRT) spectra have been characterized.[82] It was found that both CC-pol-8s*f* and HBB2 (the 2B potential of WHBB) reproduce the experimental VRT spectra "about equally well".[82]

1.3.2 Comparison to the CCSD(T) "Gold Standard"

Here, we assess the ability of the models presented in the previous section to describe the 2B and 3B water interactions of water. In this study, roughly 1400 2B interactions and 500 3B interactions were evaluated at the CCSD(T)/aug-cc-pVTZ level[83,84] and corrected for the basis-set superposition error (BSSE) using the cluster counterpoise method of Eq. A.9. These (flexible) molecular configurations were extracted from 1) classical molecular dynamics (MD) simulations of hexamers at T $\leq$ 30 K using the WHBB potential, 2) classical MD simulations of ice I_h carried out with TTM3-F at 50 K, and 3) classical MD simulations of bulk water at 298 K and experimental density using TTM3-F. Hereafter, these configurations are referred to as "low-energy" configurations. For the analysis of E3B, the CCSD(T) reference interaction energies were recomputed for "rigidified" molecules corresponding to the flexible configurations that were used in the comparison of the other models. All DFT energies were computed using the aug-def2-TZVPP basis[84,85] with the exception of BLYP-D, for which the TZV2P basis was used for consistency with the original parametrization of the model.[68,69,86] MP2 energies were computed with the aug-cc-pVTZ basis, and both DFT and MP2 interactions were corrected for BSSE. With the exception of the semi-empirical methods, all *ab initio* calculations were performed using the freely-available *ab initio* package ORCA.[87] PM3 and PM3-MAIS energies were calculated using the AMBER/SQM semi-empirical package,[88] while the SCP-NDDO energies were obtained using CP2K.[64,89] A linear regression analysis for the data presented in Figs. 1.1 and 1.2, as well as root mean

square error with respect to CCSD(T) data, are presented in the supporting material.

Figs. 1.1 and 1.2 show correlation plots for the 2B and 3B interactions calculated for all models described in Section 1.3.1 relative to the CCSD(T)/aug-cc-pVTZ energies. While most empirical pairwise force fields implicitly include nuclear quantum effects, models such as q-TIP4P/f and q-SPC/Fw were specifically parameterized for quantum simulations and, therefore, presumably provide an approximation to the actual Born-Oppenheimer PES.[49,50] As can be seen from Fig. 1.1, q-TIP4P/f deviates substantially from the true 2B potential energy surface to compensate for the neglect of higher-order interactions (Fig. 1.2). Force fields that account for higher-order terms generally provide a more accurate description of the 2B interactions than effective pairwise models. In this context, while E3B and TTM3-F/TTM4-F/AMOEBA treat higher-order interactions using different schemes, all four models give 2B interactions that are in closer agreement with the CCSD(T) results than the effective pairwise models. It is interesting to note that E3B, which does not not explicitly include induction and was not parameterized using *ab initio* data, describes the 3B contributions energies reasonably well.

The three polarizable models considered in this study (TTM3-F, TTM4-F, and AMOEBA) differ in the way they describe the variation of the molecular charge distribution. As an isolated monomer deforms, the molecular dipole moment varies in a "nonlinear" fashion with respect to the intramolecular coordinates, resulting in a "nonlinear dipole moment surface" (DMS).[90] In TTM4-F, the first-order changes of the DMS are fit to electric multipoles and polarizabilities calculated at the MP2 level. The intramolecular dependence of the atomic charges in TTM3-F was instead motivated by the observation that, while the gas phase monomer charges decrease during the homolytic dissociation, a water molecule in the condensed phase dissociates into charged ions. This argument was used to justify an empirical correction to *ab initio*-derived values, giving rise to effective charges that increase as the monomer geometry departs from equilibrium.

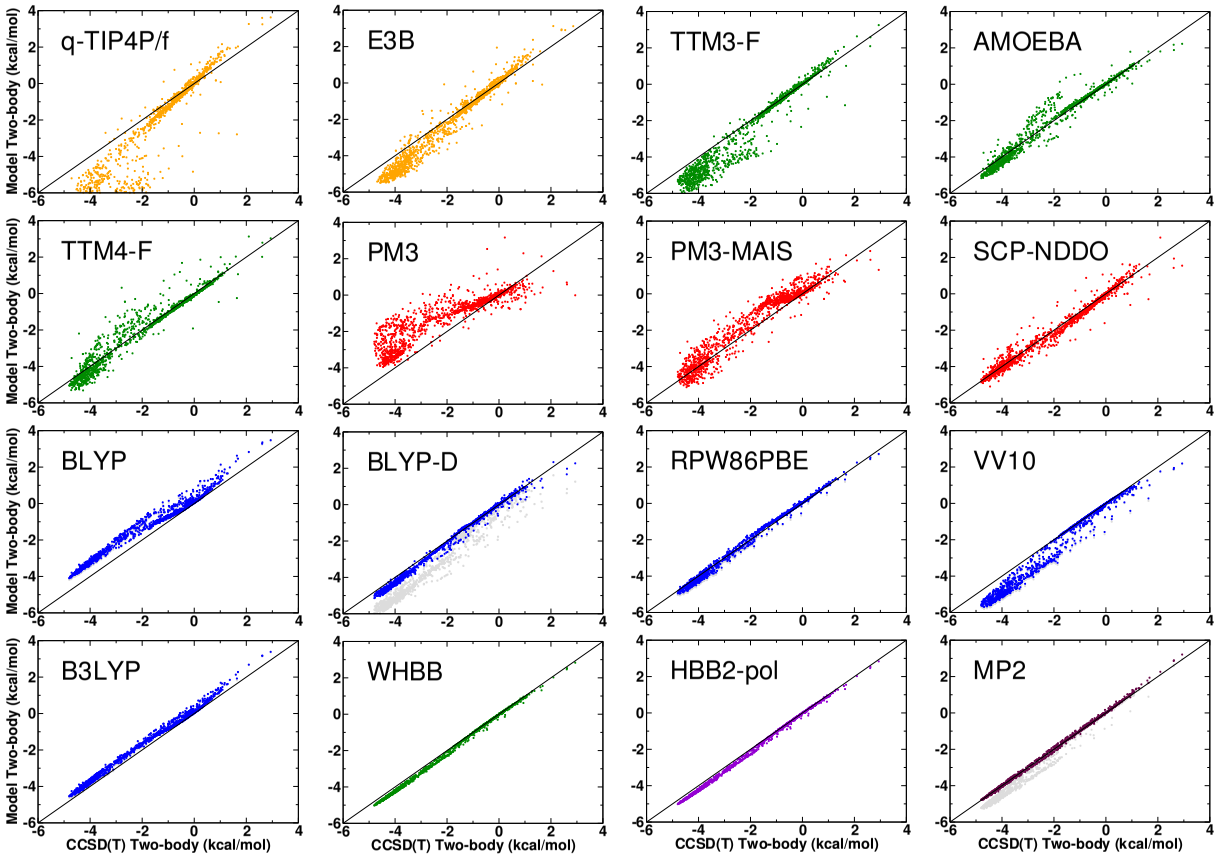

Figure 1.1: Comparison of model 2B interactions to CCSD(T). Plotted on the x-axes is the BSSE-corrected CCSD(T)/aug-cc-pVTZ energies. On the y-axes are the energies for each model. Empirically parametrized models are in orange, polarizable models in green, semiempirical methods in red, DFT methods in blue, and MP2 in maroon. For DFT and MP2, the colored dots are BSSE-corrected energies, while gray dots are BSSE-uncorrected energies. The "first principles"-based PES proposed by us, HBB2-pol, is in violet.

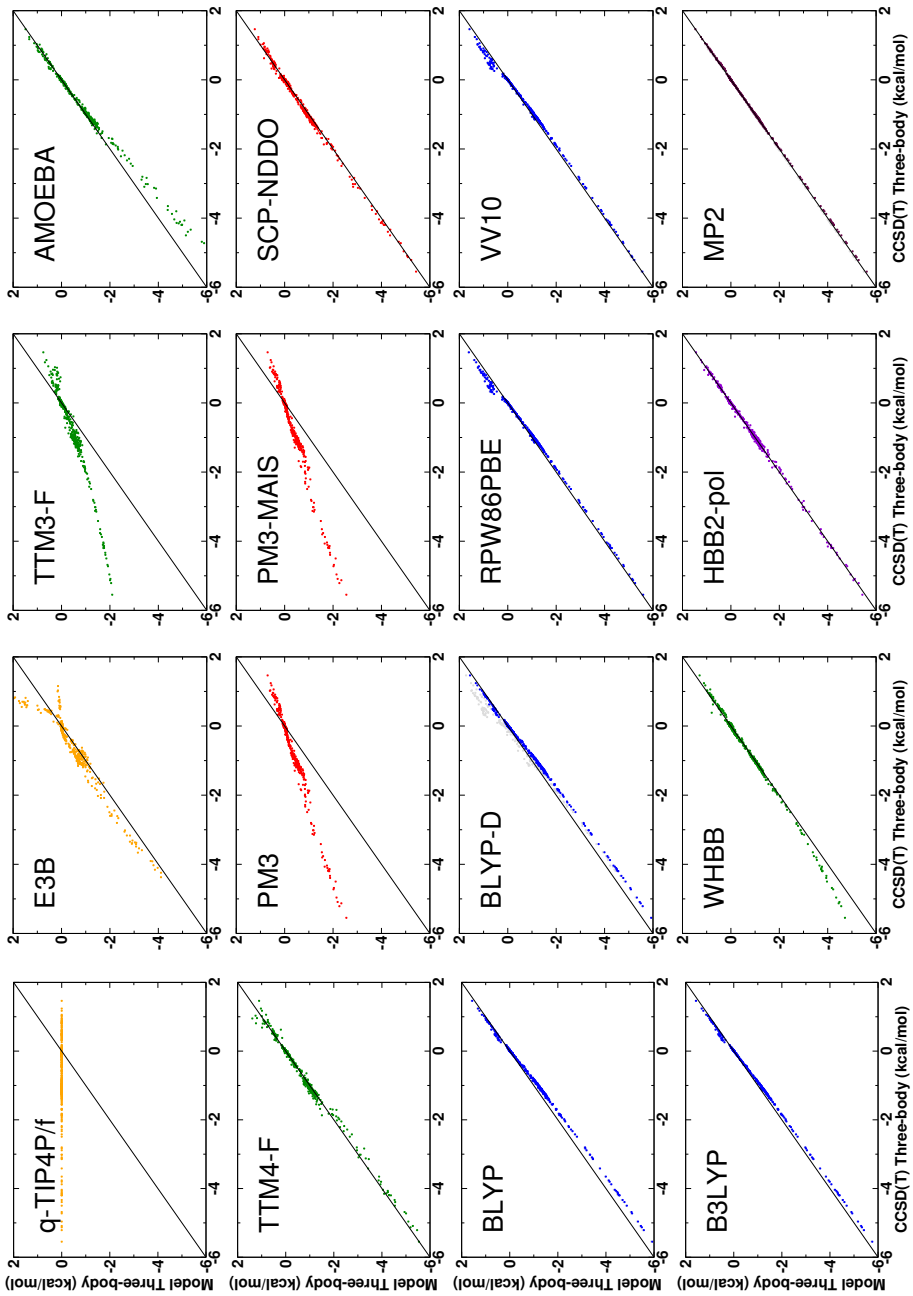

Figure 1.2: Comparison of model 3B interactions to CCSD(T). Plotted on the x-axes is the CCSD(T)/aug-cc-pVTZ energies corrected for BSSE. On the y-axes are the 3B energies for each model. The color scheme is the same as in Fig. 1.1.

By contrast, although AMOEBA takes into account intramolecular flexibility in the PES, the monomer charges are geometry independent.[60] It is also important to mention that, unlike in TTM3-F, the molecular polarizability of both AMOEBA and TTM4-F is anisotropic. It is unclear whether the inaccuracy observed in the TTM3-F 3B energies arises from its use of effective charges, isotropic molecular polarizability or both.

Among the semiempirical methods, PM3 was fitted to a wide range of experimental and *ab initio* data, while PM3-MAIS and SCP-NDDO were both fitted to *ab initio* reference data of water clusters. It is therefore not surprising that the 2B interactions of PM3-MAIS and SCP-NDDO are in better agreement with the CCSD(T) data than PM3. It is interesting, however, that SCP-NDDO shows much tighter correlation to the *ab initio* data than PM3-MAIS, even though the latter uses almost twice as many adjustable parameters as SCP-NDDO. While the two MNDO-type semiempirical methods display significant deficiencies in describing the 3B interactions, SCP-NDDO reproduces the CCSD(T) data quite accurately. These results suggest that the addition of classical polarization, as implemented in SCP-NDDO, can allow semiempirical methods to accurately describe intermolecular interactions without requiring extensive reparametrizations of the core-core terms.

At the 2B level, the GGA density functionals differ appreciably from the CCSD(T) results (see Supporting Information of Ref. 91 for PBE and PBE0 results), with BLYP systematically underestimating the interaction strength. The inclusion of the dispersion correction in BLYP-D improves the agreement with the CCSD(T) values for the 2B interactions; however, although DFT is less sensitive to basis-set incompleteness than wavefunction methods, the absence of diffuse functions in the BLYP-D basis (TZV2P) results in a large BSSE correction. This is seen in Fig. 1.1, where blue circles indicate the BSSE-corrected interactions and gray circles the BSSE-uncorrected interactions. Indeed, BSSE is so small for BLYP, B3LYP, RPW86PBE, and VV10 (using the aug-def2-TZVPP

basis) that it is barely visible in Figs. 1.1 and 1.2. While BLYP-D may be able to accurately describe the 2B interactions when a sufficiently large basis-set is used or when the energy values are corrected for BSSE, how to balance these factors in condensed phase simulations is not straightforward and is the subject of ongoing research.[71,92]

While the use of hybrid functionals, such as B3LYP,[93,94] results in a much tighter correlation to the CCSD(T) data than GGA functionals, B3LYP nonetheless inherently suffers from inadequate treatment of dispersion interactions, which leads to an incorrect long-range behavior.[95] Among recent nl-vdW functionals, VV10 appears to over-correct its parent functional, RPW86PBE, leading to over bound 2B interactions. All DFT methods perform reasonably well for the 3B interactions. It is important to note that, because the dispersion correction is pair additive, BLYP and BLYP-D provide identical 3B interactions. By contrast, nl-vdW functionals include a three-body dispersion correction, although this is almost negligible for VV10 (see Supporting Information of Ref. 91). MP2 agrees well with CCSD(T), with an RMS of 0.03 and 0.02 kcal/mol for the 2B and 3B interactions, respectively. Consistent with previous observations, the magnitude of BSSE is much smaller for 3B than 2B interactions.[96]

With the exception of MP2, WHBB provides the lowest RMS for the 2B interactions. WHBB employs a permutationally invariant polynomial with 5227 coefficients that were fit to reproduce $\sim$30000 CCSD(T)/aug-cc-pVTZ 2B interactions. To account for basis-set truncation, the reference 2B interactions were chosen as weighted averages of BSSE-corrected and BSSE-uncorrected CCSD(T) interactions.[97] Since both WHBB and CC-pol reproduce the VRT spectrum of the water dimer with comparable accuracy,[82] a similar agreement with the CCSD(T) data at the 2B level is also expected for CC-pol. The agreement of WHBB with the CCSD(T) values for the 3B interactions is less satisfactory, with WHBB increasingly underestimating the energies of the lowest-lying trimers.

1.4 Short-range many-body interactions address systematic flaws in polarizable models

Due to its rapid convergence for water, the many-body expansion of interaction energies provides a viable way to "scale up" the CCSD(T) level of accuracy to a large number of molecules. Furthermore, by accurately fitting the 1B, 2B, and 3B interactions into a relatively inexpensive function, simulations of condensed phases at an effective CCSD(T) level of accuracy become feasible. For flexible monomers, the most sophisticated effort along these lines, WHBB,[80] has indisputably proven this concept. However, WHBB is not directly applicable to bulk phase simulations due to its prohibitively expensive 3B term. Motivated by this observation, we introduced an exploratory model, termed HBB2-pol,[91,98] which sought to maintain the accuracy of WHBB while simultaneously decreasing the computational expense of the model.

As was mentioned in the previous section, WHBB obtains excellent accuracy with respect to correlated electronic structure calculations through the use of permutationally invariant polynomials.[99,100] These polynomials provide a flexible basis whose (linear) coefficients can be fitted to large datasets of electronic structure reference data. However, to capture the complexity of the many-body interactions, high degree polynomials are typically required. In the case of WHBB, the fifth degree polynomials (containing 1380 terms) entering into the 3B PES are the computational bottleneck of the model. In WHBB, the 3B interaction has the form

$$V^{3B}_{\text{WHBB}} = sV^{\text{Poly}} + (1-s)V^{\text{Ind}}_{\text{TTM3-F}}, \tag{1.8}$$

where V^{Poly} is the 3B permutationally invariant polynomial and $V^{\text{Ind}}_{\text{TTM3-F}}$ is the induced 3B interaction energy of the TTM3-F polarizable model. The contributions to the 3B

energy are range separated; when the three molecules are close together, the interaction is given by V^{Poly}. At longer ranges where at least one molecule is far from the other two, the interaction is described through the TTM3-F induced energy. The range of the interactions is controlled by the switching parameter, s, which smoothly transitions from 1 to 0 as the molecules are pulled apart from one another.

Our development builds upon this range separation by exploiting the fact that the 3B interaction in water arises primarily from induction, with all other contributions vanishing quickly as the intermolecular separation increases.[96,101] This naturally leads to the following ansatz,

$$V^{3B}_{\text{HBB2-pol}} = s_3 V^{\text{Poly}} + V^{\text{Ind}}_{\text{TTM4-F}}, \tag{1.9}$$

which, while very similar in form to Eq. 1.8, contains a few key differences. First, rather than the TTM3-F model, the significantly more accurate TTM4-F model is used to describe the induced interactions (see Fig. 1.2). More importantly, the TTM4-F induction is employed at all ranges of the 3B interaction. The permutationally invariant polynomial V^{Poly} thus represents a short-range "correction" to the underlying induction energy. The physical origins of this short-ranged "correction" lie in both the breakdown of assumptions made in the derivation of the Thole-type induction term as well as the quantum-mechanical contributions of 3B exchange-repulsion and charge transfer (see Fig. 1.3).[96,101,102]

Because the ability of the polynomial to accurately fit reference data depends largely on its degree (and, therefore, its computational cost), the different representation of the 3B interactions in the HBB2-pol model along with the improved description of the induction energies in HBB2-pol allows for an accurate fit of the CCSD(T) reference data with a lower-degree polynomial. Specifically, in HBB2-pol the V^{Poly} part is a sum

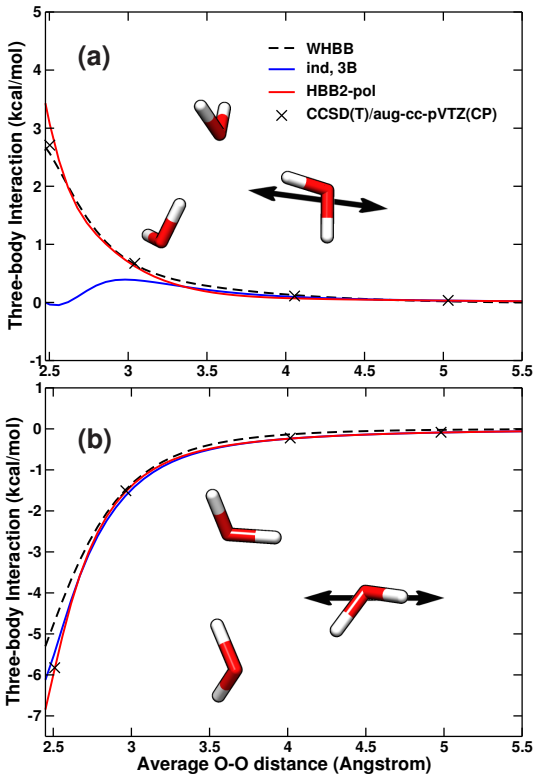

Figure 1.3: Three-body interaction at short- and long-ranges. The three-body interaction energy is plotted for two cuts through the water trimer potential energy surface: WHBB (black dashes), three-body induction (blue), HBB2-pol (red) and CCSD(T)/aug-cc-pVTZ (crosses).

of second and third degree symmetrized products of exponentials of the intermolecular separations

$$\eta_{ij} = \exp\left(-kr_{ij}\right). \tag{1.10}$$

where k is an adjustable parameter and r_{ij} is the distance between atoms i and j. Neither intramolecular distances nor the purely two-body terms – those which do not depend on the positions of all three molecules simultaneously – were included into the V^{Poly} in HBB2-pol. This results in a polynomial with 131 terms, whose linear coefficients were fit using the least squares to a dataset of 8019 trimer configurations for which the 3B energies were computed at the CCSD(T)/aug-cc-pVTZ level and corrected for BSSE.[103]

As discussed above, while the three-body interactions primarily originate from induction, for more strongly interacting clusters, effects including exchange-repulsion and charge transfer can also make a significant contribution.[96,101] As a consequence, force fields which treat only induction are inherently unable to fully describe 3B interactions. By contrast, models that only treat short-range 3B interactions are unable to describe the induction interactions that dominate at long range. This is illustrated in Fig. 1.3, where HBB2-pol is compared with WHBB, TTM4-F 3B induction, and CCSD(T)/aug-cc-pVTZ data along two representative cuts through the water trimer PES. Importantly, the CCSD(T) reference data used for this comparison were not included in the training set. This comparison clearly shows that the addition of the short-range "correction" to the induction brings the 3B interactions of HBB2-pol into close agreement with the CCSD(T) data.

1.5 Summary

In this study, the accuracy of several force fields, semiempirical methods, density functionals, and "first principles"-based potential energy surfaces was assessed against

BSSE-corrected CCSD(T)/aug-cc-pVTZ data. Our analysis of the many-body expansion of the interaction energy indicates that defects inherent to polarizable models, which are often non-negligible when molecules are close to one another, can be effectively corrected through an explicit short-range term expressed in terms of permutationally invariant polynomials. Based on these findings, we developed an exploratory model, HBB2-pol, that was derived entirely from "first principles". HBB2-pol achieves excellent accuracy with respect to CCSD(T) data. Importantly, the inclusion of explicit polarization in the three-body interaction term enables the use of relatively low-degree polynomials, which, in turn, results in a significant decrease in the computational cost associated with HBB2-pol relative to other *ab initio*-based models. This improved efficiency enabled us to perform quantum simulations of bulk water under ambient conditions.[98]

The successful development of the HBB2-pol PES represented a crucial proof of concept, where we demonstrated that a potential energy surface parametrized exclusively to correlated electronic structure reference data was capable of accurately predicting experimentally determined properties of water from small clusters to the bulk liquid without any parametrization to experiment.[98] To the best of our knowledge, this was the first time this had been done for flexible models of water; while the (flexible) WHBB many-body PES had been shown to accurately describe cluster properties, its computational expense prohibited the application of WHBB to simulations of the bulk liquid.

This dissertation focuses on the application of the many-body expansion of interactions to the calculation of vibrational spectra of condensed phase systems from "first principles." In Chapter 2, a refined version of the HBB2-pol PES, titled "MB-pol" will be presented in detail. In Chapter 3, MB-pol is combined with path integral-based quantum simulation techniques to calculate thermodynamic and dynamical properties of liquid water. A detailed analysis of the convergence of the many-body of expansion of the dipole moment and polarizability is presented in Chapter 4. Building upon that work,

a Many-Body Molecular Dynamics (MB-MD) framework for calculating vibrational spectra of condensed phase systems is developed in Chapter 5; specifically many-body models for the dipole moment and polarizability are introduced and used to calculate the infrared and Raman vibrational spectra of liquid water. This MB-MD partitioning is then used in Chapter 6 to investigate the physical origins of the IR spectrum of liquid water. This dissertation is concluded in Chapter 7 by employing the MB-MD approach to the calculation of the vibrational Sum Frequency Generation spectrum of the air/water interface and reflecting on future directions for the application of many-body expansion to the study of molecular systems.

Portions of this chapter appeared in our previously published papers, Medders, G.R.; Babin, V.B.; Paesani, F. "A critical assessment of two-body and three-body interactions in water", *J. Chem. Theory Comput.* **2013** 9, 1103–1114 and Medders, G.R.; Paesani, F. "Infrared and Raman spectroscopy of liquid water through first principles many-body molecular dynamics", *J. Chem. Theory Comput.* **2015** 11, 1145–1154. This material was reproduced with permission from the publisher.

Chapter 2

The MB-pol Potential Energy Surface

2.1 Introduction

Despite recent progress in the development of electronic structure approaches to modeling molecular interactions in increasingly larger systems, chemical accuracy remains out of reach for most applications to condensed phase systems.[63,64,104–111] In principle, accurate multidimensional potential energy surfaces (PESs) can be obtained at the coupled cluster level of theory including single, double, and perturbative triple excitations, CCSD(T), which represents the current "gold standard" of chemical accuracy. Unfortunately, the computational cost associated with CCSD(T) makes these calculations prohibitively expensive even for small molecular clusters.[31] To overcome this computational barrier while still providing an *ab initio* representation of the underlying PES, different models based on density functional theory (DFT) have been developed. However, the choice of the most appropriate functional for studying (weak) intermolecular interactions, and, particularly, hydrogen bonds in condensed phases remains the subject of ongoing research.[74,110,112,113]

As discussed in Section 1.2 the many-body expansion of interactions[114] provides

a convenient framework for the analysis and development of multidimensional PESs. Due to its rapid convergence for water and because the low-order interaction terms can be accurately calculated at the CCSD(T) level, the many-body expansion effectively enables the development of water PESs with chemical accuracy.[79,80,91,98,108]

This chapter describes the development of the many-body MB-pol water potential whose 2-body term was introduced in Ref. 115. The MB-pol potential, which represents an iterative improvement over the HBB2-pol model described in Refs. 98 and 91, takes advantage of a Thole-type model (TTM) for polarization[57] along with a pair-additive approximation of the dispersion energy to represent the intermolecular interactions in regions where the overlap between the electron densities of the individual water molecules is negligible. The complexity of the interactions at short range is additionally modeled via permutationally invariant polynomials[99,100] in quickly decaying functions of the interatomic distances. The coefficients of these polynomials are fitted to CCSD(T) reference energies. The functional form of the MB-pol potential is thus given by the following expression

$$E_N(x_1,\ldots,x_N) = \sum_i V^{1B}(i) + \sum_{i<j} V^{2B}_{\text{short}}(i,j) + \sum_{i<j<k} V^{3B}_{\text{short}}(i,j,k) + V_{\text{TTM}} + V_{\text{disp}}. \quad (2.1)$$

Here, $V^{(1\text{B})}$ is the 1-body term associated with intramolecular distortion described by the spectroscopically accurate monomer PES developed by Partridge and Schwenke,[116] V_{TTM} represents the (classical) electrostatic contributions described in detail in Ref. 115, V_{disp} represents the dispersion energy, and V^{2B}_{short} and V^{3B}_{short} are the short-range two-body and three-body terms. Both V^{2B}_{short} and V_{disp} terms were developed and validated in Ref. 115. During the development of the complete MB-pol potential, slight modifications of V^{2B}_{short} and V_{disp} were required to guarantee full integration with the three-body term, V^{3B}_{short},

which is the primary subject of this study. The article is organized as follows: Section 2.2 describes the CCSD(T) electronic structure calculations involved in the development of a representative training set of water trimers. The functional form of the MB-pol $V_{\text{short}}^{(3B)}$ term is then described in Section 2.3 along with details related to the fitting procedure. The electronic structure reference energies are used in Section 2.4 to analyze the accuracy of the three-body energies predicted by the iAMOEBA-rev5,[117] WHBB,[80] HBB2-pol,[98] and, when possible, (rigid) CCpol[81] potentials. The accuracy of MB-pol is then further assessed through comparison of the calculated third virial coefficient and relative energies of small water clusters with the corresponding (experimental and *ab initio*) data available in the literature.

2.2 Training set and reference energies

The trimer reference energies were calculated at the CCSD(T)[83] level using the augmented correlation-consistent polarized-valence triple (aug-cc-pVTZ) basis set[84,118] supplemented by an additional set of (3s,3p,2d,1f) midbond functions[119] with exponents equal to (0.9, 0.3, 0.1) for s and p orbitals, (0.6, 0.2) for d, and 0.3 for f, placed at the center of mass of each trimer configuration. The counterpoise method[103] was applied to remove the basis set superposition error (BSSE); specifically, the "site-site" variation of the counterpoise method was used, with the energies of each trimer and its constituent fragments (dimers and monomers) computed in the trimer basis.[120] Since the calculated three-body interaction energies were found to be very close to the complete basis set (CBS) limit, no explicit extrapolation was performed. All CCSD(T) calculations were carried out with Molpro[121] without assuming any symmetry constraints on the trimer structures.

The short-range three-body training set contains 12347 trimers extracted from: (a)

the low-energy subset of the HBB2-pol training set described in Ref. 91; (b) path-integral molecular dynamics (PIMD) simulations of liquid water carried out at ambient conditions (T = 298.15 K and ρ = 0.997 g/cm^3 with the HBB2-pol potential;[98] (c) PIMD/HBB2-pol simulations of small $(H_2O)_N$ clusters with $N \leq 6$ carried out in the temperature range between 30 K and 100 K;[32] and (d) constant pressure - constant temperature (NPT) molecular dynamics simulations of liquid water carried out at ambient conditions with intermediate versions of the present MB-pol potential. A large ensemble of configurations from the (b) and (c) sources described above was first formed, containing approximately an order of magnitude more trimers than were actually used in the final fitting process. These initial configurations were then filtered to remove redundant configurations, and only dissimilar geometries as determined from a root-mean squared deviation (RMSD) criterion were kept in the final ensemble.

Ideally, the perfect training set should comprise trimer configurations whose internal coordinates are defined on a multidimensional grid that uniformly covers all relevant regions of the configurational space. Unfortunately, given the dimensionality of the problem, this procedure cannot be performed for trimer configurations with flexible monomers. However, we expect that the structures extracted from PIMD trajectories and within neighborhoods of the stationary points of the trimer PES are sufficiently representative of molecular configurations with energies in the range relevant to molecular simulations of water under conditions of moderate temperature and pressure. In addition, within the protocol described above, the geometric uniformity of the training set is approximately achieved via the filtering step. The trimer geometries included in the final training set are reported in the Supporting Information along with the corresponding CCSD(T)/aug-cc-pVTZ reference energies.

2.3 Description of the MB-pol potential

The rationale behind the MB-pol functional form has roots in the observation that, due to the different physical character of the interactions between water molecules at different separations, the intermolecular potential energy can be effectively split into short-range and long-range contributions.[114] The latter arise from electrostatic interactions between permanent and induced moments of the molecular charge distributions as well as from dispersion forces. The MB-pol representation of these (long-range) contributions to the total interaction energy is described in Ref. 115. Within the many-body expansion of the interaction energy, the short-range contributions can be introduced systematically at the n-body levels ($n \geq 2$) to account for the complexity of the multidimensional potential energy surface in regions where the monomer electron densities display significant overlap. The specific functional form of the MB-pol potential at small intermolecular separations is derived by noting that, for water systems, short-range corrections to the induction energy beyond the three-body level are not necessary for achieving chemical accuracy (see Section 2.4.3).

The two-body short-range term of the MB-pol potential is described in detail in Ref. 115, with some minor modifications to further improve both accuracy and performance of the full many-body MB-pol potential in applications to water systems larger than the dimer. Specifically, three changes were adopted: (a) all $1/r^8$ terms have been removed from the expression of the dispersion energy; (b) for consistency with the three-body term, the switching function [eq (8) in Ref. 115] has been replaced by the following expression

$$s_2(x) = \begin{cases} 1 & \text{if } x < 0 \\ \cos^2 \frac{\pi}{2} x & \text{if } 0 \leq x < 1 \\ 0 & \text{if } 1 \leq x \end{cases} ; \tag{2.2}$$

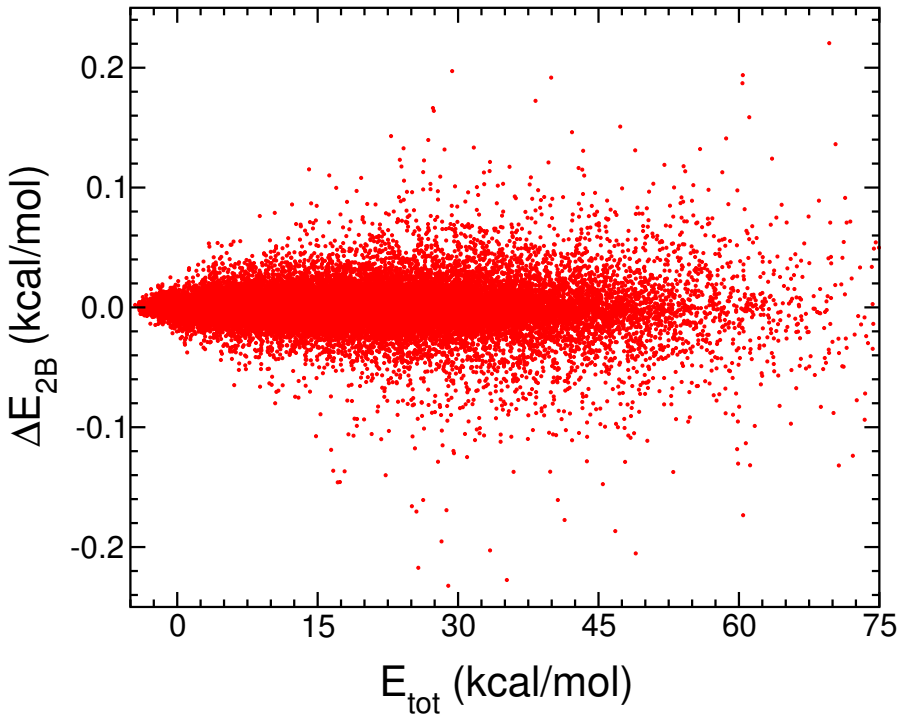

Figure 2.1: Two-body energies of MB-pol. The differences between two-body interaction energies calculated with the current and original[115] MB-pol two-body terms are plotted as a function of the dimer total energy. The RMSD over the lower 100 kcal/mol is 0.017 kcal/mol.

and (c) the corresponding cutoff distances have been reduced to $R_{i,o} = 4.5/6.5$ Å. Both long- and short- range parameters have been determined as described in Ref. 115 using the same reference energies. The new short-range fit is marginally more accurate than the original one despite the reduced cutoffs. The differences between the new and old two-body interaction energies are shown in Fig. 2.1 as a function of the dimer total energy for configurations taken from the short-range training set.

Similarly to HBB2-pol,[91] the three-body short-range term of the MB-pol potential is expressed as

$$V^{(3B)}(x_a,x_b,x_c) = V^{(3B)}_{\text{short}}(x_a,x_b,x_c) + V^{(3B)}_{\text{TTM,ind}}(x_a,x_b,x_c), \tag{2.3}$$

where $V^{(3B)}_{\text{TTM,ind}}$ is the TTM 3-body induction energy, and $V^{(3B)}_{\text{short}}$ is given by

$$V^{(3B)}_{\text{short}}(x_a,x_b,x_c) = \Big[s(t_{ab})s(t_{ac}) + s(t_{ab})s(t_{bc}) + s(t_{ac})s(t_{bc})\Big] V^{(3B)}_{\text{poly}}(x_a,x_b,x_c), \tag{2.4}$$

where the sum of the three terms in the square bracket represents a switching function that smoothly goes to zero as one of the water molecules moves apart from the other two. In Eq. 2.4,

$$s(t) = \begin{cases} 1 & \text{if } t < 0 \\ \cos^2 \frac{\pi}{2} t & \text{if } 0 \le t < 1 \\ 0 & \text{if } 1 \le t \end{cases}, \tag{2.5}$$

and

$$t_{mn} = \frac{R_{mn}}{R^{(3B)}_{cut}}, \tag{2.6}$$

where R_{mn} is the distance between the oxygen atoms of water molecules m and n, and $R^{(3B)}_{cut} = 4.5$ Å is the three-body cutoff distance. This specific value for $R^{(3B)}_{cut}$ was found to provide the optimal compromise between accuracy and computational cost.

Table 2.1: Interatomic distances used in the MB-pol three-body polynomial. The first letter of the site label denotes the atom name (O, H), the second letter distinguishes the molecules (a, b or c), the trailing digit indexes equivalent atoms within the molecule.

d_1	Ha1	Ha2
d_2	Hb1	Hb2
d_3	Hc1	Hc2
d_4	Oa	Ha1
d_5	Oa	Ha2
d_6	Ob	Hb1
d_7	Ob	Hb2
d_8	Oc	Hc1
d_9	Oc	Hc2
d_{10}	Ha1	Hb1
d_{11}	Ha1	Hb2
d_{12}	Ha1	Hc1
d_{13}	Ha1	Hc2
d_{14}	Ha2	Hb1
d_{15}	Ha2	Hb2
d_{16}	Ha2	Hc1
d_{17}	Ha2	Hc2
d_{18}	Hb1	Hc1
d_{19}	Hb1	Hc2
d_{20}	Hb2	Hc1
d_{21}	Hb2	Hc2
d_{22}	Oa	Hb1
d_{23}	Oa	Hb2
d_{24}	Oa	Hc1
d_{25}	Oa	Hc2
d_{26}	Ob	Ha1
d_{27}	Ob	Ha2
d_{28}	Ob	Hc1
d_{29}	Ob	Hc2
d_{30}	Oc	Ha1
d_{31}	Oc	Ha2
d_{32}	Oc	Hb1
d_{33}	Oc	Hb2
d_{34}	Oa	Ob
d_{35}	Oa	Oc
d_{36}	Ob	Oc

The $V_{\text{poly}}^{(3\text{B})}$ term in Eq. 2.4 is a permutationally invariant polynomial in exponentials of the interatomic distances, $d_{m=1-36}$, between all possible pairs of atoms as defined in Table 2.1. From the definition of $d_{m=1-36}$, the following variables are formed

$$\begin{aligned}
\xi_1 &= \mathrm{e}^{-k_{\text{HH,intra}}(d_1 - d^{(0)}_{\text{HH,intra}})}, \\
&\cdots \\
\xi_9 &= \mathrm{e}^{-k_{\text{OH,intra}}(d_9 - d^{(0)}_{\text{OH,intra}})}, \\
\xi_{10} &= \mathrm{e}^{-k_{\text{HH}}(d_{10} - d^{(0)}_{\text{HH}})}, \\
&\cdots \\
\xi_{36} &= \mathrm{e}^{-k_{\text{OO}}(d_{36} - d^{(0)}_{\text{OO}})}.
\end{aligned}$$

$V_{poly}^{(3\text{B})}$ is then constructed as a polynomial in ξ_i, imposing the permutational invariance with respect to permutations of the water molecules as well as to permutations of the hydrogen atoms within each molecule. $V_{poly}^{(3\text{B})}$ contains the following symmetrized monomials: (a) 13 second-degree monomials formed from all intermolecular ($\xi_{10}, \ldots, \xi_{36}$) variables; (b) 202 third-degree symmetrized monomials with at most linear intramolecular terms; and (c) 948 symmetrized fourth-degree monomials with at most linear dependence on the intramolecular variables, as well as intermolecular variables involving oxygen-oxygen and hydrogen-hydrogen distances. The complete list of all 1163 symmetrized monomials is available in the Supporting Information of Ref. 122. The coefficients ($c_{l=1-1163}$) of these terms are linear fitting parameters. In addition, $V_{poly}^{(3\text{B})}$ also contains the following 10 nonlinear parameters: $d^{(0)}_{\text{HH,intra}}$, $d^{(0)}_{\text{OH,intra}}$, $d^{(0)}_{\text{HH}}$, $d^{(0)}_{\text{OH}}$, $d^{(0)}_{\text{OO}}$, $k_{\text{HH,intra}}$, $k_{\text{OH,intra}}$, k_{HH}, k_{OH}, and k_{OO}.

Both linear and nonlinear parameters were obtained by minimizing the (regularized) weighted sum of squared residuals calculated for the short-range training set $\mathcal{S}$

described above:

$$\chi^2 = \sum_{n \in \mathcal{S}} w_n \left[V_{\text{model}}^{(3B)}(n) - V_{\text{ref}}^{(3B)}(n) \right]^2 + \Gamma^2 \sum_{l=1}^{1163} c_l^2. \tag{2.7}$$

The weights, w_n, were set to emphasize trimers with low binding energies

$$w(E) = \left(\frac{\Delta E}{E - E_{\text{min}} + \Delta E} \right)^2. \tag{2.8}$$

Here, E_{min} denotes the lowest energy in the training set (i.e., the trimer global minimum energy) and ΔE defines the range of the favorably weighted energies. ΔE = 37.5 kcal/mol was used in Eq. 2.8, which is consistent with the value used in the fit of the two-body term (25 kcal/mol).[115] The regularization parameter, Γ, was set to 1×10^{-4} in order to reduce the variation of the linear fitting parameters (larger Γ values effectively suppress any variation of the parameters) without spoiling the overall accuracy of the fit (favored by smaller Γ values). The choice of the regularization weight was further constrained by the requirement that the regularization term contributed to no more than 1 % to χ^2. The linear parameters ($c_{l=1,\ldots,1163}$) were obtained through singular value decomposition, while the simplex algorithm was used to iteratively optimize the nonlinear parameters. The optimization procedure results in a χ^2 value of 7.83 (kcal/mol)^2, corresponding to a RMSD of 0.028 kcal/mol per trimer. The RMSD for trimers with total energy within 37.5 kcal/mol of the trimer minimum energy is 0.00068 kcal/mol per trimer and the largest deviation for the same trimers is 0.16 kcal/mol. All optimized parameters are included in the reference implementation of the MB-pol potential available in the Supporting Information of Ref. 122.

2.4 Assessment of the MB-pol potential

The CCSD(T) reference energies described in Section 2.2 are used in Fig. 2.2 to investigate the accuracy of MB-pol three-body energies compared to the corresponding results obtained with the iAMOEBA-rev5,[117] WHBB,[80] and HBB2-pol[91,98] potentials, all of which have recently been shown to correctly reproduce some of the water properties under different conditions. The CC-pol model is omitted from this comparison since a flexible version of it has only been developed at the two-body level,[82] which prevents a direct comparison with the present CCSD(T) data calculated for flexible trimers. Simpler water potentials, parameterized to experimental and/or to relatively small sets of *ab initio* data (e.g., q-TIP4P/F,[49] qSPC/Fw,[50] TTM3-F,[58] and E3B[52]), are also not considered in the present study because it was previously shown that these models are unable to correctly reproduce both two-body and three-body water energies.[98]

Among the four water potentials examined in Fig. 2.2, MB-pol provides the closest agreement with the CCSD(T) reference data. Noticeable differences exist between iAMOEBA-rev5 (RMSD = 0.313 kcal/mol) and *ab initio* results, especially for three-body energies below -2 kcal mol^{-1}. Although the HBB2-pol potential (RMSD = 0.158 kcal/mol) appears to perform slightly better than WHBB (RMSD = 0.104 kcal/mol) for low-energy trimers, some important deficiencies are identified, which, based on a preliminary analysis of the liquid properties, can be attributed to trimer geometries that are associated with the low-density phase of water. For these configurations, which correspond to the "tail" deviating from the diagonal in the upper-right corner of Fig. 2.2c, HBB2-pol predicts considerably more repulsive three-body energies than CCSD(T). The energetics of these structures is analyzed in more detail in Fig. 2.3.

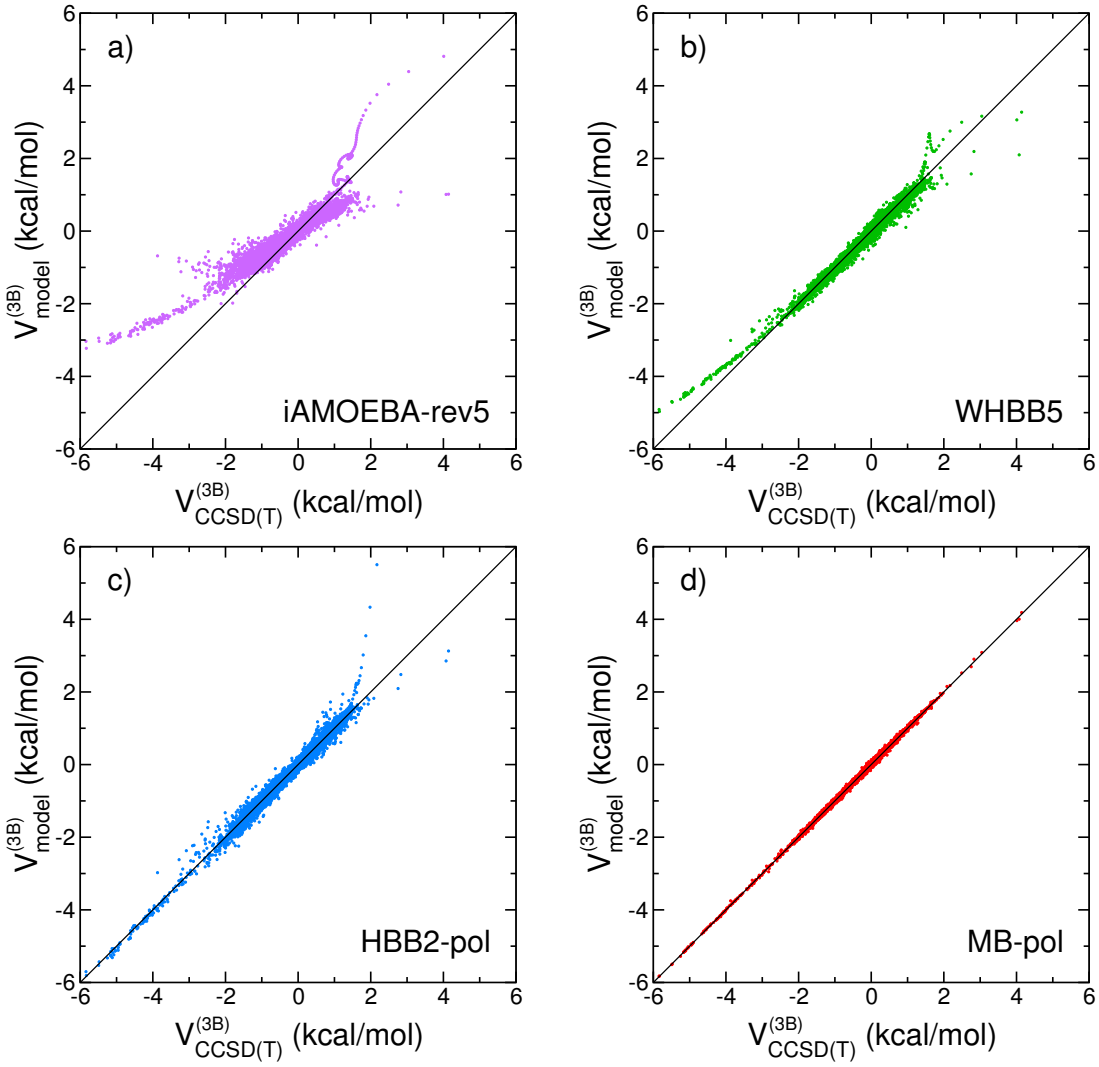

Figure 2.2: Accuracy of different models in describing the MB-pol training set. Plotted on the x-axis are the BSSE-corrected CCSD/aug-cc-pVTZ+mb values while on the y-axis are the corresponding values calculated with: a) iAMOEBA-rev5, b) WHBB5, c) HBB2-pol, and d) MB-pol. See main text for details.

Table 2.2: Binding energies of stationary points on the *ab initio* water PES. Specifically, MB-pol binding energies (E), gradient norms ($|\nabla E|_2$), and relative energies (ΔE) were calculated for the reference stationary points of the water trimer PES reported in Ref. 123. Also listed are the corresponding *ab initio* relative energies (ΔE_{ref}).[123] The RMSD of the MB-pol ΔE values relative to ΔE_{ref} is 10.68 cm^{-1}.

No.	E (cm^{-1})	$\|\nabla E\|_2$ (cm^{-1}/Å)	ΔE (cm^{-1})	ΔE_{ref} (cm^{-1})
1	-5460.57	1.07e+03	0.00	0.00
2	-5210.91	1.12e+03	249.67	268.34
3	-5016.46	1.08e+03	444.12	436.36
4	-5378.82	1.05e+03	81.75	81.09
5	-5201.36	1.09e+03	259.22	274.19
6	-4701.06	1.10e+03	759.51	752.34

2.4.1 Stationary points

The binding energies and corresponding gradients calculated at the six stationary points of the *ab initio* water trimer PES reported in Ref. 123 are listed in Table 2.2. The same quantities for trimer geometries optimized on the MB-pol PES are given in Table 2.3 along with the RMSD values calculated relative to the *ab initio* configurations from Ref. 123. Both sets of results demonstrate the high accuracy of the MB-pol trimer energies. Importantly, the MB-pol gradients associated with the reference geometries are relatively small, providing further evidence for the closeness between the the MB-pol and *ab initio* PESs. The remaining disagreement between the MB-pol and *ab initio* PESs can be attributed, at least partially, to the differences in the corresponding one-body energies. As described in Ref. 115, the monomer PES employed in MB-pol includes corrections aimed at representing mass-polarization and relativistic effects,[116] which are absent in the CCSD(T) calculations reported in Ref. 123.

In an attempt to eliminate the "bias" associated with the the different treatment of the one-body term, the comparison between MB-pol and CCSD(T) trimer interaction energies is repeated including only two- and three-body contributions $[E_3(x_1,x_2,x_3) - E_1(x_1) - E_1(x_2) - E_1(x_3)]$. For this purpose, the CCSD(T) reference values were obtained

Table 2.3: Binding energies of stationary points on the MB-pol water PES. Specifically, MB-pol binding energies (E), gradient norms ($|\nabla E|_2$), RMSD with respect to the reference *ab initio* geometries[123] and relative energies (ΔE) were calculated for the stationary points of the MB-pol water trimer PES. Also listed are the reference *ab initio* relative energies (ΔE_{ref}) for the reference stationary points of the water trimer PES reported in Ref. 123. The RMSD of the MB-pol ΔE values relative to ΔE_{ref} is 9.35 cm^{-1}.

No.	E (cm^{-1})	$\|\nabla E\|_2$ (cm^{-1}/Å)	RMSD (Å)	ΔE (cm^{-1})	ΔE_{ref} (cm^{-1})
1	-5488.26	4.67e-05	8.96e-03	0.00	0.00
2	-5239.88	5.16e-05	8.73e-03	248.39	268.34
3	-5040.52	9.14e-05	5.49e-03	447.75	436.36
4	-5405.15	9.75e-05	8.71e-03	83.11	81.09
5	-5228.84	5.01e-05	9.01e-03	259.43	274.19
6	-4729.79	8.29e-05	9.35e-03	758.47	752.34

in the complete basis set limit using a two-point extrapolation of the CCSD(T) energies calculated with aug-cc-pVTZ and aug-cc-pVQZ basis sets supplemented by either one midbond function placed at the center of mass of each trimer (Table 2.4) or four midbond functions placed at the centers of mass of the trimer and all constituent dimers (Table 2.5). After removal of the differences associated with the one-body term, the MB-pol energies for the six reference trimers lie at most within 50 cm^{-1} of the corresponding CCSD(T) energies. The agreement of MB-pol with the CCSD(T)/CBS energies is improved when larger basis sets are used in the CBS extrapolation; the RMS of the MB-pol energies with respect to the CCSD(T)/CBS energies for the trimer stationary points decreases from 42.59 cm^{-1} to 21.92 cm^{-1} when midbond functions are added to the centers of mass of both the trimer and its three constituent dimers. It is also interesting to note that the CCSD(T)/aug-cc-pVQZ+mb energy computed for the optimized trimer structure obtained on the MB-pol PES is lower than the energy associated with the optimized structure of Ref. 123.

The accuracy of the potential is further illustrated in Fig. 2.4 showing the comparison between the MB-pol and *ab initio* three-body interaction energies for a representative

Table 2.4: Trimer interaction energies with one set of midbond functions. Interaction energies ($E_3(x_1,x_2,x_3) - E_1(x_1) - E_1(x_2) - E_1(x_3)$, in cm^{-1}) are listed for the reference trimer geometries (Ref. 123). Here, "mb" corresponds to one set of midbond functions placed at the center of mass of the trimer. The RMSD of the MB-pol energies with respect to the CCSD(T)/CBS results is 42.59 cm^{-1}.

No.	CCSD(T)/aVTZ+mb (cm^{-1})	CCSD(T)/aVQZ+mb (cm^{-1})	CCSD(T)/CBS (cm^{-1})	MB-pol (cm^{-1})	Δ (cm^{-1})
1	-5436	-5596	-5713	-5662	51
2	-5167	-5322	-5435	-5402	33
3	-4965	-5113	-5222	-5185	37
4	-5349	-5505	-5620	-5574	46
5	-5158	-5312	-5425	-5390	35
6	-4715	-4841	-4933	-4883	50

Table 2.5: Trimer interaction energies with four sets of midbond functions. Interaction energies ($E_3(x_1,x_2,x_3) - E_1(x_1) - E_1(x_2) - E_1(x_3)$, in cm^{-1}) are listed for the reference trimer geometries (Ref. 123). Here, "mb^4" corresponds to four sets of midbond functions placed at the centers of mass of the trimer *a*nd its three constituent dimers. The RMSD of the MB-pol energies with respect to the CCSD(T)/CBS results is 21.92 cm^{-1}.

No.	CCSD(T)/aVTZ+mb^4 (cm^{-1})	CCSD(T)/aVQZ+mb^4 (cm^{-1})	CCSD(T)/CBS (cm^{-1})	MB-pol (cm^{-1})	Δ (cm^{-1})
1	-5514	-5608	-5677	-5662	15
2	-5259	-5349	-5415	-5402	13
3	-5073	-5159	-5221	-5185	36
4	-5431	-5523	-5591	-5574	17
5	-5251	-5340	-5405	-5390	15
6	-4724	-4801	-4857	-4883	-26

scan through the water trimer PES. The scan shown in Fig. 2.3 was investigated because it involves trimer configurations resembling those extracted from simulations of liquid water, which were found to be critical for reproducing the temperature dependence of the density at ambient pressure. Both coordinates and energies of the trimer configurations used in the scan of Fig. 2.3 are available in the Supporting Information. In the scan calculations of Fig. 2.3, all three water molecules were kept in the vibrationally-averaged geometry ($r_{\mathrm{OH}} = 1.836106337$ Bohr and $\vartheta_{\mathrm{HOH}} = 104.69^{\circ}$ as defined in Ref. 79) so that the CCpol[81] potential could also be included in the comparison. Also shown in Fig. 2.3 is a modified version (TTM4x-F) of the original TTM4-F polarizable water model,[59] which employs the same many-body polarization scheme as MB-pol.[115] In all trimer geometries used in the scan, the "central" molecule is placed on the XY plane with its oxygen atom at the origin and the HOH bisector lie along the Y axis; the "left" and "right" water molecules lie on the XZ plane with their oxygen atoms located on the Z axis at a distance of 2.7 Å from the central water oxygen. Both "left" and "right" molecules have one hydrogen atom on the Z axis so that the corresponding OH bond is oriented toward the central oxygen atom. The potential energy scan was performed by rotating the "left" and "right" molecules about the X axis by $180 - \vartheta/2$ where ϑ is the angle between the oxygen atoms of the "left", "central", and "right" water molecules. The results shown in Fig. 2.4 indicate that both TTM4x-F and HBB2-pol predict more repulsive three-body energies, while CC-pol underestimates the *ab initio* values, predicting significantly more attractive three-body energies for wider O-O-O angles. Interestingly, iAMOEBA-rev5 correctly reproduces the angular dependence of the three-body energy along the entire scan, although it underestimates the magnitude of this contribution. Good agreement with the reference CCSD(T) data is provided by both WHBB and MB-pol, which is effectively quantitative for MB-pol over the entire scan.

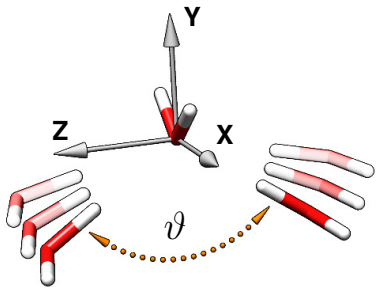

Figure 2.3: Trimer configurations used in Fig. 2.4.

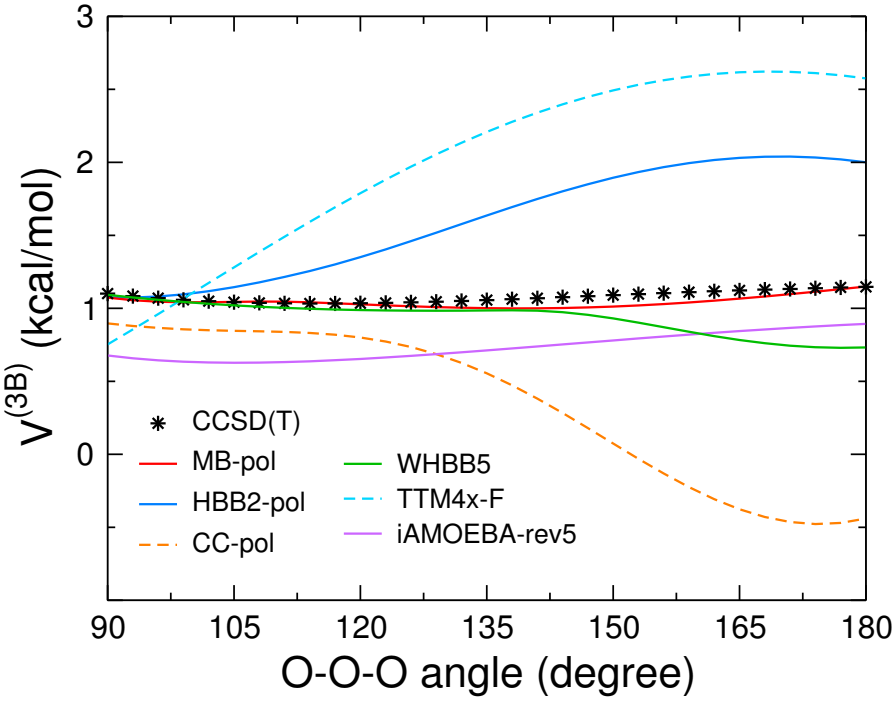

Figure 2.4: Three-body interaction energies for a trimer scan. The reference CCSD(T)/aug-cc-pVTZ values (black stars) are shown along with the MB-pol (red), HBB2-pol (blue), CC-pol (dashed orange), WHBB5 (green), TTM4x-F (dashed light blue), and iAMOEBA (violet) results.

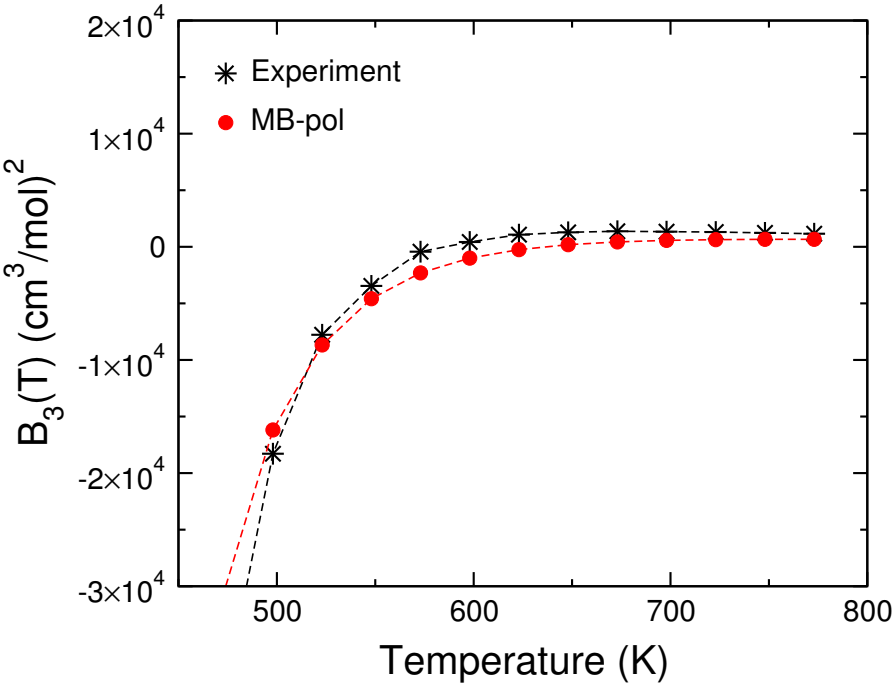

Figure 2.5: Third virial coefficient of MB-pol vs. experiment. The experiment is taken from Ref. 125.

2.4.2 Third virial coefficient

By accounting for deviations from ideal gas behavior, [124, Chapter 15] the third virial coefficient provides an indirect measure of the three-body interactions. To assess the overall accuracy of the MB-pol three-body term, the third virial coefficient was calculated at the classical level with each water monomer held fixed at the ground-state vibrationally averaged geometry ($r_{\mathrm{OH}} = 1.836106337$ Bohr and $\vartheta_{\mathrm{HOH}} = 104.69°$).[79] This classical approach is justified by the fact that the experimental data for the third virial coefficient are only available at relatively high temperatures, where quantum effects are less important. All calculations were carried out using the scheme summarized in Ref. 91. The close agreement between the MB-pol results and the available experimental data[125] shown in Fig. 2.5 provides further support for the accuracy of the MB-pol three-body term.

2.4.3 Higher-order contributions to the interaction energy

It has been shown that n-body interactions with $n \geq 4$ are relatively small for water systems.[31,38,39,42,44–46,126] By construction, these non-pair-additive interactions are described in the MB-pol potential through the $V_{\mathrm{TTM,ind}}$ induction energy term in Eq. 2.3. The analysis of the four-body interactions shown in Fig. 2.6 thus enables a quantitative assessment of the accuracy of MB-pol in describing higher-order many-body contributions. For this analysis, 2674 tetramer configurations were extracted from: 1) PIMD simulations of both cage (881 tetramers) and prism (925 tetramers) isomers of the water hexamer carried out with the HBB2-pol potential,[32] and 2) from PIMD simulations of liquid water (868 tetramers) carried out with the HBB2-pol potential.[98] In both cases, a significantly larger number of tetramer configurations was initially selected and only geometries that were separated by a uniformly weighted RMSD ≥ 0.2 Å were included in the final set. The four-body reference interaction energies were computed at the MP2 level of theory using the aug-cc-pVTZ basis set[84,118] supplemented with midbond functions[119] placed at the center of mass of each tetramer. The interaction energies were evaluated via the counterpoise method using the energies of the constituent fragments computed in the full tetramer basis.[103,120] As shown in Fig. 2.6, the mean absolute value of the four-body interaction is $\sim$0.06 kcal/mol. The RMSD associated with the MB-pol four-body energies is 0.03 kcal/mol, with the largest deviation being 0.392 kcal/mol (the mean total energy of these tetramer configurations relative to the tetramer global minimum is $\sim$42.89 kcal/mol).

2.4.4 Energetics of small water clusters

The relative energies (with respect to the corresponding global minima) of small water clusters with more than three molecules allows for a further assessment of the

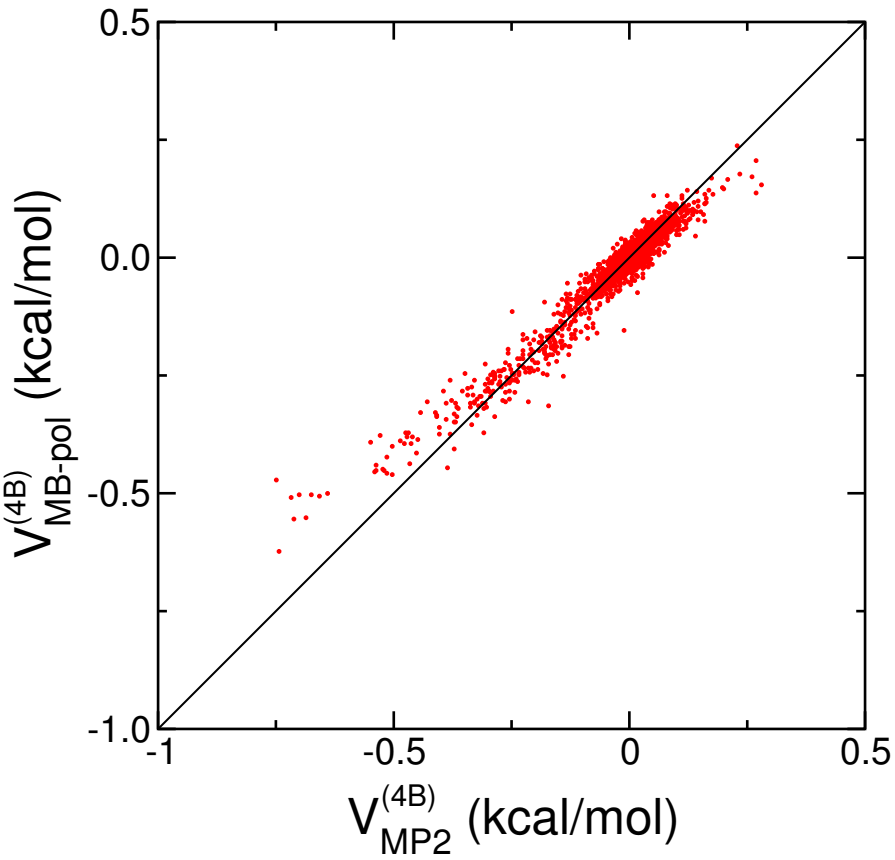

Figure 2.6: Comparison of model 4B energies to MP2. Plotted on the x-axis are the BSSE-corrected MP2/aug-cc-pVTZ+mb values while on the y-axis are the corresponding MB-pol values for the 2674 water tetramers examined.

ability of MB-pol to correctly describe n-body interactions with $n > 3$. For this purpose, the relative energies of the low-lying isomers of the water tetramer, pentamer, and hexamer calculated with the iAMOEBA-rev5, WHBB, HBB2-pol, and MB-pol potentials are compared in Fig. 2.7 with the corresponding *ab initio* values reported in the literature.[30,127] Water clusters in this size range are among the largest ones for which highly correlated electronic structure calculations are still feasible. All water potentials reproduce the *ab initio* data semi-quantitatively, with iAMOEBA-rev5, which was specifically parameterized to reproduce the relative energies of small water clusters, providing the closest agreement. Except for the CA-B isomer of the water pentamer, MB-pol reproduces (within chemical accuracy) the energy ordering of all cluster isomers as predicted by the *ab initio* calculations. In this regard, it is important to note that all *ab initio* data were calculated for geometries optimized at the MP2 level, which may also be source of some discrepancies with predictions made by potentials that were parameterized at the CCSD(T) level as WHBB, HBB2-pol, and MB-pol.

In summary, the "first principles"-based MB-pol PES has been introduced. As a model with explicit 1B, 2B, and 3B short-range interactions fitted to large datasets of highly correlated electronic structure calculations combined with a mean field description of the induced $>$ 3B interactions and asymptotically correct pairwise dispersion, MB-pol was demonstrated to correctly describe the properties of small water clusters with respect to both *ab initio* benchmark calculations and experiment. In the next chapter, the ability of MB-pol to predict both thermodynamic and dynamical properties of liquid water will be explored.

Portions of this chapter appeared in our previously published work, Babin, V.; Medders, G.R.; Paesani, F. "Development of a first principles water potential with flexible monomers. II: Trimer potential energy surface, third virial coefficient, and small clusters", *J. Chem. Theory Comput.* **2014** 10, 1599–1607. This material was reproduced with

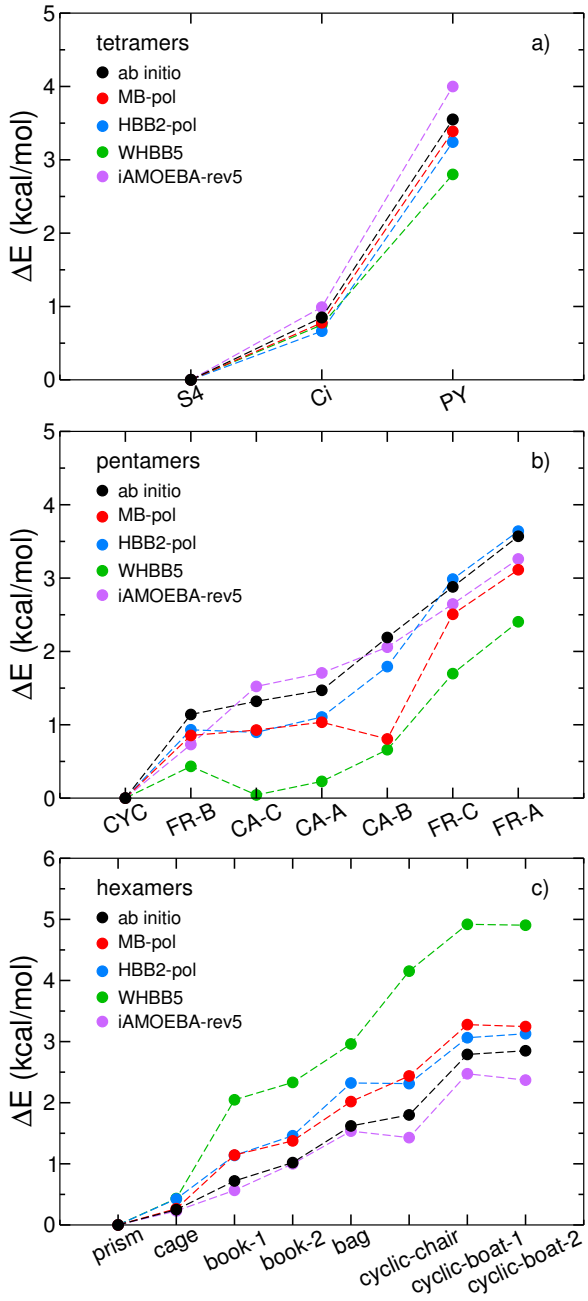

Figure 2.7: Relative energies of water tetramer, pentamer, and hexamer isomers. The relative energies are plotted in kcal/mol. The MB-pol (red), HBB2-pol (blue), WHBB5 (green), and iAMOEBA-rev5 (violet) potentials are examined, with the reference *ab initio* values shown in black having been obtained from Refs. 127 and 30.

permission from the publisher.

Chapter 3

Liquid phase properties of MB-pol

The accurate representation of intermolecular interactions combined with the development and implementation of efficient theoretical/computational approaches for simulations of large molecular systems still remain challenging problems in electronic structure theory. In many respects, these two problems are incompatible with each other because the accurate calculation of noncovalent interactions typically requires the use of highly-correlated methods whose associated computational cost effectively precludes their application to large systems. As a result, despite significant recent progress, chemical accuracy remains out of reach for most simulations of condensed phase systems.[63,64,104–111,128] This is well exemplified by the plethora of water potentials that have been developed over the years (including both force fields with different degrees of empiricism and *ab initio* models), none of which, however, is capable of correctly describing the properties of water from isolated molecules and small clusters in the gas phase to the bulk phases (e.g., see Refs. 32, 51, 112, 115, 122).

Recent studies have demonstrated that the many-body expansion of the interaction energy[129] converges rapidly for water,[31,38,39,42,44–46,126] providing a rigorous approach for the development of accurate multidimensional potential energy surfaces (PESs) for

water systems of arbitrary sizes. This rapid convergence of the many-body expansion has previously been exploited by *ab initio*-based interaction potentials, which enable a rigorous treatment of the many-body expansion of interactions through a functional form that is sufficiently flexible to accurately capture the complexity of high-quality *ab initio* reference data. Examples of such models are DPP2,[41] CC-pol,[81] WHBB,[80] and our previous water potential, HBB2-pol.[91,98] In the first two chapters of this work, we have introduced the MB-pol potential that is built upon the many-body formalism and our earlier HBB2-pol model. The accuracy of the MB-pol potential in reproducing the properties of finite-size water systems was additionally assessed in Ref. 115 through a detailed analysis of the dimer vibration-rotation tunneling spectrum, second virial coefficients, and cluster structures and energies. In this chapter we report classical and quantum dynamics simulations carried out with the MB-pol potential for liquid water at ambient conditions. Direct comparison with the corresponding experimental data demonstrates the ability (so far unique among the existing water models) of the MB-pol potential of correctly predicting the water properties from the dimer to the liquid phase entirely from "first principles".

3.1 Simulation methods

Since the MB-pol potential provides an accurate representation of the multidimensional Born-Oppenheimer PES for water,[115,122] it must be combined with simulation approaches that explicitly take into account nuclear quantum effects for a truly "first principles" description of bulk properties. In this study, path-integral molecular dynamics (PIMD)[130–133] and centroid molecular dynamics (CMD)[134] are used to calculate structural, thermodynamic, and dynamical properties of liquid water at T = 298.15 K and P = 1 atm. Since both methods and recent applications to aqueous systems have been reviewed

recently,[34,135] only a brief summary of the main concepts is given here. PIMD is based on Feynman's formulation of statistical mechanics in terms of path-integrals[136] and exploits the isomorphism between the quantum partition function of a system of *N* particles and the classical partition function of a system consisting of *N* flexible ring polymers.[130] By construction, PIMD enables the calculation of numerically exact structural and thermodynamic properties of quantum-mechanical systems.[137] The CMD formalism draws upon the prescription of distribution functions in which the exact quantum expressions are cast into a convenient phase space representation of the system dynamics.[138–144] The latter is described by classical-like equations of motion applied to the centroid coordinates corresponding to the centers of mass of the ring polymers that represent the quantum particles within the path-integral formalism. The CMD time propagation thus enables the calculation of quantum time correlation functions that can then be used to characterize the properties of condensed-phase systems within linear response theory.[145]

All simulations discussed in the following were carried out for a system consisting of 256 H_2O molecules in a periodic cubic box. Each atom was represented by a ring polymer with P = 32 beads. For comparison, simulations with a single bead (P = 1) were also performed to model the properties of water at the classical level. The PIMD simulations were carried out using the normal mode representation of the ring polymers,[133] which enables an efficient sampling of the quantum partition function.[146] All PIMD simulations were carried out in both constant temperature - constant volume (NVT) and constant temperature - constant pressure (NPT) ensembles. The equations of motion were propagated using the velocity-Verlet algorithm with a timestep $\Delta t = 0.2$ fs. The temperature was controlled via Nosé-Hoover chains (NHC) of four thermostats coupled to each degree of freedom.[147] The NPT ensemble was generated according to the algorithm described in Ref. 148.

In the CMD simulations, the centroid force was computed "on the fly" exploiting

the time-scale separation between the dynamics of the centroid degrees of freedom and the nonzero frequency normal modes.[141] The propagation of the centroid variables was carried out in the constant energy - constant volume (NVE) ensemble with NHC of four thermostats attached to each nonzero frequency normal mode. An adiabaticity parameter $\gamma = 0.1$ and a timestep $\Delta t = 0.025$ fs ensured a sufficiently large separation in time between the motion of the centroid and the nonzero frequency normal modes as well as a proper integration of the latter. In both PIMD and CMD simulations, a radial atom-atom cutoff distance of 9.0 Å was applied to the nonbonded interactions and the Ewald sum was used to treat the long-range electrostatic interactions.[149] The structural and thermodynamic properties were obtained by averaging over a 1 ns long PIMD simulations in the NPT ensemble and a 100 ps simulation in NVT. The dynamic properties were obtained by averaging over 10 independent CMD trajectories of 10 ps each. For the classical NVE simulations, two independent 200 ps simulations were performed due to the relatively long lifetime of the orientational correlation function in the classical case.

3.2 Structure of liquid water

The thermodynamic properties of liquid water at ambient conditions as predicted by the MB-pol potential are first analyzed. A comparison between the classical and quantum results along with the corresponding experimental data for the liquid density and enthalpy of vaporization is reported in Table 3.1. The density predicted by the PIMD simulations is appreciably smaller than its classical counterpart and in closer agreement with the experimental value. As discussed in previous studies (e.g., see Ref. 150) the explicit quantization of the nuclear motion results in an increase of the molecular dimensions due to zero-point energy effects and delocalization, which are thus responsible for the relatively lower density of the liquid at the quantum-mechanical

Table 3.1: Thermodynamic and dynamical properties of MB-pol water. The properties at ambient conditions from both classical and quantum simulations. Both density (ρ) and enthalpy of vaporization (H_{vap}) were calculated in the NPT ensemble, while the orientational relaxation time (τ_2) and diffusion constant (D) were calculated in the NVE ensemble. Experimental data: (a) from Ref. 151, (b) from Ref. 152, (c) from Ref. 153, and (d) from Ref. 154.

	density ($g\ cm^{-3}$)	H_{vap} ($kcal\ mol^{-1}$)	τ_2 (ps)	D ($Å^2\ ps^{-1}$)
classical	1.004(1)	10.9(2)	5.3(2)	0.12(1)
quantum	1.001(2)	10.1(4)	2.3(3)	0.22(3)
experiment	$0.997^{(a)}$	$10.5176^{(b)}$	$2.5^{(c)}$	$0.23^{(d)}$

level. Both classical and quantum results for the enthalpy of vaporization are in similar agreement with the corresponding experimental value.

The differential neutron scattering cross-sections per atom calculated from NPT simulations carried out at both classical and quantum levels are shown in Fig. 3.1 along with the available experimental data.[33,155] This comparison demonstrates that PIMD simulations with the MB-pol potential accurately represent liquid water at ambient conditions. The difference between classical and quantum results also indicate the presence of non-negligible nuclear quantum effects on the liquid structure as already discussed in the literature (e.g., see Ref. 34 for a recent review).

Classical and quantum oxygen-oxygen (O-O), oxygen-hydrogen (O-H), and hydrogen-hydrogen (H-H) radial distribution functions (RDFs) calculated in the NPT ensemble are compared in Fig. 3.2 with three recent sets of experimental data.[33,155,156] As expected from the analysis of the differential neutron scattering cross-section, the PIMD O-O RDF correctly reproduces the experimental results, predicting a relatively lower first peak compared with estimates obtained from simulations with common force fields and DFT models (see Fig. 3.3). As already observed for the HBB2-pol[98] and TTM3-F[58] potentials, some differences between MB-pol and experimental results exist for the second peak of the O-H RDF. In this regard, it is important to note that both

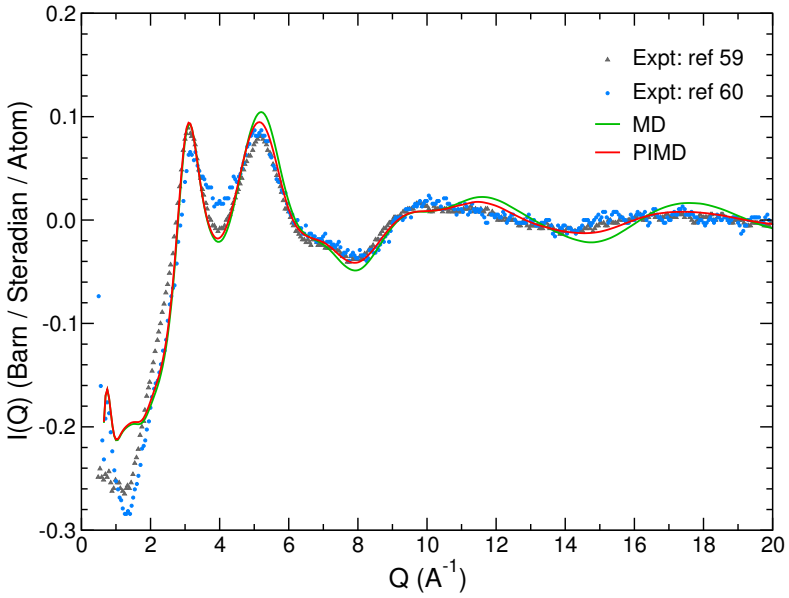

Figure 3.1: Differential neutron scattering cross-section of MB-pol water. The cross-section per atom was calculated from classical (MD, green) and quantum (PIMD, red) simulations of liquid water at T = 298.15 K and P = 1 atm. The experimental data are taken from Ref. 155 (gray) and from Ref. 33 (light blue).

position and shape of this peak, which describes the spatial correlation between O and H atoms directly involved in hydrogen bonds, are difficult to determine experimentally as demonstrated by the appreciable differences between the two sets of experimental data.

To test the ability of different water models (with flexible monomers) to reproduce the experimentally derived O-O RDFs, both classical and quantum O-O RDFs of q-TIP4P/f,[49] TTM3-F,[58] BLYP,[157] BLYP-D3[72,158] were compared against the most recent experiments and MB-pol in Fig. 3.3. These three models, whose ability to describe the many-body energies was assessed in Ref. 91, were chosen as representative of fixed-charge, polarizable, and DFT potentials that are commonly used in quantum simulations of liquid water. From the comparison shown in Fig. 3.3, q-TIP4P/f, TTM3-F, BLYP, and BLYP-D3 predict an over-structured first solvation shell, which is particularly evident in the case of BLYP. Among these four water models, BLYP-D3 is in relatively better agreement with the experimental data, displaying a slightly lower first peak in the O-O

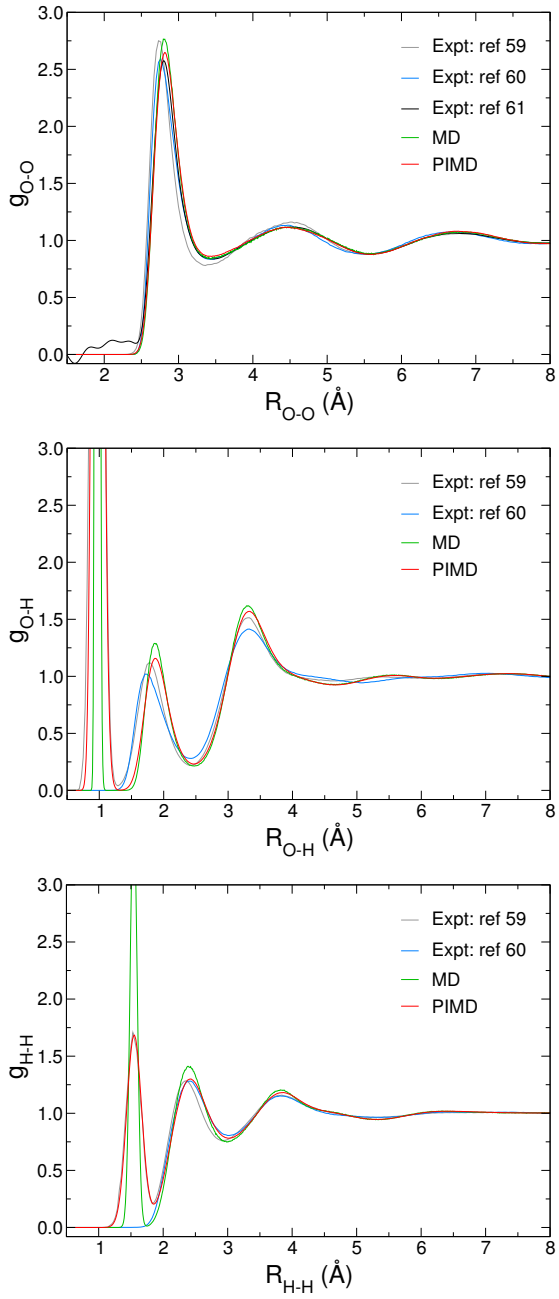

Figure 3.2: RDFs of water from classical and quantum MB-pol simulations. O-O, O-H, and H-H RDFs were calculated from both classical (MD, green) and quantum (PIMD, red) NPT simulations with the MB-pol potential at T = 298.15 K and P = 1 atm and compared with the corresponding experimental data from Ref. 155 (gray), from Ref. 33 (light blue), and from Ref. 156 (black).

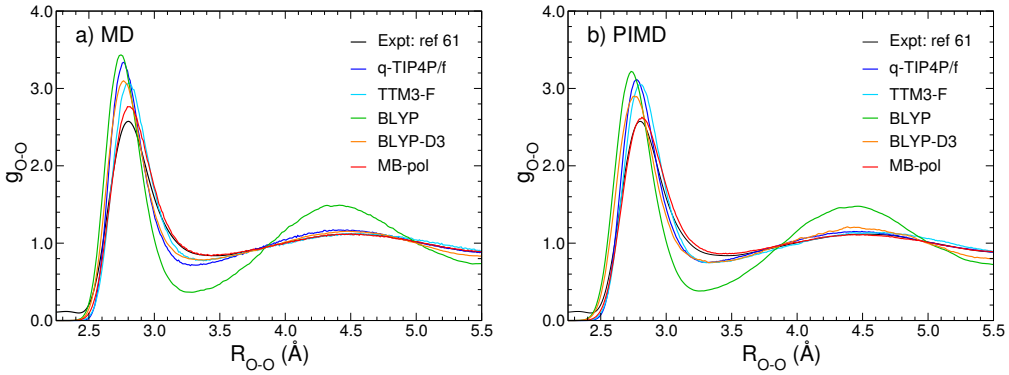

Figure 3.3: Predicted O-O RDFs of different water models. Comparison between classical (panel a, MD) and quantum (panel b, PIMD) O-O RDFs obtained from NVT simulations at the experimental density corresponding to ambiend conditions are shown for the q-TIP4P/f (blue), TTM3-F (cyan), BLYP (green), BLYP-D3 (orange), and MB-pol (red) potentials. The BLYP and BLYP-D3 results are taken from simulations carried out at 300 K in Refs. 157 and 158, respectively. The experimental results (black) are from Ref. 156.

RDF than q-TIP4P/f and TTM3-F. As mentioned above, similar deficiencies are common to the majority of water models reported in the literature. A notable exception are the results obtained with the CC-pol potential, which, however, neglects intramolecular flexibility.[81]

3.3 Dynamical properties of liquid water

The diffusion coefficient (D) and orientational relaxation time (τ_2) calculated at both classical and quantum levels are listed in Table 3.1. The diffusion constant was calculated from integration of the velocity autocorrelation function,[149]

$$D = \frac{1}{3}\int_0^\infty \langle \mathbf{v}(0)\cdot\mathbf{v}(t)\rangle. \tag{3.1}$$

The orientational relaxation time was obtained by fitting the long-time decay of the time autocorrelation function of the second-order Legendre polynomial, $C_2(t)$ in Eq. 3.2, to a

single exponential,

$$C_2(t) = \langle P_2[\mathbf{e}(0) \cdot \mathbf{e}(t)] \rangle. \tag{3.2}$$

In Eq. 3.2, $\mathbf{e}(t)$ is a unit vector along each OH bond in the body-fixed reference frame of the corresponding water molecule. The quantum results were obtained by averaging over 10 CMD trajectories of 10 ps each, while the classical results were calculated using 2 NVE trajectories of 200 ps. For both quantities, the CMD results are in quantitative agreement with the corresponding experimental values, providing further evidence of the accuracy of the MB-pol model. Interestingly, the differences between classical and quantum results are significantly larger than estimates recently reported with the q-TIP4P/f potential.[49] This difference suggests that the effects of nuclear quantization on the water properties are nontrivial even at ambient conditions and are particularly sensitive to the accuracy of the underlying PES.

In summary, the full-dimensional MB-pol potential introduced in Refs. 115 and 122 has been employed in classical and quantum simulations of liquid water at ambient conditions. The agreement between the simulation predictions and the experimental data indicates that MB-pol represents a major step toward the long-sought "universal model" capable of describing the behavior of water under different conditions and in different environments.[12] In this context, the MB-pol potential appears to be well suited for a fully "first principles" characterization of the pressure-temperature properties of water, which may help resolve some of the current controversies regarding structural, thermodynamic, and dynamical properties of bulk, interfacial, and supercooled water.

Portions of this chapter appeared in our previously published work, Medders, G.R.; Babin, V.; Paesani, F. "Development of a first principles water potential with flexible monomers. III: Liquid phase properties", *J. Chem. Theory Comput.* **2014** 10, 2906–2910. This material was reproduced with permission from the publisher.

Chapter 4

Many-body decomposition of electrostatic properties of water

4.1 Introduction

In the last decade, linear and nonlinear vibrational spectroscopy has emerged as a powerful technique that, by complementing the existing scattering methodologies, can provide direct insights into the structure and dynamics of the water HB network in different environments. As discussed in Section 1.1, within linear response theory, vibrational spectra, including infrared (IR), Raman, sum-frequency generation (SFG), and two-dimensional infrared (2D-IR) spectra, can be calculated from the Fourier transforms of appropriate single- or multi-time correlation functions involving the system dipole and/or polarizability.[27] It thus follows that an accurate representation of both the dipole and polarizability of water is necessary for the calculation of vibrational spectra that can be directly compared with the corresponding experimental measurements. This, in turn, is key to the molecular-level characterization of the structural and dynamical properties of water in different environments and conditions.

As discussed in Ref. 114, a systematic analysis of the electrostatic properties of a molecular system begins by investigating the energy of a single molecule interacting at long range with an external electric field F_α ($\alpha = x, y, z$). Up to the second-order perturbation in the electric field, the interaction is given by

$$E = E_0 - \mu_\alpha F_\alpha - \frac{1}{2} F_\alpha \alpha_{\alpha\beta} F_\beta, \tag{4.1}$$

where terms involving derivatives of the electric field have been neglected and the repeated greek indices imply sums over the corresponding cartesian coordinates. In Eq. 4.1, μ_α is the static, field-independent molecular dipole and $\alpha_{\alpha\beta}$ is the dipole-dipole polarizability. $\alpha_{\alpha\beta}$ is defined through the following sum over the system's states,

$$\alpha_{\alpha\beta} = \sum_{n\neq 0}^{\infty} \frac{\langle 0|\hat{\mu}_\alpha|n\rangle\langle n|\hat{\mu}_\beta|0\rangle + \langle 0|\hat{\mu}_\beta|n\rangle\langle n|\hat{\mu}_\alpha|0\rangle}{E_n - E_0}. \tag{4.2}$$

In some cases, higher-order static multipoles and polarizabilities can also play an important role. For example, it has been suggested that specific features of the SFG spectra of the benzene-air interface can be associated with both quadrupole moment and dipole-quadrupole polarizabilities.[159] However, this appears not to apply to water for which only the dipole and dipole-dipole polarizabilities effectively contribute to Eq. 4.1. For this reason, in the following the focus will be on the many-body convergence of μ_α and $\alpha_{\alpha\beta}$ of water, and, for the sake of simplicity, the dipole-dipole polarizability will be simply referred to as the polarizability.

Any total property p^T of an N-molecule system can be formally broken down into a many-body expansion

$$p^T_{\alpha\beta\ldots\omega}(1,\ldots,N) = \sum_i^N p^{1B}_{\alpha\beta\ldots\omega}(i) + \sum_{i<j}^N p^{2B}_{\alpha\beta\ldots\omega}(i,j) + \ldots, \tag{4.3}$$

where p^{1B}, p^{2B}, ..., refer to the one-body, two-body, ..., contributions, respectively. The 1B term is defined by the nuclear coordinates of a single monomer (which are collectively defined here by a single index, e.g., "i"). The 2B contribution depends on the nuclear coordinates of the dimer comprising monomers "i" and "j", with the greek subscripts $\alpha\beta\ldots\omega$ representing the cartesian components. In the case of tensor properties, such as μ_α and $\alpha_{\alpha\beta}$, any component of p^T is obtained by summing over the corresponding composite interactions.

In the following the 1B dipole and polarizability are defined as the corresponding monomer quantities,

$$\mu_\alpha^{1B}(i) = \mu_\alpha(i)$$
$$\alpha_{\alpha\beta}^{1B}(i) = \alpha_{\alpha\beta}(i).$$

All higher-order terms are defined recursively through Eq. 4.3. For example, the 2B dipole term is obtained by subtracting the 1B contributions from the dimer dipole,

$$\mu_\alpha^{2B}(1,2) = \mu_\alpha(1,2) - \sum_i^2 \mu_\alpha^{1B}(i), \tag{4.4}$$

and the 3B contribution is calculated by subtracting all 1B and 2B terms from the trimer dipole,

$$\mu_\alpha^{3B}(1,2,3) = \mu_\alpha(1,2,3) - \sum_{i<j}^3 \mu_\alpha^{2B}(i,j) - \sum_i^3 \mu_\alpha^{1B}(i) \tag{4.5}$$

Provided it converges relatively quickly, Eq. 4.3 thus represents a convenient framework for reconstructing the total dipole and polarizability of an N-molecule system from the sum of lower-order contributions. The many-body expansion has been employed in quantum mechanical models such as X-POL[105,160] and the effective fragment potential method.[106]

In this regard, it should be noted that a proper representation of polarizability and non-additive effects may have important implications in the description of water in heterogeneous environments, such as interfaces[161] and confined environments.[162] Furthermore, water is characterized by a relatively large molecular polarizability and it has been demonstrated that non-additive contributions to the total energy are non-negligible.[31,38,39,42–46,126] Therefore, it is expected that 2B and higher-order contributions can make a significant contribution to the total dipole moment of N-molecule systems. Perhaps less understood is how the polarizability of a gas-phase water molecule changes upon solvation. It has been suggested that promotion of electrons to some of the more-diffuse excited states appearing in Eq. 4.2 may be forbidden in the condensed phase because of the Pauli principle.[163,164] As a result, it is expected that only a subset of excited states in Eq. 4.2 are effectively allowed in the condensed phase, leading to a decrease of the molecular polarizability. In this context, it is interesting to consider how this restricted sum is manifest in water within a many-body expansion framework. If the polarizability is curtailed upon solvation, is it through the low-order (i.e., 2B and 3B) terms or through global N-body contributions that depend on the molecular configurations of the entire system?

While the total electrostatic properties of small clusters have been studied extensively, the associated many-body convergence has been explicitly examined only for a few systems. Early investigations focused on the 2B and 3B interaction-induced properties of noble gas trimers,[120,165–167] with higher-order effects being examined only for helium pentamers.[166] More recently, the many-body properties of hydrogen-bonded systems such as HF[168–171] and formaldehyde[171] chains have also been investigated. Though the specific details vary depending on the particular hydrogen-bonded system analyzed, these studies demonstrate that the higher-than-1B contributions to both dipole and polarizability are effectively dominated by 2B effects. It is important to note, however, that the dipole

hyperpolarizability appears to have a more substantial contribution from the beyond-2B terms.[169,170]

For water, *ab initio* studies of the many-body convergence of dipole and polarizability are more scarce. Several studies have focused on the electrostatic properties of the water dimer.[172–174] The polarizability of larger water clusters has been examined, where it was found that the average monomer polarizability saturates quickly with cluster size.[175–178] In Ref. 80 it was speculated that the total dipole could reasonably be approximated by considering 1B and 2B contributions only. It was also shown that the total dipole of the prism isomer of the water hexamer is effectively dominated by low-order contributions.[179] By contrast, the many-body convergence of the interaction energy for water systems has been studied extensively.[31,38,39,42–46,126] In the case of the hexamer, it was found that 2B terms represent a large fraction of the total interaction energy, although 3B terms can contribute up to 30%. All higher-order interactions contribute less than 5%.[31] Estimates of the many-body convergence in the condensed phase through the decomposition of the 21-mer cluster into 1B-4B contributions were largely consistent with the conclusions derived from the hexamer studies.[45]

In this study, a systematic analysis of the convergence of the many-body expansion of the water dipole and polarizability is reported. As mentioned above, this is the first step toward the development of "first principles" theoretical approaches to vibrational spectroscopy. In Section 4.2, the computational methods used to assess the convergence are discussed. Since the calculation of the many-body decomposition of the dipole and polarizability is computationally demanding, one of the goals of Section 4.2 is to determine the accuracy of the electrostatic properties calculated with different electronic structure methods and basis sets. The many-body convergence is then studied in Section 4.3 for small $(H2O)_N$ clusters with N=4-6 as well as for 14-mers extracted from classical simulations of liquid water and ice I_h. A summary highlighting the main

findings and discussing future directions is given in Section 4.4.

4.2 Methods

4.2.1 Calculation of Electrostatic Properties through Finite Differences or Analytical Methods

Beginning with the expression describing the interaction of a single molecule with an electric field (Eq. 4.1), Kurtz et al.[180] obtained the following equations, which are used here for finite-field calculations of both μ_α and $\alpha_{\alpha\beta}$,

$$\mu_\alpha = -\frac{1}{12F_\alpha}\Big[8\big[E(F_\alpha) - E(-F_\alpha)\big] - \big[E(2F_\alpha) - E(-2F_\alpha)\big]\Big] \tag{4.6}$$

$$\alpha_{\alpha\alpha} = \frac{1}{12F_\alpha^2}\Big[30E(0) - 16\big[E(F_\alpha) + E(-F_\alpha)\big] + \big[E(2F_\alpha) + E(-2F_\alpha)\big]\Big] \tag{4.7}$$

$$\alpha_{\alpha\beta} = \frac{1}{48F_\alpha F_\beta}\Big[-16\big[E(F_\alpha, F_\beta) - E(F_\alpha, -F_\beta) - E(-F_\alpha, F_\beta) + E(-F_\alpha, -F_\beta)\big] + \big[E(2F_\alpha, 2F_\beta) - E(2F_\alpha, -2F_\beta) - E(-2F_\alpha, 2F_\beta) + E(-2F_\alpha, -2F_\beta)\big]\Big] \tag{4.8}$$

All calculations were performed with Molpro (version 2012.1).[181] Fields of 0.005 a.u. were used to ensure numerical stability, and the accuracy was validated with respect to

the corresponding analytical calculations, which, however, are possible in Molpro only at the HF and MP2 levels.

4.2.2 Choice of the metrics for many-body convergence

In contrast to the many-body expansion of the interaction energy, the convergence of Eq. 4.3 for both dipole and polarizability is somewhat more difficult to demonstrate. The symmetry of simple systems can be exploited to examine the variation of only one component of the dipole or polarizability (e.g., the longitudinal dipole and polarizability of linear chains of HF[168]) or to eliminate off-diagonal components in the polarizability tensor (e.g., the water dimer in C_{2v} symmetry[172]). However, one is often interested in the properties of structures that do not necessarily exhibit a well-defined symmetry (e.g., the global-minimum water dimer,[174] DNA base pairs,[182] and formaldehyde chains in the crystal structure configuration[170]). Since several approaches have been proposed in the literature, three different metrics are first discussed here in the context of the N-body decomposition of the dipole. In the simplest case, the N-body contribution to the total dipole, $\Delta\mu^N$, is defined as the ratio between the magnitude of the N-body term and the magnitude of the total dipole,

$$\Delta\mu^N = \frac{||\boldsymbol{\mu}^N||}{||\boldsymbol{\mu}^T||}. \tag{4.9}$$

This metric suffers from two main problems: 1) it is not normalized ($\sum_i \Delta\mu^{iB} \neq 1$) and 2) it does not provide any indication of how the many-body components combine to define the orientation of the dipole moment. These deficiencies can be partially addressed by comparing the magnitude of the sum of the many-body contributions up to order N with the total magnitude,

$$\varepsilon_\mu^N = \frac{||\sum_{i=1}^N \boldsymbol{\mu}^{iB}||}{||\boldsymbol{\mu}^T||} \tag{4.10}$$

To study the many-body convergence of a property measured through Eq. 4.10, differences between the N and N-1 terms are calculated with the exception of the 1B contribution, ε_p^1, which is simply given by

$$\Delta p^N = \begin{cases} \varepsilon_p^1 & N = 1 \\ \varepsilon_p^N - \varepsilon_p^{N-1} & N > 1 \end{cases} \tag{4.11}$$

Using this definition, it is evident that Eq. 4.10 is normalized with respect to the total dipole. Although Eq. 4.10 represents an improvement over Eq. 4.9, it can nonetheless be ambiguous because any term with incorrect direction but correct magnitude would have no error. To address this problem, a different quantity is considered in which the magnitude of the error vector $\boldsymbol{\mu}_{err}^N$ is compared to the magnitude of the total dipole,

$$\boldsymbol{\mu}_{err}^N = \boldsymbol{\mu}^T - \sum_{i=1}^{N} \boldsymbol{\mu}^{iB} \tag{4.12}$$

$$\varepsilon_\mu^N = 1 - \frac{||\boldsymbol{\mu}_{err}^N||}{||\boldsymbol{\mu}^T||} \tag{4.13}$$

To guarantee a proper normalization, the many-body contributions of Eq. 4.13 are calculated through Eq. 4.11.

The 1B, 2B, 3B, and total dipole moments of the water trimer in the global minimum configuration are reported in the top half of Table 4.1, while the analysis of the relative contributions performed with the three metrics presented in Eqs. 4.9-4.13 is shown in the bottom half of the same table. Although Eq. 4.9 is not guaranteed to be normalized, in this case it correctly describes the relative contributions of 1B, 2B, and 3B terms. Since the 1B term has nearly the correct magnitude but the incorrect direction, Eqs. 4.10-4.11 erroneously predict the 2B dipole to have no contribution. The analysis of the Cartesian components of the many-body expansion reveals that the 2B contribution

Table 4.1: Measures of the many-body convergence of the dipole. The structure examined is the global minimum water trimer at the BSSE-corrected CCSD(T)/d-aug-cc-pVTZ level. Dipoles are reported in Debye.

Dipole moment				
	1B	2B	3B	Tot
x	0.036	-0.034	0.000	0.002
y	0.039	-0.001	-0.002	0.036
z	-0.970	-0.001	0.002	-0.969
Many-body dipole contributions, $\Delta\mu^N$				
	1B	2B	3B	Tot
Eq. 4.9	100.2%	3.5%	0.2%	104.0%
Eq. 4.10	100.2%	0.0%	-0.2%	100.0%
Eq. 4.13	96.5%	3.3%	0.2%	100.0%

instead acts to mostly preserve the magnitude of the dipole while reorienting it, an effect that is correctly captured by Eqs. 4.13 and 4.11.

The rank-two tensor nature of the polarizability poses additional complications. The many-body convergence of the average diagonal elements has often been used in the literature,

$$\alpha_{avg} = \frac{1}{3}\left[\alpha_{xx} + \alpha_{yy} + \alpha_{zz}\right]. \tag{4.14}$$

However, for ($N > 1$)-body polarizabilities, the analysis in terms of the average polarizability can be somewhat ambiguous, as will be discussed below. Similarly to the dipole moment, the many-body convergence of the polarizability can also be studied through the magnitude of an error matrix,

$$\boldsymbol{\alpha}^N_{err} = \boldsymbol{\alpha}^T - \sum_{i=1}^{N} \boldsymbol{\alpha}^{iB} \tag{4.15}$$

$$\varepsilon^N_{\alpha} = 1 - \frac{||\boldsymbol{\alpha}^N_{err}||_F}{||\boldsymbol{\alpha}^T||_F} \tag{4.16}$$

where $||\ldots||_F$ indicates the Frobenius norm.

To investigate the proposed decrease in polarizability upon solvation, the convergence of the eigenvalues of the polarizability tensor is analyzed here for a few cases. For example, to estimate the contribution of the 3B polarizability, the eigenvalues of the polarizability with up-to-3B terms ($\boldsymbol{\alpha}^{1B} + \boldsymbol{\alpha}^{2B} + \boldsymbol{\alpha}^{3B}$) can be calculated and normalized with respect to the eigenvalues of the total polarizability (Eq. 4.18),

$$\boldsymbol{\Lambda}(\alpha^N) = \boldsymbol{V}^{-1} \left[\sum_{i=1}^{N} \boldsymbol{\alpha}^{iB} \right] \boldsymbol{V} \tag{4.17}$$

$$\varepsilon^N_{\alpha;a} = \frac{\Lambda_{aa}(\alpha^N)}{\Lambda_{aa}(\alpha^T)} \tag{4.18}$$

The contribution of the 3B polarizability is then given by applying Eq. 4.11 to each eigenvalue. Table 4.2 reports the results for the many-body convergence of the polarizability calculated for the global minimum energy configuration of the water trimer. This analysis clearly shows that focusing on the average polarizability may have several drawbacks. First, the components of the many-body polarizability tensors are signed and, therefore, can cancel one another when summed. For this reason, the 2B polarizability of the water trimer appears to be nearly negligible when measured through the average polarizability. By contrast, examination of the individual eigenvalues shows that the 2B term plays a significant role by acting to reorient the trimer polarizability. Importantly, although the net effect of the 2B terms is to decrease the monomer polarizabilities, the polarizability is actually enhanced along some directions. A second drawback is that the off-diagonal elements are neglected. Though the off-diagonal elements are often small relative to the diagonal 1B components, in the case of the 2B term they can be as large as the diagonal ones. This underscores the need to calculate the full polarizability tensor when examining differences in polarizability tensors for systems where the off-diagonal elements are not zero by symmetry.

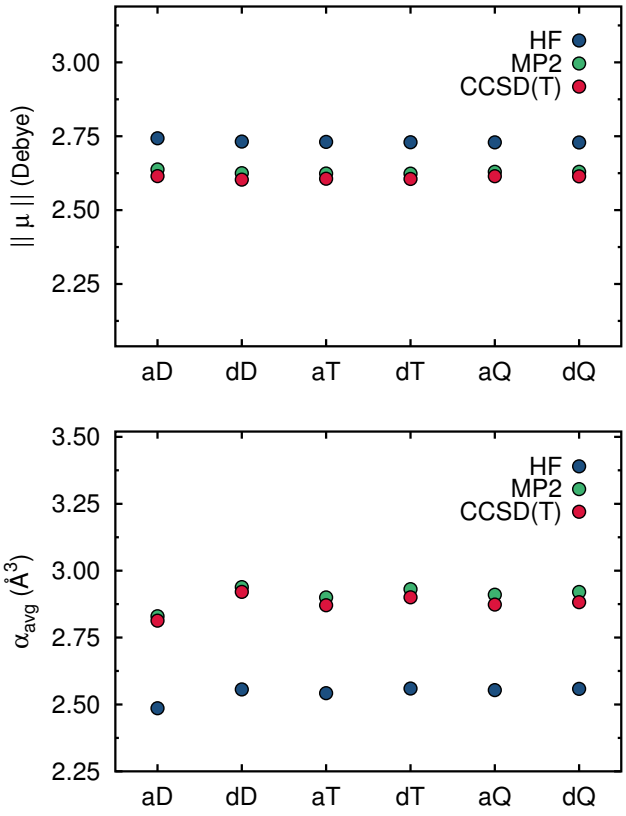

Figure 4.1: Dependence of μ and α on basis set/method. The configuration examined is the water dimer global minimum structure. In the basis set abbreviations on the x-axis, the lower-case letter refers to the size of the diffuse basis (a is "aug-" and d is "d-aug-") and the capital letter refers to the cardinal number from the cc-pVXZ basis set (X $\in$ {D,T,Q,5,...}). E.g., dD is the d-aug-cc-pVDZ basis set. To compare the dipole and polarizability, the range of the y-axes are normalized to a 25% deviation from the CCSD(T)/dQ value. The values plotted in this figure are available in the supporting information for further inspection.

Table 4.2: Measures of the many-body convergence of the polarizability. The structure examined is the global minimum water trimer at the BSSE-corrected CCSD(T)/d-aug-cc-pVTZ level. Polarizabilities are reported in Å^3.

	1B Polarizbility			2B Polarizbility			3B Polarizbility		
	x	y	z	x	y	z	x	y	z
x	4.386	0.002	-0.001	0.197	0.002	0.027	-0.017	0.000	-0.004
y	0.002	4.372	-0.070	0.002	0.179	-0.113	0.000	-0.017	0.007
z	-0.001	-0.070	4.286	0.027	-0.113	-0.482	-0.004	0.007	0.012
α_{avg}	4.348 / 101.0%			-0.035 / -0.8%			-0.007 / -0.2%		
$\Delta\alpha_a^N$	96.4%	96.1%	112.5%	4.0%	4.3%	-12.9%	-0.4%	-0.4%	0.4%
$\Delta\alpha^N$	92.6%			7.0%			0.4%		

Table 4.3: Sensitivity of the many-body decomposition to BSSE. The many-body percent contributions for the prism hexamer properties were calculated using different BSSE correction schemes. To make comparison with the 14-mer calculations presented later, these calculations are performed at RI-MP2/aug-cc-pVDZ. See Section 4.2.3 for details.

	Dipole			Polarizability		
	1B	2B	3B	1B	2B	3B
SSFC	90.3	7.4	2.1	95.0	4.0	0.9
VMFC	90.2	7.6	2.1	94.7	4.2	0.9
VMFC(3B)	90.3	7.6	2.2	94.8	4.2	0.9
VMFC(SMol)	90.7	6.7	2.1	97.0	-1.5	-0.7

4.2.3 Basis-Set Superposition Error

As is well known, the basis-set superposition error (BSSE) is an artifact associated with the incomplete nature of the basis functions that are used to calculate intermolecular interactions.[103] The most commonly used correction to this problem is the Boys-Bernardi counterpoise (CP) method.[103] The CP method was extended to the calculation of higher-order many-body interactions through the "site-site" function counterpoise (SSFC) method, which requires the N-body and lower-order interactions to be calculated in the basis-set of the N-mer system.[120] The SSFC method is advantageous in that its many-body contributions can be directly summed to give the energy of the whole N-molecule system. To achieve this, however, the SSFC method causes the M<N-body interactions to depend on the basis set of the complete N-mer (except at the complete basis set limit). A different approach to the many-body BSSE correction is the Valiron-Mayer function counterpoise (VMFC) method.[183,184] VMFC addresses the SSFC dependence on the N-mer basis set through the calculation of the 1B interaction in the monomer basis, the 2B interaction in the dimer basis, etc. The methods to correct for BSSE within the many-body expansion are reviewed in detail Appendix A for the reader's convenience.

Since both dipole and polarizability are calculated through partial derivatives of the energy with respect to external electric fields, the electrostatic properties discussed here intrinsically suffer from BSSE. For most calculations presented in this study, the SSFC method is used to correct for BSSE. For larger systems, (i.e., the water 14-mers discussed in Section 4.3), it is not feasible to perform the many-body decomposition using the full cluster basis set, and, therefore, the VMFC method is used. Because the VMFC method lacks a simple relation to the total cluster energy, the convergence of the many-body expansion can in principle be established only after computing the full many-body decomposition (up to 14B interactions).

As this is clearly impractical, the effect of approximating the true VMFC correction with the sum of 1B through 3B contributions [VMFC(3B)] is studied in Table 4.3 using the (smaller) water hexamer at the RI-MP2/aug-cc-pVDZ level. The VMFC(3B) many-body contributions are in good agreement with the true VMFC contributions as well as the SSFC results. This indicates that 1B-3B terms effectively dominate the dipole and polarizability of water independently of the system size. This analysis also demonstrates that it is inappropriate to compare the VMFC many-body contributions to the total supermolecular cluster calculation [VMFC(SMol)]. Since a consistent application of either SSFC or VMFC requires a significant amount of computational resources and considering that VMFC(3B) provides comparable accuracy, it thus appears reasonable to apply the VMFC(3B) method to the study of the 14-mer cluster.

4.2.4 Treatment of electronic correlation and the basis set

The dependence of the many-body dipole and polarizability on the electron correlation was investigated using three different methods: HF, MP2, and CCSD(T).[83,186] Dunning's correlation consistent basis sets, cc-pVXZ, with X = D, T, Q, and 5, were used in the calculations.[84] Because an accurate determination of the polarizability

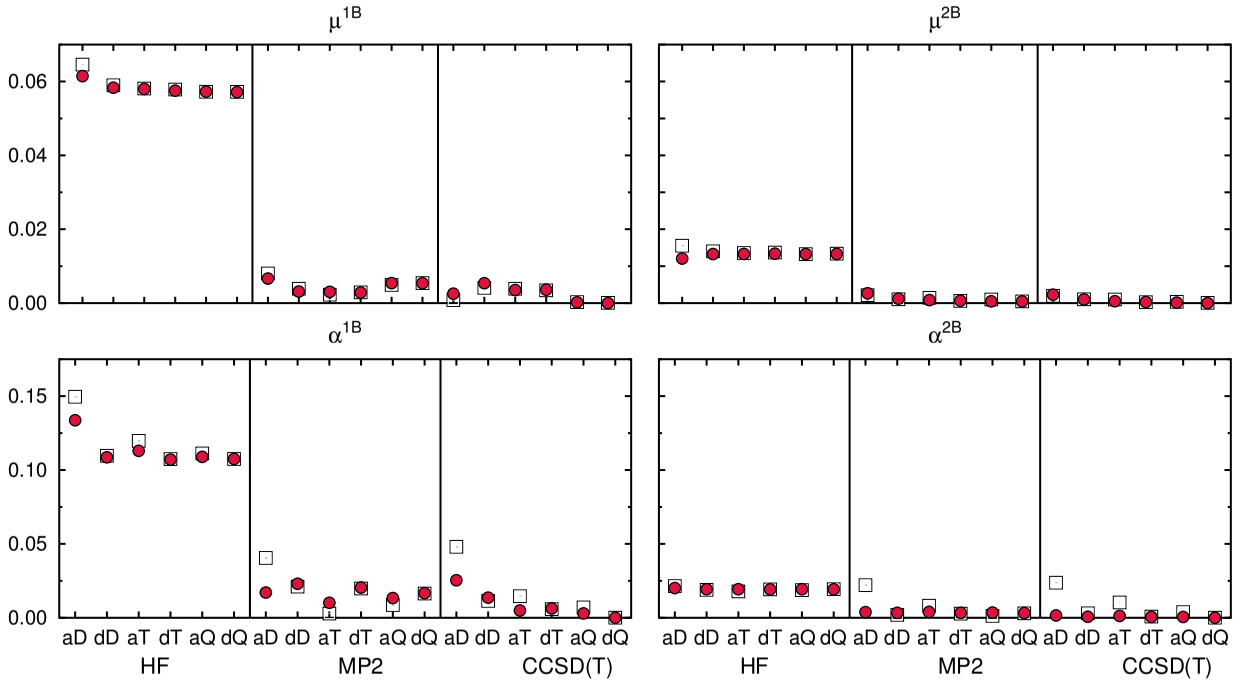

Figure 4.2: Basis set convergence of dimer electrostatics. Specifically, the basis set dependence of the decomposition of the total dipole and polarizability for the global minimum dimer into 1B and 2B contributions is presented. On the x-axis is the method/basis. The basis set abbreviations are as described in Fig. 4.1. On the y-axis, the magnitude of the differences between each method/basis and the CCSD(T)/d-aug-cc-pVQZ values are presented with respect to the magnitude of the total property (see Eq. 4.19). BSSE-corrected values (using SSFC) are plotted with solid circles, uncorrected values with open squares. The values plotted in this figure are available in the Supporting Information of Ref. 185 for further inspection.

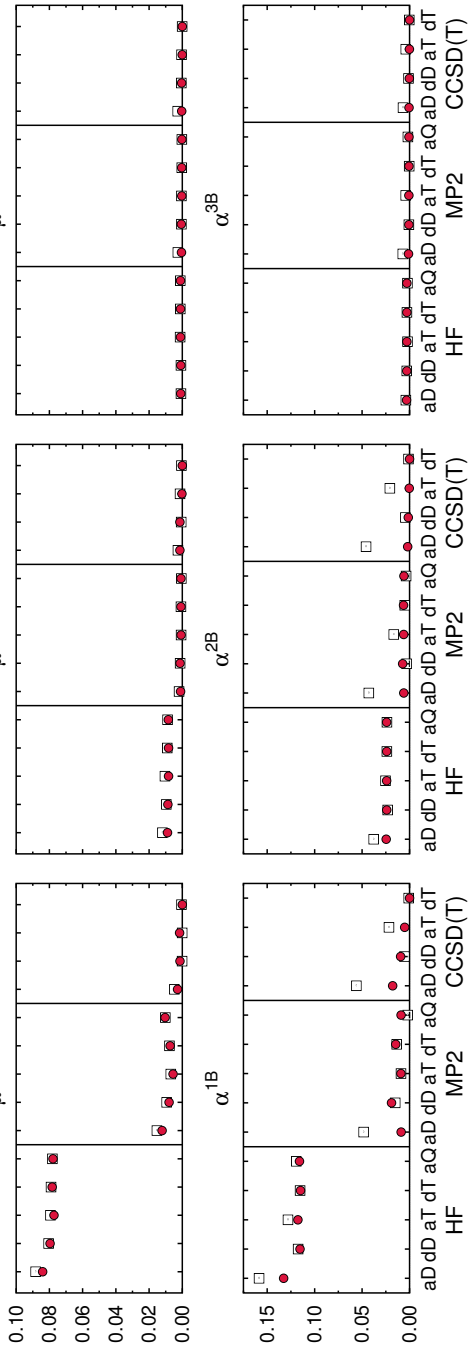

Figure 4.3: Basis set convergence of trimer electrostatics. Specifically, the basis set dependence of the decomposition of the total dipole and polarizability for the global minimum trimer into 1B, 2B and 3B contributions is presented. The values plotted in this figure are available in the Supporting Information of Ref. 185 for further inspection. See the caption of Fig. 4.2 for a more detailed description.

requires diffuse basis functions,[187] augmented (aug-) and double augmented (d-aug-) functions[118,188] were added to the cc-pVXZ basis sets. Unless otherwise stated, all calculations were corrected for BSSE using the SSFC method as described previously. Note that all calculations presented in this article use flexible monomer geometries.

In Fig. 4.1, the convergence of the total dipole and polarizability of the global minimum energy configuration of the water dimer is presented as a function of the basis set for HF, MP2, and CCSD(T). This analysis clearly show that the treatment of correlation effects is the primary source of error. In addition, as expected, the polarizability is more sensitive to the basis-set size than the corresponding dipole. To test the dependence of the decomposition of the dipole and polarizability 1B and 2B terms on the basis set and method, the many-body contributions of the water dimer global minimum geometry are presented in Fig. 4.2. The basis-set convergence of the 3B dipole and polarizability is studied using the results obtained for the global minimum energy configuration of the water trimer in Fig. 4.3. The basis set convergence is studied using

$$\sigma(p^N) = \frac{||\boldsymbol{p}^{NB}_{\text{ref}} - \boldsymbol{p}^{NB}_{\text{method}}||}{||\boldsymbol{p}^{T}_{\text{ref}}||}, \tag{4.19}$$

where p indicates the electrostatic property. These values, plotted in Figs. 4.2 and 4.3, can be interpreted as the relative error that a given N-body term contributes to the total property. Partially because both dipole and polarizability are dominated by the 1-body interactions (see Section 4.3), correlation effects and basis set truncation artifacts most significantly impact the total values at the 1B level. Similarly to the results shown in Fig. 4.1, the accuracy is primarily limited by the treatment of the electronic correlation, although the error associated with basis-set truncation can contribute up to 2.5%.

At the 1B level, the difference between the HF and CCSD(T) results may be as large as 8% for the dipole and 12% for the polarizability. At the 2B level on the

other hand, the post-HF correlation accounts for only 1% of the dipole and 2.5% of the polarizability. MP2 recovers much of the correlation effects at the 1B and 2B levels, falling within 0.1%-1% and 0.7%-2% of the CCSD(T) results for the dipole and polarizability, respectively. Correlation and finite-size basis set effects are much smaller at the 3B level, which is consistent with observations made about the 3B interaction energy.[31] Note that because the many-body expansion used here is defined by the cluster's (clamped) nuclear coordinates, our analysis includes the monomer distortion effects in the 1B term. In all cases studied, correcting for BSSE almost always improves the agreement of the specific property relative to the converged basis-set result for a particular method. As the polarizability is more sensitive to the basis-set size, it also exhibits larger BSSE than the dipole for a given method/basis. For basis sets larger than aug-cc-pVQZ, however, the BSSE becomes negligible.

4.3 Many-body decomposition of the dipole moment and polarizability of water clusters

The many-body decomposition of both the dipole and the polarizability of small $(H_2O)_N$ clusters with $N = 4 - 6$ is presented in Tables 4.4 and 4.5, respectively. The clusters, arranged from top to bottom in order of increasing binding energy, are obtained from Refs. 127 and 30. As is well known, the lowest energy isomers of the 4-mer and 5-mer are cyclic, with all oxygen atoms lying in the same plane. The hexamer is the smallest water cluster whose minimum-energy configuration has a three-dimensional hydrogen-bond network. Because the hexamer has several nearly iso-energetic yet topologically distinct isomers, characterizing the properties of the hexamer in particular has long been of interest to both theorists[29–31,35,189] and experimentalists.[190–194] Based on the results presented in Figs. 4.2 and 4.3, all calculations for these clusters were

Table 4.4: Many-body convergence of the dipole for small clusters. The percent contributions of the many-body dipoles to the total cluster dipole were computed at the RI-MP2/aug-cc-pVTZ level. BSSE was corrected through the SSFC method. All greater-than-4B dipoles are less than 0.1%. C/D refers to constructive or destructive interference of the dipoles, with the percentage being the difference between the orientation independent and dependent measures (Eqs. 4.10 and 4.13, respectively).

	Isomer	C/D	Dipole			
			1B	2B	3B	4B
4-mer	S_4	—	—	—	—	—
	C_i	—	—	—	—	—
	Pyramid	D(4%)	96.4	2.5	1.1	0.0
5-mer	Cyclic	D(6%)	95.4	4.1	0.4	0.1
	FR-B	D(7%)	93.7	5.0	1.1	0.1
	Cage-C	C	96.1	2.1	1.8	0.1
	Cage-A	C	96.4	2.0	1.4	0.2
	Cage-B	C	92.4	6.1	1.2	0.1
	FR-C	D(4%)	95.1	3.7	1.8	0.1
	FR-A	D(4%)	97.4	1.4	1.2	0.0
6-mer	Prism	D(4%)	90.3	7.4	2.2	0.0
	Cage	D(8%)	84.8	13.1	1.9	0.1
	Book	D(5%)	90.5	6.7	2.4	0.3
	Cyclic	—	—	—	—	

Table 4.5: Many-body convergence of the polarizability for small clusters. The percent contributions of the many-body polarizabilities to the total cluster polarizability were computed at the RI-MP2/aug-cc-pVTZ level. BSSE was corrected through the SSFC method. All greater-than-4B polarizabilities are less than 0.1%.

	Isomer	Polarizability			
		1B	2B	3B	4B
4-mer	S_4	91.1	8.0	0.7	0.2
	C_i	90.9	8.3	0.7	0.1
	Pyramid	96.7	2.8	0.5	0.0
5-mer	Cyclic	90.6	8.4	0.8	0.1
	FR-B	92.4	6.7	0.8	0.1
	Cage-C	95.0	4.1	0.7	0.1
	Cage-A	94.9	4.4	0.7	0.1
	Cage-B	95.2	3.8	0.9	0.1
	FR-C	94.2	4.9	0.7	0.2
	FR-A	93.0	6.2	0.7	0.2
6-mer	Prism	95.1	3.8	1.0	0.1
	Cage	93.5	5.5	0.9	0.1
	Book	92.6	6.2	1.0	0.1
	Cyclic	91.0	8.1	0.8	0.1

Table 4.6: Many-body contributions to 14-mer dipoles and polarizabilities. Calculations were performed at RI-MP2/aug-cc-pVDZ level and corrected for BSSE using the VMFC(3B) method. The global minimum configuration is from an extensive study of the AMOEBA potential energy surface.[197]

Isomer	C/D	Dipole			Polarizability		
		1B	2B	3B	1B	2B	3B
liquid1	C	62.6	35.6	1.7	96.7	1.8	1.5
liquid2	D(6%)	82.2	14.7	3.0	95.4	3.1	1.5
ice1	D(28%)	46.3	51.1	2.7	96.2	3.0	0.9
ice2	C	65.7	33.0	1.3	96.1	2.8	1.2
Global min[197]	D(7%)	90.1	6.0	3.9	93.4	4.3	2.2

performed at the BSSE-corrected RI-MP2/aug-cc-pVTZ level.[195,196] RI-MP2 was found to agree with MP2 to within 10^{-6} a.u. for the dipole and 10^{-3} a.u. for the polarizability.

The many-body decomposition of the small cluster dipoles is presented in Table 4.4. The dipole moments of the S_4 and C_i 4-mer and the cyclic hexamer are zero by symmetry. For nearly all clusters, the permanent dipole of the non-interacting water monomers in the cluster configuration comprises more than 80% of the total dipole. Surprisingly, to a good approximation, the dipole for these clusters is largely pairwise additive. While 3B interaction energies can constitute up to 30% of the interaction energy for these clusters, non-additive induced dipole effects account only for $< 2.5\%$ of the total dipole. Similarly to the dipole, the many-body decomposition of the polarizability is dominated by the 1B term. As shown in Table 4.5, the 2B polarizability contributes up to 9%, while the 3B term is responsible for roughly 1%. For both dipole and polarizability, 4B and higher contributions give rise to at most 0.2% to the total property.

In an attempt to assess the convergence of Eq. 4.3 for condensed phase water systems, the global minimum energy configuration of the 14-mer cluster (as was recently determined Kazachenko and Thakkar[197] using the Amoeba force field[60]) is also analyzed along with four 14-mer clusters that were randomly extracted from 100ps classical molecular dynamics simulations of liquid water at ambient density at 298K and ice Ih at

50K using the HBB2-pol[91,98] and TTM3-F[58] potentials, respectively. The liquid and ice 14-mer configurations can be found in the supporting information of this paper, while the configuration of the AMOEBA 14-mer global minimum structure can be found in the supporting information of Ref. 197. The 14-mer clusters were specifically chosen because they are large enough to contain at least one molecule that is completely "solvated" by the remaining molecules, yet of sufficiently small size that the many-body decomposition of both dipole and polarizability is still feasible. As discussed in Section 4.2.3, since the SSFC method becomes prohibitively expensive for large clusters, the VMFC(3B) method was used to correct for BSSE in all calculations of the 14-mer clusters that were carried out at the RI-MP2/aug-cc-pVDZ level.

The many-body convergence of the polarizability is consistent between small clusters and 14-mers, with the 1B term contributing more than 90%, the 2B term 2-5%, and 3B less than 2.5%. For the many-body convergence of the dipole, the contributions of the 1B and 2B dipoles are in competition. This behavior, which is distinct from that of the polarizability, originates in part from the fact that, while the 1B polarizability eigenvalues are always large and positive, the 1B dipole moments can combine constructively or destructively, depending on their orientation. The extent to which the many-body dipoles constructively or destructively interfere can be determined by the difference between the orientation-independent (Eq. 4.10) and orientation-dependent (Eq. 4.13) convergence metrics. When the difference is zero, the many-body dipoles are adding constructively, otherwise the many-body dipoles are destructively interfering with one another. This effect is presented in Tables 4.4 and 4.6 where the threshold for interference to be considered destructive is taken as 3%.

Although the global minimum configuration of the 14-mer clusters (as predicted by the Amoeba force field[60]) has dipole contributions that are similar to those obtained for the smaller cluster minima, the broken hydrogen-bond configurations extracted

from liquid and ice simulations exhibit different behavior. For these condensed phase configurations, it is seen that even when the 1B and 2B terms add constructively, the 2B term can still contribute substantially. Interestingly, while the many-body dipoles often destructively interfere to a small extent, in the case of the ice1 configuration the 2B term acts to reorient the 1B contribution to such an extent that the interaction-induced 2B dipole actually overcomes the monomer contribution. By contrast, the 3B dipoles make minor contributions in all cases, which is likely related in part to the rapid convergence of the many-body polarizability.

4.4 Summary

The convergence of the many-body expansion of the dipole and the (dipole-dipole) polarizability for water was examined. It was demonstrated that methods commonly used to assess the many-body convergence of these two properties can often provide ambiguous results, especially for systems of lower symmetry. Alternative measures were introduced and used to study the dependence of both dipole and polarizability on electronic structure methods, basis sets, and schemes for correcting for the basis-set superposition error. It was found that, although the average diagonal polarizability elements associated with the 2B and 3B terms appear to slightly decrease relative to the gas-phase (1B) term, an analysis of the many-body convergence based on this observation may be incomplete. By considering the convergence of the eigenvalues individually, it was observed that the 2B polarizability can play a significant role, with water systems often becoming more polarizable in some directions relative to an isolated molecule.

The many-body convergence of both dipole and polarizability of $(H_2O)_N$ clusters with $N = 4-6, 14$ was also examined. It was shown that if the total dipole and polarizability of these clusters are treated as pairwise additive, more than 97.5% of the

total value is recovered. The dipole and polarizability become accurate to within 99.8% if 3B terms are also included. This analysis suggests that a theoretical/computational strategy based on the many-body expansion of the electrostatic properties can be used to accurately determine the dipole and polarizability of condensed-phase water systems from highly correlated electronic structure calculations, as already done for the total interaction energy.[78–80,91] This, in turn, will enable the calculation of (linear and nonlinear) water vibrational spectra entirely from "first principles".

Portions of this chapter appeared in our previously published work, Medders, G.R.; Paesani, F. "Many-body convergence of the electrostatic properties of water", *J. Chem. Theory Comput.* **2013** 9, 4844–4852. This material was reproduced with permission from the publisher.

Chapter 5

Many-body models for the dipole moment and polarizability of water

Having established in the previous chapter the many-body convergence of the dipole moment and polarizability for water, we now recall from Chapter 1 that, within linear response theory and the dipole approximation, infrared and Raman spectra can be obtained through the quantum time-correlation function of the dipole moment and polarizability, respectively (see Eqs. 1.3 and 1.4). Through this time-dependent formulation of the vibrational spectra, it is thus possible to identify three different elements that are necessary for a rigorous calculation of vibrational spectra: 1) a simulation approach capable of correctly describing the quantum dynamics of the system of interest, 2) an accurate representation of the underlying (Born-Oppenheimer) potential energy surface entering the Hamiltonian that governs the system dynamics, and 3) accurate representations of the multidimensional dipole moment and polarizability surfaces as a function of the system's degrees of freedom.

By far the most common approach to solving Eqs. 1.3 and 1.4 is to approximate them through their classical counterparts. Classical correlation functions can be easily ob-

tained from classical molecular dynamics (MD) simulations. For liquid water at ambient conditions, classical spectra describe the low frequency (least quantum-mechanical) vibrations reasonably well,[198–203] but predict relatively large blueshifts for the high-frequency (more quantum-mechanical) vibrations, such as the OH stretch.[135,204]

Methods that attempt to directly obtain quantum time-correlation functions for condensed-phase systems can be classified in two main groups. The first group contains methods that treat quantum mechanically only a subset of the system's vibrations, with all remaining degrees of freedom being described at the classical level.[205–211] These mixed QM/MM methods have been extensively applied to model vibrational spectra of isotopically dilute solutions of HOD in H_2O (D_2O),[208,212,213] where, because of large frequency separations, the OD (OH) stretch of HOD can be considered to be effectively decoupled from all other system's vibrations. Although mixed QM/MM methods have been successful in reproducing the OD (OH) line shape of HOD in H_2O (D_2O), recent ultrafast spectra measured for HOD in D_2O have demonstrated that the OH stretch of HOD is not fully decoupled from the other degrees of freedom.[214]

The second group of methods attempt to treat all nuclear degrees of freedom quantum mechanically by extending Feynman's path-integral formulation of statistical mechanics to dynamical properties. One of the first and more common approaches belonging to this group is centroid molecular dynamics (CMD) which, building upon path-integral molecular dynamics (PIMD), approximates the exact quantum dynamics in terms of classical-like equations of motion for the centroids of Feynman's ring-polymers.[141,143] Depending on the shape of the underlying potential energy surface, CMD may suffer from the so-called "curvature problem", which can limit its applicability to molecular clusters and low-temperature condensed-phase systems.[215,216] However, it has been demonstrated that for water at ambient conditions the curvature problem is negligible, with IR spectra obtained from CMD being in very good agreement with those obtained

from mixed QM/MM methods.[135,216,217] With the same spirit as CMD, ring-polymer molecular dynamics (RPMD) was subsequently introduced as an *ad hoc* approach to extending PIMD to the calculation of dynamical properties.[218] Vibrational spectra of water calculated from RPMD are contaminated by spurious resonances of the ring-polymer lying within the frequency window of the OH stretch.[219] A thermostatted version of ring-polymer molecular dynamics (TRPMD) has also been proposed, which removes the spurious resonances from the spectral region of interest by thermostatting the normal modes of the ring-polymers.[220] A systematic analysis of the OH line shape of HOD in D_2O has recently shown that the line shape predicted by TRPMD is significantly broader and somewhat blueshifted relative to the corresponding mixed QM/MM results.[135]

Regardless of the simulation method used to evaluate the relevant quantum time-correlation functions, a potential energy surface (PES) must be defined. For the case of liquid water, innumerable models of varying degrees of empiricism have been developed. A systematic comparison of the accuracy with which different models describe the Born-Oppenheimer PES for water was presented in Chapter 1 through a detailed analysis of the corresponding many-body expansion of interaction energies (see Figs. 1.1 and1.2).[91] Among existing water models, our many-body potential with explicit polarization, MB-pol,[115,122] was shown in Chapter 2 to accurately reproduce both structures and energetics of small clusters,[115,122] as well as thermodynamic and dynamical properties of bulk water at ambient conditions in Chapter 3.[221] Improving upon other many-body potentials (e.g., CC-pol,[78,79,81,82,222] WHBB,[80,209] and HBB2-pol[91,98]), MB-pol was derived entirely from a large dataset of highly-correlated many-body interaction energies using machine learning techniques and employing mathematical functions that are sufficiently flexible to capture the multidimensional complexity of the reference electronic structure data.

The final element needed to solve Eqs. 1.3 and 1.4 are the system's dipole moment and polarizability surfaces. In MD simulations with empirical (nonpolarizable and polariz-

able) potentials, the dipole moment and polarizability of the water molecules can be easily calculated from the corresponding parameters defined within the model.[58,59,200,223–225] In the case of *ab initio* simulations, the system's dipole moment and polarizability can be obtained from the electronic density.[201,202,226–230] Within the MB framework, the many-body potential used in the actual MD simulations does not contain any information about the distribution of the electron density and the dependence of the energy on external electric fields. For this reason, many-body potentials cannot, in principle, be used to derive electrostatic properties, such as the system's dipole moment and polarizability, which are required for modeling IR and Raman spectra. While MB potentials employ some form of polarizable electrostatics to describe higher-order molecular interactions, the energy arising from this approximate electrostatics model is corrected with functions (e.g., permutationally invariant polynomials in the case of WHBB,[80,209] HBB2-pol,[91,98] and MB-pol[115,122,221]) that account for quantum-mechanical effects neglected by purely polarizable models, including correlation, exchange, charge penetration and charge transfer. While these effects clearly affect both the charge distribution and polarizability, the extent to which these contributions affect the infrared and Raman vibrational spectra of liquid water is less clear. It is interesting to note that the same reasoning should be extended to empirical water models, in which partial charges and/or atomic polarizabilities, which are primarily fitted to reproduce structural and thermodynamic properties of the system of interest, do not necessarily have a rigorous connection to the underlying multidimensional dipole moment and polarizability surfaces.

In this study, we introduce accurate, "first principles" representations of the multidimensional dipole moment and polarizability of water which enable molecular-level investigations of IR and Raman spectra at a fully quantum-mechanical level. Our contribution builds upon the observation that the dipole moment and polarizability of a molecular system can rigorously be calculated through derivatives of the system's

energy with respect to external electric fields,[114] which thus allows one to derive formally exact many-body expansions of the dipole moment and polarizability. Exploiting the convergence of the many-body expansion for the dipole moment and polarizability of water, in Section 5.1, we describe the theoretical and technical details associated with the development of many-body representations of the dipole moment and polarizability for water. In Section 5.2, we demonstrate that these MB representations are capable of accurately describing the dipole moment and polarizability of water systems in both gas and condensed phases. Finally, we combine the newly developed many-body dipole moment and polarizability surfaces with CMD simulations carried out with our many-body MB-pol potential to model both IR and Raman spectra of liquid water at ambient conditions. A brief summary and outlook is given in Section 5.3.

5.1 Technical details of the MB-μ and MB-α models

Within the many-body formalism, the total dipole moment of a molecular system can be expressed as

$$\mu_\alpha^{N-mer}(1,\ldots,N) = \sum_i^N \mu_\alpha^{1B}(i) + \sum_{i,j}^N \mu_\alpha^{2B}(i,j) + \cdots + \mu_\alpha^{NB}(1,\ldots,N), \tag{5.1}$$

where μ^{nB} is the n-th interaction-induced dipole term and the subscript Greek letters are the Cartesian tensor indices. An analogous expression can be written for the system polarizability

$$\alpha_{\alpha\beta}^{N-mer}(1,\ldots,N) = \sum_i^N \alpha_{\alpha\beta}^{1B}(i) + \sum_{i,j}^N \alpha_{\alpha\beta}^{2B}(i,j) + \cdots + \alpha_{\alpha\beta}^{NB}(1,\ldots,N), \tag{5.2}$$

where α^{nB} is now the n-th interaction-induced polarizability. In Ref. 185, we established the convergence of Eqs. 5.1 and 5.2 for water clusters of different sizes and demonstrated

that, for these systems, the dipole moment and polarizability are dominated by 1B and 2B contributions, with higher-order terms contributing by less than 4%. Unlike the interaction energy, where 3B terms make a substantial contribution,[31,38,39,42–46,126] the more rapid convergence of Eqs. 5.1 and 5.2 suggests that accurate representations of the dipole moment and polarizability surfaces for water can focus on 1B and 2B contributions only, allowing 3B and higher-order terms to be treated in an effective way (or neglected entirely, depending on the desired level of accuracy). Importantly, as discussed in Ref. 185, an explicit description of the 2B term is required to correctly capture electronic quantum-mechanical effects arising from exchange and charge transfer,[96,231,232] which play an important role at short-range and are neglected in purely polarizable models.

Exploiting Eqs. 5.1 and 5.2 and building upon our previous results,[185] we introduce here many-body representations of the dipole moment (MB-μ) and polarizability (MB-α) surfaces for water derived from highly-correlated electronic structure calculations. The 2B terms were fit to 2B dipoles and polarizabilities calculated for 42,495 configurations defined in Ref. 115 at RI-MP2/aug-cc-pVTZ level using the counterpoise method to correct for the basis set superposition error (BSSE).[103] It was previously demonstrated by us that this level of theory accurately describes 2B electrostatic properties of water.[185] Both dipole moment and polarizability were calculated through finite field calculations with an electric field of 0.005 a.u.[180] using Molpro (version 2012.1).[181]

MB-μ is constructed as a sum of explicit 1B and 2B terms, with higher-order terms being described through many-body polarization. Specifically, the 1B term corresponds to the LTP2011 dipole moment surface developed by Lodi et al.,[233] which was fitted to the water molecular dipole obtained from all-electron, internally contracted multireference configuration interaction calculations in the aug-cc-pV6Z basis set, including corrections for size-extensivity and relativistic effects. The 2B dipole moment is modeled through two distance-dependent contributions. In the long-range region, where two water molecules

are well separated, the total dipole moment is represented through classical electrostatic interactions ($\mu^{2B,ind}$) described by the Thole-type point polarizable dipole model of Burnham and co-workers,[59] with minor modifications described in Ref. 115 and referred to here as TTM4-Ind. At short-range, where the electron densities of the individual monomers overlap, permutationally invariant polynomials are used to correctly capture the complexity of electronic quantum-mechanical effects contributing to the 2B dipole moment ($\mu^{2B,poly}$),

$$\mu_\alpha^{2B}(i,j) = s_2\big(t(r_{OO})\big)\mu_\alpha^{2B,poly}(i,j) + \big[1 - s_2\big(t(r_{OO})\big)\big]\mu_\alpha^{2B,ind}(i,j). \tag{5.3}$$

Here, s(t(r_{OO})) is a switching function which smoothly transitions from 1 to 0 as a molecule is moved from the short-range to the long-range region, with the range of the transition region being controlled by $t(r) = \frac{r-R_i}{R_o-R_i}$. The inner and outer limits of the transition region, R_i and R_o, correspond to 5.5 and 7.5 Å, respectively. The explicit form of the switching function is given by:

$$s_2(x) = \begin{cases} 1 & \text{if } x < 0 \\ 1 + x^2(2x-3) & \text{if } 0 \le x < 1 \\ 0 & \text{if } x \le 1 \end{cases}. \tag{5.4}$$

The short-range dipole moment is represented as a sum over atom-centered effective charges multiplied by the corresponding cartesian positions of the atoms,

$$\mu_\alpha^{2B} = \sum_i^6 q^i(\xi_0^\mu, \ldots, \xi_{14}^\mu) r_\alpha^i. \tag{5.5}$$

The effective charges depend on the internal geometry of the water dimer through the auxiliary variables, ξ, which are formed from the exponentiated interatomic distances,

e^{-kr}, where k is a parameter controlling the range of the exponent and r is the interatomic distance as shown in the Supporting Information of Ref. 234. These auxiliary variables are used to form polynomials that are invariant with respect to permutation of all other like atoms and co-vary when permuted with any other like atom. A more complete description of the symmetrization procedure can be found in the appendix of Ref. 80. Importantly, to ensure invariance of the dipole moment upon changes of the origin, the sum of the effective charges was constrained to zero during the fitting. Because charge neutrality is only achieved within a given tolerance (in this study, an RMS error of 10^{-6} charge units), each dimer was positioned such that the oxygen-oxygen (O-O) center of mass was located at the origin of the reference frame. This was enforced both during the fitting and the evaluation of the 2B dipole. The charges, q^i in Eq. 5.5 are described in terms of 5th order polynomials in the auxiliary variables presented in Table S1, resulting in 1826 terms for q^{H} and 1052 for q^{O}. The linear coefficients of the polynomials were fitted with singular value decomposition, while the 6 nonlinear parameters describing the range of the auxiliary variables were optimized using the simplex algorithm, with the least squares solution solved at each iteration, as described in Ref. 115. For the 2B dipole, the 6 nonlinear parameters were those controlling the range of the auxiliary variables ($k^O_{OO}, k^O_{OH}, k^O_{HH}, k^H_{OO}, k^H_{OH}, k^H_{HH}$), where the superscript indicates the effective charge to which the variable contributes and the subscript indicates the atom pair that forms the distance. All higher-order terms of MB-μ are described by the TTM4-Ind model.[59,115] For the training set of 42,495 2B dipoles moments, the RMS error is 0.01 Debye. When restricting the set to the $\sim$32,000 dimers with binding energies less than 25 kcal/mol, the RMS error is 0.003 Debye.

Similarly to MB-μ, the many-body representation of the water polarizability (MB-α) is described as a sum of explicit 1B and 2B contributions, and a many-body term

defined as

$$\alpha_{\alpha\beta}^{N-mer}(1,\ldots,N) = \sum_{i}^{N} \alpha_{\alpha\beta}^{1B}(i) + \sum_{i,j}^{N} \alpha_{\alpha\beta}^{2B}(i,j) + \alpha_{\alpha\beta}^{MB}(1,\ldots,N). \tag{5.6}$$

Here, $\alpha_{\alpha\beta}^{1B}$ is the 1B polarizability model of Avila, which was fitted to molecular polarizabilities calculated at the CCSD level.[235] Unlike the 2B term of the dipole moment, the short-range 2B contribution to the polarizability was directly built on top of the TTM4-Ind polarizability. The 2B polarizability thus takes the form:

$$\alpha_{\alpha\beta}^{2B}(i,j) = s_2\big(t(r_{OO})\big)\alpha_{\alpha\beta}^{2B,poly}(i,j) + \alpha_{\alpha\beta}^{2B,ind}(i,j) \tag{5.7}$$

where the switching function again turns off the short-range 2B polynomial over a transition region from 5.5 to 7.5 Å. The short-range 2B term is represented by effective isotropic polarizabilities, α^i, which are cast into the lab frame by multiplication by the unit vectors, $\hat{\varepsilon}^i$, associated with the sites carrying the effective polarizabilities,

$$\alpha_{\alpha\beta}^{2B,poly} = \sum_{i}^{10} \alpha^i(\chi_0^\alpha,\ldots,\chi_{30}^\alpha)\hat{\varepsilon}_\alpha^i\hat{\varepsilon}_\beta^i. \tag{5.8}$$

This functional form has the advantage of generating a symmetric polarizability tensor. The invariance of the polarizability with respect to the choice of origin is achieved by centering the O-O centroid of the dimer at the origin. Importantly, unlike the dipole moment, where the effective charges were placed only on the atomic sites, there effective polarizabilities are positioned on both the 3 atomic sites and 2 additional lone-pair sites of each molecule. As in Ref. 115, the 2 lone-pair positions are defined by:

$$\boldsymbol{r}_L^{\pm} = \boldsymbol{r}_O + \frac{1}{2}\gamma_{\parallel}(\boldsymbol{r}_{OH_1} + \boldsymbol{r}_{OH_2}) \pm \gamma_{\perp}[\boldsymbol{r}_{OH_1} \times \boldsymbol{r}_{OH_2}] \tag{5.9}$$

The parameters $\gamma_{\parallel}$ and $\gamma_{\perp}$ determine the position of the lone pair and were optimized during the fitting process. The effective polarizabilities are represented by polynomials formed from auxiliary variables depending on the interatomic distances (Table S1), maintaining the same in-/co-variant properties as for the dipole moment. Due to the larger number of possible interatomic distances, the exponential range parameters were assumed to be the same regardless of the effective polarizability (i.e., $k_{ab}^{O} = k_{ab}^{H} = k_{ab}^{L}$). This results in the following nonlinear parameters: $k_{OO}, k_{OH}, k_{OL}, k_{HH}, k_{HL}, k_{LL}, \gamma_{\parallel}$, and $\gamma_{\perp}$. Third-degree polynomials, composed of 435 terms for α^{O}, 603 terms for α^{H}, and 588 terms for α^{L}, were found to accurately fit the reference data, with a RMS error of 0.03 Å^3 over the entire training set and 0.009 Å^3 for dimers with binding energy less than 25 kcal/mol. In the same way as for MB-μ, all higher order term in MB-α are represented through the many-body polarizability arising from the TTM4-Ind model.[59,115]

5.2 Results

5.2.1 Electrostatic properties of clusters

Since MB-μ and MB-α correctly reproduce the reference dipole moment and polarizability of the corresponding training sets, their transferability is tested here on small water clusters. Specifically, we compare the 2B dipole moment and polarizability predicted by MB-μ and MB-α for fourteen water clusters containing from 4 to 6 molecules against reference values obtained at the RI-MP2/aug-cc-pVTZ level of theory including the counterpoise correction for BSSE. Fig. 5.1 shows the RMS error in the dipole moment (panel a) and polarizability (panel b) obtained for different (electrostatics) models, including TTM4-Ind, WHBB, and the present many-body representations, MB-μ and MB-α. In these comparisons, TTM4-Ind refers to the polarizable model derived from the TTM4-F water potential which MB-μ and MB-α are built upon, while WHBB

is the model of Refs. 209 and 80 in which the total dipole is expressed as a sum of 1B and 2B contributions only. Following our previous analysis of the electrostatic properties of water clusters,[185] the RMS errors shown in Fig. 5.1 for each model were calculated as

$$||\boldsymbol{p}^{\mathrm{MP2}} - \boldsymbol{p}^{\mathrm{model}}||_{\mathrm{F}}, \tag{5.10}$$

where $\boldsymbol{p}$ is either the dipole moment or the polarizability. The absolute values of the dipole moment and polarizability calculated for each cluster at the RI-MP2/aug-cc-pVTZ level of theory are also shown in Fig. 5.1a and c, respectively, as a reference.

MB-μ and MB-α clearly provide a significant improvement upon the underlying, purely induced-electrostatics model (TTM4-ind), with an error in the 2B properties of at most 5% and 6% for the dipole moment and polarizability, respectively. Since both MB-μ and WHBB are built within the same theoretical framework, they would be expected to be able to describe the 2B dipole moment of water clusters with similar accuracy. This suggests that the relatively larger RMS errors predicted by WHBB in Fig. 5.1 compared to MB-μ likely arise from differences in the composition of the corresponding training sets as well as in the level of theory used in the calculations of the reference data. Specifically, the WHBB 2B dipole moment surface was derived from BSSE-uncorrected MP2/aug-cc-pVTZ dipole moments, while the reference data used to develop MB-μ were calculated including the counterpoise correction to remove the BSSE.

5.2.2 Comparison of two-body properties from liquid simulations

Having established that MB-μ and MB-α correctly describe the dipole moment and polarizability of water clusters, the accuracy of both many-body surfaces is assessed here for liquid water. While reference data for the energetics of water systems in periodic boundary conditions have recently become available from quantum Monte Carlo

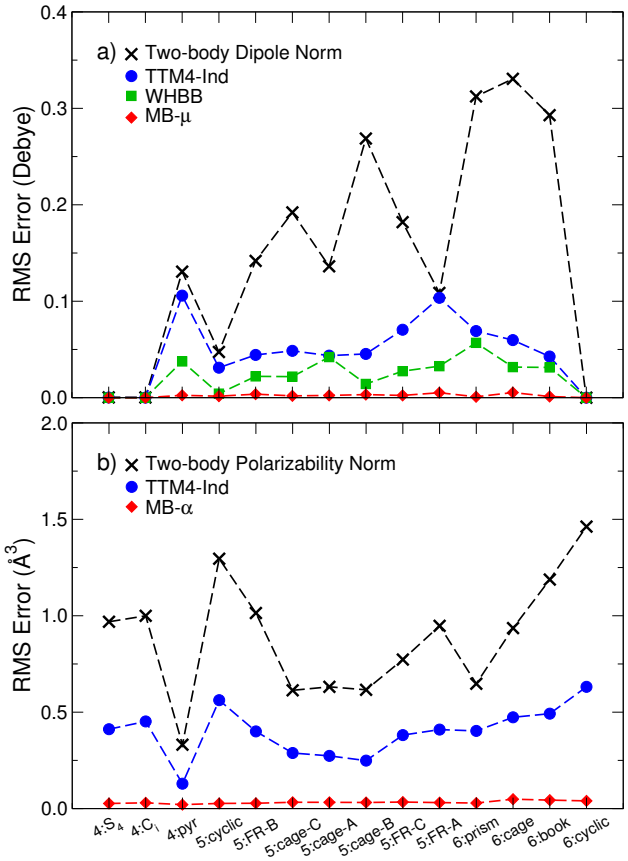

Figure 5.1: Accuracy of 2B dipole and polarizability models for clusters. The ability of different models for the dipole moment and polarizability to reproduce the two-body electrostatic properties was examined for clusters ranging from 4 to 6 water molecules. On the x-axis is the isomer index, corresponding to the clusters of Refs. 127 and 30, where the first number indicates the number of water molecules in the cluster. Panel a) shows the RMS error in the 2B dipole, while panel b) shows the RMS error in the 2B polarizability. The reference electronic structure method is RI-MP2/aug-cc-pVTZ with BSSE-correction.

calculations,[128,236] we are unaware of any benchmark calculations for the electrostatic properties of liquid water. For this reason, we investigate here the accuracy of MB-μ and MB-α in reproducing the dipole moment and polarizability of a hydrogen-bonded dimer along a 2 ps CMD trajectory extracted from a simulation of liquid water at ambient conditions.[221] During the course of the 2 ps trajectory, the distance between the oxygen atoms of the two water molecules within the dimer varies from 2.6 to 3.6 Å(with an average value of 3.15 Å), exploring a wide range of hydrogen bond strengths. For each trajectory frame, reference 2B dipoles and polarizabilities were calculated at the RI-MP2/aug-cc-pVTZ level of theory including the counterpoise correction for BSSE. It is important to emphasize that none of the 1000 dimer configurations visited during the CMD trajectory were included in the training sets used for the development of both MB-μ and MB-α.

The fractional errors ($||\boldsymbol{p}^{\mathrm{MP2}} - \boldsymbol{p}^{\mathrm{model}}||_{\mathrm{F}}/||\boldsymbol{p}^{\mathrm{MP2}}||_{\mathrm{F}}$) associated with the dipole moment and polarizability calculated with different models, including four density functionals commonly used in water simulations (BLYP, B3LYP, PBE, and PBE0) are shown in Fig. 5.2. All DFT calculations were carried out with the aug-cc-pVTZ basis and were corrected for BSSE using the counterpoise method. Given the similarity between the results obtained with BLYP and PBE, and with B3LYP and PBE0, only the PBE and PBE0 fractional errors are included in Fig. 5.2. Figures including the results of BLYP and B3LYP are reported in the Supporting Information of Ref. 234.

Fig. 5.2 clearly demonstrates that the 2B polarizability is more difficult to reproduce than the 2B dipole moment. Among all models considered here, MB-μ and MB-α provide the closest agreement with the reference data, with fractional errors of 0.4% and 2% in the 2B dipole moment and polarizability, respectively. GGA functionals (i.e., PBE and BLYP) have errors of $\sim$20% for the dipole moments and more than 60% for the polarizabilities. Compared to the GGA functionals, hybrid functionals predict

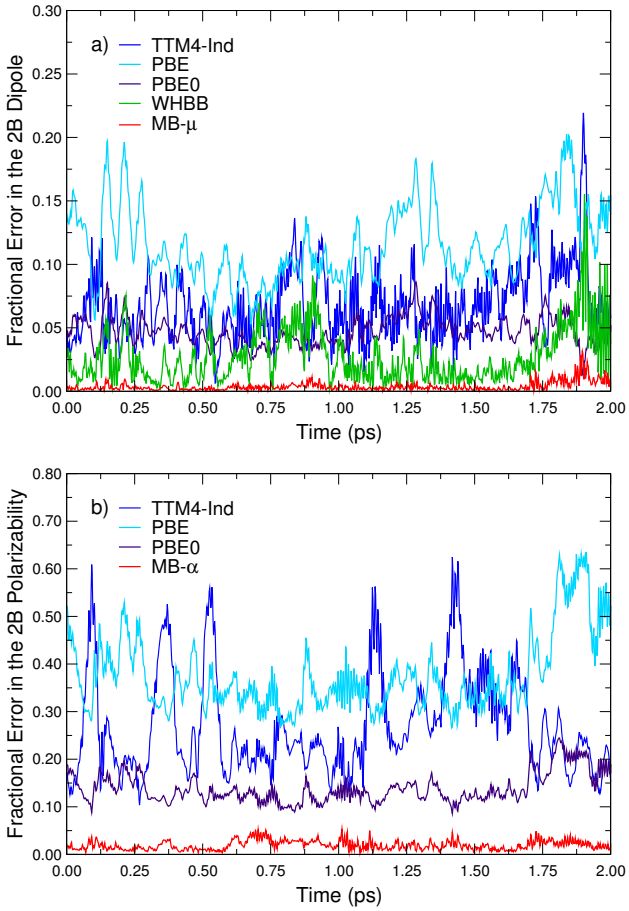

Figure 5.2: Accuracy of 2B dipole and polarizability models for liquid water. Fractional errors in the 2B dipole moments (a) and polarizabilities (b) are examined for hydrogen-bonded dimers extracted from a CMD simulation of liquid water. The reference properties were calculated at the RI-MP2/aug-cc-pVTZ level with BSSE correction. The properties obtained for PBE and PBE0 density functionals are also presented (obtained in the aug-cc-pVTZ basis with BSSE correction). Results for BLYP and B3LYP are shown in the supporting information.

relatively smaller fractional errors for all configurations visited along the CMD trajectory, which suggests that inclusion of Hartree-Fock exchange is likely important for correctly describing the electrostatic properties of liquid water.

5.2.3 Vibrational spectra of liquid water

Given the accuracy of our MB-pol, MB-μ, and MB-α models in describing the water interactions, dipole moment, and polarizability, respectively, these three surfaces were combined in CMD simulations of IR and (anisotropic) Raman spectra of liquid water at ambient conditions. The frequency dependence of the IR absorption coefficient, $\alpha(\omega)$, was calculated as

$$\alpha(\omega)n(\omega) = \left[\frac{2\omega}{3V\hbar c\varepsilon_{\text{vac}}}\right]\tanh\left(\frac{\hbar\omega}{kT}\right)\int_{-\infty}^{\infty} dt\ e^{-i\omega t}\langle\boldsymbol{\mu}(0)\boldsymbol{\mu}(t)\rangle, \tag{5.11}$$

where V is the volume of the simulation cell, c is the speed of light, $n(\omega)$ is the frequency dependent refractive index of the system, and $\langle\boldsymbol{\mu}(0)\boldsymbol{\mu}(t)\rangle$ is the quantum dipole autocorrelation function.[200] The anisotropic reduced Raman spectrum was obtained through

$$R(\omega) = \frac{2\omega}{(\omega_I-\omega)^4}\tanh\left(\frac{\hbar\omega}{kT}\right)\int_{-\infty}^{\infty} dt\ e^{-i\omega t}\langle\text{Tr}[\boldsymbol{\beta}(0)\boldsymbol{\beta}(t)]\rangle \tag{5.12}$$

where $\boldsymbol{\beta}$ is the anisotropy of the dipole-dipole polarizability, obtained by removing the isotropic component of the polarizability tensor: $\beta_{\alpha\beta} = \alpha_{\alpha\beta} - \frac{1}{3}\text{Tr}[\boldsymbol{\alpha}]$.[200] MB-$\mu$ and MB-α were used to calculate the correlation functions appearing in Eqs. 5.11 and 5.12, and the nuclear quantum dynamics was approximated by adiabatic CMD using the MB-pol potential energy surface to describe the interactions between water molecules. Specifically, 10 CMD simulations were performed, each of which was 27 ps long and was seeded from different initial conditions obtained from PIMD simulations carried out at 298.15 K. The CMD propagation was performed in the normal-mode representation,

and decoupling between the dynamics of the centroid and non-zero frequency normal modes was obtained with an adiabaticity parameter $\gamma = 0.1$.[135,237] A timestep of 0.02 fs was found to be sufficient for energy conservation. To enable the study of local solvation dynamics, instead of simulating pure H_2O, a single HOD molecule was placed in 215 H_2O molecules. In this very low concentration, the HOD molecule is expected to have negligible effect on the calculated vibrational spectra (indeed, no OD stretch or HOD bend is discernible in the calculated spectra).

The IR and Raman spectra calculated using our MB-MD approach are compared with the corresponding experimental results in Fig. 5.3. The reduced anisotropic Raman spectra (Fig. 5.3 c-d) are presented in arbitrary units, with the experimental spectrum being adapted from Ref. 238. Due to the use of arbitrary units in the experimental results, the Raman spectrum was normalized to the maximum intensity of the OH stretch, with successive levels of magnification employed for the bend and the low-frequency features. Examination of the low-frequency features of the IR and Raman indicates that MB-pol provides an accurate description of the hydrogen-bonding dynamics. The maximum of the calculated librational band (the broad feature between 400-800 cm^{-1} that is related to the hindered reorientation of molecules due to intermolecular interactions) lies within a few wavenumbers of the corresponding experimental value. In addition, the frequency of hydrogen-bond stretches ($\sim$180 cm^{-1}) is also in very good agreement with experiment. For the Raman spectra, the agreement between simulation and experiment is somewhat less good, with the position of hydrogen-bond stretches being slightly blueshifted and the tapered plateau of the librations being less defined. As is apparent in Fig. 5.3, some imbalance in the relative intensities of the stretch, bend, and librations is also present in the calculated Raman spectrum compared to the experimental results. Given the good agreement between the calculated and measured IR spectra, which implies that the underlying (quantum) molecular dynamics is correctly described, and

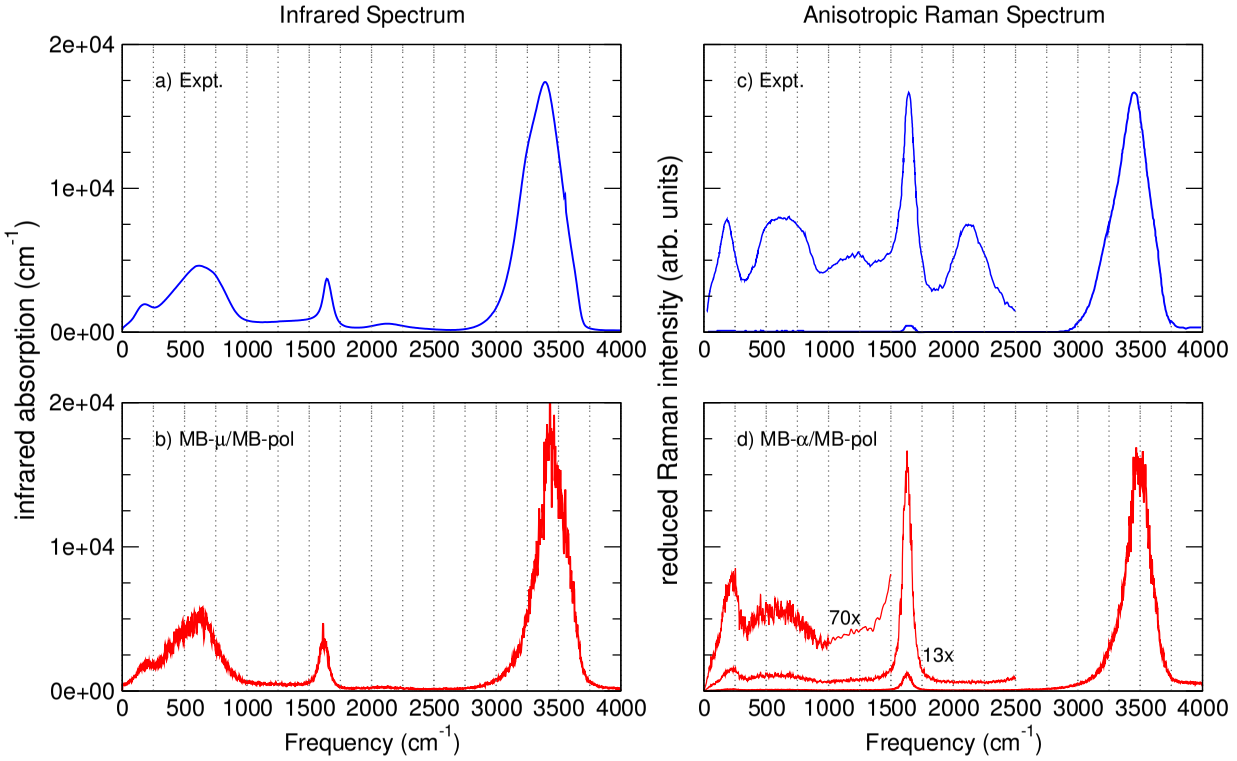

Figure 5.3: IR and Raman spectra of water: MB-MD vs. experiment. Experimental and simulated infrared (panels a and b) and reduced anisotropic Raman spectra (panels c and d) are presented for liquid water at ambient conditions. To clearly demonstrate the line shape of the lower frequency features in the Raman spectra, the features were magnified, as described in the text. The experiment in panel a) and c) are adapted from Refs. 239 and 238, respectively.

the high accuracy of the 1B and 2B polarizabilities, the differences found between the calculated and experimental Raman spectra suggests that effects neglected by MB-α may be important for the low-frequency Raman spectra, such as higher-order many-body effects, frequency-dependence of the polarizability, or higher-order polarizability. In addition, the association band around 2250 cm^{-1} (arising from the combination of librational modes with the bending vibration) is barely visible in the IR spectrum and is missing in the Raman spectrum. It has been suggested that this feature, which may partially be recovered in the local monomer approximation,[135] may not be completely captured through approximate quantum dynamics methods such as CMD.

The position of the OH stretch feature has proven quite difficult to reproduce theoretically due to its anharmonicity and highly quantum-mechanical nature. For instance, empirically parameterized fixed-charge water models typically struggle to capture the average position of the OH band and predict shape and intensity (even after inclusion of quantum correction factors) that are significantly different from the experimental measurements.[49,50] Inclusion of polarization effects improves the agreement, particularly in the intensity,[58,59,223,225] although empirical parametrizations are often still required to correctly predict the position of the OH band. More recently, *ab initio* molecular dynamics simulations have also been employed to model vibrational spectra. Classical MD simulations with GGA functionals predict strongly redshifted bands (upon which nuclear quantum effects can be expected to lead to an additional redshift of 100-150 cm^{-1}).[227–229] Somewhat better results were obtained with hybrid functionals, although both position and intensity of the OH stretch band are generally not well reproduced.[230] The position of the OH stretch band predicted by our MB-MD approach is in good agreement with the corresponding experimental results, with the intensity maxima being slightly shifted to the blue by roughly 40 and 30 cm^{-1} in the IR and Raman spectra, respectively. However, the shape of both the calculated IR and Raman OH bands is somewhat skewed. In

particular, while both the onset 2850 cm^{-1} and the position and shape of the blue side of the band are in good agreement with experiment, the low-frequency portion of the band is lacking in intensity.

To understand the origin of the missing stretch intensity, it is important to consider the interplay between the bend and stretch bands. With a maximum at $\sim$1640 cm^{-1} in both IR and Raman experimental spectra, the bend overtone lies squarely within the OH stretching region. For this reason, it has long been proposed that a Fermi resonance may exist in liquid water between the bend overtone and the OH stretch fundamental, contributing additional intensity to the signal in the strongly H-bonded stretch region. Therefore, in order to correctly reproduce the shape of the OH stretch band, 1) the bending frequency must be correctly positioned such that its overtone coincides with the stretch fundamental, and 2) the dynamical method must be able to describe the Fermi resonance between the two vibrational modes. While it is unclear the extent to which CMD can capture Fermi resonances, the present situation is undoubtedly complicated by the fact that the water bend is slightly red-shifted with respect to experiment, by 28 and 14 cm^{-1} for the IR and Raman spectra, respectively. Given the overall good agreement between the calculated and experimental bending frequency and similar baselines for the calculated and experimental OH bands, it thus seems plausible that the lack of intensity in the red portion of the calculated OH band may be due to quantum dynamical effects associated with the Fermi resonance, rather than to missing hydrogen-bonded configurations.

5.2.4 Characterizing the stretch line shape

To provide further insights into the origin of the OH line shape, the contributions of the potential energy surface and the Fermi resonance are isolated by studying the vibrationally decoupled OD stretch of the dilute HOD molecule in H_2O. Due to the difference in molecular mass, the bend overtone of HOD lies outside the OH stretch

band, removing the effect of the Fermi resonance. In addition, because the O-O radial distribution function given by MB-pol is in excellent agreement with the experimental data, this suggests that the differences in the vibrational structure between simulations and experiment are likely associated to the distribution of the hydrogen atoms, which can be efficiently probed by focusing on the OD vibrational frequencies of HOD in H2O. Therefore, this isotopically dilute system is ideally situated to disentangle dynamical effects from the underlying potential energy landscape.

To assess the accuracy of MB-pol in describing the OD vibrational frequencies of the HOD molecule in H_2O, QM/MM calculations at the MP2/cc-pVTZ level were performed for 2500 cluster configurations extracted from a 10 ps CMD simulation of HOD in H_2O. Following the procedure outlined by the Skinner group,[212] water molecules were assigned to the QM region if their oxygen atoms were within 4.2 Å of the deuterium atom of the central HOD molecule. In addition to these QM H_2O molecules, a-SPC/Fw partial charges placed on the next closest 80 molecules were included in the MM region.[240] For each cluster configuration, the fully anharmonic OD transition frequency from the v=0 to the v=1 energy level was obtained by calculating the potential energy curves associated with the OD stretch and solving the corresponding one dimensional vibrational Schrödinger equation . Analogous QM/MM calculations were carried out using MB-pol to describe the QM region. The resulting in the frequency distributions are shown in Fig. 5.4.

The comparison clearly shows that the frequencies obtained from MB-pol and MP2 are effectively identical. However, the frequency distributions cannot be compared directly to the OD IR or Raman line shape because they lack dynamical effects (e.g., motional narrowing and non-Condon effects[216,217,241]), which are known to red-shift and narrow the underlying distribution. Nonetheless, the close agreement between MB-pol and MP2 provides further evidence that the MB-pol is comparable in accuracy with

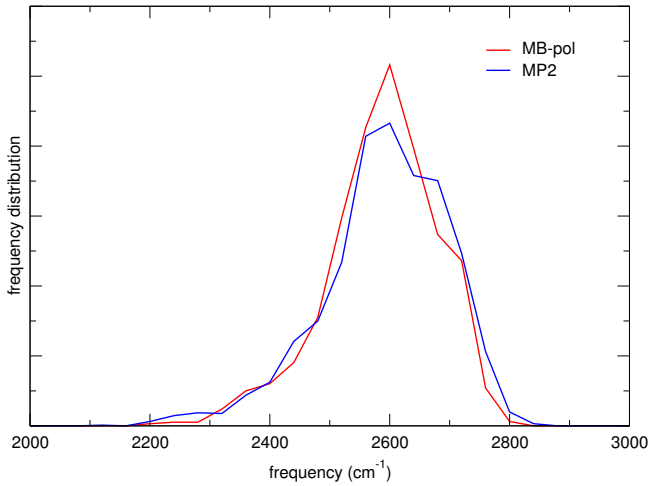

Figure 5.4: Frequency distributions of dilute HOD in H_2O: MB-pol vs. MP2. OD frequency distributions of dilute HOD in H_2O obtained from QM/MM calculations with MB-pol (red) and MP2/cc-pVTZ (blue) used in the QM region.

correlated electronic structure methods, suggesting that the origin of the differences between the calculated and experimental spectra in the OH region lies elsewhere. To enumerate possible sources of differences, we first note that the intramolecular geometry of water molecules in the liquid may be (slightly) different for MB-pol and MP2, resulting in (slightly) different OH vibrational frequencies. These differences cannot be identified in the QM/MM calculations, which probe only one of the intramolecular degrees of freedom. To test this hypothesis, CMD simulations at the MP2 [and possibly CCSD(T)] levels should be carried out. Unfortunately, despite much recent progress,[107] such simulations are currently out of reach. Second, an exact treatment of nuclear quantum dynamics is required to rigorously capture the nature of the OH stretch band. While not yet possible, this is also an active area of research.[242] Finally, the differences between the experimental and simulated line shape could arise from intrinsic limitations of current, state-of-the-art electronic structure methods from which the many-body surfaces are derived.[243,244]

5.3 Summary

Accurately modeling vibrational spectra from first principles requires a robust treatment of quantum dynamical effects, a potential energy surface that gives rise to the vibrational motion of the system, and a faithful description of the dipole moment and polarizability as a function of the system's degrees of freedom (which determine the IR and Raman activity of the underlying vibrations, respectively). In this chapter, we developed first principles models for the dipole moment (MB-μ) and polarizability (MB-α) of water by exploiting the rapid convergence of the many-body expansion of interactions. Centroid molecular dynamics simulations were used to simulate the quantum dynamics of liquid water at ambient conditions using the MB-pol PES. Using MB-μ and MB-α to evaluate the dipole and polarizability time correlation functions, it was found that the both the positions and intensities of the infrared and Raman spectral features are in good agreement with experiment, without relying on any *ad hoc* parameters. While the overall position and intensity of the OH vibrational band are accurate, the line width is somewhat too narrow, lacking intensity on the red side.

To provide insights into the origin of the OH line shape, QM/MM calculations of dilute HOD in H_2O were carried out using MB-pol and MP2 in the QM region. The close agreement between the OD frequency distributions obtained at the MB-pol and MP2 level provides further evidence for the accuracy of MB-pol. This also suggests that the differences between the calculated and experimental spectra may be due to inherent limitations in the electronic structure methods (even at the correlated level)[243,244] and/or an incomplete description of the nuclear quantum dynamics (e.g., combination bands and Fermi resonances). Given that MB-pol, MB-μ, and MB-α are built purely from "first principles" and provide good agreement with correlated electronic structure calculations, this analysis suggests that empirical models that provide the correct OH

line width for H_2O using MD or CMD simulations may benefit from unphysical compensation of effects—artificially encoding quantum dynamical effects into the empirically parametrized interactions. While improvements to electronic structure and quantum dynamics methods may be necessary to quantitatively capture subtle, quantum mechanical effects, the many-body molecular dynamics (MB-MD) approach introduced here represents a powerful, "first principles" method for the study of vibrational spectroscopy.

Portions of this chapter appeared in our previously published work, Medders, G.R.; Paesani, F. "Infrared and Raman spectroscopy of liquid water through first principles many-body molecular dynamics", *J. Chem. Theory Comput.* **2015** 11, 1145–1154. This material was reproduced with permission from the publisher.

Chapter 6

The Role of the Interplay of the Potential Energy and Dipole Moment Surfaces in Controlling the Infrared Activity of Liquid Water

From the theoretical point of view, a consistent interpretation of the vibrational spectra of water has been complicated, in part, by the existence of numerous (both *ab initio* and empirically parametrized) molecular models that can sometimes give conflicting explanations for the same spectroscopic features.[17,23] To address this issue, in the previous chapter summarized our recently introduced unified methodology for modeling molecular systems, denoted many-body molecular dynamics (MB-MD), which is built entirely upon correlated electronic structure data and includes a quantum-mechanical description of the molecular motion.[234] By construction, MB-MD thus enables "first principles" calculations of structural, thermodynamic, dynamical, and spectroscopic properties of molecular systems from the gas to the condensed phase. Through the

MB-MD methodology, empirical parameter-free representations of the two properties that are required for the calculation of IR vibrational spectra, the multidimensional potential energy surface (PES) and dipole moment surface (DMS), are independently derived from large datasets of correlated electronic structure calculations.[234] While dynamical effects, such as motional narrowing, play an important role in determining the final line shapes,[208,245,246] from a static perspective, the IR spectrum can be related to the frequencies of the vibrational modes determined by the multidimensional potential energy surface weighted by the associated squared transition dipoles. In this contribution, we seek to understand the roles of both PES and DMS in determining the IR activity of liquid water by investigating the properties of three different molecular models that have been previously applied to the study of IR spectra.

The first model, q-TIP4P/f, is a flexible point charge force field that was empirically parametrized to reproduce thermodynamic and dynamical properties of bulk water.[49] q-TIP4P/f employs fixed point charges located on the hydrogen atoms and a fictitious "M-site" located along the HOH bisector. Fixed charge models are, however, unable to correctly describe the geometry dependence of the molecular dipole of water, an effect which is important for reproducing the correct IR spectral intensities of water.[90,246] Furthermore, such models predict that molecules with the same internal geometry but which reside in different hydration environments have the same dipole moments, which limits the ability of such models to reproduce the IR absorption intensities.[200,224] Polarizable force fields, on the other hand, allow the molecular charge distributions to adapt to the surrounding electrostatic environments.[247] As a representative example of polarizable force fields, we consider the TTM3-F model,[58] which has been extensively used in simulations of vibrational spectra.[34,219,248] TTM3-F was fitted to a relatively small set of *ab initio* data and employs "Thole type" damped point polarizable dipoles to recover induced dipole effects.[57] Though its associated molecular DMS is based on

an accurate fit to correlated electronic structure data,[116] the TTM3-F model empirically adjusts the gas phase dipole moment to generate an effective DMS that partially accounts for environmental effects in the condensed phase.[58] This is similar to strategies employed in other polarizable models, where the charges were modified to reproduced spectroscopic features.[223,225]

The final model considered here is the many-body MB-pol PES and the associated MB-μ DMS, recently developed by us.[234] Both MB-pol PES[115,122,221] and MB-μ DMS[234] were independently developed by using the many-body expansion of the interaction energy (and its derivatives with respect to electric fields),

$$E_N(x_1,\ldots,x_N) = \sum_a V^{(1\mathrm{B})}(x_a) + \sum_{a<b} V^{(2\mathrm{B})}(x_a,x_b) + \cdots + V^{(\mathrm{NB})}(x_1,\ldots,x_N). \qquad (6.1)$$

Eq. 6.1 is a rigorous representation of the total interaction energy of a generic N-molecule system which is expressed as a sum of one-body (1B) monomer distortion energies, two-body (2B) pairwise interactions, up to N-body interactions.[91] The MB-pol PES relies on explicit fits to a large number of correlated electronic structure data for the 1B, 2B, and 3B interactions, while including all higher-order terms in Eq. 6.1 through many-body induction.[59,115] The even more rapid convergence of the many-body expansions for the water dipole moment and polarizability (established in Ref. 185) allowed us to introduce highly accurate multidimensional representations for these properties (called MB-μ and MB-α, respectively). These many-body representations describe the dipole moment and polarizability of an N-molecule water system in terms of explicit 1B and 2B contributions derived from correlated electronic structure data, with all higher-order interactions being included through many-body induction.[234] The MB-μ and MB-α surfaces were employed in adiabatic centroid molecular dynamics (CMD)[141,237] simulations carried out with the MB-pol PES to model the infrared and Raman spectra of liquid water at ambient

conditions.[234] Detailed comparisons with experimental measurements and electronic structure data obtained at the MP2 level demonstrated the accuracy of the MB-MD methodology in reproducing both positions and shapes all spectral features. The analysis of the stretching band calculated for pure H_2O and a dilute mixture of HOD in H_2O also highlighted some differences in the low-frequency portion of the band ($\sim$3000-3250 cm^{-1}) compared to the experimental line shape. We refer the reader to Ref. 234 for a discussion of possible sources for this missing intensity.

In this contribution, following Ref. 234, the calculation of the relevant quantum dipole autocorrelation function (Eq. 5.11) was carried out within the CMD formalism, which has been shown to provide an approximate yet accurate description of several quantum dynamical properties of liquid water at ambient conditions.[135,216] Specifically, for each model PES described above (q-TIP4P/f, TTM3-F, and MB-pol), path-integral molecular dynamics (PIMD) simulations[137] were initially performed in the NVT ensemble for systems containing 216 molecules at 298.15 K and experimental density. Using 10 different initial conditions obtained from the PIMD simulations, 27 ps adiabatic CMD simulations were then performed in the normal-mode representation. The quantum partition function was discretized using 32 quasiparticles or "beads", and an adiabaticity parameter $\gamma = 0.1$ was used to guarantee a sufficient decoupling between the dynamics of the centroid variables and the non-zero frequency normal modes.[135,237] A time step of 0.02 fs was used to ensure energy conservation and the dynamical trajectories were saved every 2 fs to enable the calculation of the quantum dipole autocorrelation function required by Eq. 5.11. The "time-reversible always stable predictor-corrector" method was used with fourth order extrapolation in the predictor to obtain the induced dipoles.[249]

We begin the analysis of the infrared spectra of water by briefly summarizing the main spectroscopic features of interest: 1) a small shoulder at $\sim$180 cm^{-1} associated with the hydrogen bond stretch corresponding to hindered translations of molecules within the

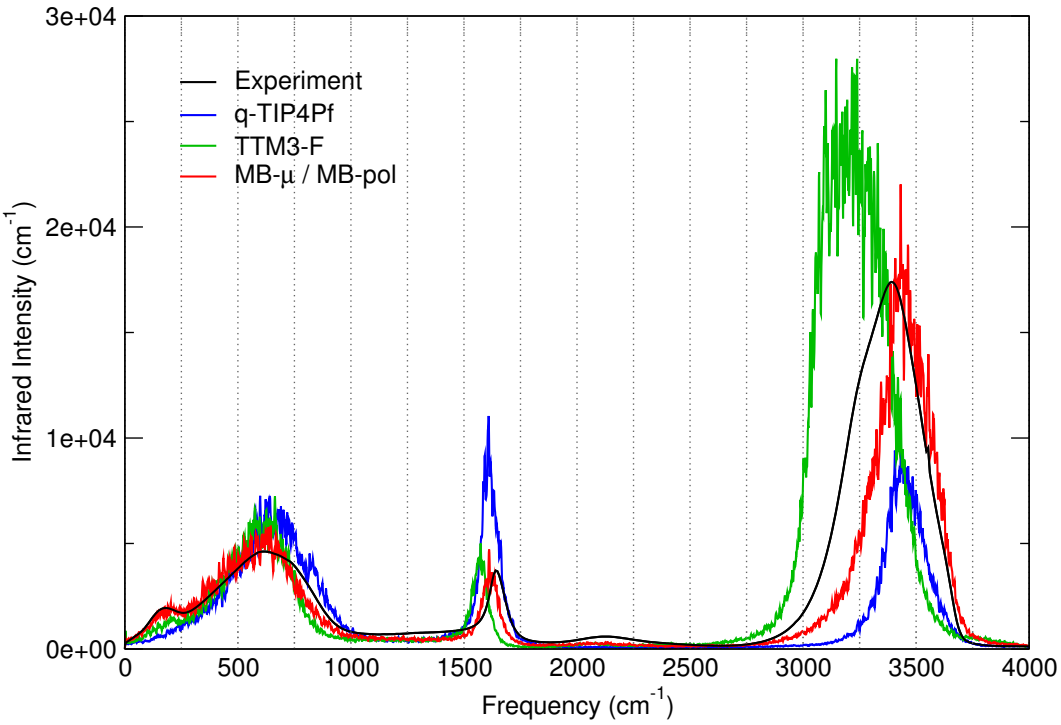

Figure 6.1: IR spectra of liquid water from different models. Specifically, the IR spectra predicted by q-TIP4P/f (blue), TTM3-F (green) and MB-μ/MB-pol (red) are presented. The spectra were calculated through Eq. 5.11, with the ensemble average obtained through centroid molecular dynamics (CMD) simulations performed at 298.15 K. The experimental data (black) are adapted from Ref. 239.

liquid; 2) a broad librational band at $<$1000 cm^{-1} associated with hindered reorientations of molecules within the hydrogen bonding network; 3) a peak at $\sim$1650 cm^{-1}, associated with the HOH bending vibration; 4) a small, broad feature near $\sim$2250 cm^{-1} corresponding to the combination band of the bending vibrations with librations; and 5) a broad feature at 3000-3750 cm^{-1}, corresponding to the OH stretching vibrations. Although the details of these features will be discussed in depth later, it is useful to make some general comments about the performance of the three different water models. As was mentioned earlier, because the q-TIP4P/f molecular dipole moment does not respond to the fluctuating electric field of the surrounding environment, it is to be expected that the infrared intensities predicted by q-TIP4P/f differ significantly from experiment. Despite the fact that the q-TIP4P/f bend and stretch features are thus predicted to have nearly the same intensity, the overall position of these features are in reasonably good agreement with experiment as a result of the empirical parameterization. TTM3-F, which includes induced dipole effects, more accurately predicts the relative intensities for the bending and stretching bands. However, while position and width of the classical TTM3-F stretch and bend features were shown to be in good agreement with experiment,[58] inclusion of nuclear quantum effects results in a significant redshift of both bands.[219] The bending and stretching spectral features predicted by the MB-pol PES in combination with the MB-μ DMS are in good agreement with experiment in terms of their positions, shapes, and intensities, although, as discussed in Ref. 234, the intensity is underestimated on the red portion of the OH stretching band.

While Fig. 6.1 is sufficient to make general statements about the accuracy of the three different water models, it does little to explain what physical aspects of a model make it more or less accurate. Since the infrared activity is related to both the vibrational structure of the energy landscape and the dipole derivatives along those vibrational coordinates, we attempt to better assess the accuracy of each model by decoupling the

effects associated with the corresponding PES and DMS. For this purpose, the dynamical trajectories obtained from the CMD simulations carried out with each model PES are combined in turn with the DMSs associated with the other models. This results in 9 different quantum dipole autocorrelation functions, leading to the 9 IR spectra shown in Fig. 6.2, each of which is labeled according to the notation "DMS / PES".

The hydrogen bond stretch at $\sim$180 cm^{-1} is a key feature to highlight polarization effects on the IR spectra. While all three model PESs are capable of describing hindered translations between hydrogen bonded molecules, the q-TIP4P/f simple point charge representation of the dipole moment, which is independent of the surrounding chemical environment, necessarily generates no IR signal from this purely translational motion.[200] This effect can be clearly seen in the top row of Fig. 6.2 (corresponding to the q-TIP4P/f, TTM3-F, and MB-pol CMD trajectories combined with the q-TIP4P/f DMS), where no distinct feature corresponding to the hydrogen bond stretch can be identified in the region between 0 and 250 cm^{-1}. When the CMD trajectories are combined with the TTM3-F DMS (middle row of Fig. 6.2), some enhancement in the 0-250 cm^{-1} region is observed, particularly in the case of MB-pol trajectories (panel *f*), although the hydrogen bond stretch is still not well resolved with any of the combinations of the TTM3-F DMS with the three PESs. On the other hand, the hydrogen bond stretch feature is found to be in good agreement with experiment when the MB-μ DMS is combined with the CMD trajectories performed with the MB-pol PES.

The different shapes of the hydrogen bond stretch features of the MB-pol PES identified in panels *f* and *i* can be traced back to (at least) two main differences between the MB-μ and TTM3-F DMSs. First, the TTM3-F DMS, which has only a single (isotropic) polarizable site on each molecule, is unable to completely capture the anisotropy of the induced dipole along the hydrogen bond stretching coordinate. Second, the TTM3-F DMS only accounts for induced dipoles, neglecting charge transfer effects

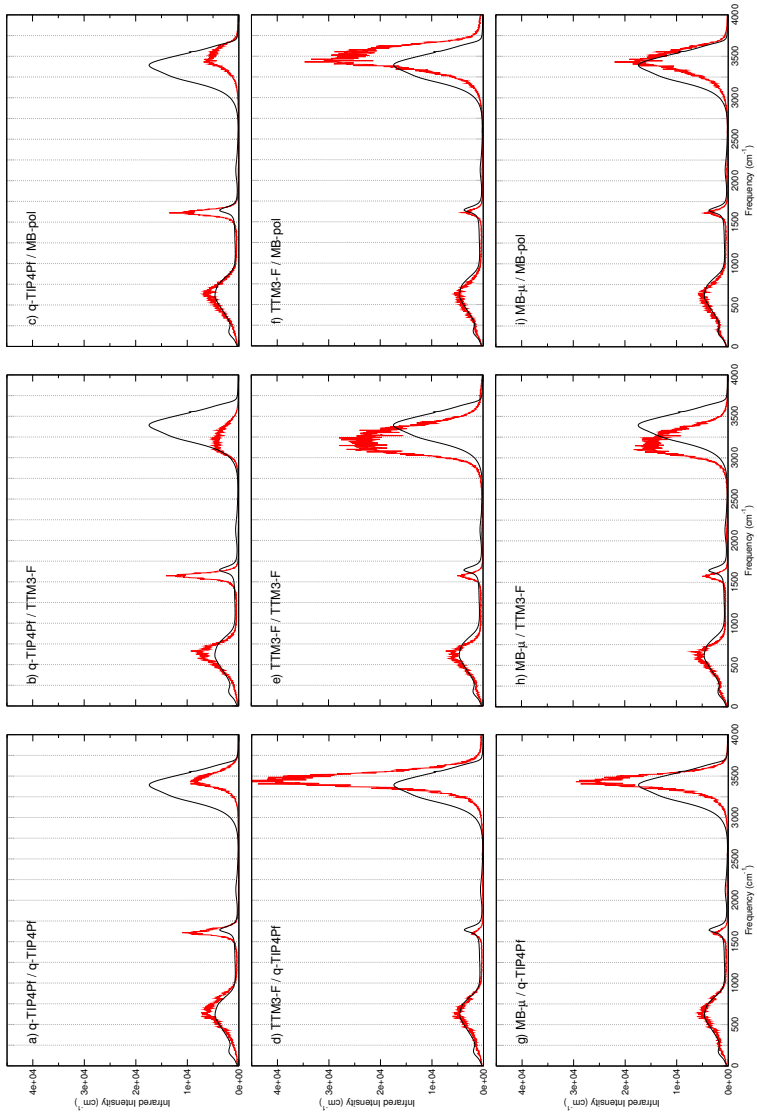

Figure 6.2: Role of the PES and DMS in controlling IR activity. The IR spectra are obtained from different combinations of PES and DMS. The panels are labeled using the notation "DMS / PES". Each row thus displays the IR spectra calculated with the same model DMS combined with CMD trajectories obtained with the three model PESs. Analogously, each column displays the IR spectra calculated with the different model DMSs combined with CMD trajectories obtained with the same model PES. For example, in panel d), the IR spectrum is calculated using the DMS of the TTM3-F model combined with the CMD trajectory obtained with the q-TIP4P/f model.

which have been proposed to be an important source of spectral intensity for the hydrogen bond stretching vibration in Raman calculations.[202] The MB-μ DMS, on the other hand, includes both the anisotropy of the 2B induced dipole and all charge transfer effects, which are accurately modeled at the MP2 level. Unfortunately, a precise decomposition of the different electronic contributions to the IR intensity associated with the hydrogen bond stretching vibration can not be obtained through the analysis of the TTM3-F and MB-μ DMSs since such analysis requires models that are capable of accurately and explicitly distinguishing charge transfer from induction. It can also be noted that the hydrogen bond stretch of both the q-TIP4P/f and TTM3-F PESs are blue shifted (by 80 and 30 cm^{-1}, respectively) relative to experiment when calculated using the MB-μ DMS, indicating that the hydrogen bonds in these models are possibly too strong. The librational band is well described by both q-TIP4P/f and MB-pol PESs when the accurate many-body MB-μ DMS is used (panels *g* and *i*), while it appears to be slightly too narrow in the case of the TTM3-F PES (panel *h*). By contrast, both q-TIP4P/f and TTM3-F DMSs tend to overestimate the intensity of the librational band, particularly near the maximum at $\sim$600 cm^{-1}.

Examining the spectral features associated with the intramolecular bending and stretching vibrations, it is clear that the energy landscape defined by the individual PESs primarily controls the position of these two bands. To understand the spuriously similar intensity of the spectral features associated with the bending and stretching vibrations in the q-TIP4P/f spectrum shown in Fig. 6.1, the MB-μ DMS is applied to the q-TIP4P/f PES in panel *g* of Fig. 6.2. The overestimated intensity of the OH stretch band when using an accurate DMS indicates that the q-TIP4P/f PES may not be capable of accurately describing the dependence on the intramolecular stretching vibrations on strength of the intermolecular hydrogen bonds. In particular, the high infrared intensity of the OH stretching band may suggest that the underlying frequency distribution of

the q-TIP4P/f OH stretch is too narrow, implying that the q-TIP4P/f OH oscillators either sample a restricted range of hydrogen bonding environments or that the two OH oscillators of each molecule do not sufficiently couple to one another to acquire partial symmetric/asymmetric splitting.[250]

At the same time, the q-TIP4P/f DMS also suffers from deficiencies. An accurate 1B representation of the DMS should have a low intensity in the OH stretching region (see Fig. 6.3). Despite the fact that it uses a purely 1B representation of the DMS, the q-TIP4P/f DMS applied to the MB-pol PES results in an artificially enhanced height of the OH stretch band. The electrostatic origin of this spuriously large intensity can be more clearly seen in the bending region. Regardless of which PES is used, the q-TIP4P/f DMS always predicts a high bend intensity. However, applying either the TTM3-F or the MB-μ DMS to the q-TIP4P/f PES brings the bend intensity into reasonable agreement with experiment. This high intensity caused by the q-TIP4P/f DMS would likely be observed in all fixed point charge models of water which represent the enhanced condensed-phase dipole through the use of artificially large effective charges. Through the analysis of the q-TIP4P/f infrared line shape, it can thus be seen that decoupling the influence of the PES from the DMS in the calculation of infrared spectra provides specific insights that are useful to assess the ability of a model to correctly reproduce the underlying physics.

To further investigate the role of the DMS in determining the position and shape of the OH stretching band, we performed a many-body decomposition of the dipole moment using the MB-μ DMS. Specifically, for the set of CMD trajectories carried out with the MB-pol PES, the total dipole moment was calculated according to three different approximations: the 1B approximation (i.e., only including all molecular gas-phase dipoles in the geometry of the liquid), the 1B+2B approximation (i.e., adding all short-ranged pairwise interaction-induced dipoles to the 1B total dipole moment), and the MB approximation using with the full MB-μ DMS (i.e., including 3B and higher-order induced

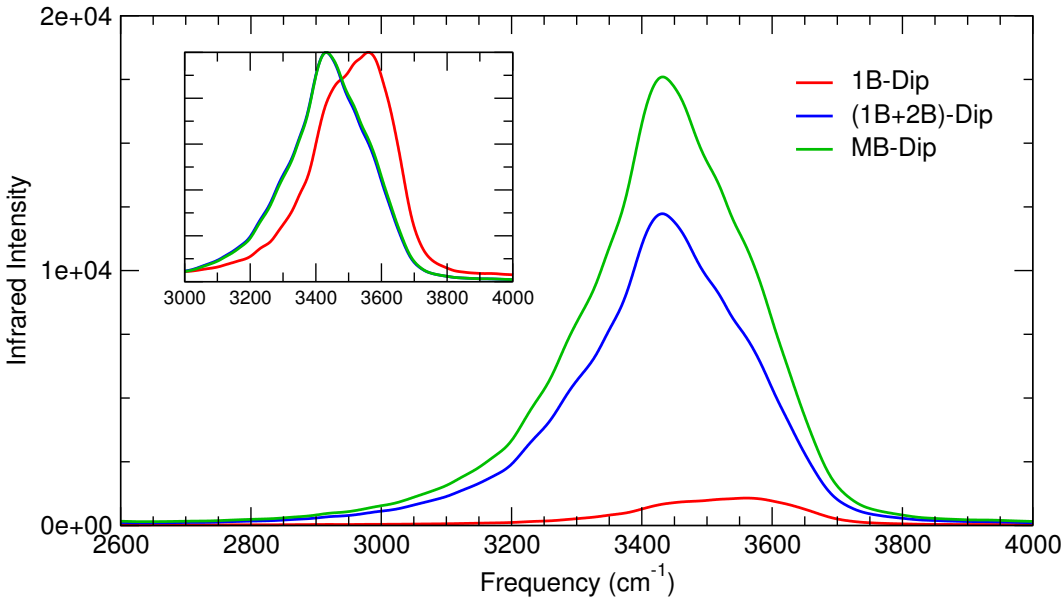

Figure 6.3: Many-body dipole contributions to the IR spectrum of water. Specifically, the decomposition of the IR spectrum obtained from CMD trajectories with the MB-pol PES was performed in terms of the many-body components of the MB-μ DMS. 1B-Dip indicates that the one-body (gas-phase monomer) dipoles were used to calculate the dipole of the molecules sampled along the MB-pol CMD trajectories, from which the IR spectrum was calculated. (1B+2B)-Dip indicates that short-ranged two-body dipoles were used in addition to the one-body dipoles. MB-Dip is the full MB-μ many-body dipole. The inset shows each curve normalized to its corresponding maximum. The spectra were smoothed to facilitate the comparison between the line shapes obtained using the different approximations.

dipoles to the 1B+2B total dipole). Focusing on the OH stretch band shown in Fig. 6.3, the 1B dipole approximation predicts a very weak OH stretch intensity. This result is not unexpected since the average dipole moment of a water molecule increases significantly upon solvation relative to its gas-phase value. Therefore, neglecting all solvation-induced effects, the 1B approximation consistently underestimates the infrared intensity. Inclusion of the short-ranged (i.e., for pairs of molecules with an oxygen-oxygen separation of < 3.7 Å) pair-wise interaction-induced dipoles results in a substantial increase in the OH stretch intensity, with the full many-body dipole contributing a slight additional enhancement and bringing the total intensity into good agreement with experiment. The many-body dipoles also have an additional effect on the position of the OH stretching band as demonstrated in the inset of Fig. 6.3 by normalizing 1B, 1B+2B, and MB bands to the corresponding maxima. While the 2B dipole enhances the overall intensity, it does so preferentially for OH stretches in more strongly H-bonded configurations. In general terms, this is consistent with the picture that the magnitude of the change of the dipole moment with respect to a vibrational coordinate is much larger when the OH is vibrating into a nearby H-bond acceptor relative to its rate of change in a weak or broken H-bond configuration. The many-body analysis shown in Fig. 6.3 thus demonstrates that this "non-Condon" effect plays an important role at the 2B dipole level,[241] with all higher-order induced dipoles having a negligible impact on the position of the OH stretching band.

It is often assumed that the underlying multidimensional PES is primarily responsible for the position (and, to a large extent, the shape) of the infrared spectrum of liquid water. There are, however, interesting cases for which this first order assumption breaks down, namely, where there exists a non-uniform distribution of dipole derivative magnitudes that can result in a preferential enhancement of some portions of the line shape with respect to others. Here, we illustrated an excellent example of this through

the analysis of the hydrogen bond stretch feature at $\sim$180 cm^{-1}. While non-existent in intensity for models which have no induced, many-body dipoles, the peak appears (although is not well-resolved) for a non-polarizable PES when combined with a many-body DMS. Ultimately, an accurate representation of both the potential energy surface and the dipole moment surface is necessary to obtain good agreement with the experimental infrared spectrum of liquid water. By examining the detailed many-body contributions of the dipole to the infrared spectrum, the line shape of the OH band is shown to be well described through a sum of molecular dipoles and short-ranged two-body dipoles. This supports an interpretation of the IR intensity of the OH band of liquid water as a probe of the local hydration environment. A question that remains to be addressed is the extent to which the underlying frequency distribution is affected by delocalized collective vibrations.

Portions of this chapter appeared in our previously published work, Medders, G.R.; Paesani, F. “On the interplay of the potential energy and dipole moment surfaces in controlling the infrared activity of liquid water”, *J. Chem. Phys.* **2015** 142, 212411. This material was reproduced with permission from the publisher.

Chapter 7

Conclusions and Future Directions

Perhaps due to its ubiquity and its simple molecular structure, it may surprise to the reader to learn that many aspects of water remain poorly understood. In fact, the literature on water properties contains numerous controversies, ranging from such fundamental topics as the structure of liquid water to the proposed existence of a critical point in supercooled water.[251–253] While theory and molecular simulation can be useful tools to explain and predict these phenomena, many of the controversies are instead exacerbated by conflicting results obtained from computational studies themselves. An intrinsic issue to results obtained from molecular simulations is the choice of the potential energy surface (PES) that governs the system. Common strategies for modeling the PES, as discussed in Section 1.3.1, involve the parametrization of effective interactions to empirical data; however, the transferability of such a strategy to water in different phases or under different conditions is clearly questionable. Alternative approaches such as *ab intio* molecular dynamics attempt to solve the electronic structure problem on the fly; however, such simulations are currently limited in accuracy both due to intrinsic deficiencies in the electronic structure method and their computational cost, which can severely affect the statistical convergence of such results.

The aim of this work is to present an alternate approach to studying systems from gas to condensed phases by the development of computationally tractable yet purely "first principles"-based models for the potential energy surface of water. Relative to conclusions derived from empirically tuned force fields, there is comparably less room to dispute results derived from highly-correlated electronic structure calculations. To circumvent the (often prohibitive) cost of calculating *ab initio* energies for condensed phase systems, the total (binding) energy can be instead expressed as a many-body expansion of the intermolecular interactions (see Section 1.2):

$$E(1,\ldots,N) = \sum_{i}^{N} V^{1B}(i) + \sum_{i<j}^{N} V^{2B}(i,j) + \sum_{i<j<k}^{N} V^{3B}(i,j,k) + \cdots + V^{NB}(1,\ldots,N). \quad (7.1)$$

In systems for which it converges quickly, such as water, the many-body expansion provides a route through which the cost of the electronic structure calculations can mitigated.

In Chapter 1, we assessed the ability of different strategies (including polarizable and non-polarizable force fields, semi-empirical methods, density functional theory models, and MP2) to reproduce thousands of CCSD(T) reference two- and three-body interaction energies,[91] leading us to develop an exploratory model, HBB2-pol, with the accuracy of CCSD(T) but with efficiency comparable to a polarizable model.[91,98] Building on that work, Chapter 2 describes our refined model, named MB-pol, that consistently employs an accurate description of dispersion and (many-body) electrostatic interactions combined with short-ranged terms at the two- and three-body, which account for short-range quantum mechanical effects.[115,122,221] Importantly, MB-pol reproduces experimental measurements of cluster properties and, as shown in Chapter 3, thermodynamic and dynamical properties of bulk water at ambient conditions, without relying on any empirical parametrization to experiment.[115,122,221]

However, unlike the electronic structure calculations to which MB-pol was fitted, the PES has no knowledge of the electron distribution and, for this reason, the many-body PES can not used to calculate electrostatic properties such as the dipole moment or polarizability, which are required to obtain the IR and Raman spectra. In Chapter 4, we demonstrated that the many-body expansions of the dipole and polarizability, defined through derivatives of Eq. 7.1 with respect to electric fields, also converge for water.[185] From this finding, in Chapter 5 we built many-body models for the dipole moment surface and polarizability surface of water, that were parametrized to large datasets of highly correlated electronic structure calculations.[234] Importantly, since IR and Raman spectra can be calculated through quantum time correlation functions of dipole and polarizability, respectively, which can be approximate calculated through centroid molecular dynamics simulations using the MB-pol PES, this many-body molecular dynamics (MB-MD) approach allows us to rigorously model vibrational spectra from first principles.[234,254] Finally, in Chapter 6, the MB-MD approach was employed to decouple the contributions of the potential energy surface and the dipole moment surface from the IR spectrum of liquid water.

While the work presented so far has been applied only to linear vibrational spectra and to systems containing only water molecules, the simulation strategies developed here are extendable to a much broader class of problems. We take this opportunity to briefly highlight ongoing work in one such direction.

7.1 Application of MB-MD to modeling vibrational sum frequency generation spectroscopy

In recent years, vibrational sum frequency generation (vSFG) spectroscopy has emerged as a potentially powerful tool for probing the molecular structure and dynamics

of interfaces. As a second-order nonlinear spectroscopy, vSFG is intrinsically sensitive to interfaces, with its signal being generated by vibrational modes that are both IR and Raman active and which reside in non-centrosymmetric environments. While vSFG provides an experimentally measurable probe of molecular motions in the first few monolayers of an interface,[23] resolving the spectral features into a molecular level picture poses substantial difficulties. Recently developed phase sensitive experimental vSFG techniques allow for the direct measurement of the imaginary component of the resonant nonlinear susceptibility, which can be calculated from the time-correlation function:[255]

$$\chi^{R,(2)}_{ijk}(\omega) = \frac{i\omega}{k_B T}\int_0^\infty dt e^{-i\omega t}\langle \alpha_{ij}(t)\mu_k(0)\rangle. \tag{7.2}$$

Here, ω is the frequency, $\chi^{R,(2)}$ is the resonant (vibrationally enhanced) SFG susceptibility, and i,j,k are the components related to the polarization conditions. In the SSP polarization, the sign of the imaginary part of Eq. 7.2 is related to the direction of the transition dipole at a given vibrational frequency.

Presented in black in Fig. 7.1 is the experimental spectrum of $\mathrm{Im}[\chi^{R,(2)}_{ijk}(\omega)]$ of the air/water interface. The narrow and positive feature at ~ 3700 cm^{-1} is referred to as the "free OH" peak, corresponding to OH stretches that are pointed out of the bulk. The broad, negative portion of the spectrum from 3200–3600 cm^{-1} correspond to the stretches of hydrogen-bonded water molecules pointed into the bulk. One particularly controversial aspect of the imaginary component of the χ^R is the explanation for the low-frequency positive-going feature in the OH stretching region of the air/water interface (between 3000 to 3200 cm^{-1}). Insights from computer simulations into the molecular level origin of this feature offer (at least) two distinct explanations: one attributes this feature to the presence of specific configurations related to three-body interactions;[23] the other attributes the positive intensity to an anisotropic induced dipole of strongly hydrogen-

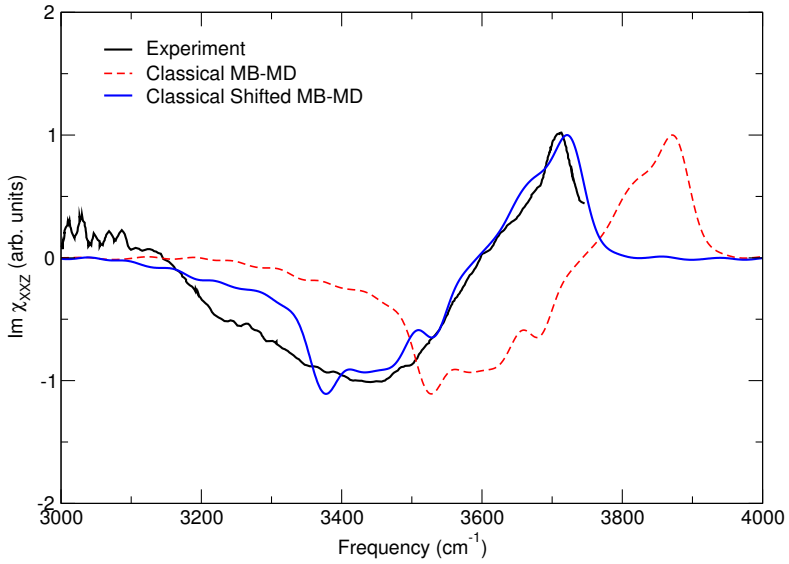

Figure 7.1: Sum frequency generation spectra of the air/water interface. The SSP polarization condition is examined. Black is the experimentally measured spectrum of Ref. 17. The red dashed line is the classical vSFG spectrum obtained using the MB-pol PES and the MB-μ and MB-α dipole and polarizability, respectively, in Eq. 7.2. The blue line is the classical result shifted by 150cm^{-1} to roughly account for nuclear quantum effects.

bonded interfacial water molecules.[17] Importantly, however, both sets of simulations employ some degree of empiricism in modeling the positive feature of the vSFG spectra. If a set of models could be developed and applied successfully to simulating the vSFG without including any *a priori* knowledge of the spectrum, such a result would perhaps provide some chance of resolving the controversies surrounding the molecular level origins of the vSFG.

For this reason, here we rely on our three "first principles"-based models that have been demonstrated accurately recover the total (binding) energy, dipole moment, and polarizability of arbitrary water systems with respect to highly correlated electronic structure methods The dashed red lines in Fig. 7.1 correspond to the SFG spectrum obtained with classical MD simulations of the MB-pol PES. In these simulations, the system was composed of 512 molecules in a slab geometry for a 26x26x100 box in periodic boundary conditions. The water molecules formed a stable slab at the center of

the cell when equilibrated in the NVT ensemble at 298.15 K. Over the course of a 2.5 ns NVT simulation, 415 initial conditions were extracted. For each initial condition, a 50 ps NVE simulation was performed. The results $\sim$ 20 ns of NVE trajectories were used to calculate the correlation function of Eq. 7.2 corresponding to the SSP polarization. The total SSP correlation function was defined as

$$\frac{\langle \alpha_{xx}(t)\mu_z(0)\rangle + \langle \alpha_{yy}(t)\mu_z(0)\rangle}{2} \tag{7.3}$$

where the z direction is perpendicular to the interface and the averaged xx and yy components of the polarizability were used to exploit the symmetry in the plane of the air/water interface. Due to the slab configuration and the signed nature of the dipole moment, the z component of the dipole for molecules in the lower half of the slab was mirrored following the protocol of Ref. 26. Finally, while the full MB-μ model was used in the calculation of the dipole, the NB contribution to the polarizability was neglected. While this approximation will affect the overall intensity of the spectral features, it is not expected to alter the lineshape.[234] Furthermore, since the SFG spectra are reported in arbitrary units, neglecting α^{NB} is expected to have negligible impact on the results.

As discussed in Chapter 5, vibrational spectra of bulk water in the OH stretching region obtained from classical MD simulations are artificially blue shifted with respect to those obtained from CMD. For both IR and Raman of bulk water obtained from MB-MD, the OH stretching band obtained from classical MD is artificially blue shifted by $\sim$ 150cm^{-1} relative to CMD. To roughly account for these missing nuclear quantum effects in the SFG results obtained with classical MD simulations, the blue line of Fig. 7.1 has been rigidly shifted 150cm^{-1}. While the net role of nuclear quantum effects on the lineshape may be different between the bulk liquid and the air/water interface, this simplistic shift at least provides an initial point of comparison between the

classical spectra and experiment. As shown in Fig. 7.1, both the MB-MD linewidth of the free OH peak and position/depth of the H-bonded region around 3400cm^{-1} are in very good agreement with experiment. While the intensity dies off more quickly than experiment in the region from 3200-3350cm^{-1}, the extent of the negative-going feature is in qualitatively good agreement with experiment.

Curiously, no positive signal is observed in the very low frequency part of the OH stretching region. This disagreement may arise from approximating the quantum spectrum as a shifted classical spectrum. CMD simulations of the air/water interface are currently underway. Importantly, however, since the models employed here were fitted only to correlated electronic structure calculations and, by construction, accurately describe the two-body induced electrostatic properties at the MP2/aug-cc-pVTZ level and the three-body intermolecular interactions at the CCSD(T)/aug-cc-pVTZ level, these simulations should ultimately shed light on the controversial hypotheses for the molecular level origins of the Im$[\chi^{R,(2)}_{ssp}(\omega)]$ of the vSFG of the air/water interface.

7.2 Future Directions

While the research presented in this dissertation has already been useful itself in probing questions ranging from the vibrational structure of liquids to the isomeric stability of small clusters, perhaps the most exciting aspect of this work is the future studies it enables. The development of empirical force fields for molecular simulation has long been fraught with difficulty. Furthermore, while great promise exists for *ab initio* molecular dynamics approaches (enabled, in part, due to ever-increasing computational power), such simulations currently afford limited predictive ability due to both inaccuracies in commonly used density functionals and current limitations in sampling ability.

On the other hand, by exploiting the convergence of the many-body expansion

of interactions, we have demonstrated the ability to predict the properties of water in environments ranging from small gas phase clusters to the liquid – without any empirically derived parameters. While the "proof-of-concept" work presented here was applied only to pure water system, these techniques are expected to be equally applicable to any system for which the many-body expansion converges. This, in turn, will open the doorway to studying problems ranging from solvation of ions to the dynamics of molecules at atmospherically relevant interfaces with predictive accuracy.

Appendix A

Basis-set superposition error in the many-body expansion

While the many-body expansion of interaction energies provides a computationally tractable approach for accurately characterizing molecular systems, the accurate calculation of the many-body terms can be somewhat tricky.[256] A primary issue involves the treatment of basis-set superposition error, which arises from an incomplete description of the wavefunction or density. In the classic example of the dimer interaction energy, Eq. 1.6, basis-set superposition has the consequence of over-stabilizing the dimer. This is because the dimer calculation effectively has access a larger basis set than the isolated monomers, resulting in a artificially lower energy within variational methods.[103] The counterpoise correction of Boys and Bernardi proposes an intuitive solution to this issue;[103] rather than calculating the monomer energies in their own basis sets, each monomer calculation is instead performed in the basis set of the dimer, eliminating any inconsistency in the size of the basis sets:

$$V^{2B}(1,2) = E(1,2) - E(1;G_2) - E(2;G_1). \tag{A.1}$$

In this notation, $E(1;G_2)$ indicates a calculation of the cluster composed of energy of fragment 1 in presence of the "ghost" orbitals of fragment 2 (but without the electrons and nuclear charges of fragment 2).

Unfortunately, however, there is no unique way to account for BSSE within the many-body expansion beyond the two-body level. For this reason, several strategies exist to treat BSSE for systems containing many fragments.

Site-site function counterpoise method:

The site-site function counterpoise (SSFC) method treats BSSE consistently at all levels of the many-body expansion by performing each energy evaluation in the basis set of the full system.[120] The many-body terms become:

$$\begin{aligned} V^{1B} =& E(i;G_{jkl\ldots}) \\ V^{2B} =& E(ij;G_{kl\ldots}) - E(i;G_{jkl\ldots}) - E(j;G_{ikl\ldots}) \\ V^{3B} =& E(ijk;G_{l\ldots}) - \sum_{i<j}^{3} V^{2B}(ij) - \sum_{i}^{3} V^{1B}(i) \\ =& E(ijk;G_{l\ldots}) - \left[E(ij;G_{kl\ldots}) + E(ik;G_{jl\ldots}) + E(jk;G_{il\ldots})\right] \\ &+ E(i;G_{jkl\ldots}) + E(j;G_{ikl\ldots}) + E(k;G_{ijl\ldots}). \end{aligned} \tag{A.2}$$

As the complete basis-set limit is approached and the effect of the ghost orbitals disappears (e.g., $E(i;G_{jkl\ldots}) \rightarrow E(i)$), the original many-body terms are obtained.

An important property of the many-body expansion is that each interaction is defined only in terms of the N-molecules required to specify the N-body interaction. In contrast, while the SSFC method systematically corrects for BSSE at all levels of the many-body expansion, it also has the unfortunate effect of making each term in the many-body expansion depend on the orbitals of the complete cluster. For example, to obtain the 1B energies of molecules embedded in a larger cluster with the SSFC method,

one must calculate the energy for each monomer in the ghost orbitals of the complete cluster, even if every monomer in the system has the same geometry!

Valiron-Mayer Function Counterpoise Method:

To avoid the issue of N-body interactions depending on more than N molecules, one could envision an alternative approach where the orbitals of all molecules forming the N-body interactions are always used in the calculation of that interaction, but where the orbitals of all other molecules are neglected. This approach, known as the "Valiron-Mayer Function Counterpoise" (VMFC) method,[183,184] can perhaps be more easily understood through an example (adapted from Ref. 31). Consider the 3B interaction from molecules 2, 4, and 6 within a hexamer; the full SSFC corrected 3B interaction (which requires that the 3B interaction be calculated in the hexamer basis), could be approximated by neglecting the effects of molecules 1, 3, and 5 and performing an SSFC corrected 3B calculation in the trimer basis formed by molecules 2, 4, and 6. Thus, within this VMFC framework, the BSSE-corrected many-body interactions are:

$$V^{1B} = E(i) \tag{A.3}$$

$$\begin{aligned} V^{2B} =& E(ij) - E(i;G_j) - E(j;G_i) \\ V^{3B} =& E(ijk) - \left[E(ij;G_k) + E(ik;G_j) + E(jk;G_i)\right] \\ &+ E(i;G_{jk}) + E(j;G_{ik}) + E(k;G_{ij}). \end{aligned}$$

$$\ldots \tag{A.4}$$

While this approach solves the "embedded many-body" problem of the SSFC method, when one attempts to reconstruct the total energy of the cluster from the VMFC approach,

however, the following is obtained (for an example trimer):

$$
\begin{aligned}
E(ijk) &= \sum_{i}^{3} V^{1B}(i) + \sum_{i<j}^{3} V^{2B}(ij) + V^{3B}(123) \qquad \text{(A.5)} \\
&= \big[E(i)+E(j)+E(k)\big] \\
&\quad + \Big[\big[E(ij)-E(i;G_j)-E(j;G_i)\big] + \big[E(ik)-E(i;G_k)-E(k;G_i)\big] \\
&\qquad + \big[E(jk)-E(j;G_k)-E(k;G_j)\big]\Big] \\
&\quad + \Big[E(ijk) - \big[E(ij;G_k)+E(ik;G_j)+E(jk;G_i)\big] \\
&\qquad + E(i;G_{jk})+E(j;G_{ik})+E(k;G_{ij})\Big].
\end{aligned}
$$

Collecting the terms by interaction order, it is easily seen that the total energy is only recovered when:

$$
\begin{aligned}
E(i)+E(j)+E(k) = &\big[E(i;G_j)+E(i;G_k)+E(j;G_i)+E(j;G_k)+E(k;G_i)+E(k;G_j)\big] \\
&- \big[E(i;G_{jk})+E(j;G_{ik})+E(k;G_{ij}\big] \qquad \text{(A.6)}
\end{aligned}
$$

and

$$
E(ij)+E(ik)+E(jk) = \big[E(ij;G_k)+E(ik;G_j)+E(jk;G_i)\big]. \qquad \text{(A.7)}
$$

These conditions are only satisfied in the CBS limit, implying that the VMFC method does not maintain the relationship between the many-body expansion and the cluster (binding) energy, except in the CBS limit. Therefore, while the VMFC method decreases the sensitivity of the many-body calculations to basis-set truncation, one must nonetheless exercise care to the many-body interactions are calculated in the CBS limit.

Cluster counterpoise method:

Recently, an alternative formulation of the counterpoise method has been intro-

duced, which preserves the connection between the many-body interactions and the total energy of the system. In this "cluster counterpoise" method, the BSSE-corrected interaction energy of an N molecule system is expressed as:

$$E_{N-mer}^{CCP}(1,\ldots,N) = E(1,\ldots,N) - \sum_{i}^{N}\left[E(i;G_{1,\ldots,N;\neq i}) - E(i)\right] \tag{A.8}$$

Using Eq. 1.5 in combination with the above to generate the BSSE-corrected many-body interactions, one obtains:

$$V^{1B} = E(1) \tag{A.9}$$

$$\begin{aligned} V^{2B} =& E^{CCP}(12) - \sum_{i}^{2} V^{1B}(i) \\ =& \Big[E(12) - \big[E(1;G_2) - E(1)\big] - \big[E(2;G_1) - E(2)\big]\Big] - E(1) - E(2) \\ =& E(12) - E(1;G_2) - E(2;G_1) \end{aligned} \tag{A.10}$$

$$\begin{aligned} V^{3B} =& E^{CCP}(123) - \sum_{i<j}^{3} V^{2B}(ij) - \sum_{i}^{3} V^{1B}(i) \\ =& \Big[E(123) - \big[E(1;G_{23}) - E(1)\big] - \big[E(2;G_{13}) - E(2)\big] - \big[E(3;G_{12}) - E(3)\big]\Big] \\ & - \Big[\big[E(12) - E(1;G_2) - E(2;G_1)\big] + \big[E(13) - E(1;G_3) - E(3;G_1)\big] \\ & \quad + \big[E(23) - E(2;G_3) - E(3;G_2)\big]\Big] \\ & - \Big[E(1) + E(2) + E(3)\Big] \\ =& \Big[E(123) - E(1;G_{23}) - E(2;G_{13}) - E(3;G_{12})\Big] \\ & - \Big[\big[E(12) - E(1;G_2) - E(2;G_1)\big] + \big[E(13) - E(1;G_3) - E(3;G_1)\big] \\ & \quad + \big[E(23) - E(2;G_3) - E(3;G_2)\big]\Big] \end{aligned} \tag{A.11}$$

While requiring more calculations than either the SSFC or VMFC approaches, this

definition may seem preferable because the V^{1B}, V^{2B}, V^{3B} expressions still revert to the correct expressions in the CBS limit and each N-body calculation naturally requires only the basis set of the N-mer. Furthermore, regardless of the system size, the BSSE-corrected binding energy of the cluster is recovered when many-body interactions are summed. This approach was used by us in our 2013 paper,[91] although the details of the method were outside the scope of that article. Around the same time, this method was independently developed and reported by the Herbert group.[257]

In summary, since the counterpoise method of Boys and Bernardi cannot be uniquely adapted to treatment of BSSE beyond the 2B level, many method exist for the correction of BSSE. To these ends, the following general guidelines may help the reader choose what variant of the counterpoise correction may be most appropriate for a given problem.

- In situations where a quick and accurate estimate of a cluster's interaction energy are desired, but where the accuracy of each individual many-body interaction is of secondary importance, the cluster counterpoise method is an appropriate choice.[257]

- If both accurate many-body interactions and the correct interaction energy are desired, the SSFC method is clearly the best choice. However, the computational expense of this method significant because the SSFC many-body interactions depend on the geometry of the total cluster. Taking a 20 molecule system as an example, the three-body interaction of molecules ABC must be calculated in the basis of the 20-mer. As a consequence, if the position of even a single atom of one of the 17 other molecules is changed, the SSFC three-body interaction of ABC is, by definition, different.[120]

- The VMFC enables the calculation of accurate many-body interactions that do not depend on the (larger) cluster of interest. However, one must be very careful when

using this approach because the summed VMFC many-body interactions do not give the correct interaction energy except in the complete basis limit.[183, 184]

Bibliography

[1] Y. Maréchal, *The Hydrogen Bond and the Water Molecule: The Physics and Chemistry of Water, Aqueous and Bio-Media*. Elsevier, Amsterdam (2006).

[2] A. P. Ault, T. L. Guasco, O. S. Ryder, J. Baltrusaitis, L. A. Cuadra-Rodriguez, D. B. Collins, M. J. Ruppel, T. H. Bertram, K. A. Prather, and V. H. Grassian, *J. Am. Chem. Soc.*, **135**, 14528–14531 (2013).

[3] D. J. Tobias, A. C. Stern, M. D. Baer, Y. Levin, and C. J. Mundy, *Annu. Rev. Phys. Chem.*, **64**, 339–59 (2013).

[4] A. J. Hopkins, C. L. McFearin, and G. L. Richmond, *J. Phys. Chem. C*, **115**, 11192 (2011).

[5] A. M. Jubb, W. Hua, and H. C. Allen, *Annu. Rev. Phys. Chem.*, **63**, 107–130 (2012).

[6] L. P. DeFlores and A. Tokmakoff, *J. Am. Chem. Soc.*, **128**, 16520–16521 (2006).

[7] S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. M. Leitner, and M. Havenith, *Proc. Natl. Acad. Sci.*, **104**, 20749–52 (2007).

[8] . Szyc, M. Yang, E. T. J. Nibbering, and T. Elsaesser, *Angew. Chem. Int. Ed.*, **49**, 3598–3610 (2010).

[9] S. M. Gruenbaum and J. L. Skinner, *J. Chem. Phys.*, **135**, 075101 (2011).

[10] J. Canivet, A. Fateeva, Y. Guo, B. Coasne, and D. Farrusseng, *Chem. Soc. Rev.*, **43**, 5594–5617 (2014).

[11] 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**, 9488–9498 (2010).

[12] F. N. Keutsch and R. J. Saykally, *Proc. Natl. Acad. Sci.*, **98**, 10533–10540 (2001).

[13] M. D. Fayer and N. E. Levinger, *Annu. Rev. Anal. Chem.*, **3**, 89–107 (2010).

[14] S. Horvath, A. B. McCoy, B. M. Elliott, G. H. Weddle, J. R. Roscioli, and M. A. Johnson, *J. Phys. Chem. A*, **114**, 1556–68 (2010).

[15] F. Perakis, J. a. Borek, and P. Hamm, *J. Chem. Phys.*, **139**, 014501 (2013).

[16] G. R. Medders and F. Paesani, *J. Phys. Chem. Lett.*, **5**, 2897–2902 (2014).

[17] S. Nihonyanagi, T. Ishiyama, T. Lee, S. Yamaguchi, M. Bonn, A. Morita, and T. Tahara, *J. Am. Chem. Soc.*, **133**, 16875–16880 (2011).

[18] L. Piatkowski, Z. Zhang, E. H. G. Backus, H. J. Bakker, and M. Bonn, *Nature Comm.*, **5**, 4083 (2014).

[19] A. Millo, Y. Raichlin, and A. Katzir, *Appl. Spectrosc.*, **59**, 460–466 (2005).

[20] W. J. Smit and H. J. Bakker, *J. Chem. Phys.*, **139**, 204504 (2013).

[21] A. B. McCoy, *J. Phys. Chem. B*, **118**, 8286–94 (2014).

[22] Y. R. Shen and V. Ostroverkhov, *Chem. Rev.*, **106**, 1140–54 (2006).

[23] J. L. Skinner, P. A. Pieniazek, and S. M. Gruenbaum, *Acc. Chem. Res.*, **45**, 93–100 (2012).

[24] I. V. Stiopkin, C. Weeraman, P. A. Pieniazek, F. Y. Shalhout, J. L. Skinner, and A. V. Benderskii, *Nature*, **474**, 192–195 (2011).

[25] M. Vinaykin and A. V. Benderskii, *J. Phys. Chem. Lett.*, **3**, 3348–3352 (2012).

[26] Y. Nagata, C.-S. Hsieh, T. Hasegawa, J. Voll, E. H. G. Backus, and M. Bonn, *J. Phys. Chem. Lett.*, **4**, 1872–1877 (2013).

[27] S. Mukamel, *Principles of Nonlinear Optical Spectroscopy*. Oxford University Press, Oxford (1995).

[28] A. Nitzan, *Chemical Dynamics in Condensed Phases*. Oxford University Press (2006).

[29] E. E. Dahlke, R. M. Olson, H. R. Leverentz, and D. G. Truhlar, *J. Phys. Chem. A*, **112**, 3976–84 (2008).

[30] D. M. Bates and G. S. Tschumper, *J. Phys. Chem. A*, **113**, 3555–3559 (2009).

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

[32] V. Babin and F. Paesani, *Chem. Phys. Lett.*, **580**, 1–8 (2013).

[33] A. K. Soper and C. J. Benmore, *Phys. Rev. Lett.*, **101**, 065502 (2008).

[34] F. Paesani and G. A. Voth, *J. Phys. Chem. B*, **113**, 5702–19 (2009).

[35] Y. Wang, V. Babin, J. M. Bowman, and F. Paesani, *J. Am. Chem. Soc.*, **134**, 11116–9 (2012).

[36] T. Markland and B. Berne, *Proc. Natl. Acad. Sci.*, **109**, 7988–7991 (2012).

[37] D. Hankins, J. W. Moskowitz, and F. H. Stillinger, *J. Chem. Phys.*, **53**, 4544–4554 (1970).

[38] S. S. Xantheas, *J. Chem. Phys.*, **100**, 7523–7534 (1994).

[39] S. S. Xantheas, *Chem. Phys.*, **258**, 225–231 (2000).

[40] A. Defusco, D. Schofield, and K. D. Jordan, *Mol. Phys.*, **105**, 2681–2696 (2007).

[41] R. Kumar, F. Wang, G. R. Jenness, and K. D. Jordan, *J. Chem. Phys.*, **132**, 014309 (2010).

[42] M. P. Hodges, A. J. Stone, and S. S. Xantheas, *J. Phys. Chem. A*, **101**, 9163–9168 (1997).

[43] L. Ojamie and K. Hermansson, *J. Phys. Chem.*, **98**, 4271–4282 (1994).

[44] J. M. Pedulla, F. Vila, and K. D. Jordan, *J. Chem. Phys.*, **105**, 11091 (1996).

[45] J. Cui, H. Liu, and K. D. Jordan, *J. Phys. Chem. B*, **110**, 18872–18878 (2006).

[46] A. Hermann, R. Krawczyk, M. Lein, P. Schwerdtfeger, I. Hamilton, and J. J. P. Stewart, *Phys. Rev. A*, **76**, 013202 (2007).

[47] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.*, **79**, 926 (1983).

[48] H. J. C. Berendsen, J. P. M. Postma, W. F. v. Gunsteren, and J. Hermans. Intermolecular Forces. In B. Pullman, editor, *Intermolecular Forces*, pages 333–342. Reidel Publishing Company, Dordrecht (1981).

[49] S. Habershon, T. E. Markland, and D. E. Manolopoulos, *J. Chem. Phys.*, **131**, 024501 (2009).

[50] F. Paesani, W. Zhang, D. A. Case, T. E. Cheatham, and G. A. Voth, *J. Chem. Phys.*, **125**, 184507 (2006).

[51] C. Vega and J. L. F. Abascal, *Phys. Chem. Chem. Phys.*, **13**, 19663–19688 (2011).

[52] R. Kumar and J. L. Skinner, *J. Phys. Chem. B*, **112**, 8311–8318 (2008).

[53] C. J. Tainter, P. A. Pieniazek, Y. Lin, and J. L. Skinner, *J. Chem. Phys.*, **134**, 184501 (2011).

[54] C. J. Tainter and J. L. Skinner, *J. Chem. Phys.*, **137**, 104304 (2012).

[55] P. E. M. Lopes, B. Roux, and A. D. Mackerell, *Theor. Chem. Acc.*, **124**, 11–28 (2009).

[56] J. Applequist, J. R. Carl, and K. Fung, *J. Am. Chem. Soc.*, **94**, 2952–2960 (1972).

[57] B. T. Thole, *Chem. Phys.*, **59**, 341–350 (1981).

[58] G. S. Fanourgakis and S. S. Xantheas, *J. Chem. Phys.*, **128**, 074506 (2008).

[59] C. J. Burnham, D. J. Anick, P. K. Mankoo, and G. F. Reiter, *J. Chem. Phys.*, **128**, 154519 (2008).

[60] P. Ren and J. W. Ponder, *J. Phys. Chem. B*, **107**, 5933–5947 (2003).

[61] J. J. P. Stewart, *J. Comput. Chem.*, **10**, 209–220 (1989).

[62] M. I. Bernal-Uruchurtu and M. F. Ruiz-López, *Chem. Phys. Lett.*, **330**, 118–124 (2000).

[63] D. T. Chang, G. K. Schenter, and B. C. Garrett, *J. Chem. Phys.*, **128**, 164111 (2008).

[64] G. Murdachaew, C. J. Mundy, G. K. Schenter, T. Laino, and J. Hutter, *J. Phys. Chem. A*, **115**, 6046–53 (2011).

[65] A. D. Becke, *Phys. Rev. A*, **38**, 3098–3100 (1988).

[66] C. Lee, W. Yang, and R. Parr, *Phys. Rev. B*, **37**, 785–789 (1988).

[67] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.*, **77**, 3865–3868 (1996).

[68] S. Grimme, *J. Comput. Chem.*, **25**, 1463–1473 (2004).

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

[70] J. L. Fulton, G. K. Schenter, M. D. Baer, C. J. Mundy, L. X. Dang, and M. Balasubramanian, *J. Phys. Chem. B*, **114**, 12926–12937 (2010).

[71] Z. Ma, Y. Zhang, and M. E. Tuckerman, *J. Chem. Phys.*, **044506**, 044506 (2012).

[72] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *J. Chem. Phys.*, **132**, 154104 (2010).

[73] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, *Phys. Rev. Lett*, **92**, 246401 (2004).

[74] K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, *Phys. Rev. B*, **82**, 081101 (2010).

[75] O. A. Vydrov and T. Van Voorhis, *J. Chem. Phys.*, **133**, 244103 (2010).

[76] J. Wang, G. Román-Pérez, J. M. Soler, E. Artacho, and M. V. Fernández-Serra, *J. Chem. Phys.*, **134**, 024516 (2011).

[77] E. Murray and G. Galli, *Phys. Rev. Lett.*, **108**, 105502 (2012).

[78] R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. van Der Avoird, *J. Chem. Phys.*, **128**, 094314 (2008).

[79] R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. van Der Avoird, *J. Chem. Phys.*, **128**, 094313 (2008).

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

[81] R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. v. d. Avoird, *Science*, **315**, 1249–52 (2007).

[82] C. Leforestier, K. Szalewicz, and A. v. d. Avoird, *J. Chem. Phys.*, **137**, 014305 (2012).

[83] K. Raghavachari, G. W. Truck, J. A. Pople, and M. Head-Gordon, *Chem. Phys. Lett.*, **157**, 479–483 (1989).

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

[85] F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, **7**, 3297–3305 (2005).

[86] A. Schafer, C. Huber, and R. Ahlrichs, *J. Chem. Phys.*, **100**, 5829–5835 (1994).

[87] F. Wennmohs and F. Neese, *Chem. Phys.*, **343**, 217–230 (2008).

[88] R. C. Walker, M. F. Crowley, and D. A. Case, *J. Comput. Chem.*, **29**, 1019–1031 (2008).

[89] *http://cp2k.berlios.de/.*

[90] C. J. Burnham and S. S. Xantheas, *J. Chem. Phys.*, **116**, 5115 (2002).

[91] G. R. Medders, V. Babin, and F. Paesani, *J. Chem. Theory Comput.*, **9**, 1103–1114 (2013).

[92] J. VandeVondele and J. Hutter, *J. Chem. Phys.*, **127**, 114105 (2007).

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

[94] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.*, **98**, 11623–11627 (1994).

[95] J. Klimes and A. Michaelides, *J. Chem. Phys.*, **137**, 120901 (2012).

[96] W. Chen and M. S. Gordon, *J. Phys. Chem.*, **100**, 14316–14328 (1996).

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

[98] V. Babin, G. R. Medders, and F. Paesani, *J. Phys. Chem. Lett.*, **3**, 3765–3769 (2012).

[99] B. J. Braams and J. M. Bowman, *Int. Rev. Phys. Chem.*, **28**, 577–606 (2009).

[100] Z. Xie and J. M. Bowman, *J. Chem. Theory Comput.*, **6**, 26–34 (2010).

[101] E. Mas, R. Bukowski, and K. Szalewicz, *J. Chem. Phys.*, **118**, 4386–4403 (2003).

[102] J. Caldwell, L. X. Dang, and P. A. Kollman, *J. Am. Chem. Soc*, **112**, 9144–9147 (1990).

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

[104] R. Car and M. Parrinello, *Phys. Rev. Lett.*, **55**, 2471–2474 (1985).

[105] W. Xie, L. Song, D. G. Truhlar, and J. Gao, *J. Chem. Phys.*, **128**, 234108 (2008).

[106] M. S. Gordon, L. Slipchenko, H. Li, and J. H. Jensen, *Annu. Rep. Comput. Chem.*, **3**, 177–193 (2007).

[107] M. Del Ben, M. Schönherr, J. Hutter, and J. VandeVondele, *J. Phys. Chem. Lett.*, **4**, 3753–3759 (2013).

[108] A. P. Bartók, M. J. Gillan, F. R. Manby, and G. Csányi, *Phys. Rev. B*, **88**, 054104 (2013).

[109] S. Wen and G. J. O. Beran, *J. Chem. Theory Comput.*, **7**, 3733–3742 (2011).

[110] C. Zhang, J. Wu, G. Galli, and F. Gygi, *J. Chem. Theory Comput.*, **7**, 3054–3061 (2011).

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

[112] A. M. g. j, A. K. Kelkkanen, K. T. Wikfeldt, J. Schiø tz, J. J. r. Mortensen, L. G. M. Pettersson, B. I. Lundqvist, K. W. Jacobsen, A. Nilsson, and J. K. Nø rskov, *J. Phys. Chem. B*, **115**, 14149–60 (2011).

[113] O. A. Vydrov and T. Van Voorhis, *J. Chem. Theory Comput.*, **8**, 1929–1934 (2012).

[114] A. J. Stone, *Theory of Intermolecular Forces*. Oxford University Press (1997).

[115] V. Babin, C. Leforestier, and F. Paesani, *J. Chem. Theory Comput.*, **9**, 5395–5403 (2013).

[116] H. Partridge and D. W. Schwenke, *J. Chem. Phys.*, **106**, 4618 (1997).

[117] L.-P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande, *J. Phys. Chem. B*, **117**, 9956–72 (2013).

[118] R. A. Kendall, T. H. Dunning, and R. J. Harrison, *J. Chem. Phys.*, **96**, 6796–6806 (1992).

[119] F.-M. Tao and Y.-K. Pan, *J. Chem. Phys.*, **97**, 4989 (1992).

[120] B. H. Wells and S. Wilson, *Chem. Phys. Lett.*, **101**, 429–434 (1983).

[121] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, 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, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and M. Wang. MOLPRO, version 2012.1, a package of ab initio programs (2012).

[122] V. Babin, G. R. Medders, and F. Paesani, *J. Chem. Theory Comput.*, **10**, 1599–1607 (2014).

[123] J. A. Anderson, K. Crager, L. Fedoroff, and G. S. Tschumper, *J. Chem. Phys.*, **121**, 11023–11029 (2004).

[124] T. Hill, *An Introduction to Statistical Thermodynamics*. Dover (1986).

[125] G. Kell, G. McLaurin, and E. Whalley, *Proc. R. Soc. Lond. A*, **425**, 49–71 (1989).

[126] R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. T. Bell, and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, **14**, 15328–15339 (2012).

[127] B. Temelso, K. A. Archer, and G. C. Shields, *J. Phys. Chem. A*, **115**, 12034–12046 (2011).

[128] M. A. Morales, J. R. Gergely, J. McMinis, J. M. McMahon, J. Kim, and D. M. Ceperley, *J. Chem. Theory Comput.*, **10**, 2355–2362 (2014).

[129] J. Mayer and M. Mayer, *Statistical Mechanics*. John Wiley & Sons Inc (1940).

[130] D. Chandler and P. G. Wolynes, *J. Chem. Phys.*, **74**, 4078 (1981).

[131] M. Parrinello and A. Rahman, *J. Chem. Phys.*, **80**, 860 (1984).

[132] B. De Raedt, M. Sprik, and M. L. Klein, *J. Chem. Phys.*, **80**, 5719 (1984).

[133] B. J. Berne and D. Thirumalai, *Annu. Rev. Phys. Chem.*, **37**, 401–424 (1986).

[134] G. A. Voth, *Adv. Chem. Phys.*, **93**, 135 (1996).

[135] M. Rossi, H. Liu, F. Paesani, J. Bowman, and M. Ceriotti, *J. Chem. Phys.*, **141**, 181101 (2014).

[136] R. P. Feynman, *Statistical Mechanics*. Benjamin, New York (1972).

[137] M. E. Tuckerman. Path Integration via Molecular Dynamics. In J. Grotendorst, D. Marx, and A. Muramatsu, editors, *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, volume 10, pages 269–298. John von Neumann Institute for Computing, Julich, vol. 10 edition (2002).

[138] J. Cao and G. A. Voth, *J. Chem. Phys.*, **101**, 6157 (1994).

[139] J. Cao and G. A. Voth, *J. Chem. Phys.*, **101**, 6168 (1994).

[140] J. Cao and G. A. Voth, *J. Chem. Phys.*, **100**, 5093–5105 (1994).

[141] J. Cao and G. A. Voth, *J. Chem. Phys.*, **100**, 5106 (1994).

[142] J. Cao and G. A. Voth, *J. Chem. Phys.*, **101**, 6184 (1994).

[143] S. Jang and G. A. Voth, *J. Chem. Phys.*, **111**, 2357 (1999).

[144] S. Jang and G. a. Voth, *J. Chem. Phys.*, **111**, 2371 (1999).

[145] B. J. Berne, *Adv. Chem. Phys.*, **17**, 63–227 (1970).

[146] R. W. Hall and B. J. Berne, *J. Chem. Phys.*, **81**, 3641 (1984).

[147] G. J. Martyna, M. L. Klein, and M. Tuckerman, *J. Chem. Phys.*, **97**, 2635 (1992).

[148] G. J. Martyna and M. E. Tuckerman, *J. Chem. Phys.*, **110**, 2810–2821 (1999).

[149] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*. Oxford (1987).

[150] F. Paesani, S. Iuchi, and G. A. Voth, *J. Chem. Phys.*, **127**, 074506 (2007).

[151] *CRC Handbook of Chemistry and Physics*. CRC Press.

[152] W. Wagner and A. Pruss, *J. Phys. Chem. Ref. Data.*, **31**, 387 (2002).

[153] K. Krynicki, C. D. Green, and D. W. Sawyer, *Faraday Discuss.*, **66**, 199 (1978).

[154] Y. L. a. Rezus and H. J. Bakker, *J. Chem. Phys.*, **123**, 114502 (2005).

[155] A. Soper, *Chem. Phys.*, **258**, 121–137 (2000).

[156] L. Skinner, C. Huang, D. Schlesinger, L. Pettersson, A. Nilsson, and C. Benmore, *J. Chem. Phys.*, **138**, 074506 (2013).

[157] M. Ceriotti, J. Cuny, M. Parrinello, and D. E. Manolopoulos, *Proc. Natl. Acad. Sci.*, **110**, 15591–15596 (2013).

[158] L. Wang, M. Ceriotti, and T. E. Markland. Private Communication (2014).

[159] K. Matsuzaki, S. Nihonyanagi, S. Yamaguchi, T. Nagata, and T. Tahara, *J. Phys. Chem. Lett.*, **4**, 1654–1658 (2013).

[160] W. Xie and J. Gao, *J. Chem. Theory Comput.*, **3**, 1890–1900 (2007).

[161] P. Jungwirth and B. Winter, *Annu. Rev. Phys. Chem.*, **59**, 343–366 (2008).

[162] J. C. Rasaiah, S. Garde, and G. Hummer, *Annu. Rev. Phys. Chem.*, **59**, 713–740 (2008).

[163] A. Morita, *J. Comput. Chem.*, **23**, 1466–1471 (2002).

[164] G. Lamoureux and B. Roux, *J. Phys. Chem. B*, **110**, 3308–3322 (2006).

[165] J. J. Perez, J. H. R. Clarke, and A. Hinchliffe, *Chem. Phys. Lett.*, **104**, 583–586 (1984).

[166] J. Waite and M. G. Papadopoulos, *Theor. Chem. Acta*, **75**, 53–65 (1989).

[167] C. Aleman, J. J. Perez, and A. Hinchliffe, *Int. J. Mass Spectrom. Ion Processes*, **122**, 331–336 (1992).

[168] B. Skwara, W. Bartkowiak, and D. L. Silva, *Theoret. Chem. Acc.*, **122**, 127–136 (2009).

[169] A. Baranowska, A. Zawada, B. Fernández, W. Bartkowiak, D. Kedziera, and A. Kaczmarek-Kedziera, *Phys. Chem. Chem. Phys.*, **12**, 852–862 (2010).

[170] A. Zawada and W. Bartkowiak, *Comput. Theor. Chem.*, **967**, 120–128 (2011).

[171] A. Zawada, A. Kaczmarek-Kdziera, and W. Bartkowiak, *J. Mol. Model.*, **18**, 3073–3086 (2012).

[172] G. Karlstrom and A. J. Sadlej, *Theor. Chem. Acta*, **61**, 1–9 (1982).

[173] G. Maroulis, *J. Chem. Phys.*, **113**, 1813–1820 (2000).

[174] A. Zawada, A. Kaczmarek-Kdziera, and W. Bartkowiak, *Chem. Phys. Lett.*, **503**, 39–44 (2011).

[175] T. K. Ghanty and S. K. Ghosh, *J. Chem. Phys.*, **118**, 8547–8550 (2003).

[176] M. Yang, P. Senet, and C. Van Alsenoy, *Int. J. Quantum Chem.*, **101**, 535–542 (2005).

[177] A. Krishtal, P. Senet, M. Yang, and C. Van Alsenoy, *J. Chem. Phys.*, **125**, 34312 (2006).

[178] J. R. Hammond, N. Govind, K. Kowalski, J. Autschbach, and S. S. Xantheas, *J. Chem. Phys.*, **131**, 214103 (2009).

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

[180] H. A. Kurtz, J. J. P. Stewart, and K. M. Dieter, *J. Comput. Chem.*, **11**, 82–87 (1990).

[181] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, and Others. MOLPRO, version 2012.1, a package of ab initio programs (2012).

[182] Z. Czynikowska, R. W. Góra, R. Zaleśny, W. Bartkowiak, A. Baranowska-czkowska, and J. Leszczynski, *Chem. Phys. Lett.*, **555**, 230–234 (2013).

[183] J. C. White and E. R. Davidson, *J. Chem. Phys.*, **93**, 8029–8035 (1990).

[184] P. Valiron and I. Mayer, *Chem. Phys. Lett.*, **275**, 46–55 (1997).

[185] G. R. Medders and F. Paesani, *J. Chem. Theory Comput.*, **9**, 4844–4852 (2013).

[186] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*. Dover Publications, Mineola (1996).

[187] A. Roztoczyska, A. Kaczmarek-Kdziera, R. W. Góra, and W. Bartkowiak, *Chem. Phys. Lett.*, **571**, 28–33 (2013).

[188] D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, **100**, 2975–2988 (1994).

[189] J. J. Foley and D. A. Mazziotti, *J. Phys. Chem. A*, **117**, 6712–6716 (2013).

[190] R. N. Pribble and T. S. Zwier, *Science*, **265**, 75–79 (1994).

[191] K. Liu, M. G. Brown, C. Carter, R. J. Saykally, J. K. Gregory, and D. C. Clary, *Nature*, **381**, 501–503 (1996).

[192] K. Nauta and R. E. Miller, *Science*, **287**, 293–295 (2000).

[193] C. Steinbach, P. Andersson, M. Melzer, J. K. Kazimirski, U. Buck, and V. Buch, *Phys. Chem. Chem. Phys.*, **6**, 3320–3324 (2004).

[194] C. Perez, M. T. Muckle, D. P. Zaleski, N. A. Seifert, B. Temelso, G. C. Shields, Z. Kisiel, and B. H. Pate, *Science*, **336**, 897–901 (2012).

[195] D. E. Bernholdt and R. J. Harrison, *Chem. Phys. Lett.*, **250**, 477–484 (1996).

[196] M. Feyereisen, G. Fitzgerald, and A. Komomicki, *Chem. Phys. Lett.*, **208**, 359–363 (1993).

[197] S. Kazachenko and A. J. Thakkar, *J. Chem. Phys.*, **138**, 194302 (2013).

[198] W. B. Bosma, L. E. Fried, and S. Mukamel, *J. Chem. Phys.*, **98**, 4413 (1993).

[199] S. Saito and I. Ohmine, *J. Phys. Chem.*, **102**, 3566 (1995).

[200] S. Iuchi, A. Morita, and S. Kato, *J. Phys. Chem. B*, **106**, 3466–3476 (2002).

[201] M. Heyden, J. Sun, S. Funkner, G. Mathias, H. Forbert, M. Havenith, and D. Marx, *Proc. Natl. Acad. Sci.*, **107**, 12068–73 (2010).

[202] Q. Wan, L. Spanu, G. A. Galli, and F. Gygi, *J. Chem. Theory Comput.*, **9**, 4124–4130 (2013).

[203] P. Hamm, *J. Chem. Phys.*, **141**, 184201 (2014).

[204] R. Ramírez, T. López-Ciudad, P. Kumar, and D. Marx, *J. Chem. Phys.*, **121**, 3973–83 (2004).

[205] T. L. C. Jansen, W. Zhuang, and S. Mukamel, *J. Chem. Phys.*, **121**, 10577–98 (2004).

[206] J.-H. Choi, S. Hahn, and M. Cho, *Int. J. Quantum Chem.*, **104**, 616–634 (2005).

[207] V. Buch, T. Tarbuck, G. L. Richmond, H. Groenzin, I. Li, and M. J. Shultz, *J. Chem. Phys.*, **127**, 204710 (2007).

[208] J. L. Skinner, B. M. Auer, and Y.-s. Lin. Vibrational Line Shapes, Spectral Diffusion, and Hydrogen Bonding in Liquid Water. In *Adv. Chem. Phys.*, volume 142, pages 59–103 (2009).

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

[210] Y. Wang and J. M. Bowman, *J. Chem. Phys.*, **136**, 144113 (2012).

[211] J.-H. Choi and M. Cho, *J. Chem. Phys.*, **138**, 174108 (2013).

[212] B. Auer, R. Kumar, J. R. Schmidt, and J. L. Skinner, *Proc. Natl. Acad. Sci.*, **104**, 14215–20 (2007).

[213] C. J. Tainter, Y. Ni, L. Shi, and J. L. Skinner, *J. Phys. Chem. Lett.*, **4**, 12–17 (2013).

[214] L. De Marco, K. Ramasesha, and A. Tokmakoff, *J. Phys. Chem. B*, **117**, 15319–27 (2013).

[215] A. Witt, S. D. Ivanov, M. Shiga, H. Forbert, and D. Marx, *J. Chem. Phys.*, **130**, 194510 (2009).

[216] F. Paesani and G. A. Voth, *J. Chem. Phys.*, **132**, 014105 (2010).

[217] F. Paesani, S. S. Xantheas, and G. A. Voth, *J. Phys. Chem. B*, **113**, 13118–13130 (2009).

[218] I. R. Craig and D. E. Manolopoulos, *J. Chem. Phys.*, **121**, 3368–73 (2004).

[219] S. Habershon, G. S. Fanourgakis, and D. E. Manolopoulos, *J. Chem. Phys.*, **129**, 074501 (2008).

[220] M. Rossi, M. Ceriotti, and D. E. Manolopoulos, *J. Chem. Phys.*, **140**, 234116 (2014).

[221] G. R. Medders, V. Babin, and F. Paesani, *J. Chem. Theory Comput.*, **10**, 2906–2910 (2014).

[222] U. Góra, W. Cencek, R. Podeszwa, A. v. d. Avoird, and K. Szalewicz, *J. Chem. Phys.*, **140**, 194101 (2014).

[223] P. K. Mankoo and T. Keyes, *J. Chem. Phys.*, **129**, 034504 (2008).

[224] T. Ishiyama and A. Morita, *J. Chem. Phys.*, **131**, 244714 (2009).

[225] T. Hasegawa and Y. Tanimura, *J. Phys. Chem. B*, **115**, 5545–53 (2011).

[226] P. L. Silvestrelli, M. Bernasconi, and M. Parrinello, *Chem. Phys. Lett.*, **277**, 478–482 (1997).

[227] M. Sharma, R. Resta, and R. Car, *Phys. Rev. Lett.*, **95**, 187401 (2005).

[228] H.-S. Lee and M. E. Tuckerman, *J. Chem. Phys.*, **126**, 164501 (2007).

[229] W. Chen, M. Sharma, R. Resta, G. Galli, and R. Car, *Phys. Rev. B*, **77**, 245114 (2008).

[230] C. Zhang, D. Donadio, F. Gygi, and G. Galli, *J. Chem. Theory Comput.*, **7**, 1443–1449 (2011).

[231] K. Szalewicz, S. J. Cole, W. Kolos, and R. J. Bartlett, *J. Chem. Phys.*, **89**, 3662–3673 (1988).

[232] R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. T. Bell, and M. Head-Gordon, *J. Phys. Chem. A*, **111**, 8753–8765 (2007).

[233] L. Lodi, J. Tennyson, and O. L. Polyansky, *J. Chem. Phys.*, **135**, 034113 (2011).

[234] G. R. Medders and F. Paesani, *J. Chem. Theory Comput.*, **11**, 1145–1154 (2015).

[235] G. Avila, *J. Chem. Phys.*, **122**, 144310 (2005).

[236] D. Alfè, A. P. Bartók, G. Csányi, and M. J. Gillan, *J. Chem. Phys.*, **138**, 221102 (2013).

[237] T. D. Hone and G. A. Voth, *J. Chem. Phys.*, **121**, 6412 (2004).

[238] M. H. Brooker, G. Hancock, B. C. Rice, and J. Shapter, *J. Raman. Spectrosc.*, **20**, 683–694 (1989).

[239] J. E. Bertie and Z. Lan, *Appl. Spectrosc.*, **50**, 1047–1057 (1996).

[240] K. Park, W. Lin, and F. Paesani, *J. Phys. Chem. B*, **116**, 343–352 (2012).

[241] J. R. Schmidt, S. A. Corcelli, and J. L. Skinner, *J. Chem. Phys.*, **123**, 044513 (2005).

[242] N. Makri, *J. Chem. Phys.*, **141**, 134117 (2014).

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

[244] J. ezáč and P. Hobza, *J. Chem. Theory Comput.*, **9**, 2151–2155 (2013).

[245] R. Kubo, *Adv. Chem. Phys.*, **15**, 101 (1969).

[246] H. Ahlborn, X. Ji, B. Space, and P. B. Moore, *J. Chem. Phys.*, **111**, 10622 (1999).

[247] P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn, *Nature*, **282**, 459–464 (1979).

[248] F. Paesani, *Phys. Chem. Chem. Phys.*, **13**, 19865–19875 (2011).

[249] J. Kolafa, *J. Comput. Chem.*, **25**, 335–42 (2004).

[250] C. Zhang, R. Z. Khaliullin, D. Bovi, L. Guidoni, and T. D. Kühne, *J. Phys. Chem. Lett.*, **4**, 3245–3250 (2013).

[251] D. T. Limmer and D. Chandler, *J. Chem. Phys.*, **135**, 134503 (2011).

[252] D. T. Limmer and D. Chandler, *J. Chem. Phys.*, **138**, 214504 (2013).

[253] J. C. Palmer, F. Martelli, Y. Liu, R. Car, A. Z. Panagiotopoulos, and P. G. Debenedetti, *Nature*, **510**, 385–388 (2014).

[254] G. R. Medders and F. Paesani, *J. Chem. Phys.*, **142**, 212411 (2015).

[255] A. Morita and J. T. Hynes, *J. Phys. Chem. B*, **106**, 673–685 (2002).

[256] R. M. Richard, K. U. Lao, and J. M. Herbert, *Acc. Chem. Res.*, **47**, 2828–2836 (2014).

[257] R. M. Richard, K. U. Lao, and J. M. Herbert, *J. Phys. Chem. Lett.*, **4**, 2674–2680 (2013).