Extension of the Many-Body Expansion (MBE) to Periodic Systems: Developing Tools to Analyze and Improve Models of Intermolecular Interactions

Kristina Herman

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Sotiris S. Xantheas, Chair

Anne B. McCoy

Cody W. Schlenker

Program Authorized to Offer Degree:
Department of Chemistry

University of Washington

Abstract

Extension of the Many-Body Expansion (MBE) to Periodic Systems: Developing Tools to Analyze and Improve Models of Intermolecular Interactions

Kristina Herman

Chair of the Supervisory Committee:
Sotiris S. Xantheas
Chemistry

Intermolecular interactions govern the structure and dynamics of molecular systems, which collectively give rise to their physical properties across scales from the nano- to the meso- and the macro-scale. These collective properties are often sensitive to the level of theoretical description manifested either by the accuracy of a classical interaction potential or the level of electronic structure theory used to describe the fundamental interactions at the molecular level. Due to the prohibitive scaling of electronic structure calculations with the system size (oftentimes $\mathcal{O}(N^6)$–$\mathcal{O}(N^7)$ for the "gold-standard" coupled cluster methods, where N is the number of basis functions), we must seek alternative ways of evaluating the properties of these complex molecular systems without compromising accuracy. The many-body expansion (MBE), a fragmentation approach, partitions the full system into a set of smaller subsystems and combinatorically represents the properties of the full system (i.e., binding energy, forces, dipole moment, etc.). This approach offers a powerful alternative to address the "scaling curse" of accurate electronic structure methods with system size. The following will highlight efforts to unravel the nature of many-body effects in aqueous ionic systems, develop transferable classical models to accurately describe those intermolecular interactions at a reduced cost, and extend the many-body expansion (MBE) formalism to periodic systems.

The MBE was applied to investigate the influence of "structure-making" and "structure-

breaking" ions in the Hofmeister series on the energetics of aqueous cations and anions (SO_4^{2-}, ClO_4^-, Ca^{2+}, NH_4^+). Significant differences in the many-body terms were identified for the structure-making ions, which exhibit the strongest ion-water and the weakest water-water interactions. Conversely, the structure-breaking ions exhibit weaker ion-water interactions and stronger water-water interactions, demonstrating the intricate balance of interactions governing the energetics of these systems. The trend demonstrating the anti-correlation between the ion-water and water-water interactions persisted across 13 different ion-water systems and further quantified the role of ions (and the identity of said ion) in affecting the water-water interactions.

Having established the strong many-body character in ion-water systems, the next step was to develop interaction potentials to describe many-body effects in aqueous ionic systems. A classical induction model using a detailed description of the field due to a charge distribution using distributed multipoles and the response of the charge distribution to an external field using distributed polarizabilities was developed to model 3- and 4-body interactions. The induction energies were benchmarked against 3,120 *ab initio* 3-body energies for 13 different ion-water-water and water-water-water systems. The induction model was subsequently improved by developing and implementing geometry-dependent distributed multipole and polarizability surfaces. For water, the induction model augmented with 3-body dispersion was compared against results from an existing database of 43,844 3-body and 3,603 4-body CCSD(T) energies. The model was found to reproduce the 3- and 4-body interactions with mean absolute errors of 0.054 and 0.026 kcal/mol, respectively. These findings suggest that the developed classical model with zero adjustable parameters yields an accuracy that is on-par with models fit to tens of tens of thousands of CCSD(T) energies using thousands of adjustable parameters. This physics-based approach provides a simple, fast, and most importantly transferable way of modeling many-body effects in aqueous molecular systems, eliminating the need to perform tens of thousands of expensive electronic structure calculations for each change in solvent or solute identity.

A novel approach was developed to extend the MBE to periodic systems and subsequently used to decompose the lattice energies of seven polymorphs of ice into their constituent many-body terms. The sum of the many-body terms (1-body through 4-body) was shown to match the value obtained with periodic boundary conditions using the minimum image convention and an Ewald summation. A resulting three-way relationship was established, demonstrating the correlation between the many-body terms, the local tetrahedral order, and the density of the ice polymorphs. Specifically, ice polymorphs existing at low pressures were shown to have strong cooperative effects, near perfect tetrahedral order, and 5- and 6-membered hydrogen bond cycles. Conversely, high pressure ice polymorphs exhibited weak cooperative effects, low tetrahedral order, and a mixture of hydrogen bond cycle sizes. This lends valuable insight into the structure-energy relationship that governs the complex phase diagram of ice.

Besides its utility of gaining insight into the nature of intermolecular interactions, the MBE for periodic systems is currently being used to circumvent the poor scaling of electronic structure methods by computing 2- and 3-body corrections to the correlation energy. This approach is trivially parallelizable and is founded on the preference of performing several CCSD(T) calculations for much smaller systems (dimers and trimers) therefore reducing the size (N) and corresponding cost (N^7) of the requisite CCSD(T) calculations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Number Page

ACKNOWLEDGMENTS

Thank you to the many collaborators that I have had the privilege to work with. I learned valuable things from each collaboration that have helped me become a better scientist.

Thank you to Anne McCoy for her mentorship and helpful discussions. I would also like to thank members of the McCoy group (past and present) for their camaraderie and their feedback on communicating research effectively to different audiences.

I would like to thank all the past and present members of the Xantheas research group (Joe Heindel, Garrett Santis, Max Hoffman, Madisen Hallsten, Joani Mato, Roy Irving) for supporting me throughout graduate school, especially Joe Heindel for helping me navigate the first few years of research and Garrett Santis for helpful research (and other) conversations over the last 4+ years.

Finally, thank you to my graduate advisor, Sotiris Xantheas. I appreciate all the opportunities given to me during the last 5 years. These experiences have helped me to become an independent scientist and I am very grateful for that.

DEDICATION

To my parents (Fred and Karen), my soon-to-be husband (Nic), and my siblings (Elizabeth, Emily, and Eric)

Chapter 1

INTRODUCTION

The potential energy function of a molecular system, $V(\mathbf{R}_n, \mathbf{r}_e)$, where $\mathbf{R}_n$ and $\mathbf{r}_e$ define the positions of the nuclei and electrons respectively, is central to chemistry. The Born-Oppenheimer approximation adiabatically separates the nuclear and electronic degrees of freedom and defines the potential energy surface as a function of the nuclear coordinates, $V(\mathbf{r}_e; \mathbf{R}_n)$. The potential energy surface is thus defined by the intramolecular geometry and the intermolecular interactions. Statistical mechanics provides a route to obtain nanoscale and macroscopic properties from these atomic-level interactions. Importantly, these intermolecular interactions can be shown to govern the structure and dynamics of systems, the relative stability of molecular crystals and gas-phase clusters, and the interaction of systems with light, among other important physical properties, highlighting the vital importance of understanding and accurately modeling these interactions at the atomic level.

However, a significant challenge lies in the development of an accurate potential energy surface for a system, which is inherently high-dimensional, while also ensuring that the developed potential is computationally tractable, since many simulation methods requires millions of energy evaluations (or many more!). The remainder of the chapter outlines methods to represent the potential energy of molecular systems, followed by a brief summary of simulation methods that can be used to obtain macroscopic properties via statistical mechanics.

1.1 Representations of the potential energy surface

The intermolecular interactions of a system can be described *ab initio* or by using classical methods. While electronic structure methods can be used directly to obtain energies and forces *ab initio*, the prohibitive scaling of these methods prevents their routine use in large-scale simulations. Conversely, classical methods do not include an explicit representation of electrons but instead are parameterized using *ab initio* or experimental results.

1.1.1 Ab initio methods

First principles methods incorporate an explicit representation of the electronic degrees of freedom in the calculation of the energy and forces. The goal of these methods is to solve Schrödinger's equation

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

where E is the electronic energy of the system and ψ is the wavefunction, which by definition contains all information about the system. $\hat{H}$ is the Hamiltonian which is defined as the sum of the potential and kinetic energies of the nuclei (n) and electrons (e) in the system,

$$\begin{aligned}\hat{H} &= \hat{T}_e + \hat{V}_{n-n} + \hat{V}_{e-n} + \hat{V}_{e-e} \\ &= \frac{1}{2}\sum_i \nabla_i^2 + \sum_{I<J}\frac{Z_I Z_J}{r_{IJ}} - \sum_{i,I}\frac{Z_I}{r_{Ii}} + \sum_{i<j}\frac{1}{r_{ij}},\end{aligned} \tag{1.2}$$

where capital I, J, ... represent the indices of the nuclei and lowercase i, j, ... represent those of the electrons. Z_I is the nuclear charge of nucleus I and r_{ij} indicates the distance between electrons i and j. Note that the kinetic energy term of the nuclei is not included in this expression due to the Born-Oppenheimer approximation. Equation 1.1 contains two unknowns, namely the energy and the wavefunction, and can be solved analytically only for systems with a single electron. For many-electron systems, several approximations need to be made in order to obtain the energy via the variation principle. To this end, many theories exist to approximate the wavefunction in order to solve the Schrödinger equation numerically. The presence of the $1/r_{ij}$ term in the Hamiltonian (Equation 1.2), where i, j are the coordinates of any two electrons, is responsible for the primary challenge in describing the electron-electron correlation. These theories can broadly be classified as wavefunction-based or density-based methods.

Wavefunction-based methods

Wavefunction-based methods use the Hartree-Fock (HF) theory as the foundation. While HF theory incorporates anti-symmetry with respect to the exchange of electrons (fermions), it does not include explicit electron correlation and formally scales as $\mathcal{O}(N^4)$, where N is the number of basis functions. Rather, it computes the interaction of the electrons in the average field of all the electrons in the system. Post-HF methods are meant to include some degree of the missing electron correlation (originating from the $1/r_{ij}$ term in the Hamiltonian) to the wavefunction and the electronic energy. Various methods in this category include Møller-Plesset (MP) perturbation theory (with its most popular variant being MP2, which is truncated at the second order), coupled cluster (CC), multi-configuration self-consistent field (MCSCF) and variants of the configuration interaction (CI) theories, which formally scale from $\mathcal{O}(N^5)$ to $\mathcal{O}(N^9)$. These methods consider the contribution from excitations from the occupied to virtual molecular orbitals to describe the dynamic electron correlation. In practice, coupled cluster with single, double and perturbative triple excitations, CCSD(T), has been shown to provide sufficient agreement with experimental estimates of binding energies, equilibrium geometries, and rotational constants, among other properties.[7–10] Its success and reliability has led to its reputation as the current "gold-standard" for benchmark values. However, these methods can be extended to higher order excitations. For instance, CC can include full triple (CCSDT) or quadruple (CCSDTQ) and so on excitations, whereas CI can start from just the single and double excitations from a HF (HF+1+2) or an MCSCF (MCSCF+1+2) wavefunction and go up to including all possible (single, double, triple, quadruple, etc.) excitations from the occupied to the virtual orbitals (Full CI or FCI). The latter, however, scales as $\mathcal{O}(N!)$, making it practically intractable for all but the smallest systems.[11–13]

Density-based methods

Density functional theory (DFT) is founded on the premise that the ground state electron density determines the ground state properties of a system. DFT, in general, describes the wavefunction as a functional of the ground state electron density. These functionals can have many different forms. "Jacob's ladder"[14,15] organizes the functionals according to their complexity, with the simplest functionals (i.e., local density approximation) including only a representation of the electron density at all points in space and the most complicated functionals (i.e., double hybrid) including contributions from higher order derivatives of the electron density and 2nd order perturbation theory. DFT scales formally as $\mathcal{O}(N^3)-\mathcal{O}(N^4)$ depending on whether it includes exchange contributions, but the accuracy is often found to be either system-dependent or not systematically improvable upon moving up "Jacob's ladder". A recent study[8] benchmarked the performance of various DFT functionals against CCSD(T) to describe important low-lying minima of the benzene dimer potential energy surface (PES) and demonstrated the promise of certain DFT functionals to achieve the same accuracy of CCSD(T) at a significantly reduced cost. However, it also highlighted the caveat that this accuracy is often system- and functional-specific. Importantly, ascending Jacob's ladder (increasing functional complexity) does not necessarily guarantee improvement in the accuracy. The availability of accurate, high-level theory benchmarks, such as the ones obtained at the CCSD(T) level of theory, therefore constitutes a powerful tool for assessing the accuracy of lower scaling methods, such as DFT. These lower scaling methods can be subsequently used for evaluating the energy of a system multiple times, i.e. during a molecular dynamics simulation, a task that would not be possible with these higher order methods.

Basis sets

Parallel to the hierarchy of electronic structure theories, another important component of electronic structure calculations is the basis set, which is used to represent the molecular orbitals (through a linear combination of the atomic basis functions). For variational methods, an improved description of the wavefunction will lower the energy, although no lower than the true energy according to the variational principle. In principle, in the infinite, complete basis set (CBS) limit with FCI, the *true* energy and wavefunction of a system can be obtained. In reality, this calculation cannot be performed. The CBS limit can in principle be estimated as the limit of a series of systematically increasing basis sets. To this end, Dunning and co-workers[16–23] developed the family of "correlation consistent" pVNZ basis sets (N = D, T, Q, 5 for basis sets of double, triple, quadruple, quintuple etc. quality) by grouping together functions that contribute approximately the same correlation energy to the atomic ground states. It was subsequently shown[24–28] that by using the hierarchy DZ → TZ → QZ $\cdots$ of these basis sets it was possible to extrapolate certain molecular properties, such as equilibrium bond lengths, binding energies and/or harmonic frequencies, to their respective CBS limits by employing heuristic extrapolation schemes based on the cardinal number N of the basis set ($N = 2, 3, 4, \ldots$ for the DZ, TZ, QZ, ...).

The above computational challenges limit the system sizes and timescales that are accessible with *ab initio* methods but do provide a path towards selecting a lower scaling level of theory to perform simulations. That being said, especially for long simulations it is often desirable to parameterize classical interaction potentials (no electronic degrees of freedom) that are based on simpler and faster analytical forms of the energy as a function of the coordinates of the atoms in the system.

1.1.2 Classical interaction potentials

Classical potentials are designed to be much faster than *ab initio* methods. These classes of potential functions can be either fitted to experimental data, developed using principles from classical physics, or fitted to high-level *ab initio* data. They can also incorporate the underlying physics in a hierarchical fashion.

Pairwise additive potentials

The simplest potentials are pairwise additive, considering interactions between just all pairs of atoms in the system. One example is the 4-site transferable interaction potential (TIP4P), which has the functional form:

$$V = \sum_{i}^{N} \sum_{j>i}^{N} -\frac{q_i q_j}{4\pi\epsilon_0 R_{ij}} + 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{R_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}}\right)^{6}\right] \tag{1.3}$$

This incorporates a pairwise coulombic interaction between all intermolecular partial charges (q_i) separated by distance, R. It also includes a Lennard-Jones potential to model van der Waals interactions. The parameterization for these types of potentials are typically empirical, fitting the atomic charges q_i (and thus dipole moment), ϵ, and σ to reproduce experimental observables, such as the radial distribution function (RDF), specific heat, densities, and phase transitions. Due to the simple functional form, these potentials can be used to simulate systems with thousands of atoms and long timescales (up to ms). However, these potentials do not incorporate the polarizability of molecules which inhibits their transferability to different system sizes[29] and conditions.

Polarizable potentials

Another class of potentials have been developed to reflect the polarizability of molecules, meaning that their charge distribution is influenced by an external field (due to other molecules/ions) and, thus, is inherently environment-dependent. This polarizability has been implemented into potentials such as the Thole-type model (TTM),[30–37] AMOEBA potentials,[38–42] MB-UCB potential,[43] and is the basis for accurate classical induction models.[44,45] In general, these potentials show great promise due to their ease of transferability across chemical space, oftentimes relying on parameters derived from a single *ab initio* calculation of the constituent monomer.

Data-driven potentials

Recently, many data-driven potentials have been developed by either correcting deficiencies in the short-range interactions for existing polarizable potentials (using a Δ-machine learning approach) or by training a ML model on *ab initio* data directly. Potentials such as MB-pol[46–48] and WHBB[49] belong to the former, while the q-AQUA[50] potential belongs to the latter category. These potentials have achieved a high level of accuracy[29,46,47,50–54] but exhibit limited transferability across chemical space and, sometimes, configurational space, depending on the robustness of the training set. Naturally, when an additional molecule is considered (not in the original dataset), additional data must be generated to re-fit or develop a new model. For complex chemical systems, this process is not sustainable due to the computational expense associated with producing high-fidelity training data.

1.2 Simulation methods to obtain macroscopic properties

Computational chemists rely on both lower scaling first principles (DFT) and classical potentials to sample the configuration space of complex systems in order to obtain a molecular-level understanding of several processes and to predict macroscopic properties of complex systems. Statistical mechanics bridges atomic-level interactions, V($\mathbf{r}$), and the classical world, providing a way to calculate macroscopic properties from atomistic simulations. More specifically, the partition function,

$$Z = \frac{1}{h^3} \int e^{-\beta H(\mathbf{p},\mathbf{q})} d\mathbf{q}^3 d\mathbf{p}^3, \tag{1.4}$$

represents a statistical ensemble, where $\beta = \frac{1}{k_B T}$, $H(\mathbf{p}, \mathbf{q}) = P + V$ is the Hamiltonian of the system, $\mathbf{p}$ is the positions of the atoms, and $\mathbf{q}$ is the momenta of the atoms. The partition function is effectively a temperature-dependent weighted probability of the microstates of a system, including contributions from translational, rotational, vibrational, and electronic degrees of freedom. Many physical observables such as the heat capacity, internal energy, Gibbs free energy, and entropy can be derived directly from the partition function.

In most cases, it is impractical to obtain all microstates in the system and compute the observables analytically. Therefore, it is necessary to instead sample the relevant statistical ensembles and compute the observables numerically. This can be achieved through Monte Carlo (MC) or molecular dynamics (MD) methods.[55] According to the ergodic hypothesis, both methods should yield the same results either through a Boltzmann-weighted configuration-average in the case of MC or by a time-average (in the long time limit) in the case of MD.

1.2.1 Monte Carlo methods

MC is a stochastic method to sample a statistical ensemble.[55,56] The Metropolis algorithm[57] is used to sample microstates through a random walk process in which microstates are randomly perturbed to sample new microstates. If the new microstate is lower in energy than the previous microstate, the move is accepted. If it is higher in energy, the move is accepted with probability p:

$$p = e^{-\Delta E/(k_B T)} \tag{1.5}$$

where ΔE is the difference in energy between the previous microstate and the current microstate, k_B is the Boltzmann constant, and T is the temperature. This helps to maintain the representative Boltzmann distribution of the ensemble of interest. MC methods require only the energy (and not the forces) of a microstate. Time-independent correlations, such as the pair correlation function (radial distribution function), $g(r)$, can be obtained which can be compared to the structure factor, $S(q)$, derived from X-ray and neutron diffraction experiments.[58–61] Naturally, the dynamics of a system over time and any time correlations cannot be accessed using this method. However, MC methods are straightforward to parallelize, improving the sampling efficiency.

1.2.2 Molecular dynamics

MD is a serial method that propagates a system in time using Newton's equations of motion.[55] More specifically, the velocity-Verlet algorithm is commonly used to minimize the error in the positions of the atoms after time step, Δt.[62] Thermostats and/or barostats are implemented to sample particular NVT (constant number of particles, pressure, temperature) or NPT (constant number of particles, pressure, temperature) ensembles.[63–65] Importantly, in the long time limit (for an ergodic simulation), the time-averaged properties should be equivalent to the Boltzmann-weighted microstate average obtained from MC simulations.

However, since this method propagates the positions of atoms in time, it can naturally be used to obtain a mechanistic understanding of dynamics processes. It can also be used to obtain valuable time correlations which can be compared to experimental observables. For example, the Fourier transform of the dipole autocorrelation function,

$$C(\tau) = \langle \mu(t) \cdot \mu(t+\tau) \rangle \tag{1.6}$$

where $\mu(t)$ is the molecular dipole moment at time t and τ is a time delay, yields the infrared (IR) spectrum of the system.[66–68]

1.2.3 Quantum statistics

To obtain quantum statistics, path integral (PI)[69] methods can be applied to either MC or MD methods.[70] The intent of these methods is to incorporate nuclear quantum effects which are responsible for the delocalization of light atoms.[71] Depending on the temperature of the simulation, these effects may also become important in the description of heavier elements as well.[72] This can be accomplished in the centroid[73–75] or ring-polymer (RP) formalisms[76,77] by representing each respective atom as a ring of "beads" that each interact with the other atoms (in the usual manner) and also with the other "beads" in the same ring through a harmonic potential. The number of beads necessary depend on the mass of the element and the temperature of the simulation. Naturally, the increase in particles/beads increases the cost of the simulation while also requiring a smaller time step.[78] PIMD simulations have been applied to water clusters to study hydrogen bond rearrangements[79–81] and in the simulation of water, ice, and other aqueous systems to probe the impact of nuclear quantum effects on hydrogen bonding.[78,82–87]

1.3 Summary

The above discussion highlights the connection that statistical mechanics provides to macroscopic properties of molecular systems along with the computational tools that can be used to sample relevant statistical ensembles. In addition, it emphasizes that the potential energy function, returning the energy and gradient from the atomic coordinates, is truly at the heart of this connection. Therefore, to obtain reliable predictions of macroscopic properties, it is prudent to accurately represent the interactions of molecules at the atomic-level. The representation of the potential energy can be achieved by many different approaches, ranging from slow, expensive, albeit often reliable *ab initio* methods to fast, effective pairwise or many-body potentials. The computational cost of the theoretical method determines the different timescales that can be accessed.

1.4 Outline of thesis

Chapter 2 will introduce some fragmentation methods, such as the many-body expansion (MBE), which partition molecular properties into contributions from smaller subsystems or physical interactions. The subsequent chapters will focus on the utility of the MBE to understand the nature of interactions, develop interactions potentials, and accelerate *ab initio* calculations. In Chapter 3, the application of the MBE to aqueous ionic clusters will be discussed, focusing on the interplay between ion-water and water-water interactions and how these are modulated by ion identity. Chapters 4 and 5 will outline efforts to use the MBE as a guide to develop accurate, physics-based models for many-body effects in aqueous systems. While the MBE for gas-phase, finite systems is straightforward, Chapter 6 accounts how the MBE can be extended to periodic systems using a novel method. Chapter 7 outlines an analysis of several ice polymorphs, defining novel correlations between the density of the polymorph, the cooperativity of the interactions, and the local tetrahedral order. Finally, chapter 8 presents conclusions with a discussion of current work that leverages the MBE for periodic systems to obtain benchmark lattice energy calculations at a reduced computational cost.

Chapter 2

FRAGMENTATION METHODS

Fragmentation methods partition molecular properties into contributions from smaller subsystems or from different physical interactions. These methods can be used to:

- understand the nature of interactions,[1–3]
- guide the development of accurate interatomic potentials,[43–47, 50, 88] and
- circumvent the "scaling curse" of electronic structure methods,

making them powerful computational tools.[88]

These methods include, among others, symmetry-adapted perturbation theory (SAPT), energy decomposition analysis (EDA), the molecular tailoring approach (MTA), and the many-body expansion (MBE). The MBE will be the focus of the remaining chapters and for this reason it will be introduced in detail. Other methods (MTA, SAPT, EDA) will be introduced briefly and contrasted with the MBE.

2.1 Many-body expansion

The MBE represents properties of the full system combinatorically through its contributions from smaller subsystems. These subsystems are produced from combinations of non-overlapping fragments. While its origin lies in combinatorial mathematics (and the inclusion-exclusion principle), Hankins et al.[89] applied this method for the first time in the chemical physics domain in 1970, examining the nonadditivity of interactions in the water trimer. Specifically, they reported the 3-body contribution to the HF energy of the double donor, double acceptor, and sequential water trimer configurations; however, the ring water trimer global minimum was not considered in that study. Subsequently, Clementi and co-workers[90] considered the non-additivity of the SCF interaction in several configurations of the water trimer and reported that it was “fairly small”. They also pointed out that the dominant contribution to the non-additive energy was the long-range induction energy. In the mid-1990’s the MBE was first applied to the MP2 optimized structures and subsequent binding energies of water clusters up to the hexamer.[91,92] Subsequently, the non-additive terms in gas phase water clusters were contrasted to those in liquid water configurations,[93] whereas the effect of the hydrogen bonding network on the magnitude of the cooperative effects was also investigated.[94] These initial studies served as the stepping-stone for the development of interaction potentials for water by fitting to *ab initio* data, which were appropriate for nuclear statistical mechanical (rather than classical) simulations. These efforts resulted in the development of the flexible, polarizable potentials for liquid water (TTM, TTM2-R, TTM2-F, TTM2.1-F, TTM3-F)[30–32,35–37] that led to the first path integral simulations of liquid water with a flexible, polarizable model.[78] These initial efforts have fueled subsequent development by others of many-body flexible, polarizable potentials such as MB-pol[46–48] HBB2-pol,,[95] WHBB,[49,96] and q-AQUA.[50]

In general, quantity $\mathcal{O}$ (i.e., binding energy, forces, dipole moment, etc.) may be cast as:

$$\mathcal{O} = \sum_{i}^{N} \Delta\mathcal{O}_i + \sum_{i}^{N} \sum_{j>i}^{N-1} \mathcal{O}_{ij} + \sum_{i}^{N} \sum_{j>i}^{N-1} \sum_{k>j>i}^{N-2} \mathcal{O}_{ijk} + \cdots \tag{2.1}$$

where N is the number of fragments (or “bodies”) in the system. The fragments are typically taken to be either ions or molecules interacting through non-covalent bonds, such as hydrogen bonds, but it should be noted that recent works have explored the use of atoms (breaking covalent or metal-metal bonds) or elementary particles as fragments.[97–99] For an exact representation of the property $\mathcal{O}$, the expansion extends to N order, where N is the number of fragments in the system. However, it is often the case that the expansion may be truncated at a lower order with minimal loss in accuracy. The various terms in the MBE are defined as follows:

$\Delta\mathcal{O}_i$ is the 1-body term of the i-th fragment defined as:

$$\Delta\mathcal{O}_i = \mathcal{O}_i - \mathcal{O}_{i,ref} \tag{2.2}$$

where $\mathcal{O}_i$ is the quantity for the i-th fragment and $\mathcal{O}_{i,ref}$ is the reference quantity for the non-interacting i-th fragment. $\Delta\mathcal{O}_{ij}$ is the 2-body term of the i and j fragments defined as:

$$\Delta\mathcal{O}_{ij} = \mathcal{O}_{ij} - \mathcal{O}_i - \mathcal{O}_j \tag{2.3}$$

where O_{ij} is the quantity for the dimer of fragments i and j. $\Delta\mathcal{O}_{ijk}$ is the 3-body term defined as:

$$\Delta\mathcal{O}_{ijk} = \mathcal{O}_{ijk} - \Delta\mathcal{O}_{ij} - \Delta\mathcal{O}_{jk} - \Delta\mathcal{O}_{ik} - \mathcal{O}_i - \mathcal{O}_j - \mathcal{O}_k \tag{2.4}$$

where O_{ijk} is the quantity for the trimer of fragments i, j, and k. The higher-order terms are defined analogously to the 3-body term. Since the number of n-body fragments (for a system of N molecules) grows as the binomial coefficient,

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}, \tag{2.5}$$

the MBE is particularly useful if it converges at a low-order in the expansion. This convergence, however, has been shown to be dependent on the system and property of interest, respectively.[1–3,97,98] Further, unlike some of the energy decomposition techniques outlined below, the physical origin of the many-body terms is oftentimes unclear.

2.2 Molecular tailoring approach

The MTA is very similar to the MBE but with a few notable approximations/modifications. It defines the fragments so that they are overlapping and typically consist of multiple molecules or atoms, thereby incorporating short-range higher order many-body effects. The many-body effects are assumed to be short-range, which is oftentimes a reasonable assumption. In addition, the short-range interactions are prioritized and only nearby combinations of overlapping fragments are typically considered. In this way, only the most salient interactions are included, reducing in some instances the overall computational cost (in comparison to the full MBE).

Consider an example system of molecules A, B, and C that are arranged linearly. The overlapping fragments AB and BC may be selected, defining the total energy of the system, ABC, as:

$$\begin{aligned} E_{ABC} &\approx E_1 + E_2 - E_{1\cap 2} \\ &= E_{AB} + E_{BC} - E_B \end{aligned} \tag{2.6}$$

In comparison to the MBE, it can be shown that this expression assumes that the 3-body energy (ΔE_{ABC}) and the 2-body energy (ΔE_{AC}) are zero. While this is a simple example, it demonstrates the approximations that the MTA applies to the MBE. The MTA has been used to develop benchmarks for the binding energy of large molecular clusters at a reduced computational cost,[100, 101] optimize the geometries of large molecules and clusters,[102, 103] and estimate the strength of intramolecular hydrogen bonds.[104–106]

2.3 Symmetry-adapted perturbation theory

SAPT[107] divides the n-body terms of the MBE into physically-meaningful contributions. The basis for SAPT is the treatment of intermolecular interactions as a perturbation to the individual fragments:

$$E_{AB} = E_A + E_B + W \tag{2.7}$$

In this way, the calculation of the interaction will be free from the basis set superposition error (BSSE). The interaction energy (W) is obtained by calculating the potential between the electrons of one fragment and the nuclei of another fragment (and vice versa) and also the intermolecular contributions to the electron correlation. The 2-body interaction is divided into electrostatic, exchange, induction, and dispersion components. In addition, there are additional corrections that account for the antisymmetry of the wave function for some of the components (denoted with "-exchange"). A series of interaction potentials for carbon dioxide,[108] water,[109] benzene,[110] and ammonia[111] have been developed using the 2-body SAPT as a guide. This formalism can indeed be extended to higher orders. However, the computation of the 3-body contribution is much more expensive, a fact that has hindered its widespread use.

2.4 Energy decomposition analysis

Like SAPT, EDA divides the n-body terms into physically-meaningful contributions. EDA[112, 113] divides the energy of a system into contributions termed "frozen", "polarization", and "charge transfer", using a series of steps to optimize the molecular orbitals of a system.

$$E_{AB} = E_A + E_B + E_{frz} + E_{pol} + E_{CT} \tag{2.8}$$

The most popular method utilizes the absolutely localized molecular orbitals (ALMO) formalism. The frozen term (E_{frz}) arises from the interaction of two monomers (with the monomer wave functions) at the dimer geometry. The polarization term (E_{pol}) accounts for intramolecular relaxation of the orbitals due to the external field. The charge transfer term (E_{CT}) arises from the complete intermolecular relaxation of the molecular orbitals. This method has been used to develop the MB-UCB potential for water[43] which has subsequently been extended to select ion-water systems.[114]

Chapter 3

THE MANY-BODY EXPANSION FOR AQUEOUS SYSTEMS REVISITED: III. HOFMEISTER ION - WATER INTERACTIONS

Reproduced in part with permission from [Kristina M. Herman, Joseph P. Heindel, Sotiris S. Xantheas. The many-body expansion for aqueous systems revisited: III. Hofmeister ion–water interactions; *Phys. Chem. Chem. Phys.* 23, 11196-11210]. Copyright [2021] Royal Society of Chemistry

3.1 Introduction

The Hofmeister series was originally established in the 1880s by Franz Hofmeister to order cations and anions based on their efficacy in precipitating proteins from an aqueous solution containing egg-white.[115,116] This same ordering was later found to be related to protein and hydrocarbon stability in solution as well as other properties that do not involve proteins, like surface tension and viscosity.[117–120] While some research has examined the interplay of ions with the protein backbone and amino acid residues to explain the salting-out properties of ions,[121–136] recent physical chemistry research[137–157] has been focused on understanding the dynamical and structural impacts that specific ions may have on liquid water and the physical mechanism underlying these effects.

The Hofmeister series orders cations and anions, respectively, in the following manner:

$$\begin{aligned} CO_3^{2-} > SO_4^{2-} > H_2PO_4^- > F^- > Cl^- > Br^- > NO_3^- > I^- > ClO_4^- > SCN^- \\ Ca^{2+} > Li^+ > Na^+ > K^+ > Cs^+ > NH_4^+ > N(CH_3)_4^+ \end{aligned} \tag{3.1}$$

The ions at the opposite ends of the Hofmeister series are termed kosmotropes (left) and chaotropes (right), or “structure-makers” and “disorder-makers”/“structure-breakers”, respectively. Kosmotropes efficiently salt proteins out of solution, stabilize the 3-dimensional configuration of proteins, and increase the surface tension and viscosity of solutions, while chaotropes behave in the opposite manner. Although the terms "structure-makers" and "disorder-makers" seem to imply the ion's effect on water molecules, the role ions play in water and the extent to which they can influence distant solvation shells is highly contested. Perhaps the most widely debated aspect of this research is determining what it means for an ion to be structure-making, how it manifests itself, and how this ability can be measured or observed. This ambiguity has led to many different interpretations and hypotheses regarding the Hofmeister effect.

Several different metrics have been used to gain insight into the Hofmeister effect, some focusing on the range of the ion's influence on the surrounding water molecules and others on the structural changes in the hydrogen bonding network as a result of the ion. Thermodynamic studies on the entropy and Gibbs energy of solvation[141] and terahertz echo experiments[138] have both suggested that Hofmeister ions cause structural changes in the hydrogen bonding network which result in different entropic quantities (as a measure of order) and decay times within the water–water modes of aqueous systems. The presence of the "dangling O–H" in aqueous nanodroplets,[142–144] red- and blue-shifting of the bending mode of water,[145] deviations in neutron diffraction reminiscent of water's behavior at high pressures,[146, 158] and orientational order measured by femtosecond elastic second harmonic scattering experiments[137] have supported a relatively long-range influence of kosmotropes, extending to or exceeding three solvation shells. However, the hindrance of water rotations in solely the innermost solvation shell using femtosecond pump–probe spectroscopy,[147] the negligible changes in the relaxation time of the bulk hydrogen bond network of various anionic Hofmeister systems from dielectric relaxation spectroscopy experiments,[148] and diminished hydrogen bonding of water molecules only in the first solvation shell relative to bulk water[149] suggested that water–water interactions are not actually disrupted or influenced beyond the first solvation shell. X-ray emission spectroscopy found contradicting results, finding some ions with no effect on water and others with a marginal effect. Importantly, these results did not follow the ordering of the Hofmeister series.[150] Computationally, molecular dynamics is often employed because of the structural and dynamical insights provided. For instance, molecular dynamics simulations have been used to observe the partitioning of the ion between the bulk and air–water interface,[120, 151] to measure tetrahedrality in the solvation shell (as a measure of order),[152] to count the average number of hydrogen bonds,[153] to examine the patterning and order of solvation shells using the pair correlation function,[154–156] and to calculate water reorientation time and ion–water hydrogen bonding lifetime using the velocity autocorrelation function.[155,157] Importantly, by using these different metrics and interpretations of "structure-making" properties, different conclusions have been reached regarding the ion's "structure-making" or "disorder-making" ability.

Despite these strides in understanding the Hofmeister effect, a general theory that unifies the observations listed above is far from being complete. Although the Hofmeister series was developed originally to track the effect of the ions on the salting-in and salting-out of proteins in an aqueous solution, recent literature may have implied that the ions (without the protein) may affect the long-range structure of liquid water depending on where they reside in the series.[137,138,142–146,158] Here, as a new perspective in the effort to gain molecular level insight into the Hofmeister effect, *ab initio* electronic structure calculations for aqueous ionic clusters were used to analyze the effect of various cations and anions on either side of the Hofmeister series (3.1) on the interactions between water molecules in the immediate surroundings of these ions. We selected aqueous ionic clusters containing singly and doubly charged cations and anions residing on the two opposite ends of the Hofmeister series to be able to investigate the effect of both the charge (nominal value, sign) and structure making/breaking ability of the ion on the interactions of the surrounding hydrogen bonded water network. We obtained cluster minima of the kosmotropes Ca^{2+} and SO_4^{2-} and the chaotropes NH_4^+ and ClO_4^- with 9 water molecules and performed a many-body expansion (MBE) of the ion–water and water–water interactions in order to be able to directly compare the results with the ones previously reported for pure water[2] and monatomic alkali metal/halide aqueous clusters[1] with the same number of fragments. As in our previous study, for each ion we considered two structural arrangements: one with the ion located inside and the other with the ion on the outside of a water cluster network. This choice of cluster size was deemed large enough to describe artifacts due to BSSE and it can produce realistic conformations mimicking fully solvated (ion inside) and interfacial (ion outside) solvation, while at the same time being computationally tractable.

The MBE, a combinatorial approach, was used to partition the binding energy of the system into its constituent n-body terms[89, 92, 159] where the "fragments" or "bodies" (B) refer to the individual water molecules and ions.

$$D_{e,cluster} = \sum \Delta E_I + \sum \Delta E_{IJ} + \sum \Delta E_{IJK} + \cdots + \sum \Delta E_{IJK \ldots N} \tag{3.2}$$

The 1-B term (ΔE_I) is the energetic penalty[160, 161] for each fragment to deform from its gas-phase geometry to the one it adopts in the cluster geometry due to its interaction with the rest of the "bodies". The 2-B term (ΔE_{IJ}) is typically the largest energetically favorable term that reflects the difference between the energy of each dimer and the energy of the monomers that comprise the dimer. The 3-B term (ΔE_{IJK}) describes how the energy of each dimer and monomer is perturbed with respect to the other monomers present in the system. The successively higher order terms are defined analogously and tend to decrease in magnitude in hydrogen-bonded systems.[93, 94]

This type of energy analysis has been recently applied to pure water clusters of varying sizes[2] as well as monatomic (alkali metal cation and halide anion) aqueous clusters.[1] It was shown that the MBE of pure water clusters converges at the 4-B term while the MBE of monatomic cations/anions is comprised of especially large 2-B interactions with a repulsive 3-B term, in contrast to the pure water clusters, which exhibit an attractive 3-B term at the cluster minima considered in these previous studies. In addition, an unexpected linear anti-correlation between the total 2-B ion–water and the total 2-B water–water interaction was observed for both the alkali metal cation and the halide anion clusters. The MBE of Hofmeister ion aqueous clusters, to the best of our knowledge not reported to date, contributes a novel perspective to the discussion of the Hofmeister effect by breaking down the energetics of the whole system, analyzing the ways and the extent that a specific ion impacts these interactions, and comparing this breakdown to the one for ions in a different position of the Hofmeister series. Furthermore, the effect of the cations/anions and kosmotropes/chaotropes on the interactions of the surrounding water molecules can be assessed independently. Being able to examine the physics of one ion at a time is a useful way to reduce the problem, since it has been shown that pairs of ions often lead to different behaviors due to solvent separated, solvent shared, and contact ion pairing.[162, 163] In addition, the relative impact of cations and anions[140, 155, 164] has also been discussed. By studying each independently, differences in the energetics of cationic and anionic systems can be analyzed and quantified. To the best of our knowledge, this type of energy analysis has not been reported earlier for these systems. The paper is organized as follows: in Section II we outline the computational details and a brief overview of the methodology we adopted in our theoretical study. In Section III we present the results for the magnitude of the various many-body (MB) interactions and the contribution of the constituent ion–water (I–W) and water–water (W–W) interactions. We further attempt to identify correlations between the total (I–W) and (W–W) interactions in an effort to investigate how ions on different ends of the Hofmeister series (chaotropes/kosmotropes) affect the W–W interaction. In that section we also investigate the profile of the 2-B BSSE correction as a function of the interfragment distance for the different ions in an effort to integrate the current results with the ones previously reported for (W–W) and alkali metal/halide–water interactions.[1,2]

3.2 Computational details

The kosmotropes Ca^{2+} and SO_4^{2-} and the chaotropes NH_4^+ and ClO_4^- ions were studied as part of water clusters containing nine water molecules. Note that these singly and doubly charged ions reside close to the opposite ends of the Hofmeister series (3.1) for both cations and anions. The size of these clusters (ion plus 9 water molecules) was selected because it was large enough to allow the formation of configurations in which the ion resides both in the inside and the outside of a water network and allow a direct comparison of the results with pure water clusters and monatomic aqueous clusters of the same size.[1,2] The intent is to initially investigate the dependence of the results on the different conformational isomers and coordination numbers, albeit with this limited sampling of the configuration space that explores "bulk-like" and "interface-like" environments. The same number of "bodies" (10) was used in order to directly compare the current results with the previous ones for the $(H_2O)_{10}$ and $Z^{+/-}(H_2O)_9$, Z = Li^+, Na^+, K^+, Rb^+, Cs^+, F^-, Cl^-, Br^-, I^-, clusters.

Some of these structures were taken from published works[165–168] and, when not available in the literature, were built from a z-matrix. The $NH_4^+(H_2O)_9$ structure with the ion on the inside was built entirely from a z-matrix, constraining dihedral angles of 120° and tetrahedral binding angles of 109.5° to keep the ion on the inside of the cluster. The remaining structures are local minima on the potential energy surface. The structures of the clusters analyzed via the MBE are shown in Figure 3.1.

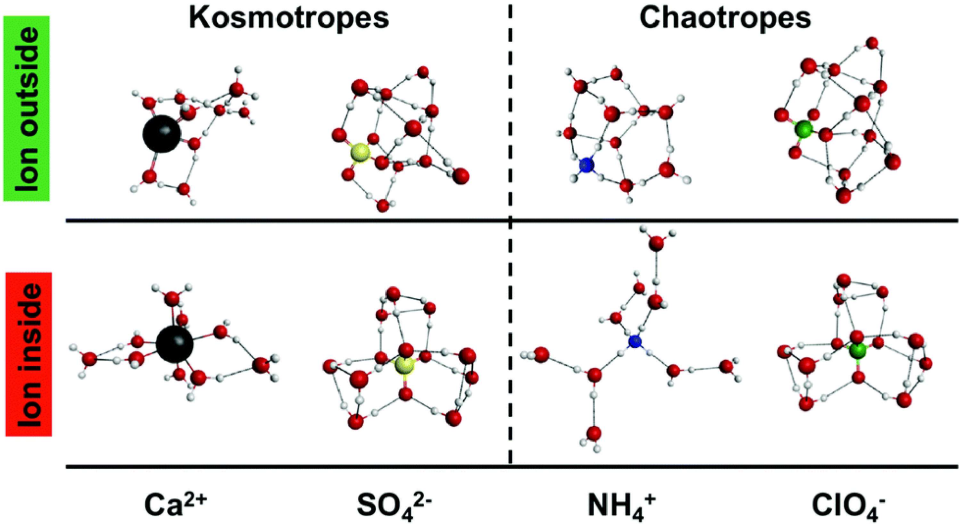


Figure 3.1: The geometries of the aqueous ionic clusters with the ion outside (top row) and inside (bottom row) a water cluster network.

The MBE, equation 3.2, was carried out to the 10th order (1024 energies computed) for each cluster, viz.

$$D_{e,cluster} = \sum_{I}^{10} \Delta E_I + \sum_{I<J}^{10} \Delta E_{IJ} + \sum_{I<J<K}^{10} \Delta E_{IJK} + \cdots + \sum_{I<J<K<\cdots<N}^{10} \Delta E_{IJK\cdots N} \quad (3.3)$$

The 1-body term (equation 3.4) is the molecular relaxation term describing the energetic difference between the monomers in the cluster geometry relative to the reference (relaxed) geometry:

$$\Delta E_I = E_I - E_{I,ref} \quad (3.4)$$

The two-body term (equation 3.5) is the energetic difference between each dimer and the monomers that comprise the dimer:

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J \quad (3.5)$$

The three-body term (equation 3.6) is defined as the energy difference between each trimer and the monomers comprising that trimer while also considering the lower order n-body terms (2-body terms of each dimer that makes up the trimer):

$$\Delta E_{IJK} = E_{IJK} - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK} - E_I - E_J - E_K \tag{3.6}$$

The higher order terms (4- to 10-B) in the expansion are defined analogously.

All calculations were performed at the MP2 level with the aug-cc-pVDZ basis set[16] for all atoms except calcium, for which the Stuttgart RSC 1997 ECP[169] was used. Importantly, the resulting n-body terms of the MBE were corrected for the basis set superposition error (BSSE) (additional 1023 energies for the various fragments in the full cluster basis for each ion and inside/outside configuration considered), which is the result of the artificial lowering of the energy of the complex due to borrowing of basis functions from other fragments in proximity. This effect is present in all binding energy calculations and, ultimately, is an artifact of an incomplete, finite basis set. We corrected for BSSE using the function counterpoise correction as outlined by Boys and Bernardi.[170] During our earlier studies,[1,2] it was shown that the MP2/aug-cc-pVDZ calculations sufficiently describe the values of the 3-B, 4-B and higher terms at the complete basis set (CBS) limit when BSSE corrections are taken into account. Naturally the 2-B term and, in turn, the total cluster binding energy (whose largest component is the former) are not converged at that level; however, the focus of the present study is on the magnitude of the higher order terms and their correlations, which are accurately accounted for. The MBE for the $Ca^{2+}(H_2O)_9$ clusters was also performed with the aug-cc-pVTZ basis set up to the 6-body term to ensure that the behavior of the MBE with the basis set for an ionic cluster encompassing a doubly charged cation follows the same pattern as for a monatomic (positive or negative) ion. The calculations were performed with zero linear dependencies to guarantee that the full number of basis functions were utilized. All calculations were performed with the NWChem 6.8 electronic structure package.[171]

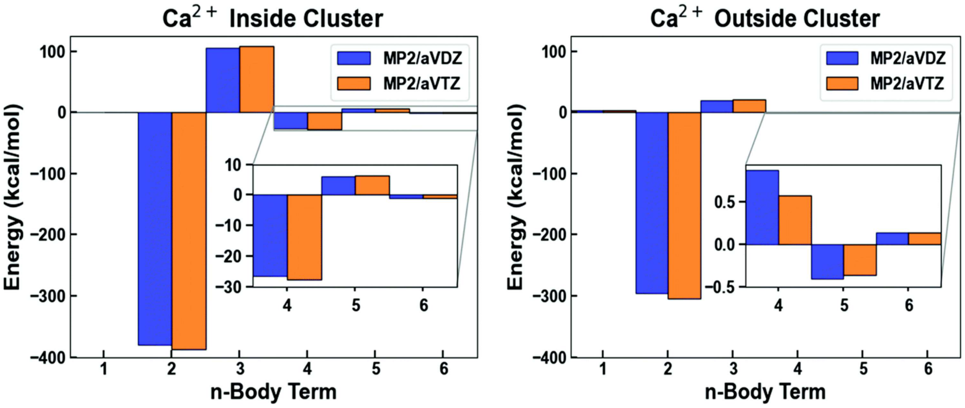


Figure 3.2: Relative magnitudes of the M-B terms up to 6-B for the two isomers of the $Ca^{2+}(H_2O)_9$ cluster at the MP2 level of theory including BSSE corrections with the aVDZ (blue) and aVTZ (orange) basis sets.

3.3 Results and discussion

3.3.1 Magnitude of many-body (MB) terms

We first investigate the behavior of the MBE for a doubly charged ion, since this is the first instance that we attempt decomposition for such a system. For the two isomers (ion inside and ion outside) of the $Ca^{2+}(H_2O)_9$ cluster (cf. Figure 3.1) the MBE was performed at the MP2 level up to the (complete) 10th order with the aVDZ basis set and up to the 6th order with the aVTZ basis set. The BSSE-corrected results with these two basis sets are summarized in Table 3.1. The BSSE-corrected MP2 1-B to 6-B terms are plotted in Figure 3.2 with the aVDZ (blue) and aVTZ (orange) basis sets. At this point we are interested in the basis set dependence of the various terms and we will discuss the differences in the magnitudes of the individual terms between the two isomers later. There is no significant qualitative difference in the resulting MB terms between these two basis sets. There is a noteworthy quantitative difference in the 2-body term with the largest difference between the calculated MB terms with different basis set sizes being 8.8 kcal mol^{-1}. This is consistent with the work reported earlier[1] on the monatomic aqueous clusters, utilizing basis set sizes up to the aug-cc-pV5Z. From these results we ascertain that the MBE for an aqueous doubly charged ion behaves similarly to that for a singly charged ion and this result further justifies the level of theory and protocol (MP2/aVDZ including BSSE corrections) we have adopted for the MBE analysis in this study.

	Ion inside		**Ion outside**	
k	**MP2/aVDZ**	**MP2/aVTZ**	**MP2/aVDZ**	**MP2/aVTZ**
1-B	0.80	0.88	3.416	3.837
2-B	-379.547	-387.367	-294.908	-303.757
3-B	104.649	107.794	18.935	21.050
4-B	-26.580	-27.612	0.879	3.581
5-B	6.074	6.378	-0.411	-0.370
6-B	-0.996	-1.070	0.133	0.136

Table 3.1: Basis set dependence of the MBE terms up to 6-B for the $Ca^{2+}(H_2O)_9$ clusters at the MP2 level with the aVDZ and aVTZ basis sets including BSSE corrections.

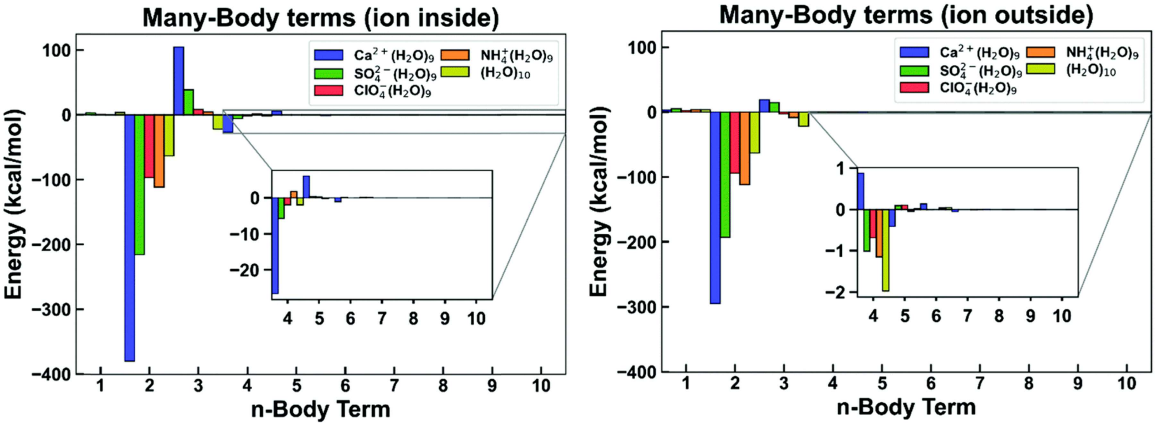


Figure 3.3: Many-body terms of each ion cluster $X^{+/-}(H_2O)_9$ and $(H_2O)_{10}$ for the ion inside (left panel) and ion outside (right panel) geometries.

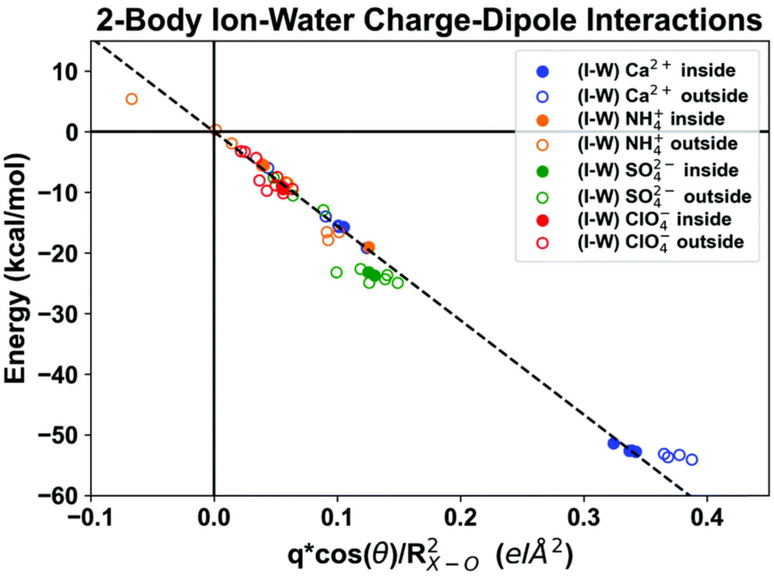


Figure 3.4: Individual 2-B (I-W) contributions for the various ions as a function of $q{*}\cos(\Theta)/R^2_{X-O}$. The linear fit (broken line) has a slope of -155.60 ($R^2 = 0.9777$).

	$Ca^{2+}(H_2O)_9$		**$NH_4^+(H_2O)_9$**		**$SO_4^{2-}(H_2O)_9$**		**$ClO_4^-(H_2O)_9$**	
k	**Ion inside**	**Ion outside**	**Ion inside**	**Ion outside**	**Ion inside**	**Ion outside**	**Ion inside**	**Ion outside**
1-B	0.81 (-0.3)	3.42 (-1.3)	0.41 (-0.4)	3.82 (-3.2)	3.24 (-1.8)	5.11 (-2.9)	1.08 (-1.2)	2.58 (-2.7)
2-B	-379.55 (128.4)	-294.91 (108.4)	-111.40 (106.5)	-111.97 (94.9)	-215.42 (120.5)	-193.41 (110.9)	-96.76 (108.7)	-94.30 (99.1)
3-B	104.65 (-35.4)	18.94 (-7.0)	4.74 (-4.5)	-8.65 (7.3)	38.73 (-21.7)	14.84 (-8.5)	8.36 (-9.4)	-2.80 (2.9)
4-B	-26.58 (9.0)	0.88 (-0.3)	1.80 (-1.7)	-1.15 (1.0)	-5.70 (3.2)	-1.01 (0.6)	-1.95 (2.2)	-0.68 (0.7)
5-B	6.07 (-2.1)	-0.41 (0.2)	-0.15 (0.1)	-0.05 (0.0)	0.3 (-0.2)	0.10 (-0.1)	0.27 (-0.3)	0.10 (-0.1)
6-B	-1.00 (0.3)	0.13 (-0.0)	-0.001 (0.0)	0.04 (0.0)	0.11 (-0.1)	-0.01 (0.0)	-0.03 (0.0)	-0.003 (0.0)
7-B	0.07 (-0.0)	-0.05 (0.0)	0.001 (0.0)	-0.01 (0.0)	-0.04 (0.0)	0.000 (0.0)	0.003 (0.0)	0.000 (0.0)
8-B	0.01 (0.0)	0.01 (0.0)	0.000 (0.0)	0.000 (0.0)	0.01 (0.0)	0.000 (0.0)	0.000 (0.0)	0.000 (0.0)
9-B	-0.002 (0.0)	0.000 (0.0)	0.000 (0.0)	0.000 (0.0)	-0.001 (0.0)	0.000 (0.0)	0.000 (0.0)	0.000 (0.0)
10-B	0.002 (0.0)	0.000 (0.0)	0.000 (0.0)	0.000 (0.0)	0.000 (0.0)	0.000 (0.0)	0.000 (0.0)	0.000 (0.0)
Total	**-295.52**	**-272.00**	**-104.60**	**-117.97**	**-178.77**	**-174.38**	**-89.03**	**-95.11**

Table 3.2: MBE (kcal mol^{-1}) terms for the various isomers of the clusters in Fig. 3.1 at the MP2/aVDZ level of theory including BSSE corrections. The numbers in parentheses correspond to the percentage of the total energy.

The magnitude of the various MBE terms (in kcal mol^{-1}) up to the 10th order for the ion inside and ion outside isomers of the four ionic clusters considered in this study (cf. Figure 3.1) at the MP2/aVDZ level of theory including BSSE corrections is listed in Table 3.2. The numbers in parentheses in Table 3.2 correspond to the percentage contribution of each term to the total cluster binding energy. These results suggest that terms above the 4-B term (save the ion inside isomer of the $\mathrm{Ca}^{2+}(\mathrm{H}_2\mathrm{O})_9$ cluster) are negligible, contributing <0.3% to the total binding energy. The BSSE-corrected many-body terms of $(\mathrm{H}_2\mathrm{O})_{10}$ will be used as a reference,45 providing a direct point of comparison in this section. The full MBE of the aqueous ionic clusters $\mathrm{Ca}^{2+}(\mathrm{H}_2\mathrm{O})_9$, $\mathrm{NH}_4^{+}(\mathrm{H}_2\mathrm{O})_9$, $\mathrm{SO}_4^{2-}(\mathrm{H}_2\mathrm{O})_9$, and $\mathrm{ClO}_4^{-}(\mathrm{H}_2\mathrm{O})_9$ and that of the $(\mathrm{H}_2\mathrm{O})_{10}$ water cluster are collected together in Figure 3.3. The aqueous ionic clusters exhibit an increased 2-B term relative to that of pure water due to the strong ion–water (charge–dipole) interaction, which is naturally larger for the divalent ions. This is consistent with what was observed with the MB expansion of water clusters containing monatomic ions.[1] The percentage contribution of the 2-B term to the total cluster binding energy is noticeably larger for the kosmotropes (calcium and sulfate) than the chaotropes (perchlorate and ammonium). This is not surprising given that kosmotropes are typically more charge dense[172–176] and the 2-B terms are dominated by the charge–dipole interactions between the ion and water. Figure 3.4 shows the individual 2-B (I–W) contributions to the total BSSE correction for the various ions considered in this study as a function of $q * cos(\Theta)/R^2_{X-O}$, where q is the magnitude of the charge, R is the distance, and Θ the angle between the ion and the molecular dipole of the nearest water molecule. The 72 points follow a linear trend (broken line) with slope -155.60 quite closely ($\mathrm{R}^2 = 0.9777$). When these interactions are not scaled for the total charge (i.e., the x-axis is just $cos(\Theta)/R^2_{X-O}$, four linear trends emerge for $q = -1, -2, +1, +2$. The 2-B terms are larger when the ion is on the inside because the ion–water distances are, on average, shorter for this cluster configuration thus resulting in stronger charge–dipole interactions. Despite the large magnitude of the attractive 2-B terms, the 3-B terms of the aqueous systems are repulsive with the exception of perchlorate and ammonium on the outside of the cluster. This contrasts the attractive 3-B term for pure water clusters, which further strengthens the hydrogen bonding network. The two kosmotropes (calcium and sulfate) have a large repulsive 3-B term, which becomes increasingly more favorable as the ion becomes more chaotropic, with perchlorate and ammonium exhibiting a favorable 3-B term when these ions are located on the outside of the cluster. The 4-B term varies significantly based on the location of the ion in the water cluster and may depend on the sign of the ionic charge (anion or cation). When the ions are located on the inside, the kosmotropes exhibit increased favorability of the 4-B term relative to pure water, the perchlorate anion behaves similarly to pure water, whereas the ammonium cation has a slightly repulsive term. However, when the ion resides on the outside of the cluster, the 4-B term is weakly attractive for all ions, save calcium. The anions have a consistently favorable 4-B term, while the cations have a strongly attractive or repulsive term depending on their location in the cluster. The 4-B term of chaotropes is noticeably more negative when the ions reside on the outside of the cluster. The opposite is true for the kosmotrope ions considered in this study.

The expansion of the pure water cluster converges at the 4-B term.[2] However, the expansion of the aqueous clusters extends beyond the 4-B term, to varying degrees. Notably, calcium, when located on the inside of the cluster, exhibits a 5-B term of 6.07 kcal mol^{-1} (2.1% of the total cluster binding energy) and a 6-B term of -1.00 kcal mol^{-1} (0.3% of the total cluster binding energy). Conversely, ammonium has a small 5-B term of -0.15 kcal mol^{-1} (0.1%) and just -0.05 kcal mol^{-1} when residing on the inside and outside of the aqueous cluster, respectively. The 5-B term of ammonium is noteworthy, but not nearly as significant as the one for the two kosmotropes. In general, the MBE appears to converge more quickly for the chaotropes than for the kosmotropes. Further, when the ion is on the inside of the cluster, the MBE tends to have larger higher order terms in the expansion. This is likely due to increased proximity of the ion to the water molecules, resulting in stronger (I–W) interactions and a further disruption and subsequent weakening of the (W–W) interaction. Considering the MB terms altogether, the expansion of the chaotropic ions more closely resembles that of pure water than the kosmotropes do.

Examining the total magnitude of the MB terms provides valuable insight, as discussed previously. However, other variables such as geometry and coordination number likely play a role, making it difficult to extract general trends solely by examining the many-body terms. The subsequent sections aim at dissecting the MB terms in order to provide a better understanding of the ion's role in affecting these MB terms and how the (W–W) interactions are impacted as a result of the ion's presence.

3.3.2 Ion–water (I–W) and water–water (W–W) contributions to the MBE terms

In this section we examine the contribution of the total (I–W) and (W–W) interactions to the total binding energy as well as to the individual MB terms. Each term in the MBE can be split into these two contributions as

$$D_e = \sum \Delta E_I + \sum \Delta E_W + \sum \Delta E_{I-W} + \sum \Delta E_{W-W} + \sum \Delta E_{I-W-W} + \sum \Delta E_{W-W-W} + \cdots \tag{3.7}$$

	$Ca^{2+}(H_2O)_9$				$NH_4^+(H_2O)_9$			
	Ion inside		**Ion outside**		**Ion inside**		**Ion outside**	
k	**I-W**	**W-W**	**I-W**	**W-W**	**I-W**	**W-W**	**I-W**	**W-W**
1-B	0.00 (0.0)	0.81 (100.0)	0.00 (0.0)	3.42 (100.0)	0.13 (32.4)	28 (67.7)	0.88 (22.9)	2.95 (77.1)
2-B	-398.78 (105.1)	19.23 (-5.1)	-284.61 (96.5)	-10.30 (3.5)	-103.21 (92.7)	-8.18 (7.4)	-69.55 (62.1)	-42.43 (37.9)
3-B	106.76 (102.0)	-2.11 (-2.0)	16.34 (86.3)	2.59 (13.7)	1.63 (34.4)	3.11 (65.6)	4.21 (-48.6)	-12.86 (148.6)
4-B	-26.96 (101.4)	0.38 (-1.4)	0.73 (82.8)	0.15 (17.3)	1.76 (97.8)	0.04 (2.2)	0.26 (-22.2)	-1.40 (122.2)
5-B	6.08 (100.1)	-0.01 (-0.1)	-0.41 (100.9)	0.004 (-0.9)	-0.14 (91.3)	-0.01 (8.7)	-0.05	0.007
6-B	-0.99 (99.1)	-0.008 (0.9)	0.14 (103.2)	-0.004 (-3.2)	-0.001	0.000	0.03	0.009
7-B	0.06	0.002	-0.05	0.000	0.001	0.000	-0.006	0.001
8-B	0.01	0.000	0.01	0.000	0.000	0.000	0.000	0.000
9-B	-0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10-B	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Total	**-313.82 (106.2)**	**18.30 (-6.2)**	**-267.86 (98.5)**	**-4.14 (1.5)**	**-99.83 (95.4)**	**-4.77 (4.6)**	**-64.24 (54.5)**	**-53.73 (45.5)**

	$SO_4^-(H_2O)_9$				$ClO_4^-(H_2O)_9$			
	Ion inside		**Ion outside**		**Ion inside**		**Ion outside**	
k	**I-W**	**W-W**	**I-W**	**W-W**	**I-W**	**W-W**	**I-W**	**W-W**
1-B	0.58 (17.9)	2.66 (82.1)	0.93 (18.3)	4.18 (81.8)	0.11 (10.4)	0.96 (89.6)	0.76 (29.4)	1.82 (70.6)
2-B	-210.22 (97.6)	-5.20 (2.4)	-174.71 (90.3)	-18.71 (9.7)	-82.93 (85.7)	-13.83 (14.3)	-64.54 (68.4)	-29.76 (31.6)
3-B	46.00 (118.8)	-7.27 (-18.8)	19.46 (131.1)	-4.62 (-31.1)	15.89 (190.0)	-7.53 (-90.0)	6.00 (-213.8)	-8.80 (313.8)
4-B	-6.08 (106.7)	0.38 (-6.7)	-1.07 (105.5)	0.06 (-5.5)	-2.14 (109.8)	0.19 (-9.8)	-0.19 (28.3)	-0.49 (71.7)
5-B	0.33 (110.7)	-0.03 (-10.7)	0.07	0.03	0.28 (103.9)	-0.01 (-3.9)	0.06	0.05
6-B	0.11 (98.0)	0.002 (2.0)	-0.005	-0.003	-0.03	0.001	-0.004	0.001
7-B	-0.04	0.000	0.000	0.000	0.003	0.000	0.000	0.000
8-B	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9-B	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10-B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Total	**-169.32 (94.7)**	**-9.46 (5.3)**	**-155.32 (89.1)**	**-19.06 (10.9)**	**-68.82 (77.3)**	**-20.21 (22.7)**	**-57.92 (60.9)**	**-37.18 (39.1)**

Table 3.3: Decomposition of the MBE (1-B to 10-B) terms into ion–water (I–W–$\cdots$–W) and water–water (W–W–$\cdots$–W) contributions (kcal mol^{-1}) for the clusters of Fig. 3.1 at the MP2/aVDZ level of theory including BSSE corrections. The numbers in parentheses indicate the percentage of each contribution to the total magnitude of the MB terms listed in Table 3.2.

The total (I–W) and (W–W) interactions are the sum of the ion-containing and pure water terms, respectively, from each MB term in the expansion. Table 3.3 lists the total (I–W) and (W–W) interactions as well as their contributions to the individual 1-B to 10-B terms for the two configurations (ion inside/ion outside) of the four ions considered in this study. The numbers in parentheses indicate the percentage contributions of each term [(I–W) and (W–W)] to either the total cluster binding energy or to the individual MB terms from 1-B to 10-B (the two numbers in parentheses add up to 100% for each MB term as well as the total).

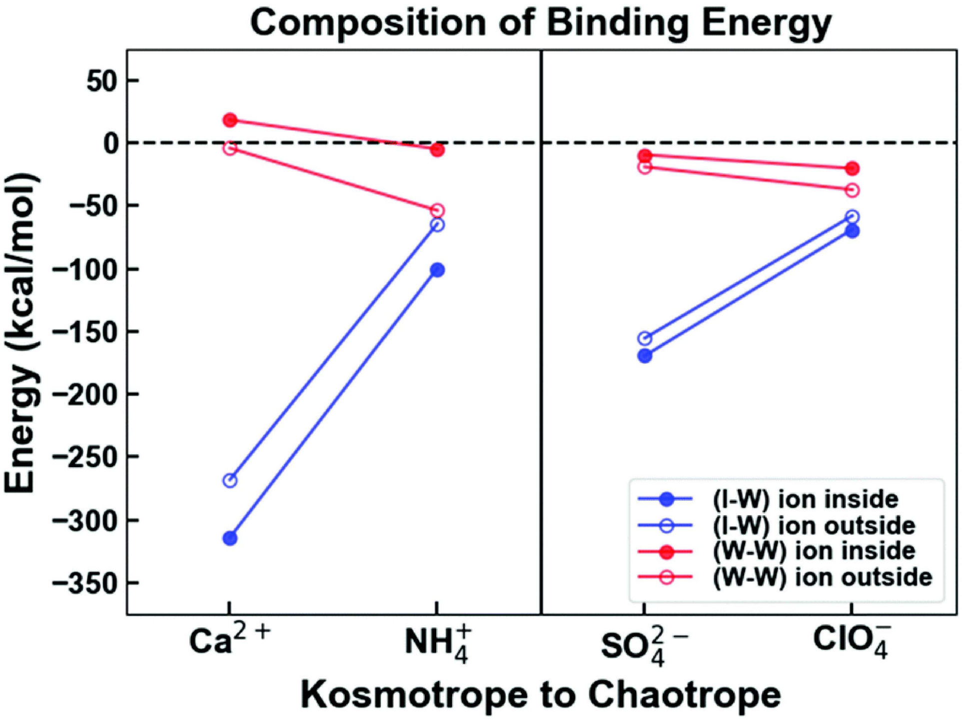


Figure 3.5: The composition of the total cluster binding energy in terms of the (I–W) and (W–W) interactions for the configurations with the ions residing on the outside (open circles) and the inside (filled circles) of the aqueous cluster. All numbers are corrected for BSSE.

Composition of the total binding energy. The total (I–W) and (W–W) interactions and their variation for each ion are shown in Figure 3.5 (note that their sum is the total binding energy for each cluster). Importantly, the cations and anions are separated, ensuring that only ions of the same sign are compared to one another depending on their position in the Hofmeister series (3.1). Within the cation (left) and anion (right) sections of the plot, the leftmost ions are the kosmotropes (Ca^{2+}, SO_4^{2-}) and the rightmost ions are the chaotropes (NH_4^+, ClO_4^-). The open circles denote the results when the ion is on the outside of the cluster whereas the closed circles the ones when the ion is in the inside. Figure 3.5 clearly demonstrates the anti-correlation between the (I–W) and (W–W) interactions, previously reported for the alkali metal and halide aqueous clusters: the stronger the (I–W) interaction, the weaker the (W–W) interaction. It is important to point out that the kosmotropes considered in this study have significantly larger total binding energies than the corresponding chaotropes, due to the difference in the magnitude of the charge ($q = \pm 2$ vs. $q = \pm 1$). However, even within the doubly charged kosmotropes (Ca^{2+}, SO_4^{2-}), the stronger (I–W) interaction (in Ca^{2+}) results in a weaker (W–W) interaction, whereas the weaker (I–W) interaction (in SO_4^{2-}) results in a slightly stronger (W–W) interaction than for Ca^{2+}. The same trend, albeit not as pronounced as in the kosmotropes, is observed for the chaotropes. For the kosmotropes, the binding energy is largely described by the (I–W) interactions since these amount from 89% to 106% of the total cluster binding energy. However, for the chaotropes the (W–W) interaction comprises a much larger percentage (55% to 95%, cf.Table 3.3) of the total cluster binding energy. Furthermore, the (I–W) interactions vary more significantly than the (W–W) interactions based on the identity of the ion. This is interesting because the ion identity affects the (W–W) interactions to a smaller degree than one would expect based on the total MB terms. Understanding these contributions to the individual many-body terms will be the subject of the following two sections.

Composition of individual MB terms. In this section we discuss the contributions of the (I–W) and (W–W) interactions to the individual 1-B to 10-B terms, which are listed in Table 3.3. The 2-B to 4-B interactions are the primary focus because they are the largest contributors to the total binding energy. Figure 3.6 shows the individual ion–water (left panel) and water–water (right panel) contributions to the 2-B to 4-B terms for the kosmotropic (leftmost ions within each panel) and the chaotropic ions (rightmost ions within each panel) plotted according to the convention used for Figure 3.5. Open circles represent

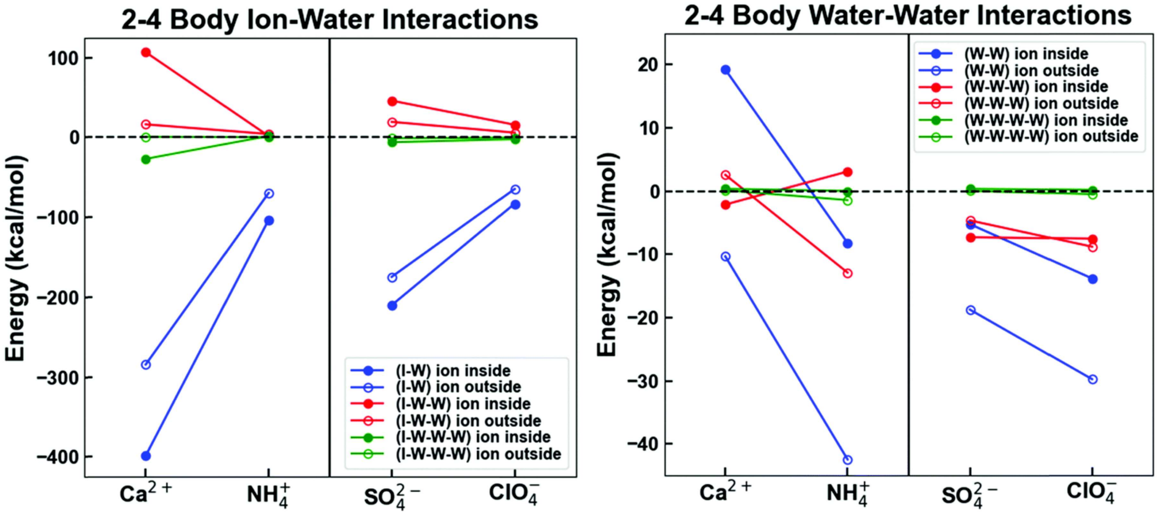


Figure 3.6: Ion–water (left panel) and water–water (right panel) contributions to the 2-, 3-, and 4-B terms of the Hofmeister ions considered in this study. The same plotting conventions as in Figure 3.4 are used. All numbers are corrected for BSSE.

the values calculated when the ion is on the outside whereas closed circles represent the values for the clusters with the ion on the inside of the cluster.

We first examine the ion–water terms (left panel of Figure 3.6). The variation of the ion–water terms is more regular than the water–water terms. Indeed, the individual (I–W), (I–W–W) and (I–W–W–W) terms oscillate in sign with their magnitude being larger for the kosmotropic ions. However, we see a more consistent trend in the 4-B (I–W–W–W) term than in the total 4-B term. The 4-B term is relatively large and favorable (attractive) for the kosmotropes, while it is practically zero for the chaotropes. Interestingly, the 4-B terms for all systems with the ion on the outside remain relatively close to zero. The 3-B and 4-B terms for the chaotropes approach zero, leaving the majority of the ion–water contributions described by the 2-B term. Most significantly, the 3-B (I–W–W) term for the ammonium systems does not exceed 4.2 kcal mol^{-1} whereas that of the calcium system is greater than 100 kcal mol^{-1}. Further, the kosmotropes, particularly calcium, exhibit higher order ion–water terms, extending up to a 6-B term that is ∼1 kcal mol^{-1}. In general, the ion–water contributions are much smaller when the ion resides on the outside of the aqueous cluster. Since the ion is, on average, located farther from the water molecules, this result is expected.

With regard to the water–water contributions in the MBE (right panel of Figure 3.6), they are more favorable for the chaotropes for each term examined (2-B to 4-B). This is consistent with the increased percentage contribution of the water–water interaction to the total cluster binding energy discussed in the previous subsection for the chaotropes. The water–water contributions are much closer to zero for the kosmotropes, reflecting the dominance of the ion–water interaction in nearly all terms of the MBE. The convergence of the water–water contributions of these aqueous clusters is similar to that for the pure water clusters, which converge at the 4-body term. However, in terms of the magnitude and sign of the terms, there are considerable differences. When the chaotropes are on the outside of the cluster, the MBE exhibits water–water 2- to 4-B body terms that resemble that of pure water. For comparison, the 2-, 3-, and 4-B terms of water are -62.76 (76.2%), -21.69 (26.3%) and -1.97 (2.4%) kcal mol^{-1}, while those of ammonium on the outside of the cluster are -42.44 (79.0%), -12.86 (23.9%) and -1.40 (2.6%) kcal mol^{-1} and perchlorate are -29.76 (80.0%), -8.80 (23.7%) and -0.49 (1.3%) kcal mol^{-1}, respectively. All other ionic clusters exhibit at least one of the 2- to 4-B terms that is energetically unfavorable. Interestingly, when the chaotropes are on the inside and the kosmotropes are on the outside, the water–water terms are similar. When the ion is in the center of the cluster, we would expect maximal disruption of hydrogen bonding. Thus, the chaotropes disrupt the hydrogen bonding network when in the inside in a similar manner to when the kosmotropes are on the outside. Further, we notice that strong kosmotropes perturb the water–water many-body terms differently. Calcium on the inside has repulsive 2-B and 4-B terms, while sulfate has only a repulsive 4-B term. The fact that calcium has a repulsive water–water 2-B term originates from the strong structuring of the water molecules in the first solvation shell around the ion resulting in the majority of the water dimers to be unfavorably oriented with the oxygen atoms (lone pairs) facing one another. Importantly, none of these aqueous clusters exhibit stronger water–water interactions than the pure water clusters. These results suggest that all ions weaken the water–water interactions, likely in their varying ability to orient nearby water molecules around themselves. The amount that they are impacted, however, depends on the identity of the ion. Kosmotropic ions exhibit weaker water–water interactions than chaotropic ones (see also the discussion about the anti-correlation between (I–W) and (W–W) interactions in the previous subsection).

3.3.3 Trends in the MBE terms across the Hofmeister series

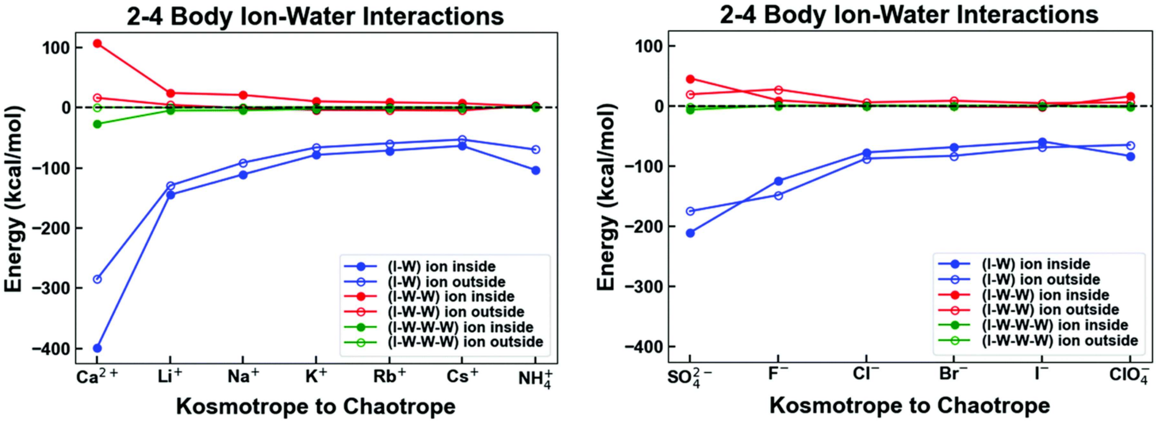


Figure 3.7: Ion–water contributions to the 2-B (blue), 3-B (red), and 4-B (green) terms of the MBE. Previously reported monatomic ions in the middle of the Hofmeister series are included.[1] The same plotting conventions as in Figure 3.5 are used. All numbers are corrected for BSSE.

By combining the results of the present study with those reported earlier for the monatomic alkali metal and halide ionic aqueous clusters,[1] we can examine the variation of the different MBE terms across the Hofmeister series. Figure 3.7 depicts the variation of the 2- to 4-B ion–water interactions for the cations (left panel) and anions (right panel) across the Hofmeister series (3.1). The corresponding plot for the water–water interaction is shown in Figure 3.8 for the cations (left panel) and anions (right panel). Both figures are drawn using the same plotting conventions as in Figure 3.5. The ion–water 2-B, 3-B and 4-B terms decrease in magnitude (the last two towards zero) when going from the kosmotropes to the chaotropes for both the cation and anion series (Figure 3.7). Specifically, the 3-B and 4-B terms decrease in magnitude (toward zero) as the center of the Hofmeister series is approached (K^{+} and Cl^{-}). However, there is no discernable difference between the ions in the center of the Hofmeister series and the chaotropes by looking at solely the ion–water contributions. All these ions have small 3-B and 4-B terms with similarly large 2-B terms.

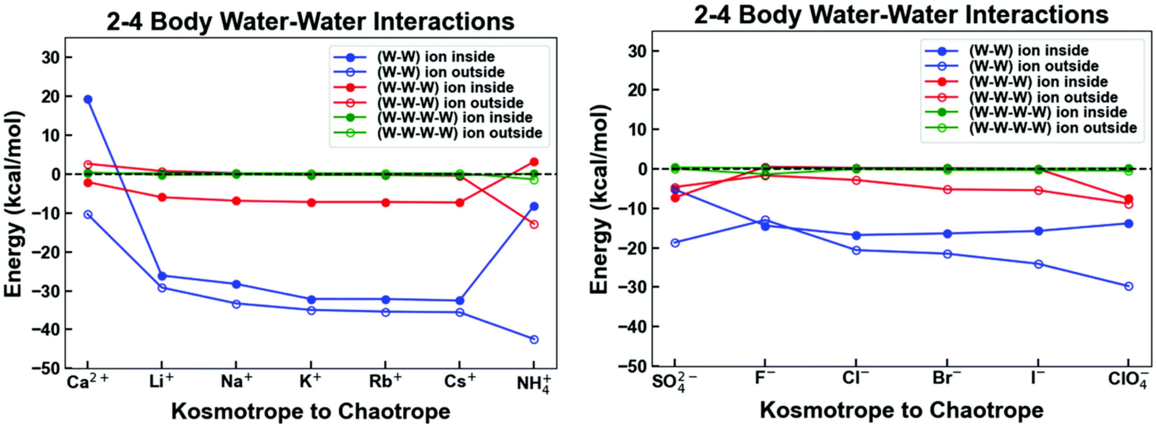


Figure 3.8: Water–water contributions to the 2-B (blue), 3-B (red), and 4-B (green) terms of the MBE. Previously reported monatomic ions in the middle of the Hofmeister series are included.[1] The same plotting conventions as in Figure 3.5 are used. All numbers are corrected for BSSE.

In contrast, the trends are not as smooth for the water–water interactions (Figure 3.8). Upon comparing the water–water interaction for ions located in the middle of the Hofmeister series in Figure 3.8, we notice different behaviors for the cations and anions. Anions, especially the halides, tend to have weaker water–water interactions than the corresponding cations (Li^+, Na^+, K^+). Additionally, each monatomic ion has a favorable (attractive) 3-B term for configurations having the anion on the outside and the cation in the inside. However, this pattern does not extend to the polyatomic ions considered in the present study. Further, we see increased favorability of the water–water interactions from the kosmotropes to the chaotropes. While that trend exists in the anions, the difference between the ions is relatively small. The Hofmeister cations calcium and ammonium generally exhibit the most extreme behavior.

Figure 3.9 and 3.10 show correlation plots between (I–W), (W–W), (I–W–W), and (W–W–W) interactions for the aqueous clusters in this study and include the results for the monatomic aqueous clusters reported earlier.[1] While the alkali metal and halide ion–water clusters exhibit clear trends (indicated by the solid lines) for each set of plots, the only Hofmeister ion that fits with these established trends is calcium. Since calcium is also a monatomic ion like the alkali metal and halide ions, it is likely that the difference in hydrogen

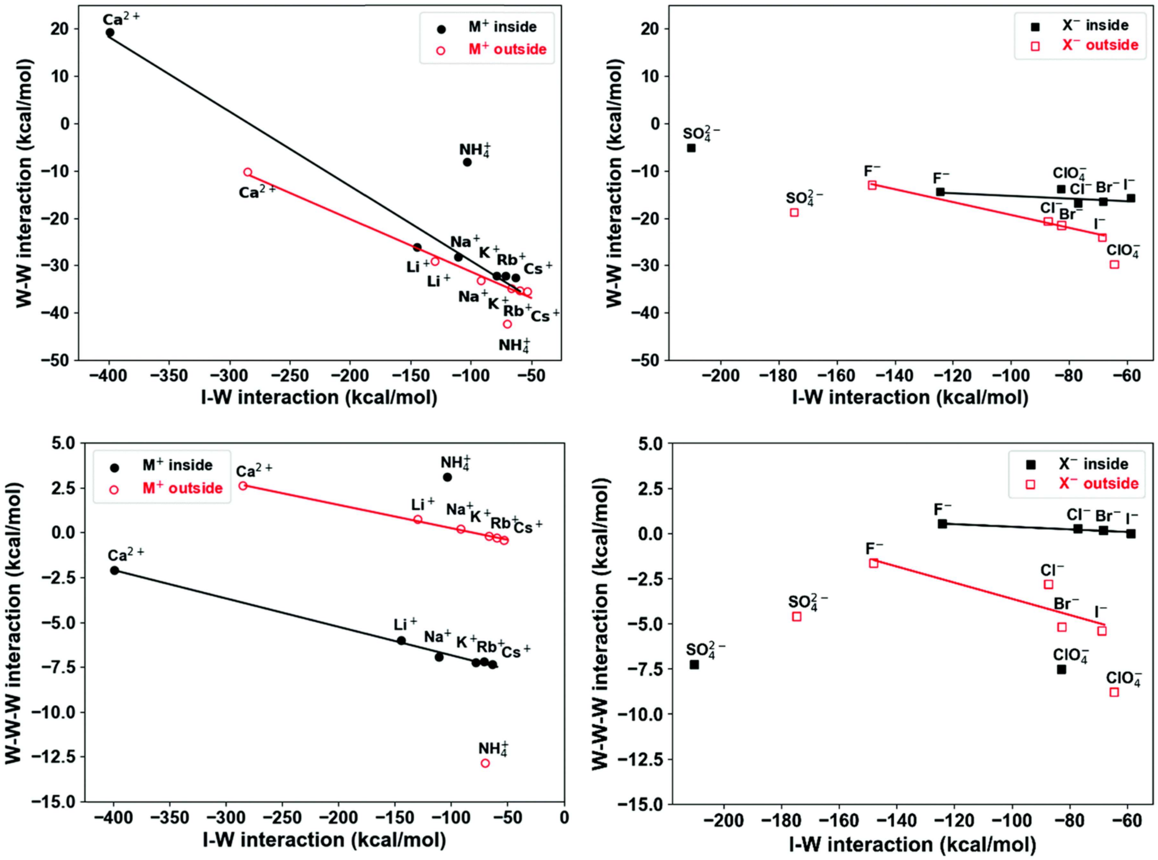


Figure 3.9: Correlations between various interactions for $Z^{+/-}(H_2O)_9$, where Z = NH_4^+, SO_4^{2-}, ClO_4^-, F^-, Cl^-, Br^-, I^- (right panel): (W–W) vs. (I–W) (top panels) and (W–W–W) vs. (I–W) (bottom panels).[1] Linear trends are shown for monatomic ions.

bonding character between monatomic and polyatomic ions complicates these correlations. Despite the polyatomic and monatomic ions not fitting on the same linear trend, we still see a consistent general anticorrelation between the (I–W–W) vs. (I–W) interactions. The ion systems with the strongest (I–W) interactions have the most repulsive (I–W–W) terms. This appears to be a quintessential characteristic of these aqueous clusters, independent of ion identity. Interestingly, ammonium often exhibits the opposite trend as the rest of the cations, viz. the point representing the interactions for the ion on the outside more closely fits the trend for the other ions when on the inside and vice versa. This is true for all correlation plots with the exception of (W–W) vs. (I–W). A further study including more

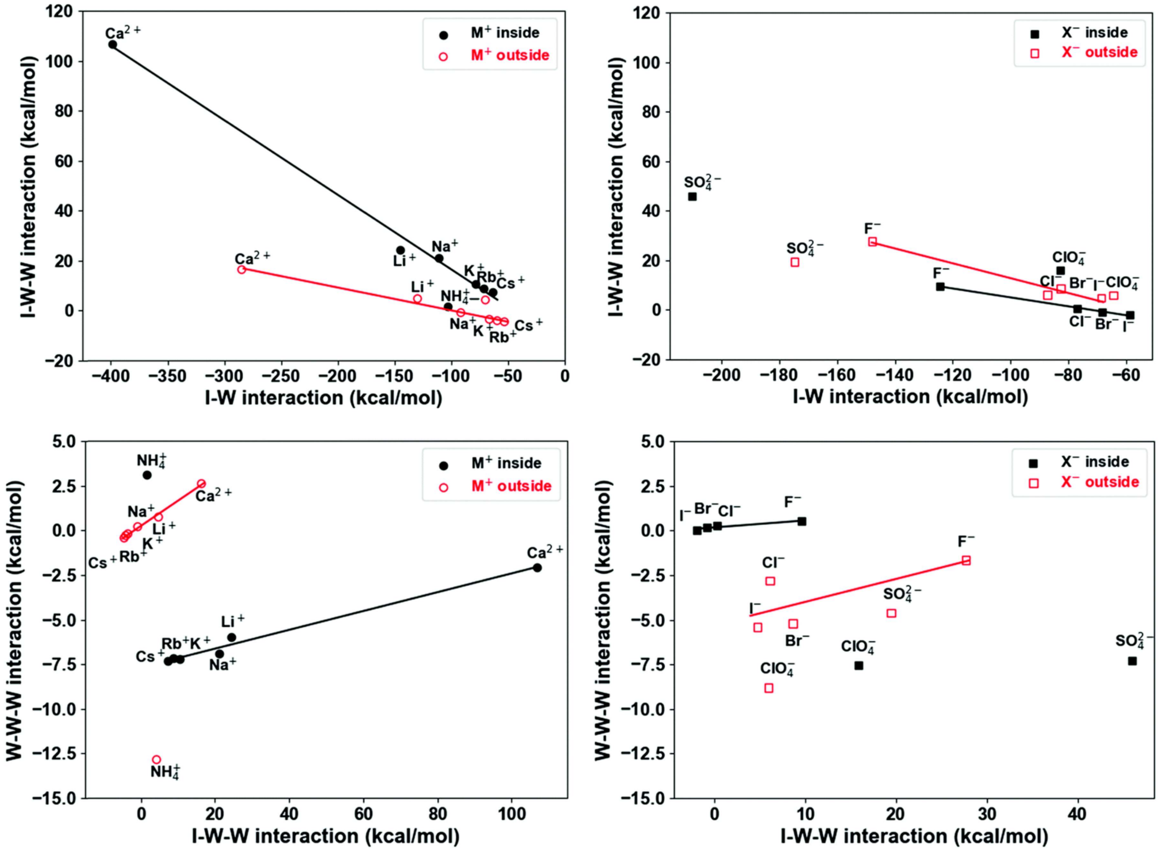


Figure 3.10: Correlations between various interactions for $Z^{+/-}(H_2O)_9$, where Z = NH_4^+, SO_4^{2-}, ClO_4^-, F^-, Cl^-, Br^-, I^- (right panel): (I–W–W) vs. (I–W) (top panels) and (W–W–W) vs. (I–W–W) (bottom panels).[1] Linear fits are shown for monatomic ions.

ions in the Hofmeister series is warranted.

3.3.4 Profile of 2-body BSSE corrections

As mentioned earlier, BSSE corrections are important for a more accurate description of the terms in the MBE. Heindel and Xantheas[1,2] have previously reported that the 2-B contribution to the total BSSE correction is the most substantial and have suggested an analytic formula for its estimate based on the distance between the fragments via a fit to an error function:

$$\Delta E_{2B-BSSE} = a[1 + erf(-b \cdot R_{ij})] \tag{3.8}$$

where a and b are empirical parameters and R_{ij} is the distance between the fragments (oxygen–oxygen or ion–oxygen distance).

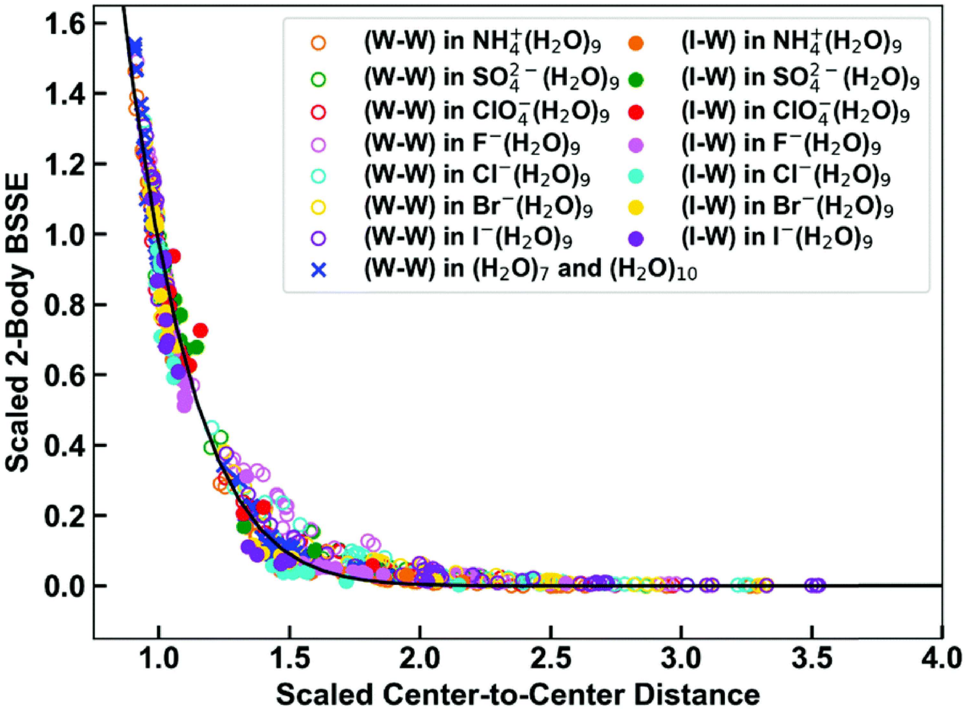


Figure 3.11: Scaled 2-B (W–W) and (I–W) BSSE corrections for $Z^{+/-}(H_2O)_9$, where Z = NH_4^+, SO_4^{2-}, ClO_4^-, F^-, Cl^-, Br^-, I^-, and scaled 2-B (W–W) BSSE corrections for $(H_2O)_7$ and $(H_2O)_{10}$ fit to eqn (3.8) ($a = 14.11$, $b = 1.29$, $R^2 = 0.9840$).[1,2]

Figure 3.11 shows the 2-B contribution to the total BSSE correction profile versus the distance between fragments (taken as either the ion–oxygen or the oxygen–oxygen distance) for the ions considered in this study (except calcium), the halide ions,[1] and pure water clusters.[2] Note that this is for 570 water dimer and 126 ion–water pairs (for the aqueous clusters of the F^-, Cl^-, Br^-, I^-, SO_4^{2-}, ClO_4^- and NH_4^-+ ions), which have been evaluated in the full cluster basis to account for the 2-B contribution to the total BSSE. As in our earlier study,[1] the x- and y-axes are scaled[177–179] by the values of the gas phase dimer (ion–water or water–water) equilibrium distance and BSSE correction of the binding energy, respectively.

The scaled 2-B BSSE corrections for the various aqueous systems, except calcium, follow a decaying trend with the scaled intermolecular distances that fits quite well to eqn (3.8) (a = 14.61, b = 1.30, R^2 = 0.9856), as displayed by the solid line in Figure 3.11. Calcium was excluded from Figure 3.11 because it exhibits a 2-B BSSE profile that decays to zero more rapidly than the other ions as the distance between pairs of molecules increases, behaving similarly to $Li^+(H_2O)_9$.[1]

3.4 Conclusions

The effect of the Hofmeister ions on the structure of the solvent is an active topic in physical chemistry research. In an effort to provide a quantitative understanding of how ions in the Hofmeister series affect the interaction between the surrounding water molecules we have performed a MBE analysis using aqueous ionic clusters as models. We have examined aqueous clusters of both anionic and cationic kosmotropes (Ca^{2+} and SO_4^{2-}) and chaotropes (NH_4^+ and ClO_4^-) with 9 water molecules to compare with previous water, alkali metal and halide aqueous clusters of the same size. The alkali metal and halide ion–water MBE analyses reported previously have been useful data for comparison, given that those ions are located in the middle of the Hofmeister series. In agreement with our previous results for these other systems we have found that the MBE converges to practically the 3-B term, with terms above 4-B being negligible, contributing <0.3% to the cluster binding energy. As observed in monatomic aqueous systems,[1] the 2-B term in the expansion is accentuated relative to that of pure water, even more so with kosmotropes. Similarly, the 3-B term is repulsive, in contrast to pure water clusters, with the exception of chaotropes residing on the outside of the water cluster. The expansion of the kosmotropic systems, in general, is largely dominated by the ion's contributions to the many-body terms with the (W–W) terms lying relatively close to zero. In contrast, the chaotropic systems, exhibiting much weaker (I–W) interactions, had more significant contributions from water–water interactions but not surpassing those found in a pure water cluster.[2] This prominent anticorrelation between (I–W) and (W–W) interactions suggests that kosmotropes, which exhibit stronger (I–W) interactions, sacrifice the favorability of the water–water interactions, likely by preferentially orienting the water molecules around the ion to maximize the ion–water interactions. Further, systems with stronger ion–water interactions tend to have larger higher-order terms in the many-body expansion. This aligns with what has been previously reported for the $Li^+(H_2O)_9$ system.[1] The expansion of chaotropic systems, exhibiting relatively weak ion–water interactions, converges more quickly and more closely resembles that of water.

Interestingly, the expansion of kosmotropic systems on the outside looks very similar to that of chaotropic systems on the inside of the cluster. Since (I–W) interactions are smaller when the ion is on the outside of the cluster, this demonstrates that kosmotropes and chaotropes interact in a similar way with water. However, due to their difference in (I–W) interactions, kosmotropes have an increased ability to diminish water–water interactions, which is evident when the ion is centrally located in the cluster. When the chaotropes reside on the outside of the cluster, the expansion more closely resembled that of pure water. Importantly, there is no evidence suggesting that water–water interactions are enhanced in the presence of either kosmotropic or chaotropic ions. Rather, the many-body expansion of increasingly chaotropic ions approaches the behavior of water, as the ion presents a weaker influence on the surrounding water molecules. This is true for both the water–water contributions and the overall many-body terms. These results suggest that all ions disrupt and weaken water–water hydrogen bonding in the short-range. These observations are consistent with other experimental and theoretical studies supporting a relatively local disruption or weakening of hydrogen bonds.[147,173,176,180,181]

Lastly, we found that the 2-B contribution to the total BSSE correction profile with intermolecular separation follows the trend ($R^2 = 0.9856$) reported previously for water, alkali metal and halide aqueous clusters.[1,2] However, as previously observed for some other alkali metal ion–water systems, namely $Li^+(H_2O)_9$, the 2-B BSSE profile for calcium decays more rapidly as the distance between the ion and the water molecule increases. While some of the monatomic cations have a different profile, ammonium aligns quite well with the anion–water and pure water clusters.

The many-body expansion for aqueous ionic clusters is more complex than that for the pure water clusters and converges at different ranks of the expansion depending on the identity of the ion and its position within a water network. This is consistent with the fact that the development of ab initio based, many-body polarizable ion–water classical potentials[6, 182] has not yet attained the accuracy of the pure water potentials[183] developed with the same fitting protocol. The static cluster configurations used as models in this study naturally fail to capture any dynamical effects that kosmotropes and chaotropes may impart upon the surrounding hydrogen bonding network. Nevertheless, they provide the stepping-stone for future studies that will examine the effect of geometric conformations and temperature effects on the magnitude of the respective MBE terms for those systems. A more detailed and quantitative analysis of the structural patterns induced by the disparate MB interactions of these different ions is warranted as it can provide a useful perspective into the elusive Hofmeister effect.

Chapter 4

A CLASSICAL MODEL FOR 3-BODY INTERACTIONS IN AQUEOUS IONIC SYSTEMS

Reproduced in part with permission from [Kristina M. Herman, Anthony J. Stone, Sotiris S. Xantheas. A classical model for three-body interactions in aqueous ionic systems; *J. Chem. Phys.* 157, 024101.]. Copyright [2022] AIP Publishing

4.1 Introduction

The strong nonadditive interactions in aqueous systems[94,160,184–188] present a challenge in the development of accurate potential energy surfaces compared to systems exhibiting largely pairwise interactions. The many-body expansion (MBE) for aqueous systems[1–3,89,92,94,160,189–201] has manifested itself a valuable guide in fitting *ab initio*-based classical potentials to accurately reproduce the many-body terms of these systems.[30,31,46,47,49,202–204]

The MBE partitions the total energy of a N-body system as a sum of its constituent fragments or "bodies", which refer to the ions and water molecules,

$$E_N = \sum_{p=1}^{N} E_N^{(p)} \tag{4.1}$$

$$E_N^{(1)} = \sum_{i=1}^{N} E_i - E_{ref} \tag{4.2}$$

$$E_N^{(2)} = \sum_{i<j}^{N} E_{ij} - E_i - E_j \tag{4.3}$$

$$E_N^{(3)} = \sum_{i<j<k}^{N} E_{ijk} - E_{ij} - E_{ik} - E_{jk} + E_i + E_j + E_k \tag{4.4}$$

...

where E_{ref} is the energy of the isolated (non-interacting) fragment.

While these explicit many-body potentials are typically quite accurate, some of them usually require tens of thousands of high-level *ab initio* calculations to fit against. Further, this process must be repeated for each interaction type within a new system of interest. For certain terms in the expansion (e.g. the 2-body term), this procedure may be necessary for the desired accuracy. However, for higher order terms in the expansion (which are smaller in magnitude), it may be worthwhile to consider alternative classical representations of these interactions. Exploring a classical representation of the higher order terms may lead to a more efficient framework that is transferable across different ionic aqueous systems.

The 2-body term is the largest contributor to the total binding energy of aqueous systems and is strongly attractive.[1–3, 189, 190] While arguably the most important component of the binding energy, the 2-body term is notably challenging to fit due to the sensitivity of its accuracy to the basis set superposition error (BSSE), the basis set size, and the level of electron correlation.[2] For example, to arrive at an accurate 2-body potential for water, tens of thousands of CCSD(T) 2-body energies have been used to fit the term using permutationally invariant polynomials (PIPs).[205] More specifically, the 2-body term of the WHBB potential was fit to 30,000 CCSD(T)/aVTZ values[49] whereas the 2-body term of MB-pol was fit to 40,000 CCSD(T)/Complete Basis Set (CBS) values.[46] The RMSEs for these 2-body fitted potentials were determined to be 0.15 kcal/mol and 0.05 kcal/mol, respectively, when tested on 42,394 CCSD(T)/CBS dimer energies.[206]

The 3-body term for aqueous systems is the second largest contributor (only behind the 2-body term) to the binding energy, comprising between 2.9-35.4% of the binding energy for $X^{+/-}(H_2O)_9$[1, 3] and $(H_2O)_{10}$[2] systems. In pure water systems, the sum of the 3-body interactions is energetically favorable (attractive). The 3-body term of the WHBB potential was fit to 30,000 MP2/aVTZ calculations[49] whereas the 3-body term of the MB-pol potential was fit to 12,000 CCSD(T)/aVTZ calculations.[47] In ion-water systems, the 3-body interactions are often found to be repulsive.[1, 3] While the 2-body terms are sensitive to several factors mentioned earlier, the 3-body energy has been shown to contain only a negligible amount of electron correlation energy, as the majority of the electron correlation energy is present in the 2-body term;[1–3] in addition, the 3-body term is not sensitive to the basis set and/or BSSE.[1–3] That said, it may not be necessary to fit the 3-body term to numerous 3-body energies obtained from high level electronic structure calculations. Rather, it may be desirable to represent the 3-body interaction classically. It has been established that the vast majority of nonadditive interactions in water is due to polarization.[199] However, the accuracy of a classical representation of the 3-body term has not been explored thoroughly for different monatomic and polyatomic ion systems.

The goal of this paper is to examine the accuracy of a classical 3-body induction model in reproducing individual *ab initio* 3-body terms for a variety of monatomic and polyatomic ion systems. We report the performance of this model on numerous monatomic (cations: Ca^{2+}, Li^+, Na^+, K^+, Rb^+, Cs^+, anions: F^-, Cl^-, Br^-, I^-) and polyatomic ions (cation: NH_4^+, anions: SO_4^{2-}, ClO_4^-) using induction models both without ion polarization and including ion polarization. By examining a variety of ion systems, we test the transferability of this model to differing ion strengths, polarizabilities, and hydrogen bonding arrangements. This work aims to provide an understanding of the underlying physics that governs the dominant nonadditive interactions in aqueous ionic systems and test its performance compared to *ab initio* results for these systems. Such an effort will inform future development of accurate and efficient many-body potentials for aqueous ionic systems.

4.2 Theoretical and computational details

4.2.1 The classical 3-body induction model

The pair interaction energy or 2-body energy correction U_{ij} of molecules i and j can be defined as their energy E_{ij} in the geometry in which they occur in the cluster, but in the absence of all other molecules, less the energies of the isolated molecules i and j. Similarly the 3-body correction for molecules i, j and k can be defined as the energy of that triple in the absence of all other molecules, less the energies of the isolated molecules and the 2-body corrections U_{ij}, U_{ik} and U_{jk}. This view is equivalent to the definitions in eqns. (4.1–4.4).

The principal 3-body contributions to the energy of an aqueous ionic hydrogen bonded network arise from the induction energy, that is from the polarization of each molecule by the electric field of its neighbours. The contribution from three-body dispersion is much smaller.[207] In the hydrogen bonded network of water, arrangements are favoured in which the dipole moment induced in each molecule by the electric field of its neighbours enhances its static dipole moment, which in turn enhances its electric field. This effect favours

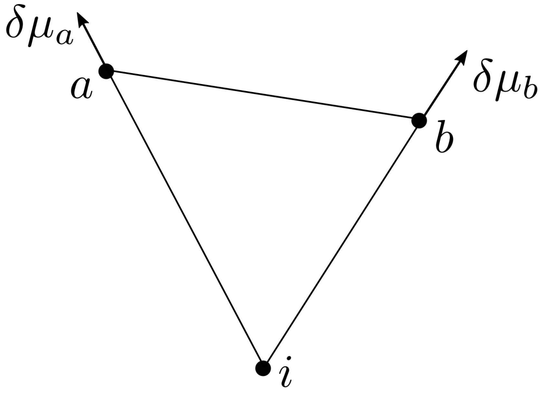


Figure 4.1: Induced moments on two water molecules a and b in the field of an ion i.

nose-to-tail cyclic arrangements of the hydrogen bonds in small water clusters and leads to a cooperative many-body effect on the induction energy. In the presence of an ion, however, this cooperativity is disrupted. To see how this arises, consider the arrangement in Figure 4.1.

Here we have two water molecules a and b and an ion i. The electric field of the ion induces dipoles $\delta\mu_a$ and $\delta\mu_b$, additional to the static dipole moments, not shown. There is a change to the induction energy of molecule a, but this is part of the U_{ia} pair energy, and likewise there is a change to the U_{ib} pair energy. However there is a repulsion between the induced dipoles $\delta\mu_a$ and $\delta\mu_b$, and this is a 3-body effect. An estimate using reasonable values for the parameters suggests that the energy is positive and of the order of a few kJ mol^{-1} if the molecules and ion are close to contact. Moreover the field of the ion will tend to orient the static dipole moments of the neighbouring water molecules in the direction away from or towards the ion (depending on the ion's sign) and this will tend to disrupt the cooperative network of hydrogen bonds, reducing the magnitude of the negative 3-body energies. This is a greatly over-simplified picture, but it provides a useful qualitative idea of the nature of the ion–water–water (I-W-W) 3-body effect. For the calculations described below, a much more detailed and accurate description was used, involving distributed multipoles and distributed polarizabilities.[207] In this treatment, the change ΔQ_t^a to multipole moment t at site a of molecule A depends on the electrostatic fields at sites a' on A due to the polarized moments

u on all sites b of the other molecules B:

$$\Delta Q_t^a = \sum_{B \neq A} \alpha_{tt'}^{aa'} T_{tu}^{a'b} (Q_u^b + \Delta Q_u^b), \tag{4.5}$$

where $\alpha_{tt'}^{aa'}$ is the polarizability relating the induced moment t at site a to the electrostatic field component t' at site a', and $T_{tu}^{a'b}$ is the general interaction function describing the electrostatic energy between multipole moments $Q_t^{a'}$ and Q_u^b. Summation over repeated suffixes is implied. The distributed polarizabilities are calculated by the CamCASP program.[208] In practice we use localized polarizabilities, which are non-zero unless $a' = a$; these are derived from the general distributed polarizabilities using the Orient program.[209] The set of coupled equations in eq. (4.5) is solved iteratively to obtain the ΔQ_t^a for all molecules A, and then the induction energy is

$$E_{\text{ind}} = \frac{1}{2} \sum_A \sum_{B \neq A} \Delta Q_t^a T_{tu}^{ab} Q_u^b, \tag{4.6}$$

which again is calculated by the Orient program. This calculation is carried out for all the dimer and trimer subsystems in the cluster to obtain the 3-body energy, and can be used to obtain 4-body and higher terms if required. The calculation is very fast – under a second to obtain the 3-body energy for any of the clusters considered here.

4.2.2 Details of the calculations

The optimized ion-water cluster geometries were obtained from previous MP2 calculations.[1,3] In those works, MP2 calculations were also carried out to evaluate the 3-body energies, both for the set of water-water-water (W-W-W) trimer subsystems and the set of ion-water-water (I-W-W) trimers. These calculations were corrected for basis set superposition error (BSSE) using the counterpoise method of Boys and Bernardi.[170] For all (W-W-W) and most (I-W-W) calculations, Dunning's aug-cc-pVDZ basis set[16,210] was used. The Stuttgart RSC 1997 ECP[169,211] was used for calculations containing Ca^{2+}, Rb^+, and Cs^+ while the aug-cc-pVDZ-PP basis set[212,213] was used for all I^- calculations. All MP2 calculations were performed with the NWChem 7.0.2 electronic structure suite.[214]

The induction energy contributions to the 3-body energy can be explored accurately and in detail using the Orient program.[209] The cluster geometries for Orient were derived from the MP2 optimized geometries by replacing each H_2O molecule by an H_2O with the O in the same position and the H atoms in the same plane but with the bond lengths and bond angle as in the equilibrium isolated molecule.

The detailed classical description of the induction energy uses electrostatic distributed multipoles up to hexadecapole and distributed polarizabilities up to quadrupole–quadrupole on the O and H atoms of water, taken from the ASPW4 intermolecular potential[215] for the water dimer. The monatomic ions were described by a point charge, and by dipole–dipole polarizabilities taken from Li *et al.*,[216] which were evaluated experimentally for ions in aqueous solution. For the polyatomic ions, distributed multipoles up to hexadecapole and distributed polarizabilities up to quadrupole–quadrupole were calculated using CamCASP[208] with the aug-cc-pVTZ basis set.[16,210] The induction interactions were damped at short range by the procedure used for the ASPW4 water dimer potential. The effect of varying the damping parameter away from the value of 1.41 used in the ASPW4 potential has been explored.

Second order Moller–Plesset (MP2) perturbation theory calculations were also used to evaluate the 3-body terms for these geometries, using the basis sets listed above, in order to obtain more direct comparisons between the two approaches.

4.2.3 Dataset of (I-W-W) and (W-W-W) trimers

Ion-water-water (I-W-W) trimer subsystems containing F^-, Cl^-, Br^-, I^-, Li^+, Na^+, K^+, Rb^+, Cs^+, Ca^{2+}, ClO_4^-, SO_4^{2-}, or NH_4^+, and water-water-water (W-W-W) trimer subsystems were obtained from ion–$(H_2O)_9$ clusters from previous works.[1–3] Two clusters for each ion system were considered, one cluster containing the ion inside the cluster and one containing the ion on the outside of the cluster. In the original study,[1,3] the 3-body terms for each of these subsystems (72 (I-W-W) and 168 (W-W-W) for each ion system) were computed at the cluster optimized geometries at the MP2 level of theory. This amounts to a total of 936 (I-W-W) and 2,184 (W-W-W) trimers for all systems combined. To compare the proposed induction model with the *ab initio* benchmarks on an equal footing, MP2 results were evaluated at the frozen monomer geometry (i.e., when the 1–B = 0) for each subsystem, as used in Orient. As an additional test of the (W-W-W) classical model, we have utilized the dataset of 12,347 CCSD(T)/aVTZ 3-B energies used to train MB-pol.[47] Because the classical induction model considered in this work is implemented as a rigid model, we have focused on the 1,744 trimers (out of the 12,347 trimers) for which the monomers have R_{OH} values within 0.04 Å of equilibrium (0.96 Å) and θ_{HOH} values within 1.5° of the equilibrium (104.5°) for a more fair comparison.

4.3 Results and discussion

4.3.1 Establishing accurate ab initio reference values for the dataset of (I-W-W) and (W-W-W) trimers

The aqueous clusters that the trimer subsystems were taken from are shown in Figure 4.2. The purpose of sampling two different cluster configurations for each ion system is to include a wider range of (I-W-W) trimer arrangements, representing surface-like and bulk-like configurations. Each ionic cluster contains 36 (I-W-W) and 84 (W-W-W) trimers. It has been previously established that the total 3-body term in aqueous systems is weakly dependent on the basis set size.[2] Figure 4.3 shows a comparison of the MP2/aVDZ results for (Li^+-W-W) against those with two larger basis set sizes, namely aug-cc-pVTZ and aug-cc-pVQZ, as well as for (Ca^{2+}-W-W) against those with the aug-cc-pVTZ basis set. The difference between the results obtained with different basis set sizes is small, with Root-Mean-Square-Errors (RMSEs) not exceeding 0.06 kcal/mol for the 3-body terms in the $Li^+(H_2O)_9$ or $Ca^{2+}(H_2O)_9$ systems. This verifies that the individual 3-body terms are also insensitive to basis set size, a result that further validates the use of MP2/aVDZ results for the other ion systems.

Furthermore, the basis set superposition error (BSSE) has been found to affect mainly the 2-body interaction.[2] Correlation plots between the uncorrected and BSSE-corrected MP2/aVDZ results are shown in Figure 4.4. The correlation plots for each of the individual ion systems are included in the Supporting Information (SI). For the individual halide-water systems, the RMSEs do not exceed 0.04 kcal/mol (0.029 kcal/mol for the halides altogether). Similarly, for the individual polyatomic systems, the RMSEs do not exceed 0.04 kcal/mol

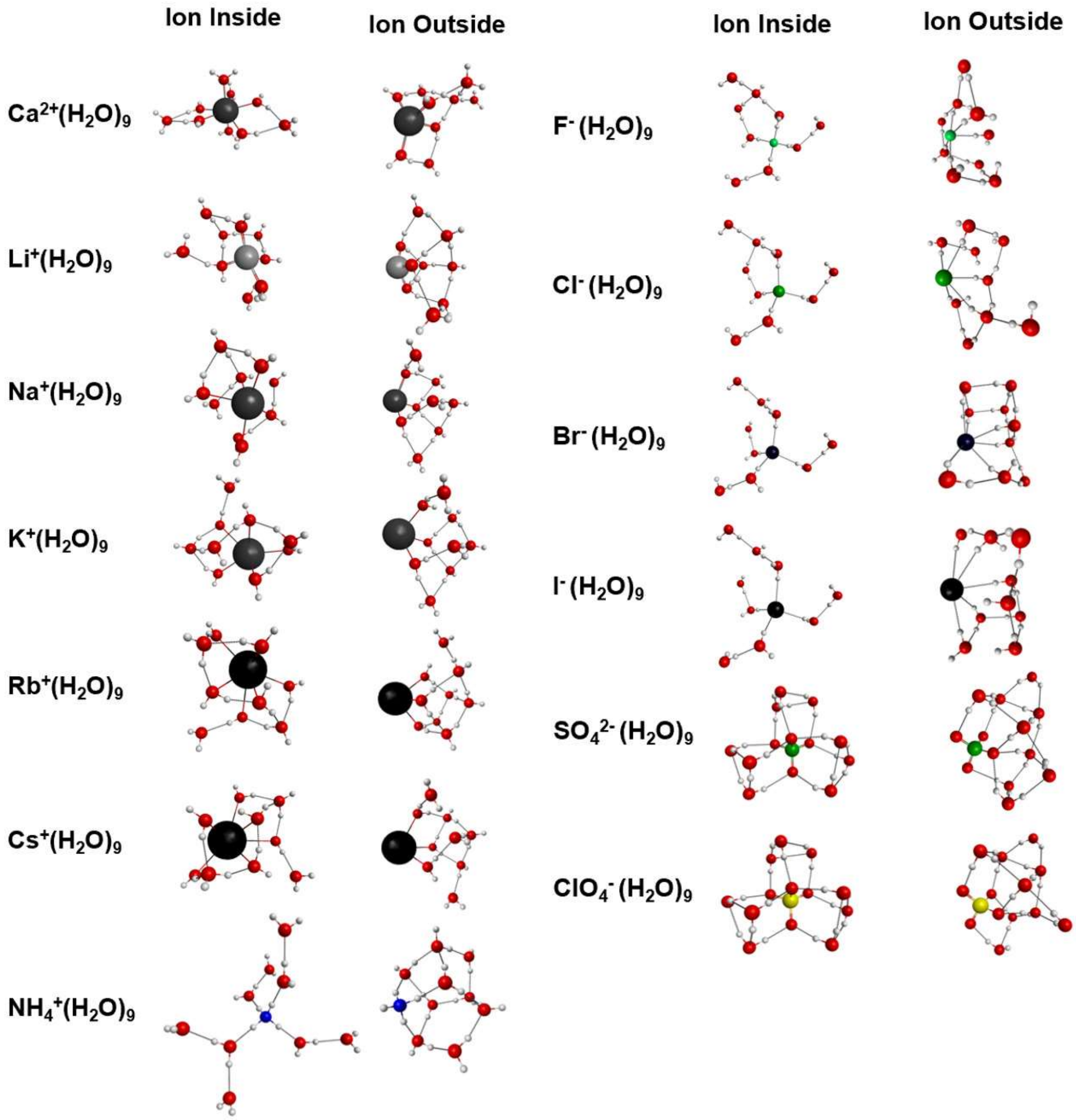


Figure 4.2: The cation (left column) and anion (right column) water cluster geometries that the trimers were taken from. For each ion system, two different cluster geometries were considered: one with the ion centrally located and the other with the ion on the edge of the cluster.[1,3]

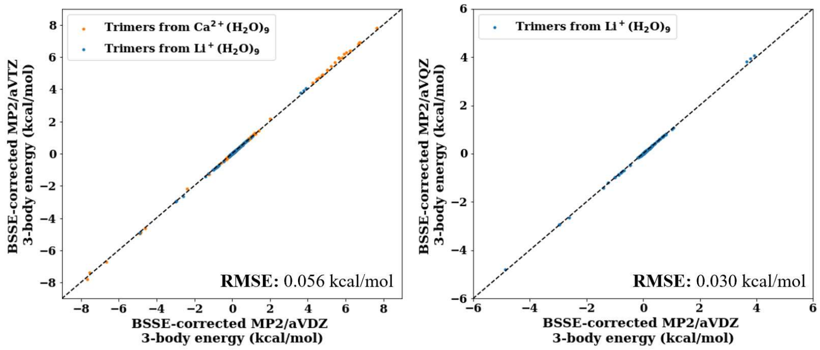


Figure 4.3: A comparison of different basis set sizes (aug-cc-pVDZ vs. aug-cc-pVTZ and aug-cc-pVQZ) on the individual 3-body terms. These 3-body terms were computed at the cluster optimized geometry with the lithium ion (blue) and calcium ion (orange).[1,3]

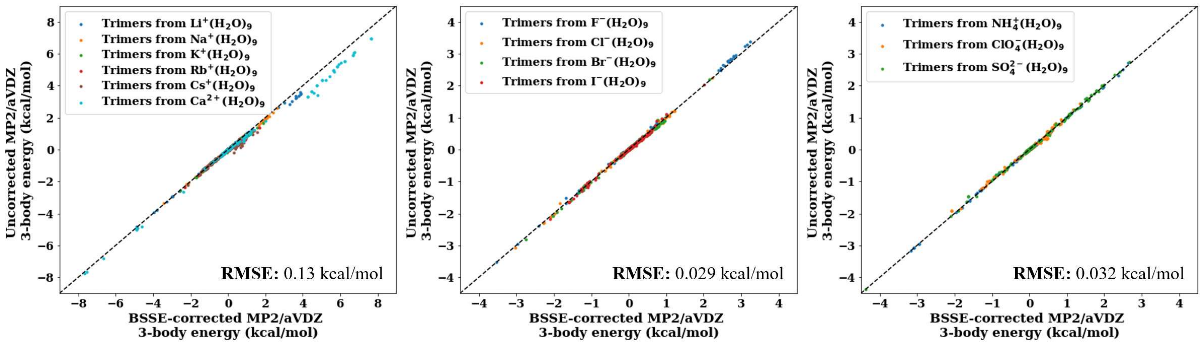


Figure 4.4: A comparison of the uncorrected and BSSE-corrected 3-body energies computed at the MP2/aVDZ level of theory at the cluster optimized geometries.[1,3]

(0.032 kcal/mol for the polyatomic ions altogether). The RMSEs for the different cation systems range from 0.04-0.28 kcal/mol (0.13 kcal/mol for the cations altogether). This shows that the individual 3-body terms are generally insensitive to the BSSE but this is system-dependent. The largest deviation tends to appear in the strongly repulsive 3-body terms for the trimers of Li^+ and Ca^{2+}. For this reason, the uncorrected MP2/aVDZ values are used for all system calculations except Li^+, Ca^{2+}, and Cs^+, for which the results were corrected for BSSE. Altogether, these results show that the 3-body terms in these systems are insensitive to basis set size and BSSE, thus validating the use of uncorrected MP2/aVDZ 3-body calculations for most (except some cation) systems.

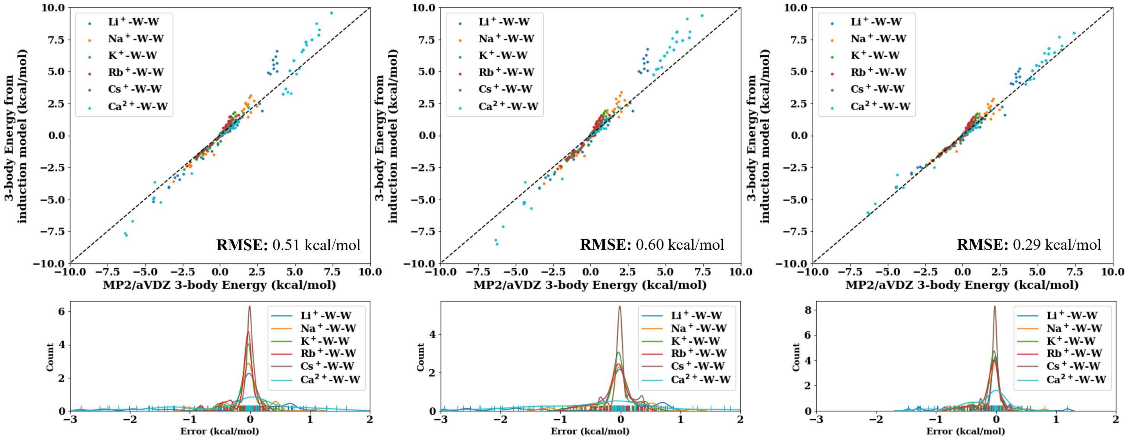
4.3.2 Classical induction model for 3-body (I-W-W) interactions

Now that the computational protocol used to compute the reference 3-body interactions has been validated, let us turn our attention to the performance of the classical induction model in reproducing these interactions. For each set of ions (monatomic cations, monatomic anions, and polyatomic ions), we will discuss the accuracy of the classical induction model in representing these interactions and specifically the role of ion polarizabilities in these interactions. Further, we will provide a comparison of this model to other many-body ion-water potentials in the literature, when available. Note that we will focus on the MP2 results at the frozen (relaxed) intramolecular geometry to compare to the induction model on an equal footing.

Monatomic cations

Figure 4.5 shows the accuracy of the classical induction model without ion polarization (left panel), with ion polarization (middle panel), and with ion polarization with increased damping (right panel). The distributions of the errors ($E_{\mathrm{MP2}} - E_{\mathrm{model}}$) are shown below the respective correlation plots. The corresponding RMSEs and MAEs for the individual ion systems are organized in Table 4.1. As expected, there is not a significant qualitative difference between the models with and without ion polarization because the cations are not very polarizable. However, by adjusting the damping parameter ($1.41 \rightarrow 1.15$) the induction model aligns more closely with the MP2 reference values (RMSE: 0.29 kcal/mol). The largest errors lie with the (Li$^+$-W-W) and (Ca^{2+}-W-W) trimers, for which the RMSE is 0.41 kcal/mol for both systems. This is mainly due to deviations in the repulsive region (positive values) of the correlation plots.

Table 4.2 organizes a comparison of the total 3-body (I-W-W) energies using the proposed induction model, CCSD(T)(-F12b), i-TTM,[6] and AMOEBA 2009[217] for small $X^+(H_2O)_n$ clusters, as published previously.[6] Note that for the $n = 2$, $n = 3$, and $n = 4$ results, there are 1, 3, and 6 individual (I-W-W) calculations, respectively, that are added together. The i-TTM and AMOEBA 2009 models consider electrostatic, induction, repulsion, and dispersion interactions. The proposed model in this work considers only the induction energy and agrees quite well with the *ab initio* benchmarks.



Without ion polarization
Damping parameter: **1.41**

With ion polarization
Damping parameter: **1.41**

With ion polarization
Damping parameter: **1.15**

Figure 4.5: A comparison of the MP2/aVDZ (I-W-W) energies (BSSE-corrected for the Li$^+$, Cs$^+$, Ca^{2+} systems) against the induction model without ion polarization (left, RMSE: 0.51 kcal/mol), including ion polarization (middle, RMSE: 0.60 kcal/mol), and including ion polarization with increased damping (right, RMSE: 0.29 kcal/mol).

		Without ion pol.		**With ion pol. (DF=1.41)**		**With ion pol. (DF=1.15)**	
Trimer	Number of points	RMSE	MAE	RMSE	MAE	RMSE	MAE
Li^+-W-W	72	0.76	0.40	0.80	0.42	0.41	0.22
Na^+-W-W	72	0.34	0.21	0.42	0.27	0.26	0.16
K^+-W-W	72	0.24	0.15	0.30	0.20	0.23	0.14
Rb^+-W-W	72	0.22	0.13	0.28	0.19	0.22	0.14
Cs^+-W-W	72	0.13	0.080	0.17	0.11	0.11	0.069
Ca^{2+}-W-W	72	0.88	0.61	1.06	0.75	0.41	0.26
Combined	432	0.51	0.26	0.60	0.32	0.29	0.16

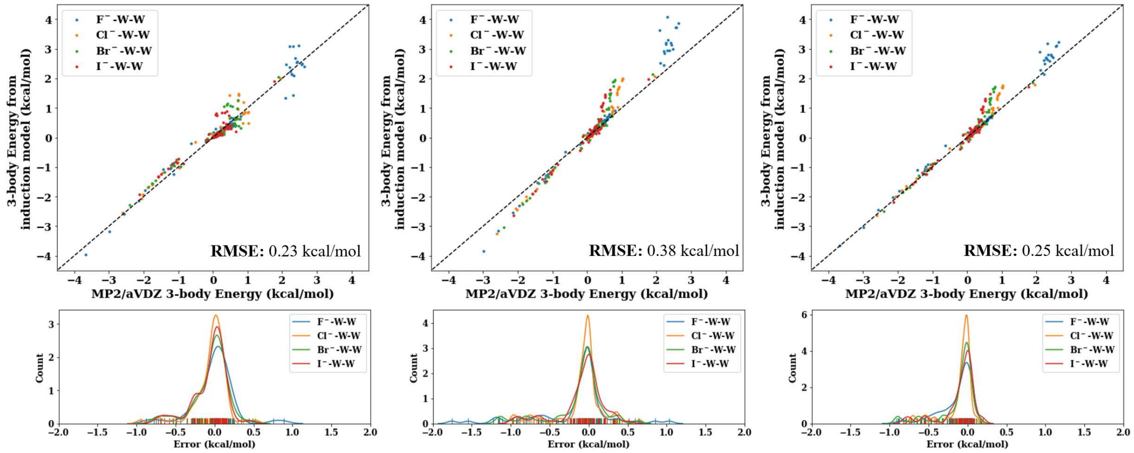
Table 4.1: The performance (RMSE and MAE, kcal/mol) of the classical induction model on the (I-W-W) 3-body interactions for the monatomic cations.

Monatomic anions

Figure 4.6 shows the accuracy of the induction model neglecting ion polarization (left panel) and including ion polarization (middle panel). Without including ion polarization, we see good agreement in the attractive region for the monatomic anions. When ion polarization is included, the induction model overestimates these interactions slightly. In the repulsive region of the plot, we see that the model without ion polarization causes more scatter in the correlation plot. However, when ion polarization is included, we see that the deviation becomes less scattered and we can see a clear trend based on the relative ionic radii. By decreasing the damping parameter from 1.4 to 1.2 (Figure 4.6, right panel), thus increasing the damping, we see an improvement in the repulsive region of the correlation plot and an improvement in the repulsive region. The RMSE decreased from 0.38 kcal/mol to 0.25 kcal/mol upon increasing the damping (Table 4.3). However, the classical induction model is still overestimating the repulsive interactions to varying degrees. Interestingly, the classical induction model performs best for the monatomic cations and anions with a decreased damping parameter of approximately 1.2 (increased damping). For the monatomic anions and cations, we found it necessary to also change the damping parameter to improve the agreement whereas the polyatomic systems performed well without any additional changes. In Figures 5 and 6, the middle panel shows the correlation trend between the model including ion polarization (with the original damping parameter) and the ab initio references. In these

		A	B	C	D	E	F
n	symmetry	**Li^+-W-W contribution to 3-body energy**					
2	D_{2d}	3.36	4.53	4.53	**3.72**	6.10	3.62
3	D_3	11.34	16.27	16.01	**13.12**	20.81	12.72
4	S_4	23.90	32.61	32.09	**26.44**	40.18	25.30
		Na^+-W-W contribution to 3-body energy					
2	D_{2d}	1.64	1.82	2.60	**2.16**	2.53	1.62
3	D_3	5.46	6.77	7.96	**6.85**	8.41	5.37
4	C_2	3.14	3.96	5.01	**4.81**	5.77	2.65
		K^+-W-W contribution to 3-body energy					
2	D_{2d}	1.38	0.80	1.69	**1.45**	1.47	1.26
3	C_2	-1.35	-1.13	-1.45	**-0.79**	-1.12	-1.64
4	C_2	1.44	0.89	1.97	**2.21**	2.01	0.94
		Rb^+-W-W contribution to 3-body energy					
2	D_{2d}	1.14	0.55	1.54	**1.35**	1.17	1.12
3	C_2	-1.63	-1.27	-1.78	**-1.13**	-1.44	-1.84
4	C_2	0.74	0.18	1.35	**1.64**	1.07	-0.50
		Cs^+-W-W contribution to 3-body energy					
2	C_s	0.76	0.81	0.48	**0.55**	0.81	0.73
3	C_2	-1.91	-1.32	-2.08	**-1.45**	-1.72	-2.07
4	C_2	0.56	-0.26	0.94	**1.27**	0.55	0.21

Table 4.2: A comparison of the classical induction model presented in this work, both without ion polarizability (B) and with it (damping factors 1.41 (C) and 1.15 (D)) to the 3-body (I-W-W) interactions in $X^+(H_2O)_n$ obtained with CCSD(T)(-F12b) (A), i-TTM (E), and AMOEBA 2009 (F), as previously published.[6] The bold values highlight the classical induction model that performed most accurately when compared to the MP2 (I-W-W) 3-body energies in this work.



Without ion polarization
Damping parameter: **1.41**

With ion polarization
Damping parameter: **1.41**

With ion polarization
Damping parameter: **1.20**

Figure 4.6: A comparison of the MP2/aVDZ (I-W-W) energies against the induction model without ion polarization (left, RMSE: 0.23 kcal/mol), including ion polarization (middle, RMSE: 0.38 kcal/mol), and including ion polarization with increased damping (left, 0.25 kcal/mol).

cases, the error is larger than the leftmost panel (without ion polarization). However, this is due to what appears to be a change in slope in the correlation plot. This is especially noticeable in the attractive region which is why the rightmost panel shows the correlation after both including ion polarization and tweaking the damping parameter. Regardless, the RMSDs for all systems improve from the leftmost panel (no ion polarization) to the rightmost panel (including ion polarization and tweaking damping parameter if necessary) except for the monatomic anions in which we see nearly the same RMSD. The attractive region is improved upon for the monatomic anions by including ion polarization, but the repulsive region, while forming straighter lines, deviate more from the line. This implies that there may be an additional physical interaction that could be added to the induction model to improve the agreement in the repulsive region for the monatomic anions, specifically.

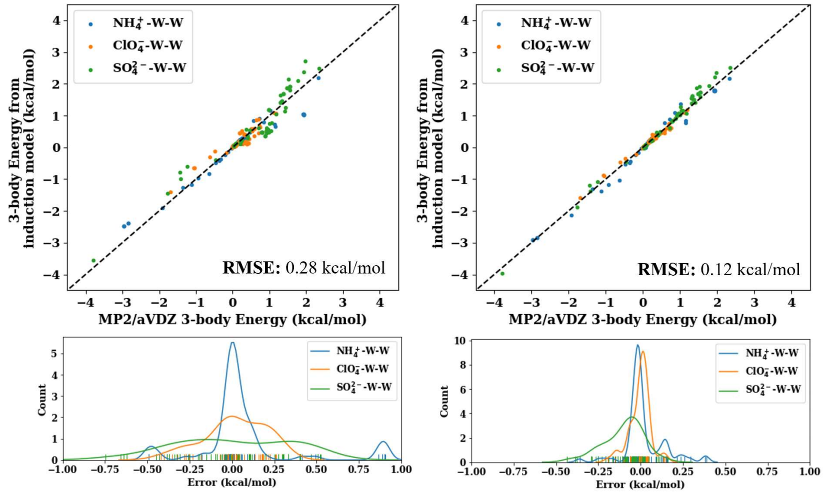
Polyatomic ions

Figure 4.7 shows the accuracy of the induction model neglecting ion polarization (left panel) and including ion polarization (right panel). While we already see good agreement using the

		No ion pol.		**Ion pol. (DF=1.41)**		**Ion pol. (DF=1.20)**	
System	Number of points	RMSE	MAE	RMSE	MAE	RMSE	MAE
F$^-$-W-W	72	0.26	0.17	0.50	0.29	0.24	0.14
Cl$^-$-W-W	72	0.22	0.14	0.31	0.18	0.22	0.12
Br$^-$-W-W	72	0.23	0.15	0.36	0.21	0.28	0.15
I$^-$-W-W	72	0.22	0.14	0.30	0.19	0.24	0.14
Combined	288	0.23	0.15	0.38	0.22	0.25	0.14

Table 4.3: The performance (RMSE and MAE, kcal/mol) of the induction models on the (I-W-W) 3-body interactions for the monatomic anions.

induction model without ion polarization, including ion polarization improves the agreement further. While other ion systems benefited from tweaking the damping, these systems did not require any further adjustments to the model. Table 4.4 outlines the RMSEs and MAEs for the individual ion systems with and without ion polarization included in the model. Further, by examining the error distribution, we see that the maximum error in the 3-body term does not exceed 0.5 kcal/mol. This demonstrates the promise of a classical induction model in accurately representing 3-body interactions for other aqueous polyatomic ionic systems.



Without ion polarization
Damping parameter: **1.41**

With ion polarization
Damping parameter: **1.41**

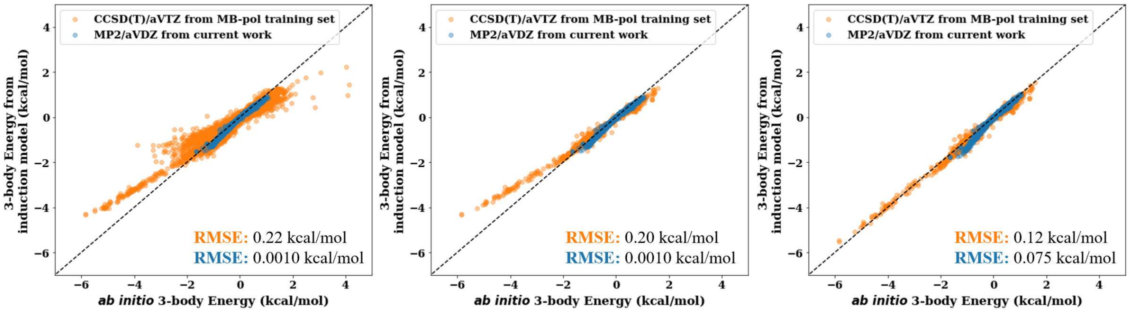
Figure 4.7: A comparison of the MP2/aVDZ (I-W-W) energies against the induction model without ion polarization (left) and including ion polarization (right).

		No ion pol.		**Ion pol. (DF=1.41)**	
System	Number of points	RMSE	MAE	RMSE	MAE
NH_4^+-W-W	72	0.30	0.16	0.12	0.074
ClO_4^--W-W	72	0.18	0.14	0.061	0.042
SO_4^{2-}-W-W	72	0.33	0.29	0.15	0.12
Combined	216	0.28	0.20	0.12	0.078

Table 4.4: The performance (RMSE and MAE, kcal/mol) of the induction models on the (I-W-W) 3-body interactions for the polyatomic ions.

4.3.3 Classical induction model for 3-body (W-W-W) interactions

The induction model was tested on two different datasets of 3-body interactions. The set of 2,184 MP2/aVDZ 3-body terms from the ionic and water clusters and the set of 12,347 CCSD(T)/aVTZ 3-body terms that were used to train the MB-pol potential.[47] Note that the MP2 values calculated in this work are at the frozen monomer geometries, for which



Damping parameter: **1.41**
Full CCSD(T)/aVTZ dataset

Damping parameter: **1.41**
Subset of CCSD(T)/aVTZ dataset

Damping parameter: **1.60**
Subset of CCSD(T)/aVTZ dataset

Figure 4.8: Comparison of the induction model against the MP2/aVDZ (W-W-W) reference energies from this work (blue) and a subset of CCSD(T)/aVTZ (W-W-W) energies from the MB-pol training set (orange).

the 1-B term is zero. However, this is not the case for the values in the MB-pol training set so the comparison of the induction model (rigid monomers) to the full dataset may be less direct. The accuracy of the induction model in computing these 3-body interactions is shown in Figure 4.8 (left panel). Because the MB-pol training set consists of a wide range of monomer geometries, we have also filtered out all geometries for which $R_{\mathrm{OH}} < 0.92\,\text{Å}$, $R_{\mathrm{OH}} > 1.00\,\text{Å}$, $\theta_{\mathrm{HOH}} < 103^\circ$, or $\theta_{\mathrm{HOH}} > 106^\circ$ to provide a fair comparison to the induction model (rigid). After filtering out more distorted molecules, there are 1,744 trimers remaining for the comparison (Figure 4.8, middle panel).

		Ind. model (DF=1.41)		**Ind. model (DF=1.60)**	
Dataset	Number of points	RMSE	MAE	RMSE	MAE
(W-W-W) from clusters	2184	0.0011	0.024	0.075	0.033
Mb-pol train set^{a} - Full	12347	0.22	0.12	0.19	0.11
MB-pol train set^{a,b} - Subset	1744	0.20	0.10	0.12	0.074

a. CCSD(T)/aVTZ 3-body energies used to train MB-pol, Babin et al.[47]

b. Only trimers for which $R_{\mathrm{OH}} < 0.04$ Å and $\theta_{\mathrm{HOH}} < 1.5^\circ$ from equilibrium.

Table 4.5: The performance (RMSE and MAE, kcal/mol) of the induction model on the MP2/aVDZ and CCSD(T)/aVTZ datasets for the (W-W-W) interactions.

By increasing the damping parameter (decreasing the damping), the fit of the strongly attractive 3-body terms improves greatly (Figure 4.8, right panel). The resulting RMSE is 0.12 kcal/mol. This accuracy is on par with other explicit many-body potentials fitted to *ab initio* 3-body terms, such as HBB2-pol and WHBB. However, because the proposed model is rigid, we are currently limited in our ability to describe 3-body terms for which the monomers are significantly distorted from equilibrium monomer geometry. This will be the subject of a future study.

4.4 Conclusions

In this work the physical origin of the 3-body interaction in aqueous ionic systems was explored using a classical induction model. By modeling the 3-body term as an interaction between induced multipoles for various (I-W-W) (13 different ion systems) and (W-W-W) subsystems, we find excellent agreement with the established MP2 and CCSD(T) 3-body reference values. More specifically, the RMSEs for monatomic cations, monatomic anions, and polyatomic ions were 0.29 kcal/mol, 0.25 kcal/mol, and 0.12 kcal/mol, respectively, when compared to a total of 936 MP2/aVDZ reference 3-body terms. Further, the classical induction model returned a RMSE of 0.12 kcal/mol on 1,744 CCSD(T)/aVTZ 3-body (W-W-W) energies used to train MB-pol. Note that this was the subset of the data for which all monomers had R_{OH} values within 0.04 angstrom of equilibrium and θ_{HOH} values within 1.5 degrees of equilibrium to compare the performance of the induction model (rigid, at equilibrium geometry) on an equal footing. This is on par with the performance of other explicit many-body potentials fit to *ab initio* data. The success of this model demonstrates that the 3-body term for aqueous systems can be accurately modeled classically, without fitting to tens of thousands of high-level *ab initio* calculations. More importantly, this work provides a fast, accurate and efficient method to model 3-body effects that is transferable across systems with different ions.

In general, we find that including ion polarization improves the description of the 3-body term for all ion systems considered. For the monatomic anions and cations, we found it necessary to also change the damping parameter to improve the agreement whereas the polyatomic systems performed well without any additional changes. The only adjustable parameter that was varied in this study was the damping factor. It was found that slightly decreasing the damping led to the best agreement for the monatomic ions systems while slightly increasing the damping led to the best agreement for the pure water systems. The induction model performed well on the polyatomic ion systems with the initial damping factor of 1.41. However, this damping factor was selected based on the performance on 72 (I-W-W) trimers cut from two different hydrogen bonding networks. Further tuning could be performed on a larger set of trimers, if desired. Nonetheless, the small amount of parameterization needed makes this an attractive option for other ion-water systems of interest. The induction damping could benefit from further investigation. The damping parameter is a reciprocal measure of the distance at which the damping becomes significant and can be expected to depend on the nature of the interacting molecules. However, treating it as an adjustable parameter produces satisfactory results.

This model could be further improved upon to allow for flexible monomers (and the associated change in multipoles with different intramolecular geometries). Because the intramolecular geometry of the water molecules affects the resulting interactions, this adjustment would need to be included to accurately predict the 3-body energies of systems with highly distorted intramolecular geometries. This induction model could be extended further to include 4-body and higher terms through a similar approach. This type of approach would allow all many-body effects to be described classically through induction, greatly simplifying the process of fitting accurate many-body potentials for new systems.

Chapter 5

ACCURATE CALCULATION OF MANY-BODY ENERGIES IN WATER CLUSTERS USING A CLASSICAL GEOMETRY-DEPENDENT INDUCTION MODEL

Reproduced in part with permission from [Kristina M. Herman, Anthony J. Stone, Sotiris S. Xantheas. Accurate Calculation of Many-Body Energies in Water Clusters Using a Classical Geometry-Dependent Induction Model; *J. Chem. Theory Comput.* 19 (19), 6805–6815].

5.1 Introduction

The notoriously strong many-body interactions in aqueous systems[1–3,89,92–94] necessitate the use of complex and sophisticated force fields in order to reproduce their magnitude and variation with hydrogen bonding environment including those in the presence of solutes such as ions.[6,30,31,33,38,39,41,42,46,47,49,50,202,203,218–224] These developed force fields are considered classical (rather than quantum mechanical) potential energy surfaces because the methods evaluate energies and forces without calculating a wave function, escaping the poor scaling experienced by *ab initio* methods. Initially, most of the classical interaction potentials developed for water were pairwise additive,[225–231] truncating the many-body expansion at the second order into an effective term that was designed to fold in the missing higher order terms. This was mainly achieved through the use of an enhanced static dipole moment that was larger than the gas phase monomer value (1.84 Debye).[225–231] It was soon realized that an environment-dependent dipole moment, rather than an enhanced static one, was essential for ensuring the transferability of the potentials from small to larger water clusters (n=2-25).[29] Furthermore, many-body effects have been shown to be critical in describing the structure of water[232] and, in particular, its local structure near an ion.[233] In order to capture the many-body nature of the interactions, a model must either be fit to high level *ab initio* data to yield the 3-body (and higher) energy terms explicitly or it must be polarizable to account for the change in the charge distribution due to the field from other neighboring molecules, thus implicitly incorporating many-body effects to infinite order into the model. In this paper, we incorporate geometry-dependent distributed multipole and polarizability surfaces into an induction model for the 3- and 4-body terms of water and demonstrate that this physically motivated model describes these many-body terms with an accuracy that is on par with models fit to a large database of 43,844 3-body terms and 3,603 4-body terms obtained from high level *ab initio* calculations.

A popular method of accurately representing many-body effects in classical potentials is through permutationally invariant polynomials (PIPs)[234] that are fit to a large number (typically in the tens of thousands) of high level *ab initio* calculations. This approach was pioneered by Bowman and co-workers in the WHBB[49] and q-AQUA[50] potentials, and has also been used extensively by Paesani and co-workers in the fitting of the MB-pol potential.[46–48, 235] These classical potentials were developed by correcting the deficiency in describing shorter-range interactions by fitting either a short-range correction or a standalone potential to extensive *ab initio* data. More specifically, the WHBB potential was fit to roughly 40,000 MP2/aug-cc-pVTZ calculations to add a 3-body short-range correction to the TTM3-F potential,[31] while the MB-pol potential was fit to >12,000 BSSE-corrected CCSD(T)/aug-cc-pVTZ calculations to obtain a 3-body short-range correction to the TTM4-F potential.[33] Recently, the q-AQUA potential was fit to >45,000 Basis Set Superposition Error (BSSE)-corrected CCSD(T)-F12a/aug-cc-pVTZ calculations to develop a standalone 3-body potential. The q-AQUA potential was also the first developed classical potential with an explicitly fitted 4-body term, which was accomplished using over 3,000 CCSD(T)/aug-cc-pVTZ 4-body energies derived from high level *ab initio* data. Inarguably, this approach has led to the construction of very accurate potential energy surfaces.[29, 54, 236] However, there are a few notable disadvantages of this strategy. These consist of the high level of parameterization (hundreds to thousands of parameters) and the large number of very accurate (and very costly) *ab initio* calculations required to populate the database that is used to fit the 3- and 4-body terms in these potentials. As one extends to higher orders of the MBE (i.e., the 3- and 4-body terms), the *ab initio* calculations for each geometry become more expensive due to the increase in system size N and the number of associated geometries that are needed to ensure sufficient sampling of the conformational space also increases due to the surging of the degrees of freedom. The need for high accuracy also commands the use of electronic structure methods like CCSD(T) that scale as N^7. Most importantly, incorporation of solutes requires that the process be repeated for each many-body term containing the new species. That said, if a classical, transferable model could be used to model many-body effects (3-body and higher), it would leave only the 1- and 2-body terms to be either fit to *ab initio* results (i.e. using PIPs) or calculated on-the-fly using high level *ab initio* methods.

Polarizable models, on the other hand, implicitly incorporate many-body effects. These potentials are typically founded on the multipole expansion, which expands the charge distribution of an atom or molecule in spherical harmonics. This expansion can yield multipoles at a single site or it can be distributed across multiple sites (often selected to be the atom sites) to improve the performance for shorter-range interactions.[237–239] The interactions between multipoles are strictly pairwise. However, allowing these multipoles to change or polarize in response to the field of neighboring molecules gives rise to many-body effects. A multipolar description of the molecular charge distribution of the equilibrium geometry of water has already been implemented in force fields to varying degrees.[30,33,38,39,41,43,202,203,218–223,240] However, it is expected that the distributed multipoles and polarizabilities do change as a function of the intramolecular geometry. The necessity of representing this charge flux in potentials has been previously discussed by Dinur.[241] The TTM2.1-F revision of the TTM2-F potential,[30] the AMOEBA+CF[42] potential, and the flexible MDCM potential[242] incorporate, to some extent, the redistribution of charge density with intramolecular geometry by including a geometry-dependent correction to the atomic monopoles, as has been done for geometry-dependent point charges for studies of photodissociated CO in myoglobin.[243,244] Very recently, the SCME-f model[245] introduced a single-site multipolar description of water which has a dipole and a quadrupole moment, both dependent on the intramolecular geometry. In addition, the FFLUX model recently used Gaussian process regression (GPR)[246] to predict geometry-dependent distributed atomic multipoles up to hexadecapole for water. However, this method[246] predicts the multipoles in the presence of an implicit solvent, rather than computing the induced multipoles explicitly. To the best of our knowledge, the impact of distributed multipole and distributed polarizability surfaces on the many-body interactions of a molecular system has not been examined. However, it should be noted that dipole moment surfaces and dipole and quadrupole polarizability surfaces[247] have been obtained in order to compute the infrared and Raman spectra of water, respectively. Notably, these surfaces, like SCME-f, yield a single molecular dipole moment or dipole-dipole polarizabilities rather than distributed multipoles and polarizabilities at each atom site. By distributing the multipoles and polarizabilities across atom sites, a better representation of the charge redistribution can be achieved.[237–239]

The success of a classical representation of the 3-body term with a detailed induction model, albeit based on moments and polarizabilities at the equilibrium geometry, has recently been demonstrated, including its transferability across different water-water-water (W-W-W) and ion-water-water (I-W-W) subsystems with only a single adjustable parameter.[44] While it had been anticipated that the induction energy would be the main contributor to many-body effects, this previous study demonstrated the quantitative accuracy of the induction model in doing so against accurate *ab initio* benchmarks, hinting at the potential of this model to rival the accuracy of fitted 3-body terms using tens of thousands of PIPs. Furthermore, the minimal parameterization (a single damping parameter) and the naturally transferable nature of this route makes it particularly attractive. However, the previous model was limited to a multipolar and polarization description that was static in nature (only computed at the equilibrium monomer geometry) and did not include the effect of 3-body dispersion. The description of distorted monomer geometries is crucial for modeling liquid water. For instance, during a molecular dynamics simulation the monomer geometries sampled can be far from the monomer gas phase equilibrium geometry. Consequently the model must capture the variation of the distributed multipoles and polarizabilities with these highly distorted intramolecular geometries.

The goal of this work is to expand our previous induction model, which described the interactions between water molecules by implementing geometry-dependent distributed multipoles (dm) and polarizabilities (dp), to incorporate a 3-body dispersion term and to extend it to the 4-body term. We have obtained distributed multipole and distributed polarizability surfaces as functions of the monomer intramolecular coordinates θ_{HOH}, R_{ave} (average of the two R_{OH} distances), and ΔR (difference between the two R_{OH} distances) (Figure 5.1) over a wide range of intramolecular geometries. The linearly interpolated surfaces are then used to re-examine the 3-body terms from our previous paper[44] and demonstrate the improvement of the flexible water molecules in the 3-body description of (W-W-W) 3-body terms. The contribution from the triple-dipole-dispersion interaction to the total 3-body interaction is also examined using the Axilrod-Teller-Muto potential.[248, 249] We also extend the developed classical induction model to the 4-body term in water. The results of this work demonstrate the practicality of the classical induction approach to accurately model the 3- and 4-body terms in water. More broadly, it shows that an induction model is capable of achieving an accuracy that is on a par with potentials explicitly fitted to tens of thousands of *ab initio* calculations with PIPs (containing over ten thousand parameters). Most importantly, it also addresses the issue of transferability upon including additional, different solvent molecules without the need to perform tens of thousands of high level electronic structure calculations for each system studied.

5.2 Computational details

5.2.1 a. Description of the classical model

We consider two contributions to the 3-body energy, arising from induction and dispersion, respectively. As for the 4-body energy, the only contribution considered in this study is from the induction energy, which is is computed as previously described.[44] The induced moment t at site a on molecule A due to the multipole u on site b of molecule B is evaluated self-consistently using the following expression:

$$\Delta Q_t^a = -\sum_{B \neq A} \alpha_{tt'}^{aa'} T_{t'u}^{a'b} (Q_u^b + \Delta Q_u^b) \tag{5.1}$$

where $\alpha_{tt'}^{aa'}$ is the polarizability of the multipole and $T_{t'u}^{a'b}$ is the interaction tensor. Once the induced moments are converged, the induction energy of molecule A is calculated as:

$$E_{\mathrm{ind}}^A = \tfrac{1}{2} f_{\mathrm{TT}} \sum_{B \neq A} \Delta Q_t^a T_{tu}^{ab} Q_u^b \tag{5.2}$$

where f_{TT} is the Tang-Toennies damping function:[250]

$$f_{\mathrm{TT}}(R) = 1 - \sum_{k=0}^{6} \frac{(\beta R)^k}{k!} e^{-\beta R} \tag{5.3}$$

In our previous work we chose β (units of bohr^{-1}) to be a single adjustable parameter.[44] In this work we set β as the Born-Mayer coefficient

$$\beta = -0.31(r_1^{\mathrm{vdW}} + r_2^{\mathrm{vdW}}) + 3.43 \tag{5.4}$$

where r^{vdW} is the van der Waals radius of each of the pairs of atoms (from CCSD, O: 3.07 bohr, H: 2.63 bohr).[251] Because of this choice, the decay of the damping function with respect to the interatomic distance is slightly different for each pair of sites. The above equation yields β values of 1.66, 1.53, and 1.80 for the O$\cdots$H, O$\cdots$O, and H$\cdots$H pairs of sites, respectively. In our previous work[44] we found that a single damping parameter of β=1.60 worked quite well (RMSE: 0.12 kcal/mol) for a subset of Database A, which contained trimers with intramolecular geometries close to the equilibrium ones. That value of β used in the previous study is very close to the calculated value of 1.66 used for the O$\cdots$H in the current study. A single damping parameter worked sufficiently well previously, because most of the trimers examined in that earlier study did not have two oxygen atoms or two hydrogen atoms very close to one another. However, in applying this model to the full (more extensive) Database B (see below), we find it more appropriate to use the above 3 values of β for the different pairs of sites instead. This has the additional benefit that the model has no adjustable parameters.

The multipole expansion was truncated at the hexadecapole (H) with the respective multipoles distributed on the atom sites of H_2O. Dipole-dipole (D–D), dipole-quadrupole (D–Q), and quadrupole-quadrupole (Q–Q) distributed polarizabilities are used to represent the polarizabilities of the multipoles in the presence of an electric field. Rather than using the distributed multipoles and polarizabilitites computed at the equilibrium geometry (and placed at the equilibrium configuration), we explore geometry-dependent distributed multipole and polarizability representations of the water monomers. These distributed multipole and polarizability surfaces were computed at a 3-dimensional grid of points for H_2O. The θ_{HOH} coordinate was scanned every 5° from 60–145° (18 points), the R_{ave} coordinate was scanned every 0.05 Å from 0.81–1.21 Å (9 points), and the ΔR coordinate was scanned every 0.04 Å from 0.00–0.40 Å (11 points), see Figure 5.1 for the definition of these coordinates. The total number of grid points for which the multipoles and polarizabilities were computed amounted to 1,782. The multipoles at each point were derived using CamCasp6.0[208] from the HF/aug-cc-pVTZ[16] wavefunction calculated using Psi4.[252] Note that we also compare the results of the model using multipoles and polarizabilities derived from HF/aug-cc-pVTZ to those obtained with PBE0/aug-cc-pVTZ in Section III.d. The 3-dimensional grids were linearly interpolated to yield distributed multipole and polarizability surfaces for each of the atom sites of H_2O.

The Axilrod-Teller-Muto (ATM)[248, 249] 3-body dispersion energy was evaluated using the computed C_9 coefficient at each trimer configuration as:

$$E^{ijk}_{\text{disp}} = C^{ijk}_9 \frac{(1 + 3\cos\phi_i \cos\phi_j \cos\phi_k)}{R^3_{ij} R^3_{ik} R^3_{jk}} \tag{5.5}$$

where R_{ij}, R_{ik}, and R_{jk} are the pairwise distances between the indicated molecules and ϕ_i, ϕ_j, and ϕ_k are the angles between R_{ij} and R_{ik}, R_{ij} and R_{jk}, and R_{ik} and R_{jk} vectors, respectively.

The O-O-O C_9 triple-dipole-dispersion coefficient of 303.67 (hartree)·$(bohr)^9$ was calculated from the isotropic dipole-dipole polarizabilities at imaginary frequencies for the water molecule, using a single site at the oxygen atom. This is close to the value of 308.2 (hartree)·$(bohr)^9$ evaluated from pseudo-spectral dipole oscillator strength distributions by Margoliash *et al.*[253]

5.2.2 b. Description of the CCSD(T)-quality benchmark databases for the 3- and 4-body energies

Database A: 12,260 3-body terms. The original database comprising 12,347 trimers was developed by Babin et al. to fit the 3-body PIP in the MB-pol potential.[47] These 3-body terms were evaluated at the CCSD(T) level of theory with the aug-cc-pVTZ basis set supplemented with midbond functions and were also corrected for basis set superposition error (BSSE). We have opted to remove trimer configurations from that database for which at least one of the following conditions is true:

1. the trimer contains a hydronium like molecule (three or more O-H distances ≤ 1.2 Å),
2. there exists an O$\cdots$O distance that is ≤ 2.4 Å,
3. at least one of the intramolecular geometries is outside the range of the monomer scans used to obtain the surfaces, namely $60^\circ \leq \theta_{\mathrm{HOH}} \leq 145^\circ$, 0.81 Å $\leq R_{\mathrm{ave}} \leq$ 1.21 Å and $\Delta R \leq 0.40$ Å.

We chose to remove trimers that are hydronium-like because, in these cases, it is not straightforward to divide the system into fragments (hydronium vs. water). Moreover the distributed multipoles and polarizabilities would be different if treated as hydronium rather than a neutral water molecule. The above constraints result in the removal of only 87 trimers from the original database of 12,347.

Database B: 43,844 3-body terms. The full 3-body dataset compiled by Bowman and co-workers[50] comprises 45,332 three-body terms evaluated at the CCSD(T)-F12a/aug-cc-pVTZ level of theory and corrected for BSSE. This dataset was used to fit the recently developed q-AQUA potential[50] and also encompasses the datasets that were previously used to train the MB-pol[47] and WHBB[49] potentials, respectively. That is, database A is a subset of database B. This dataset contains very accurate energies for an extensive sampling of the configurational space with the maximum O$\cdots$O distance being 9.5 Å. Using the same criteria described above for Database A, we removed 1,488 trimers from this dataset, leaving 43,844 (out of 45,332) trimer configurations to benchmark our proposed induction model against. The majority (1,401) of the trimers removed contained at least one molecule whose intramolecular geometry was outside the range of our scans. The remaining 87 trimers contained a molecule deemed "hydronium-like" based on their O$\cdots$O or O$\cdots$H distances.

Database C: 3,603 4-body terms. Database C comprises 3,603 water tetramer structures and their corresponding CCSD(T)/aug-cc-pVTZ 4-body terms. The full dataset of 3,692 4-body terms was used to fit a 4-body PIP for the q-AQUA potential.[50] We removed 29 tetramers that have at least one molecule with an intramolecular geometry outside the range of our scan and 60 tetramers due to small O$\cdots$O or O$\cdots$H values following the same criteria used for the 3-body term.

5.2.3 c. Comparison with other many-body classical potentials

The MB-pol calculations were performed with the MBX distribution of the code.[46–48, 235, 254] The TTM2.1-F[203] calculations were performed using the open-source distribution from the Pacific Northwest National Laboratory (https://sites.uw.edu/wdbase/files/2019/01/pot_ttm−1p9hi7d.zip). While the TTM2.1-F is not fitted using PIPs, it represents a simpler classical model, only using flexible monopoles and inducible dipoles at each atom site to implicitly incorporate many-body effects.

5.3 Results and Discussion

5.3.1 a. Distributed geometry-dependent multipole and polarizability surfaces

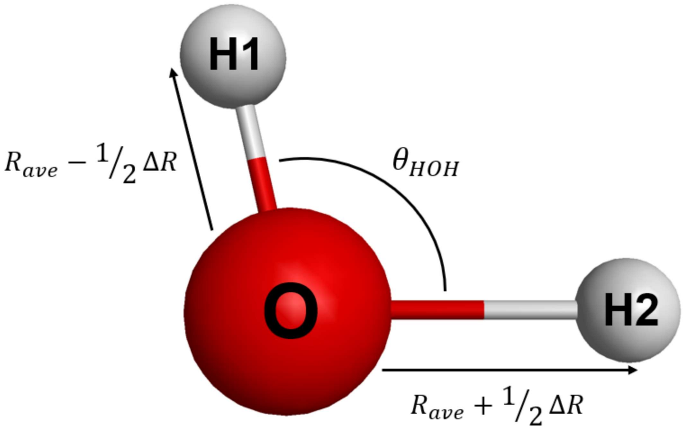


Figure 5.1: The 3 coordinates describing the intramolecular geometry of a water molecule, used as a basis for the calculation of the distributed multipole and polarizability surfaces.

The water monomer coordinate system (θ_{HOH}, R_{ave}, ΔR) used to evaluate the multipoles (up to hexadecapole) and polarizabilities (up to quadrupole-quadrupole) is shown in Figure 5.1. Details of the grid size, the process for selecting the range of values, and the level of theory used to compute the multipole and polarizability surfaces were described earlier in the computational details section. Representations of the 3-dimensional surfaces depicting the value of each multipole or polarizability as a function of the (θ_{HOH}, R_{ave}, ΔR) coordinates are included in Figures S3-S4 and are available via GitHub (https://github.com/kmherman/dm-dp-surfaces/H2O). In general, these surfaces show smooth, continuous changes in value as the intramolecular geometry changes. For the polarizability surfaces, there appears to be a more abrupt change in distributed polarizabilities for $R_{\mathrm{OH}} \geq 1.2$ Å. As a quantitative test of the linearly interpolated surfaces, 100 points were randomly selected from a uniform distribution of the 3-coordinates within the ranges of the scans. The *ab initio* distributed multipoles and polarizabilities were computed and compared to the linearly interpolated value.

The interpolated multipole and polarizability surfaces allow for the description of the charge distribution and its polarizability at a wide range of intramolecular geometries. In the following, we will focus on benchmarking the induction model with a flexible multipolar/polarizability description on the following extensive datasets of 3- and 4-body energies, while comparing its performance to the version of the model based on static (equilibrium) distributed multipoles and polarizabilities.

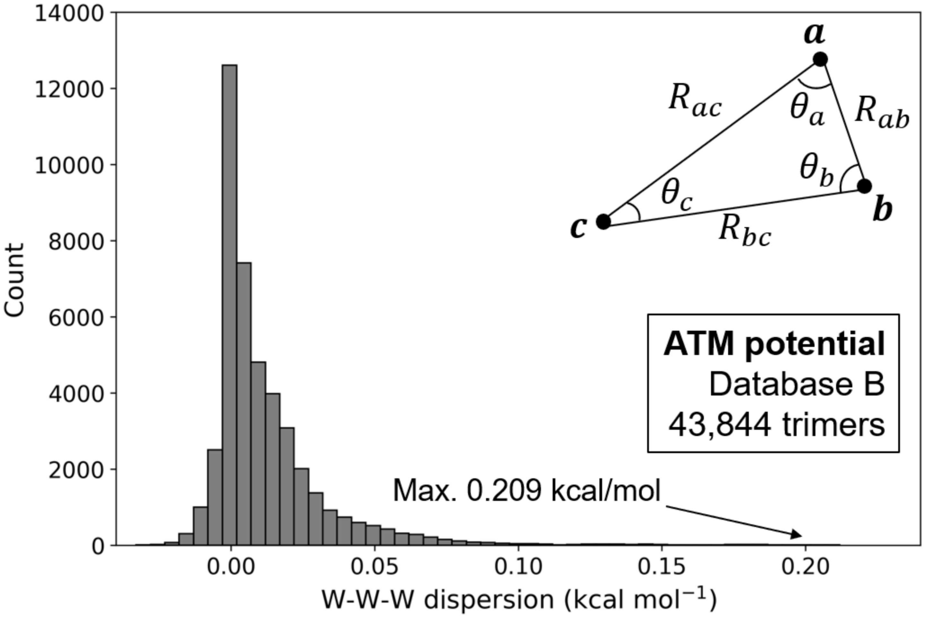


Figure 5.2: Contributions from 3-body dispersion (kcal/mol) to the total 3-body energies for the 43,844 trimers in Database B estimated with the Axilrod–Teller–Muto (ATM) potential.

5.3.2 b. Performance of the classical model for the 3-body energies

We also consider contributions from 3-body dispersion, estimated from the Axilrod-Teller-Muto (ATM) potential.[248, 249] Figure 5.2 shows the distribution of the individual 3-body dispersion energies for each of the 43,844 trimers in Database B. This demonstrates that the contribution from 3-body dispersion is typically quite small in magnitude and usually repulsive. The maximum 3-body dispersion energy, albeit for a very small number of trimers in Database B, is found to be ~0.2 kcal/mol (cf. Figure 5.2).

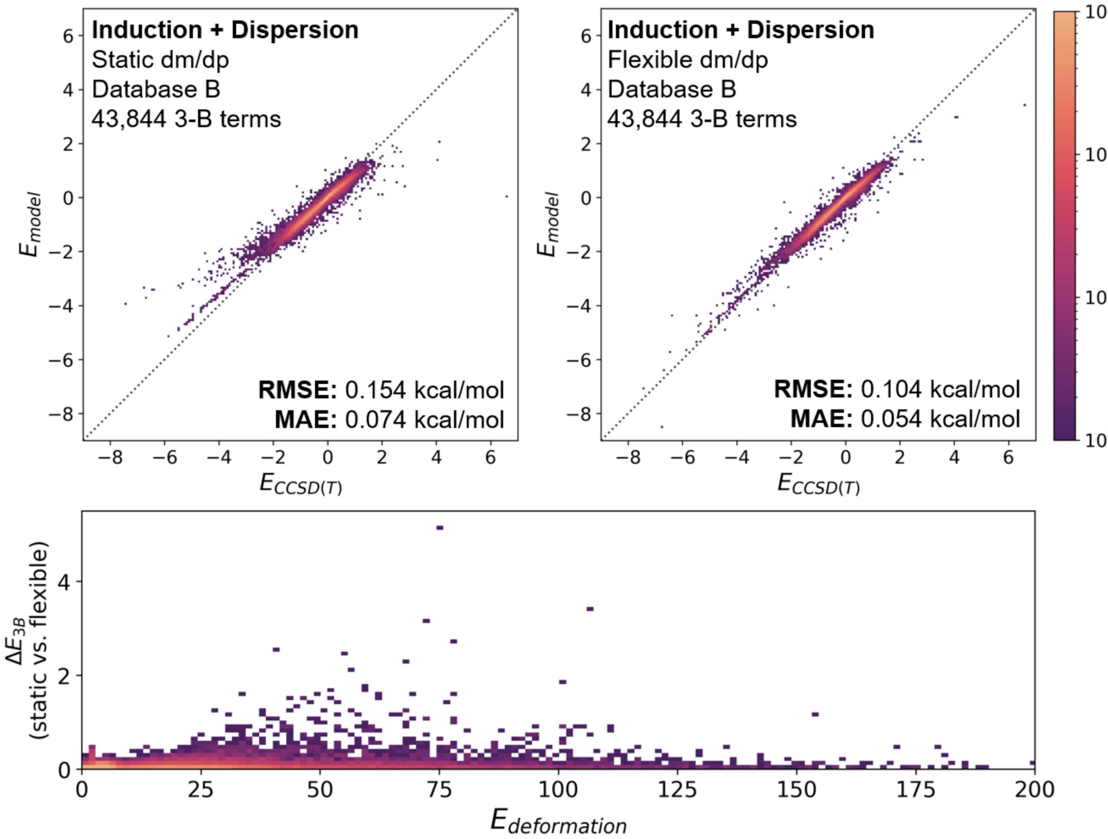


Figure 5.3: Comparison of the 3-body energies (Database B, kcal/mol) of the classical model (E_{model}) including induction and dispersion using static multipoles and polarizabilities at the equilibrium geometry (top left) and the geometry-dependent multipoles and polarizabilities (top right) vs. the *ab-initio* results. The bottom panel shows the difference between these two versions of the model as a function of the total deformation energy (kcal/mol) of the trimers.

The top two panels of Figure 5.3 show the correlation plots comparing the accuracy of the “static dm/dp” (distributed multipoles and polarizabilities at the equilibrium geometry, left panel) vs. the “flexible dm/dp” (geometry-dependent distributed multipoles and polarizabilities, right top panel) versions of the induction model on Database B. The ”static” model yields an RMSE of 0.152 kcal/mol while the ”flexible” model yields 0.104 kcal/mol (see also Table 5.1), indicating a clear improvement with the geometry-dependent multipoles and polarizabilities. Comparing the static and flexible versions of the model more directly, the bottom panel of Figure 5.3 shows the difference in the calculated energy using the equilibrium multipoles and polarizabilities (at the distorted geometries) versus the flexible multipoles and polarizabilities from the interpolated surface as a function of the total trimer deformation energy, which was computed with the Partridge-Schwenke potential energy surface.[255] As one might expect, the use of the geometry-dependent multipoles and polarizabilities is more critical for the trimers whose constituent monomers are more distorted from the gas phase equilibrium geometry. However, it is sometimes the case that both representations yield similar 3-body energies, due to a cancellation of errors. Nevertheless, the maximum difference between the computed 3-body terms is ∼5 kcal/mol, which is quite substantial given the range of 3-body energies in the database (roughly -8 to 6 kcal/mol). That said, the inclusion of geometry-dependent multipoles and polarizabilities is paramount in improving the accuracy of the induction model for the 3-body term. Additionally, while the contribution from the 3-body dispersion is very small compared to the range of induction energies (-8 to 6 kcal/mol), incorporating it in the classical model improves the agreement with the *ab initio* benchmarks. More specifically, the RMSE decreases from 0.089 to 0.084 kcal/mol on Database A and 0.106 to 0.104 on Database B (see Table 5.1). Given that the 3-body dispersion estimates from the ATM potential are very fast to evaluate since they take into account just the pairwise distances and angles between oxygen atoms, it is worthwhile to include it for quantitative accuracy.

It should be emphasized again that the classical model based on geometry dependent multipoles and polarizabilities has no adjustable parameters. All values used in the potential

		Database A 12,260 trimers	**Database B** 43,844 trimers	Number of fitted parameters	Size of training set
Induction only	Static dm/dp	0.140 (0.076)	0.152 (0.074)	0	n/a
	Flexible dm/dp	0.089 (0.050)	0.106 (0.056)	0	n/a
Induction + Dispersion	Static dm/dp	0.142 (0.077)	0.154 (0.074)	0	n/a
	Flexible dm/dp	**0.084 (0.048)**	**0.104 (0.054)**	0	n/a
Potentials fit to PIPs	HBB2-pol	0.158^a	n/a	132	~500
	WHBB5	0.104^a	n/a	1,380	~40,000
	MB-pol	0.028^b	0.116 (0.035)	1,173	12,347
	q-AQUA	n/a	0.026^c, 0.032^d	14,412	45,332

[a] Reported RMSE on entire 12,347 trimers in the MB-pol training set.[47]

[b] Reported RMSE fitting error on training set.[47]

[c] Reported RMSE fitting error on short-range portion of the q-AQUA training set.[50]

[d] Reported RMSE fitting error on long-range portion of the q-AQUA training set.[50]

Table 5.1: The root mean square error (RMSE) and mean absolute error (MAE, in parentheses) in kcal/mol of the classical model for the 3-body terms of databases A and B compared to potentials that were fit to PIPs.

(multipoles, polarizabilities, damping parameters, C_9 coefficients) are derived *ab initio*. The root mean squared errors (RMSEs) of the model are 0.084 and 0.104 kcal/mol on Databases A and B (cf. Table 5.1), respectively, demonstrating its transferability over a wide range of trimer configurations. Table 5.1 organizes the RMSEs and MAEs for the classical model (this work) against other interaction potentials (q-AQUA,[50] MB-pol,[46,47] HBB2-pol,[183] WHBB5[49]) that were fitted using PIPs to *ab initio* trimer calculations. On Database A, the classical model (induction+dispersion) yields a lower RMSE than HBB2-pol and WHBB5 (note that the RMSE of HBB2-pol and WHBB5 potentials was reported on the entire dataset of 12,347 trimers while the classical model on Dataset A, in which we removed 87 trimers from the original dataset for reasons stated above). The reported excellent accuracy for MB-pol on Database A (left top panel of Fig. 5.4) is in reality a training/fitting error as the entire database A was used for the fitting of that model using 1,173 parameters (1,163 linear and 10 non-linear). However, that potential performs worst for the larger Database B, as can be seen from Figure 5.4 (top right panel) and Table 5.1. For the larger number of trimers in Database B, the RMSD/MAE of MB-pol increases by ∼4.1x / ∼1.8x with respect to the fitting errors for Database A (0.116/0.035 kcal/mol for Database B vs. 0.028/0.019 kcal/mol for Database A). It is interesting that the TTM2.1-F potential (bottom panels of Figure 5.4), which uses geometry-dependent distributed monopoles and inducible point dipoles (5 parameters total) and is fitted to ∼45 *ab-initio* dimer and no trimer points produces RMSD/MAE values of 0.210/0.110 kcal/mol, respectively. This is to be compared with the corresponding RMSD/MAE values of 0.104/0.054 kcal/mol for the classical induction + dispersion model based on geometry-dependent multipoles and polarizabilities suggesting that it outperforms both MB-pol and TTM2.1-F on Database B, which contains database A (MB-pol training set). Importantly, these results highlight the natural transferability of the classical model to a very wide range of trimer configurations.

To better understand this difference, Figure 5.5 shows the cumulative RMSE as a function of the total deformation (1-body) energy, $E_{deformation}$, of the trimer, evaluated using the Partridge-Schwenke potential energy surface.[255] We observe that the MB-pol potential performs very accurately for small $E_{deformation}$ values. However, it produces errors for highly distorted trimer configurations that are slightly larger than the flexible classical induction + dispersion model. While it is true that the highly distorted, high energy structures are less relevant for many applications, it highlights the critical need for extensive sampling to build a training set to train the PIPs, which is a nontrivial task. On the other hand, the classical model is naturally transferable to more distorted intramolecular geometries with a clear improvement when geometry-dependent distributed multipoles and polarizabilities (classical model - flexible) are used, performing extremely well even for high-energy (large deformation) structures. However, this can be a challenge for PIPs because, as with most supervised machine learning techniques, the accuracy of the functions in regions outside their training data is not guaranteed.

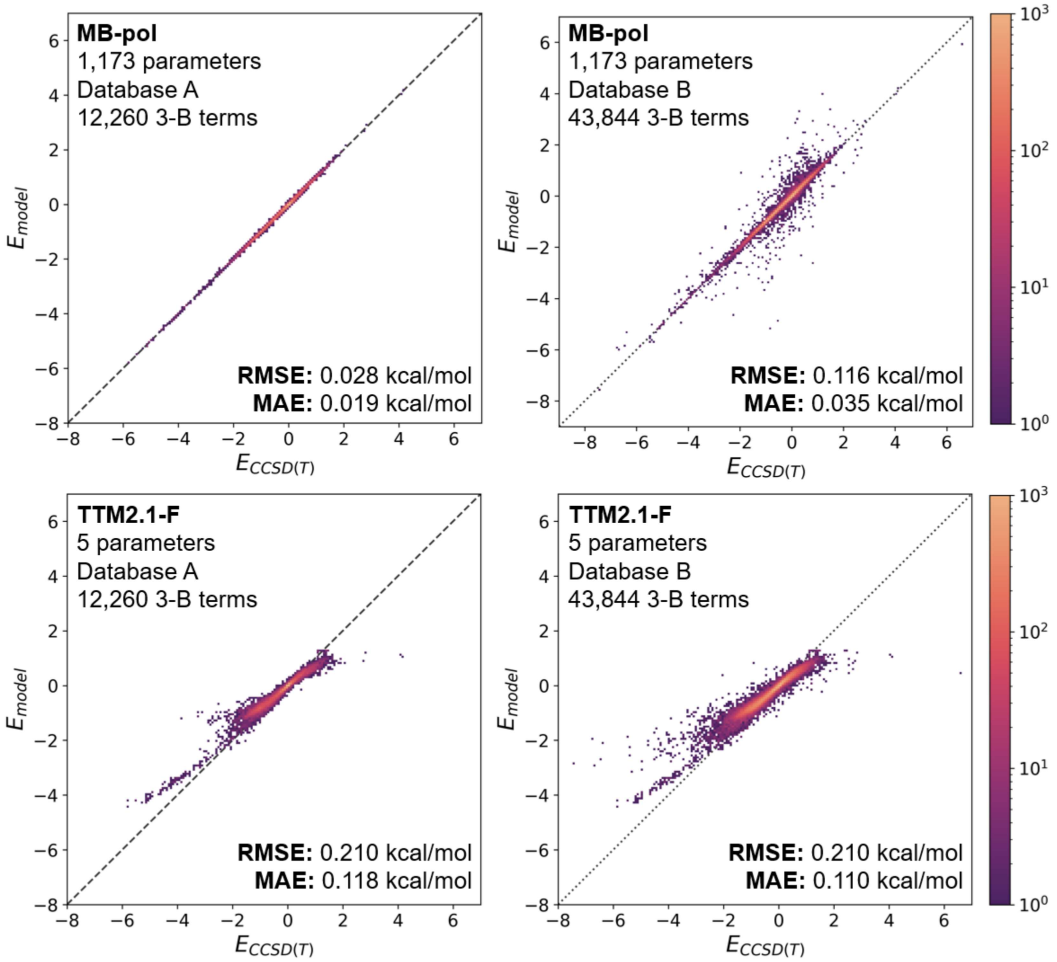


Figure 5.4: Comparison of the 3-body energy (kcal/mol) calculated using the MB-pol (top panels) and TTM2.1-F (bottom panels) potentials against the corresponding benchmark values ($E_{\mathrm{CCSD(T)}}$) for Database A (left) and Database B (right).

5.3.3 c. Performance of the classical model for the 4-body energies

Having established the success of the classical model for the 3-body interactions in water clusters, we now turn our attention to the 4-body term. Figure 5.6 depicts a correlation plot demonstrating the performance of the classical model on Database C. Table 5.2 lists the RMSE and MAE values for the induction model (this work) against the q-AQUA potential, which is the only potential to date that includes an explicitly fitted 4-body term.[50] Surprisingly, we find that the model with static multipoles and polarizabilities yields a slightly

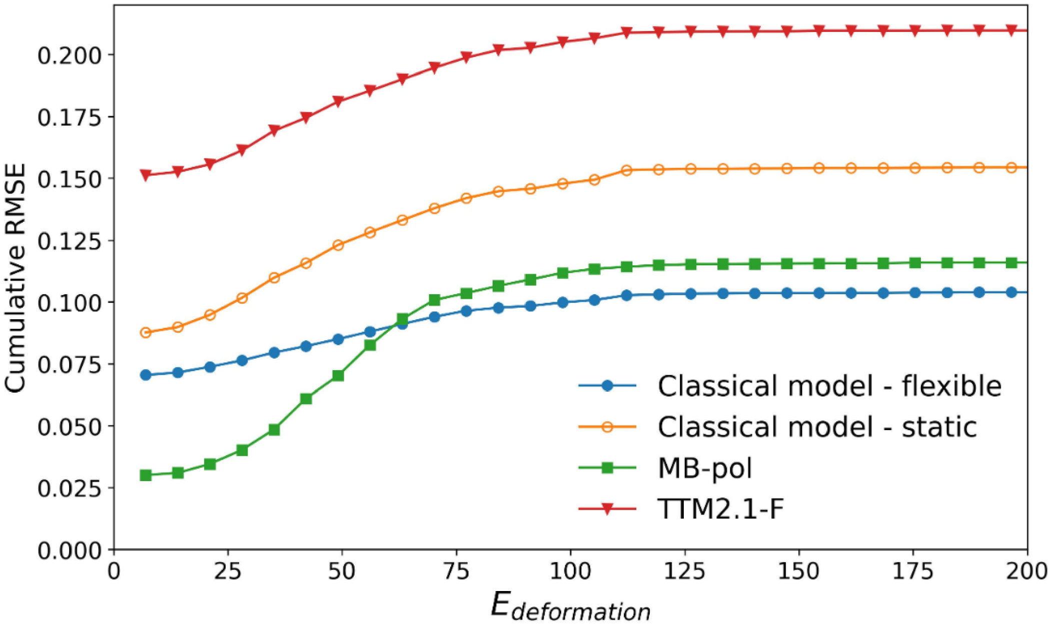


Figure 5.5: The cumulative RMSE (kcal/mol) as a function of the total deformation energy of the trimer ($E_{\text{deformation}}$, kcal/mol).

		Database C 3,603 tetramers	Fitted parameters	Size of training set
Induction only	Static dm/dp	0.036 (0.020)	0	n/a
	Flexible dm/dp	**0.058 (0.026)**	0	n/a
Potentials fit to PIPs	q-AQUA	0.021[a]	200	3,692

[a] Reported RMSE fitting error on entire 3,692 tetramers for the q-AQUA 4-body training set.[50]

Table 5.2: The root mean square error (RMSE) and mean absolute error (MAE, in parentheses) in kcal/mol of the classical model for the 4-body terms of Database C and a comparison with the reported fitting error of the q-AQUA potential.

better RMSE than the flexible model. The largest differences are due to the strongly attractive 4-body terms, which we find are over-estimated by the model with flexible moments. In general, these outliers have small interatomic O···O distances ($\leq$ 2.45 Å) and O···H distances

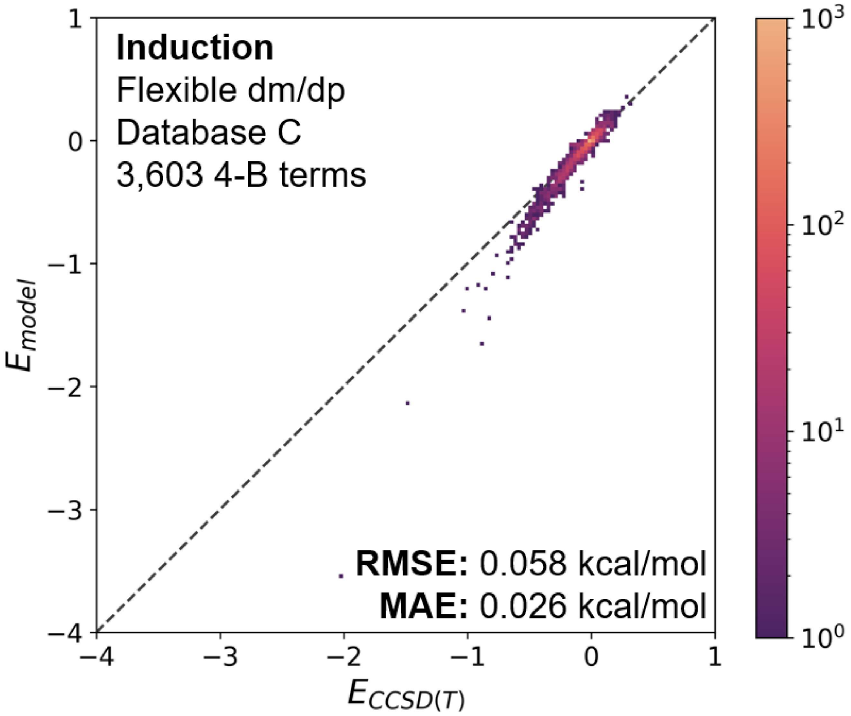


Figure 5.6: Comparison of the 4-body energy (kcal/mol) calculated using the classical induction model (E_{model}) against the benchmark CCSD(T) values ($E_{\mathrm{CCSD(T)}}$) in Database C.

which suggests that these tetramer systems overestimate the 4-body induction energy due to overpolarization at small interatomic distances. As mentioned earlier, we do not consider any trimers or tetramers in the database that have an $\mathrm{O}\cdots\mathrm{O} \leq 2.4$ Å. An RMSE value of 0.023 kcal/mol is observed for tetramers with a minimum $\mathrm{O}\cdots\mathrm{O}$ of ≥ 2.7 Å and an RMSE value of 0.042 kcal/mol for tetramers with a minimum $\mathrm{O}\cdots\mathrm{O}$ of ≥ 2.5 Å. That said, for the configurations most relevant to bulk water and water clusters, this should not pose a significant issue. However, it is possible for this to be improved by modifying the damping scheme[256] to prevent overpolarization at small interatomic distances.

Again, it is difficult to compare the accuracy of the model to q-AQUA since that model is fitted to Database C using ~200 parameters and we have removed 89/3,692 tetramers from the original database used to fit q-AQUA. However, we anticipate that the induction model will be quite transferable to a wide range of configurations, as observed for the 3-body term. Furthermore, the induction model is naturally capable of computing higher-order contributions (5-body and higher) to the induction energy through the same formalism.

5.3.4 d. Sensitivity analysis

In this section we report the performance of the model upon varying the details of electronic structure theory (level of theory and basis set) used to obtain the multipoles and polarizabilities as well as the effect of truncating the multipole and polarizability expansions.

While it has been established that a multipolar representation up to the hexadecapole (H) was necessary to obtain converged molecule dipole moments in ice Ih,[257] we find essentially no change in the accuracy of the 3- and 4-body terms by extending the model beyond the quadrupole moments (Q). For Database A, inclusion of D–D, D–Q, and Q–Q polarizabilities (RMSE: 0.083 kcal/mol) significantly improves the accuracy of the model over using only D–D (RMSE: 0.122 kcal/mol) or D–D and D–Q polarizabilities (RMSE: 0.238 kcal/mol). The same trend, viz. expansion up to (Q) and inclusion of D–D, D–Q, and Q–Q) is observed for the 4-body term (database C) While the difference in RMSE between the D-D model (RMSE: 0.122 kcal/mol) and the D-D, D-Q, Q-Q model (RMSE: 0.083 kcal/mol) is not huge, the former results in a systematic underestimation in the energetically favorable 3-body interactions.

The sensitivity of the distributed multipoles and polarizabilities with respect to the level of theory (PBE0,[258] HF, MP2) and basis set (aug-cc-pVTZ and aug-cc-pVQZ)[16] was also examined. We find that increasing the basis set size from aug-cc-pVTZ to aug-cc-pVQZ leads to small differences in the resulting distributed multipoles and polarizabilities. The PBE0 and MP2 multipoles and polarizabilities are in excellent agreement with noticeable differences from those of HF.

We also compare the results of the HF/aVTZ and PBE0/aVTZ surfaces on the databases of 3-body and 4-body term. Specifically, the HF/aVTZ surfaces result in RMSEs of 0.084, 0.104, and 0.058 kcal/mol on Databases A, B, and C, respectively. The DFT/aVTZ surfaces result in RMSEs of 0.089, 0.103, and 0.084 kcal/mol. That said, because the accuracy is overall better with the HF/aVTZ surfaces, we have chosen to report those values. The similarity in the accuracy of the results aligns with findings from a previous work that state that there is little electron correlation in the 3- and 4-body terms of water, making HF an acceptable level of theory at which to carry out those calculations.[2]

5.4 Conclusions and Outlook

This work demonstrates the success of a classical model in representing 3-body and 4-body interactions for a wide range of water trimer and tetramer configurations spanning a wide range of inter- and intra-molecular geometries. In general, a flexible (geometry-dependent) representation of the distributed multipoles and polarizabilities significantly improves the accuracy of the model and naturally becomes more important as intramolecular geometries become more distorted from their equilibrium geometry. To the best of our knowledge, the role of geometry-dependent multipole and polarizability surfaces on the 3- and 4-body term have not been previously reported. In addition to the utility of geometry-dependent multipoles and polarizabilities, we find that multipoles up to the quadrupole moment (Q) and polarizabilities up to quadrupole-quadrupole (Q–Q) are necessary to achieve an accuracy that is on par with potentials fitted to tens of thousands of high level *ab initio* calculations. In our opinion, the lack of geometry-dependent multipoles and polarizabilities and the absence of dipole-quadrupole (D–Q) and quadrupole-quadrupole (Q–Q) polarizabilities have hindered the ability of existing classical polarizable potentials to describe the many-body terms in water with high accuracy. Furthermore, we find that incorporating estimated contributions from 3-body dispersion, which are typically very small and repulsive for the systems considered in this study, marginally improves the accuracy of the classical model when compared to *ab initio*. Importantly, we show that the classical model produces 3- and 4-body energies with an accuracy that is on a par with the quantitative accuracy of many-body potentials fitted to tens of thousands of *ab-initio* calculations using thousands of parameters.

The quantitative accuracy of the classical model demonstrates that induction (and dispersion for the 3-body) is indeed sufficient to describe accurately the many-body interactions in water. The classical model introduced in this work is naturally transferable to a wide range of highly distorted configurations, an attribute that is not guaranteed with fitted PIPs. In addition, this approach is transferable to new chemical systems in a natural and straightforward manner. More specifically, it would only be necessary to derive distributed multipoles and polarizabilities from the monomer wavefunction of additional solvent molecules or ions, a process that is quite fast for an individual monomer and needs to be carried out only once. If the molecule is suspected to undergo large changes in intramolecular geometry, it is recommended to develop a geometry-dependent description of the charge distribution and polarizability. For systems with only a few degrees of freedom of the respective monomers (like water), this can be achieved by performing multiple monomer calculations at a range of intramolecular geometries. For systems with more degrees of freedom, such as polyatomic solvents or ions, it might be useful to follow a recently developed approach that uses a neural network to predict distributed multipoles for a wide range of systems and geometries.[259] Furthermore, with this approach, one need not be concerned with building a very large and sufficiently diverse training set from costly *ab-initio* calculations as required for fitting using PIPs. Nevertheless, we do acknowledge the usefulness of being able to benchmark our classical model against existing benchmark databases consisting of tens of thousands of high level *ab initio* points. We hope that the use of an accurate classical model based on induction and dispersion to describe many-body effects in hydrogen bonded systems can simplify the process of developing highly accurate potentials for multi-component systems by eliminating the need to perform computationally expensive high level *ab initio* calculations.

Chapter 6

A FORMULATION OF THE MANY-BODY EXPANSION (MBE) FOR PERIODIC SYSTEMS: APPLICATION TO SEVERAL ICE PHASES

Reproduced in part with permission from [Kristina M. Herman and Sotiris S. Xantheas. A Formulation of the Many-Body Expansion (MBE) for Periodic Systems: Application to Several Ice Phases; *J. Phys. Chem. Lett.* 14 (4), 989–999]. Copyright [2023] American Chemical Society

6.1 Introduction

The many-body expansion (MBE) is a useful tool for partitioning physical properties of a complex system into the components of its constituent smaller non-overlapping subsystems. These quantities can include, among others, the binding energy (D_e),[1–3, 92, 94] forces,[201, 260] harmonic frequencies[261] or the electron density.[262] The MBE has been previously applied to aqueous clusters, providing both a reliable computational framework and benchmark reference values for the magnitude of the many-body effects in water and ion-water clusters.[1–3] Further, some of the most accurate classical potentials for water developed thus far are built on the premise of the MBE by fitting the many-body terms (up to the 3-body or 4-body) to high-level *ab initio* data.[46, 47, 50] The success of the MBE as the foundation for the development of these potentials stems from its convergence at the 4-body term for water.[2, 94] Recently, the MBE has been extended to break covalent bonds (rather than hydrogen bonds) and used in that context to examine the alleged first row anomaly[98] for the XH_n systems (X=C, Si, Ge, Se, n=1-4) as well as to examine its convergence and applicability to small metal clusters.[97] The widespread use of the MBE for numerous properties and a variety of molecular systems is motivated by numerous computational advantages. These benefits include the trivial parallelism of the calculations, the potential to combat the poor scaling of electronic structure methods,[201] and the ease of utilizing a hybrid approach to apply different levels of theory to evaluate the many-body terms.[44]

However, the extension of the MBE for infinite periodic systems is less straightforward than the one for finite (cluster) systems and as such has not been fully developed yet. The complication arises from the fact that periodic boundary conditions (PBCs) induce interactions of fragments with their periodic images in an undesired manner. Furthermore, an attempt to enumerate all possible subsystems of an infinite system will involve an infinite number of calculations. The MBE has been previously applied to various organic crystals[263–268] to estimate their lattice energy. In the CrystaLattE code,[264] this is accomplished by enumerating all possible subsystems (dimers, trimers, tetramers of molecules) within a large supercell and passing these subsystems through a series of filters that involves the nuclear repulsion energy (NRE), root mean squared deviation (RMSD), and a comparison of the eigenvalues of the Coulomb matrix to remove redundant calculations and track the number of times these interactions occur within the larger supercell. Then, the remaining n-body terms of the unique subsystems (dimers, trimers, etc.) are calculated within a given cutoff distance and divided by the number of molecules in the subsystem. However, the success of that approach is dependent upon a robust screening protocol and the careful tracking of the number of times subsystems occur within a larger supercell, all of which needs to be repeated for each new system or geometry.

In this work, a formulation of the MBE for infinite periodic systems is introduced, in which the unique subsystems necessary to compute are determined solely through the inherent translational symmetry of the system, without any screening procedures. While the ultimate goal is to decompose the lattice energy into its many-body terms, thus obtaining information about the convergence of the MBE for infinite periodic systems, we will start by examining the many-body expansion for finite systems with translational symmetry. Namely, we first focus on finite gas phase supercells, which comprises replicated unit cells. We will use these systems as a proof of principle, showing that the method computes the unique interactions according to translational symmetry, scales them appropriately (using a simple analytic expression based on the dimensions of the supercell), and exactly recovers the full n-body terms using a fraction of the calculations.

After establishing the success of the method for finite gas phase supercells, we will extend the formulation for infinite periodic systems. For those systems, we show that the translationally unique subsystems, as determined by this method, have the same scaling factor in the limit of an infinite cell and, in that sense, contribute equally to the lattice energy. The proposed method is validated by showing that the lattice energies computed via the MBE (up to the 4-body) reproduce the result obtained using periodic boundary conditions (PBCs) with an Ewald summation for 7 different polymorphs of ice (Ih, II, VIII, IX, XIII, XIV, XV). This is the first time that the MBE has been applied to decompose the lattice energies of various ice polymorphs, partially due to the difficulty of handling the strong long-range and many-body interactions in these systems. Importantly, this systematic extension of the MBE to infinite periodic systems opens the door for the general use of this method to expand other molecular properties, combat the scaling of electronic structure methods, and benefit from the trivial parallelism of the MBE.

6.2 Computational details

The unit cell structures for ice Ih,[269] ice II,[270–272] ice VIII,[273–275] ice IX,[276–278] ice XIII,[279] ice XIV,[279] ice XV[280] are optimized with the PBE0[281] density functional at the experimental density (see the reference for optimization details).[282] The 3x3x3 proton-disorder supercells of ice Ih were generated using GenIce2[283, 284] and comprises 27 unit cells of differing hydrogen bonding arrangements, ensuring that the Bernal-Fowler rules[285] of hydrogen bonding are obeyed and that the net dipole moment of the supercell is zero. The water molecules in the ice Ih supercell were set at the average configuration of the water molecules in the unit cell (R_{O-H} = 0.9815 Å; Θ_{H-O-H} = 106.38°). The standard deviation of R_{O-H} and Θ_{H-O-H} were quite small (0.0002 Å and 0.12°, respectively). Additionally, the density of the supercell was set to the same density as the unit cell (0.93 g/cm^3).

The MBE on the unit cells were performed (without PBCs) with the TTM2.1-F and MB-pol classical interaction potentials, and also with MP2. We acknowledge that the comparison between the MP2 and the classical potential results for the gas phase unit cells takes into account only the short-range (intracell) interactions. However, extending the ab initio calculations to larger supercells is accompanied by a steep computational cost. The MBE of the gas phase unit cells were carried out up to the 5-body term (4-body term for ice Ih and ice VII, 3-body term for ice XIII) at the MP2 level of theory with Dunning's aug-cc-pVDZ[16] basis set. In addition, the MBE for some of the smaller unit cells (ice VIII and ice XV) were carried out with the Dunning aug-cc-pVTZ[16] basis set. These MP2 results were corrected for the basis set superposition error (BSSE) following the counterpoise method outlined by Boys and Bernardi.[170] Note that the 3-body term of ice XIII has not been corrected for BSSE because of the large unit cell size. From previous studies,[2] it was shown that BSSE-corrected MP2/aug-cc-pVDZ calculations accurately described the 3-body term and higher order terms of the MBE relative to the MP2/CBS values for aqueous systems. All MP2 calculations were performed with the NWChem 7.0.2 electronic structure package.[214] The MB-pol[46, 47] calculations were carried out with the legacy MB-pol distribution of the code which has been shown to reproduce the individual many-body terms of ice quite accurately relative to 2- and 3-body terms computed at the CCSD(T)/aVTZ level of theory.[5] The TTM2.1-F[30] calculations were performed using the open-source distribution from the Pacific Northwest National Laboratory (available upon request).

The MBE calculations for the ice systems were performed with an in-house python code. To test whether the proposed method yields the correct lattice energy for the potential, the results are compared to the lattice energy estimate using the implementation of the potential with PBCs. The MB-pol lattice energies were available at these geometries.[5] However, since the lattice energy of ice Ih was computed using a single proton-ordering[5] despite being a proton-disordered phase, we have used the MBX distribution of the MB-pol potential to compute the lattice energy of a proton-disordered 9x9x9 supercell of ice Ih with a cutoff of 9 Å. The TTM2.1-F lattice energies were computed using 9x9x9 supercells of the ice phases (accounting for proton-disorder in ice Ih, as described earlier) with a cutoff of 9 Å.[30]

6.3 Results and discussion

The unit cells of the 7 ice polymorphs considered in this study (Ih, II, VIII, IX, XIII, XIV, XV) are shown in Figure 6.1. First, the performance of the many-body classical interaction potentials TTM2.1-F and MB-pol on the many-body terms for the gas phase ice unit cells (i.e., no PBCs) is benchmarked against results obtained at the MP2 level of theory. The MBE is carried out in the typical manner, computing $\binom{N}{n}$ number of n-body terms for each order n of a system of N molecules in the unit cell. In the full MBE, the many-body terms

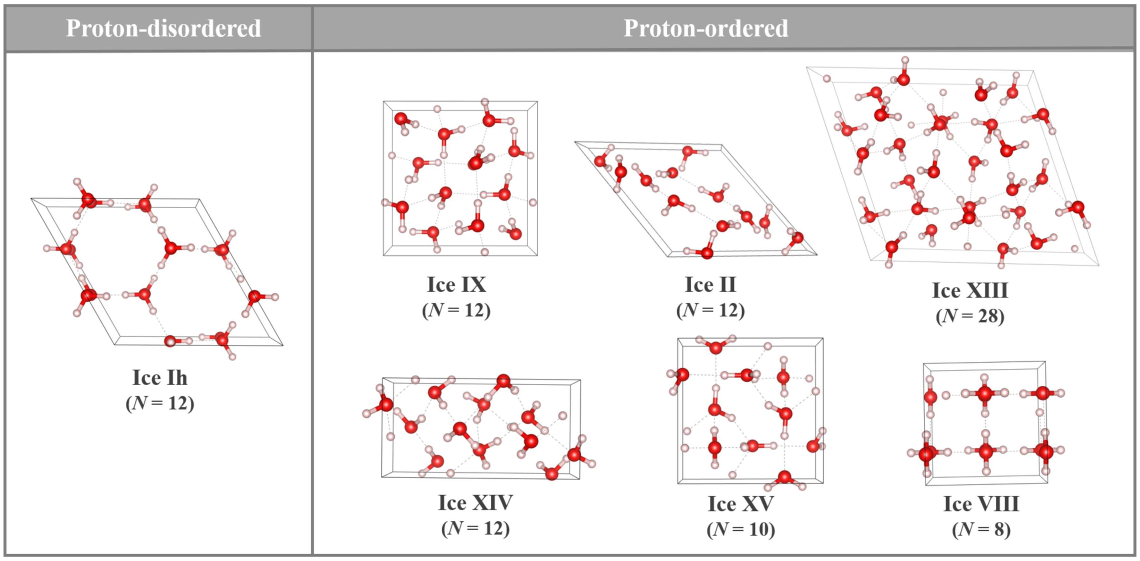


Figure 6.1: Unit cell structures of ice Ih, II, VIII, IX, XIII, XIV, and XV (unit cell visualizations produced using the software VESTA[4]). The number of molecules (N) in each unit cell is indicated in parentheses.

formally scale on the order of N^n, where N is the number of fragments and n is the n-body term of the MBE. As the system size increases, the number of higher order terms in the MBE grows steeply with N. The total n-body contribution, E(n), to the unit cell binding energy (D_e) is computed as

$$E(n) = \sum_{i=1}^{n} (-1)^{n-i} \binom{N-i-1}{n-i} \sum_{j=1}^{\binom{N}{i}} E_j(n) \tag{6.1}$$

where E$_j(n)$ is the individual n-body term of the jth n-mer and N is the number of fragments in the system.

The many-body terms of the gas phase ice unit cells (without PBCs) computed at the MP2 level of theory (BSSE-corrected), the MB-pol,[46,47] and the TTM2.1-F[30] potentials are shown in Table 6.1. It has been previously established that the BSSE-corrected MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ 2-body terms underestimate the MP2/CBS 2-body term,[2] with the latter being closer to the CBS limit than the former. As regards the 3-body and higher terms in the MBE, the BSSE-corrected MP2/aug-cc-pVDZ results have been shown

	Ih (N=12)			**IX** (N=12)			**II** (N=12)			**XIII** (N=28)		
	MP2	MB-pol	TTM2.1-F	MP2	MB-pol	TTM2.1-F	MP2	MB-pol	TTM2.1-F	MP2	MB-pol	TTM2.1-F
1B	5.02	8.95	8.95	3.37	7.13	7.13	2.45	5.84	5.84	8.08	15.51	15.51
2B	-36.24	-43.39	-46.79	-38.65	-48.94	-53.30	-21.46	-29.28	-30.93	-135.28	-167.11	-178.36
3B	-4.18	-4.02	-3.48	-3.49	-2.99	-2.59	-8.73	-8.91	-7.15		-30.24	-27.09
4B	-0.17	-0.19	-0.23	0.16	0.18	0.28	-1.00	-0.85	-0.96		-0.24	-0.35
5B	-0.01	-0.01	0.01	-0.05	-0.04	-0.03	-0.07	-0.02	-0.10		0.10	0.17
D_e (1B-5B)	-35.59	-38.65	-41.55	-38.66	-44.66	-48.51	-28.80	-33.22	-33.29		-181.98	-190.12
D_e	-35.59	-38.65	-41.55	-38.65	-44.65	-48.51	-28.80	-33.21	-33.30	-161.31	-181.99	-190.13

	XIV (N=12)			**XV** (N=10)			**VIII** (N=8)		
	MP2	MB-pol	TTM2.1-F	MP2	MB-pol	TTM2.1-F	MP2	MB-pol	TTM2.1-F
1B	3.39	6.50	6.50	2.58 (3.83)	5.03	5.03	0.88 (1.58)	2.28	2.28
2B	-39.89	-48.87	-50.41	-38.73 (-43.48)	-47.15	-47.80	-26.36 (-29.42)	-32.59	-26.51
3B	-3.59	-2.88	-3.04	-6.59 (-6.54)	-5.51	-4.66	-3.36 (-3.27)	-2.47	-2.35
4B	0.22	0.26	0.20	-0.19 (-0.20)	-0.13	-0.18	0.03 (0.03)	0.11	0.03
5B	-0.02	-0.02	0.00	-0.01 (-0.01)	0.01	0.00	-0.02 (-0.01)	0.00	0.00
D_e (1B-5B)	-39.88	-45.02	-46.75	-42.94 (-46.39)	-47.73	-47.59	-28.82 (-31.09)	-32.67	-26.55
D_e	-39.88	-45.01	-34.74	-42.94 (-46.39)	-47.73	-47.59	-28.82 (-31.09)	-32.67	-26.56

Table 6.1: MP2/aug-cc-pVDZ Results (MP2/aug-cc-pVTZ Results in Parentheses for Ice XV and VIII) for the MBE of the Gas-Phase Ice Unit Cells Compared to the Results with the Many-Body Classical Potentials TTM2.1-F and MB-pol.

to align closely with the corresponding MP2/CBS values,[2] which is further evidenced by the close agreement between the aug-cc-pVDZ and aug-cc-pVTZ basis sets (in parentheses) for ice XV and ice VIII. In fact, a recent study has shown that the 3-body term can be accurately evaluated even classically using multipoles and polarizabilities.[44] The results of Table 6.1 indicate that the MBE up to 5-body, D_e(1B-5B), reproduces the gas phase unit cell's total binding energy, D_e, to within < 0.02 kcal/mol with both the MP2 and the classical potentials, suggesting that the MBE can be safely truncated at the 5-body term even for unit cells with much larger number of molecules such as ice XIII (N=28). Additionally, the truncation at the 3-body term yields results within 1.06 kcal/mol (or within 3.2%), while the truncation at the 4-body term produces binding energies that are within 0.17 kcal/mol (or within 0.3%) of the gas phase unit cell's total binding energy. These findings establish important thresholds as regards the required accuracy in the energetics as a function of the rank in the MBE. We also find that in general both classical potentials describe the magnitude of the many-body interactions in the gas phase unit cells quite well. In a previous study, the MB-pol potential has been shown to accurately reproduce the individual many-body terms of ice relative to 2- and 3-body terms computed at the CCSD(T)/aVTZ level of theory,[5] whereas the TTM2.1-F potential tends to overestimate the total 2-body term and underestimate the total 3-body term. For this reason, we will hereafter focus on the results using the MB-pol potential. Note that the binding energies with both classical potentials are larger than the one obtained at the MP2/aug-cc-pVDZ or MP2/aug-cc-pVTZ levels of theory, mainly due to the previously discussed underestimation of the 2-body term in the latter. While the results for the gas phase unit cells help to validate the use of classical potentials for the subsequent calculations of the infinite periodic systems, the results for the gas phase unit cells of ice are naturally far from the description of the solid-state energetics. That said, we will now turn our attention to the systematic extension of the MBE, first to finite systems with translational symmetry (gas phase supercells) and then to infinite systems with translational symmetry (unit cells with periodic boundary conditions).

A schematic of the MBE for finite systems with translational symmetry (finite supercells) is outlined in Figure 6.2. While gas phase supercells may be of limited interest, the validation of the proposed MBE scheme for these systems makes the transition from finite to infinite systems more intuitive and instructive. Further, the development of the MBE for finite periodic systems facilitates bridging the computational path from small unit cell systems to large supercells to infinite ones, a desirable capability for studying molecular aggregates at the nanoscale. The goal of extending the MBE to the finite supercells is to recover the many-body terms of the full MBE using only the translationally unique subsystems (monomers, dimers, trimers, etc.). To facilitate this, we must first compute the translationally unique subsystems and scale them by the number of times they occur in the finite supercell. We focus on leveraging only the translational symmetry, since this is the only symmetry inherent to periodic systems. It should be noted that this methodology could be further extended to include additional symmetry operations within the unit cell. This will not change the result but will rather reduce the number of necessary calculations. In order to maximize the benefit of translational symmetry, the many-body terms can be cast in the contributions from the intra- and inter-cell components according to

$$D_e = \sum_{intracell} E_{1B} + \sum_{intracell} E_{2B} + \sum_{intercell} E_{2B} + \sum_{intracell} E_{3B} + \sum_{intercell} E_{3B} + \cdots \tag{6.2}$$

This is because, for example, all dimers in a supercell will be either in the same unit cell (intracell) or in two different cells (intercell). That said, rather than finding the translationally unique dimers, trimers, tetramers, etc. of molecules in the system, we can instead focus on the unit cell level, i.e., finding translationally unique dimers, trimers, tetramers, etc. of unit cells, to directly use the translational symmetry of the systems. This is because if a pair, or subset, of unit cells is equivalent according to translational symmetry (i.e., it can be translated directly on top of one another), the interactions between the molecules in those unit cells will also be identical. In the following we use the Latin symbols a, b, c to denote the crystallographic axis directions and the Greek symbols α, β, γ to indicate the number of times that a unit cell is reproduced in the a, b, c directions; for instance, the 3x3x3 supercell in Figure 6.2 will have $\alpha = \beta = \gamma = 3$, i.e., it consists of 3 unit cells along the a-, b-, and c-directions for a total of 27 unit cells. Utilizing the translationally unique sets of cells increases the generalizability of the method because the translationally unique sets of unit cells (indicated using indices in the a-, b-, and c-directions) will be identical for all periodic systems, regardless of unit cell size or shape. Therefore, by finding the unique pairs of unit cells and computing all possible dimers between those cells, the intercell 2-body term can be obtained. For the intercell 3-body term, the 3-body terms between pairs of unit cells (one molecule in one unit cell and two in the other, and vice versa) and sets of three unit cells (one molecule in each cell) must be computed. Using the same logic, all 4-body terms must be computed for tetramers residing in the same cell or up to 4 different cells, using the translationally unique sets of dimers, trimers, and tetramers of unit cells to enumerate all possible 4-body terms. For each set of X cells with N molecules in the unit cell, there will be exactly

$$\sum_{i=0}^{X} (-1)^i \binom{X}{X-i} \binom{(X-i)\cdot N}{n} \tag{6.3}$$

n-body terms for which there is at least one molecule in each of the X cells.

For finite supercells, the translationally unique subsystems can occur multiple times in the supercell system. Therefore, the computed n-body terms must be scaled by the number of times that the subsystem occurs within the larger supercell. This can be easily done using the cell dimensions along the a-, b-, and c-directions of the unit cells (see Figure 6.2) within the subsystem of unit cells (a, b, and c will revert to the normal x, y, and z directions for an orthorhombic unit cell). The scaling factor is

$$(\alpha - \Delta\alpha)(\beta - \Delta\beta)(\gamma - \Delta\gamma) \tag{6.4}$$

where α, β, and γ are the dimensions of the finite supercell in the a-, b-, and c-directions and $\Delta\alpha$, $\Delta\alpha$, and $\Delta\gamma$ are the maximum displacements (in unit cells) along the a, b, and c directions between the set of unit cells containing the subsystem (see Figure 6.2).

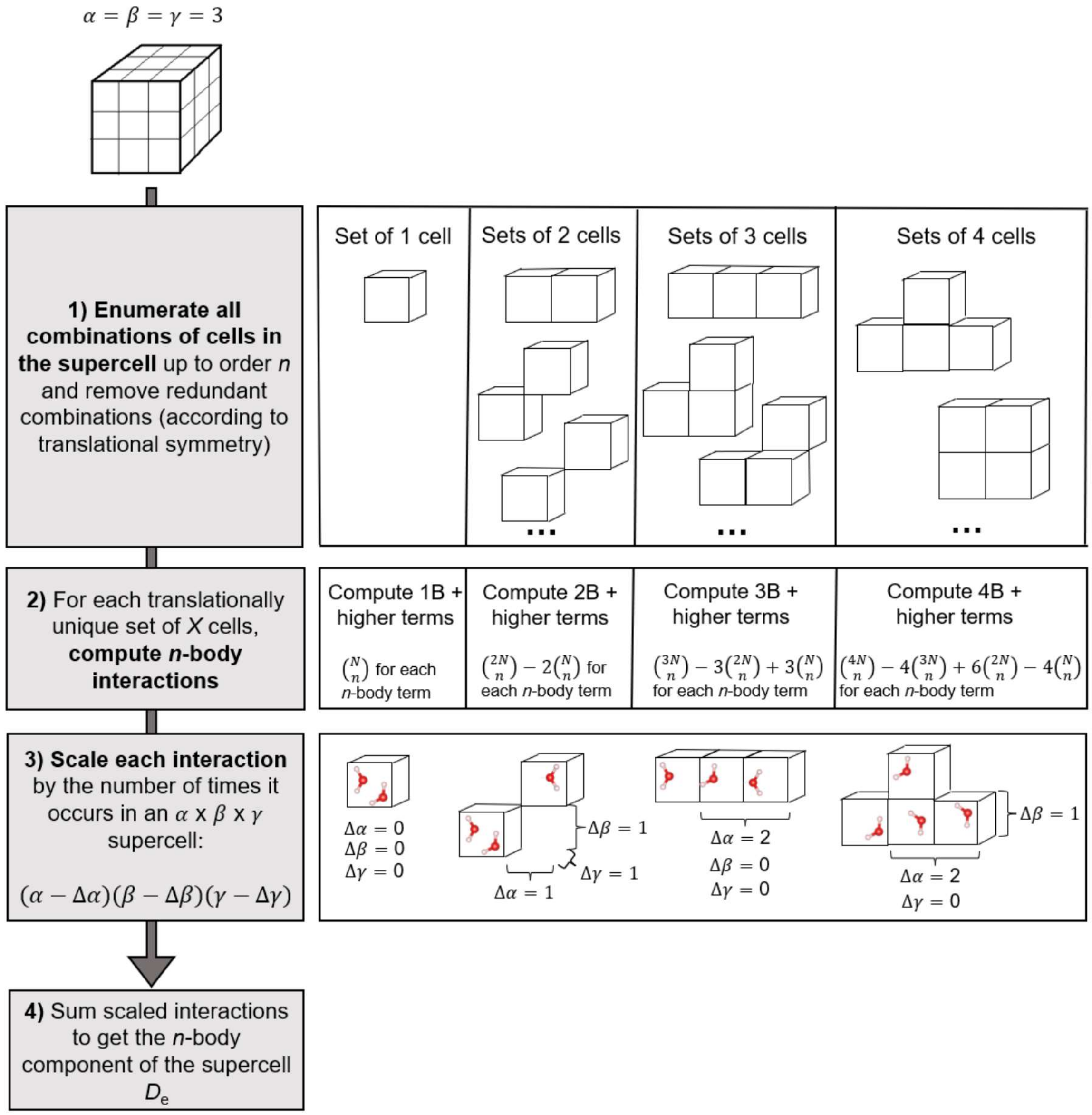


Figure 6.2: A schematic diagram demonstrating how to obtain the many-body components of the binding energy, D_e, of a 3x3x3 finite supercell ($\alpha = \beta = \gamma = 3$) using only the translationally unique subsystems up to order n. Note that only a subset of the translationally unique sets of 2, 3, and 4 unit cells are shown.

	3x3x3 supercell				**5x5x5 supercell**				**7x7x7 supercell**			
2-body	Energy (kcal/mol)	% error (vs. full)	Number of dimer calculations	Fraction of full calculation	Energy (kcal/mol)	% error (vs. full)	Number of dimer calculations	Fraction of full calculation	Energy (kcal/mol)	% error (vs. full)	Number of dimer calculations	Fraction of full calculation
Shell=0	-879.995	-55.25	28	0.12	-4074.052	-61.21	28	0.006	-11179.198	-63.46	28	0.001
Shell=1	-1958.051	-0.44	860	3.70	-10417.312	-0.83	860	0.17	-30286.991	-1.00	860	0.02
Shell=2	**-1966.646**	**0.00**	3996	17.21	-10497.372	-0.06	3996	0.80	-30561.159	-0.10	3996	0.11
Shell=3	–	–	–	–	-10503.318	-0.01	10972	2.20	-30586.192	-0.02	10972	0.29
Shell=4	–	–	–	–	**-10504.137**	**0.00**	23324	4.67	-30591.496	-0.01	23324	0.62
Shell=5	–	–	–	–	–	–	–	–	-30592.899	0.00	42588	1.13
Shell=6	–	–	–	–	–	–	–	–	**-30593.235**	**0.00**	70300	1.87
Full	-1966.646	0.00	23220	100.00	-10504.137	0.00	499500	100.00	-30593.235	0.00	3763396	100.00
3-body	Energy (kcal/mol)	% error (vs. full)	Number of trimer calculations	Fraction of full calculation	Energy (kcal/mol)	% error (vs. full)	Number of trimer calculations	Fraction of full calculation				
Shell=0	-66.667	-59.68	56	0.003	-308.644	-69.53	56	0.00003				
Shell=1	-159.602	-3.47	28408	1.72	-961.101	-5.11	28408	0.02				
Shell=2	**-165.332**	**0.00**	581304	35.10	-1010.850	-0.20	581304	0.35				
Shell=3	–	–	–	–	-1012.562	-0.03	4311416	2.59				
Shell=4	–	–	–	–	**-1012.868**	**0.00**	19345720	11.64				
Full	-165.332	0.00	1656360	100.00	-1012.868	0.00	166167000	100.00				
4-body	Energy (kcal/mol)	% error (vs. full)	Number of tetramer calculations	Fraction of full calculation								
Shell=0	3.047	-94.76	70	0.0001								
Shell=1	55.557	-4.45	532886	0.60								
Shell=2	**58.142**	**0.00**	45995046	52.15								
Full	58.142	0.00	88201170	100.00								

Table 6.2: Proof of Principle for the Introduced Method Applied to 3×3×3, 5×5×5, and 7×7×7 Finite Supercells of Ice VIII.

As proof of principle, we demonstrate the success of this method by numerically testing the resulting many-body terms using the described method against the full MBE (evaluated by computing all possible subsystems). Table 6.2 contains the MBE results for the 3x3x3, 5x5x5, and 7x7x7 supercells of ice VIII using only the translationally unique subsystems. The 2-, 3-, and 4-body terms are computed including increasing "shells" of interactions, a shell referring to the maximum distance (in unit cells) between the sets of unit cells. More specifically, shell=0 refers to only the intracell interactions and shell=2, 4, and 6 consist of all relevant interactions for 3x3x3, 5x5x5, and 7x7x7 supercells, respectively. At the maximum shell level, max(α, β, γ)-1, all unique interactions are included and the full MBE (calculating all possible $\binom{\alpha\beta\gamma N}{n}$ subsystems) matches exactly the results of the introduced method (using only translationally unique subsystems and scaling them using Eqn. 6.4). This demonstrates that the method successfully evaluates the translationally unique subsystems and scales them appropriately to yield the total n-body terms of the binding energy (D_e) of an arbitrarily large supercell. In comparison to the full MBE, this method performs fewer calculations for a given supercell size because it removes translationally identical subsystems. We also note that increasing beyond shell=2 for the 2-body term, only accounts for $\sim 0.1\%$ of the total 2-body term in the 7x7x7 supercell. Similarly, interactions beyond shell=2 account for $\sim 0.2\%$ of the 3-body interaction computed with the full MBE. The significance of the relatively local interactions further motivates the use of a cutoff distance or shell.

Now that the success of the method has been established for systems with translational symmetry (finite supercells), we will examine the formulation of the MBE in the limit of an infinite cell. Since our ultimate goal is to obtain the lattice energy (binding energy per molecule), we are interested in computing the total interactions unit per cell. That said, the scaling factor for each of the n-body terms will be exactly

$$\frac{(\alpha - \Delta\alpha)(\beta - \Delta\beta)(\gamma - \Delta\gamma)}{\alpha \cdot \beta \cdot \gamma} \tag{6.5}$$

In the limit of an infinite system (α, β, $\gamma \to \infty$), the scaling factors (6.5) for the translationally unique interactions will go to 1 according to

$$\lim_{\alpha,\beta,\gamma\to\infty} \frac{(\alpha-\Delta\alpha)(\beta-\Delta\beta)(\gamma-\Delta\gamma)}{\alpha\cdot\beta\cdot\gamma} = 1 \tag{6.6}$$

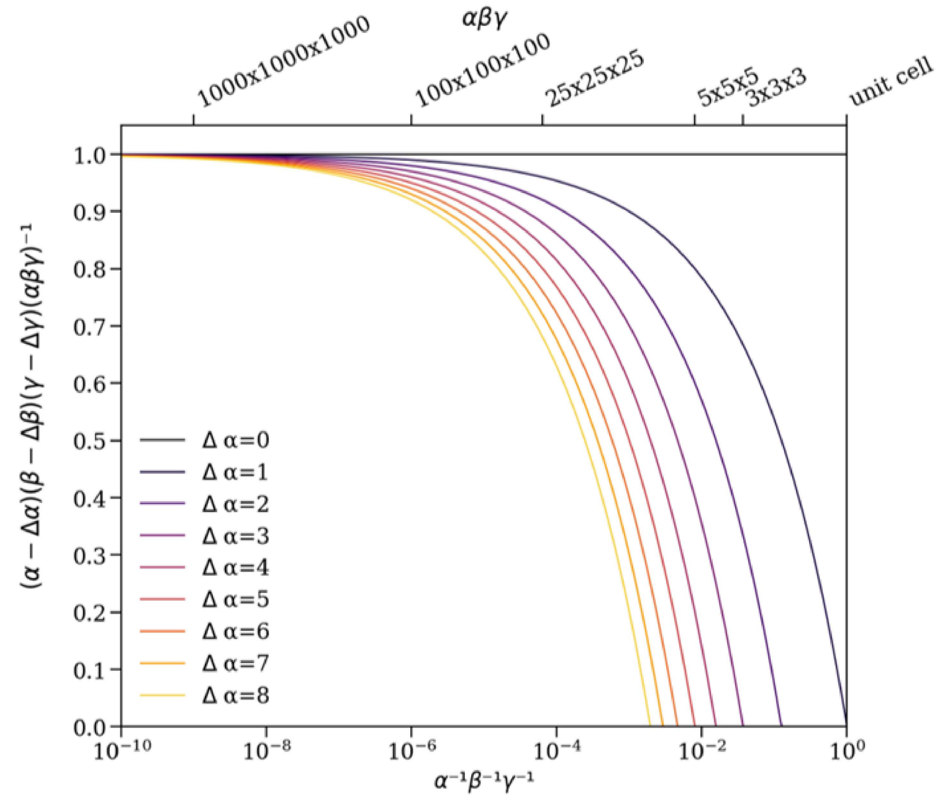


Figure 6.3: Variation of the scaling factor per unit cell (Eq. 5) for the translationally unique interactions with the dimensions α, β, γ of the supercell, with the individual lines representing different $\Delta\alpha$ values ($\Delta\beta = \Delta\gamma = 0$). In the limit of an infinite system (α, β, $\gamma \to \infty$) the scaling factors go to 1 for all $\Delta\alpha$ values.

Therefore, to compute the total binding energy per unit cell for an infinite cell, we simply calculate the sum of the translationally unique n-body interactions. Importantly, the main difference between the different two extremes of this approach (finite vs. infinite systems) is that for finite systems the translationally unique interactions are weighted by different amounts (depending on the supercell size), whereas for infinite systems the translationally unique interactions are weighted the same (Figure 6.2, step 3). The connection between the MBE for finite periodic systems and infinite periodic systems is demonstrated in Figure 6.3.

This plot clearly shows that the scaling factor of $((\alpha - \Delta\alpha)(\beta - \Delta\beta)(\gamma - \Delta\gamma\ )/(\alpha\beta\gamma)$ goes to 1 as the system size increases, under the assumption that $\Delta\alpha << \alpha$. Because we impose cutoff distances in our calculations accounting for relatively local interactions, this is a valid assumption. The results in Figure 6.3 highlight the generalizability of this approach in describing smaller finite systems, while at the same time having a straightforward extension in the limit of an infinite system.

However, since there is formally an infinite number of n-body terms, we impose distance cutoffs out to which the n-body terms will be computed. Note that the many-body terms are computed for all molecules in the unit cell and the cell-averaged lattice energy is reported as

$$E_{lattice} \simeq \frac{1}{N}[\sum E_{1B} + \sum E_{2B} + \sum E_{3B} + \sum E_{4B}] \tag{6.7}$$

i.e., the MBE is truncated at the 4-body term. We average the lattice energy over all molecules in the ice unit cells because, as discussed in a recent study by Nanayakkara et al., the water molecules within the ice unit cells can be in multiple different environments which means that it is possible that the energetic contributions may differ as well.[286]

To increase the speed of the method, the translationally unique combinations of unit cells are precomputed. This is done because combinations of unit cells are only defined by indices indicating the displacements in the a-, b-, and c-directions, making them identical for all unit cell systems regardless of unit cell size and shape. This reduces the computations by enumerating all possible many-body terms within only the relevant combinations of unit cells, based on the specified cutoff distance. It should be emphasized that no screening procedures are used in the calculations (i.e., NRE, RMSD, etc.), aside from applying a distance cutoff. The absence of screening procedures makes the possibility of optimizing the cells or driving dynamics with the MBE more practical.

To validate the proposed method, we computed the lattice energies of 7 ice polymorphs (Ih, II, VIII, IX, XIII, XIV, XV) and compared the results with the ones obtained using PBCs including the Ewald summation to account for long-range interactions. Since the MBE for infinite systems formally encompasses an infinite number of many-body terms, it is necessary to examine the convergence of the many-body components of the lattice energies with respect to a distance cutoff. In general, the maximum cutoff distances considered here are quite generous, typically extending over the length of multiple unit cell lengths.

The convergence of the 2-body component of the lattice energy with respect to the cutoff O-O distance for each of the ice polymorphs is shown in Figure 6.4a. The ice polymorphs considered in this study show a similar convergence pattern producing a value for the 2-body energy obtained with a cutoff of ∼12 Å that is within 1% of the value obtained at the maximum cutoff distance of 45 Å, except ice XIV and ice VIII for which the cutoff distance must be extended to ∼14 Å and ∼19 Å. Beyond that point, the 2-body term fluctuates by a very small amount. For comparison, a cutoff of ∼8 Å yields values within 5% of the total 2-body interaction.

For the 3-body and higher order terms (panels (b) and (c) in Figure 6.4), it is found that the product of the pairwise distances between the molecules in a subsystem serves as a more effective indicator of the magnitude of the many-body term than either the maximum distance or the sum of all pairwise distances. Not only is this representation of the cutoff more effective, but it also has a clear physical motivation because of the proportionality between the energy and $1/(R_{12} * R_{13} * R_{23}\ )$ as is seen in the Axilrod-Teller-Muto potential[248, 249] and in the interaction tensors used to describe the interaction between multipoles of different molecules,[207] since it has been shown that the 3-body terms in aqueous systems can be described very accurately using a classical induction model.[44] Borca et al. used a similar strategy to prioritize the calculation of many-body terms that were anticipated to be larger in magnitude.[264] The convergence of the 3-body term with respect to the cutoff $(R_{12} * R_{13} * R_{23})$ is shown in Figure 6.4b. The 3-body term fluctuates with respect to the cutoff for all polymorphs, except ice Ih, for which the 3-body value is converged to within 5% of the full value by a cutoff value of ∼120 Å^3. This is likely due to the differences in the proton-ordered versus proton-disordered systems. Because there is no long-range order in the proton-disordered ice system, the convergence of the 3-body term is much smoother (shorter range) than the one in the proton-ordered phases.

The convergence of the 4-body term with respect to the distance cutoff $(R_{12} * R_{13} * R_{14} * R_{23} * R_{24} * R_{34})$ is shown in Figure 6.4c. Note that while the 2- and 3-body terms

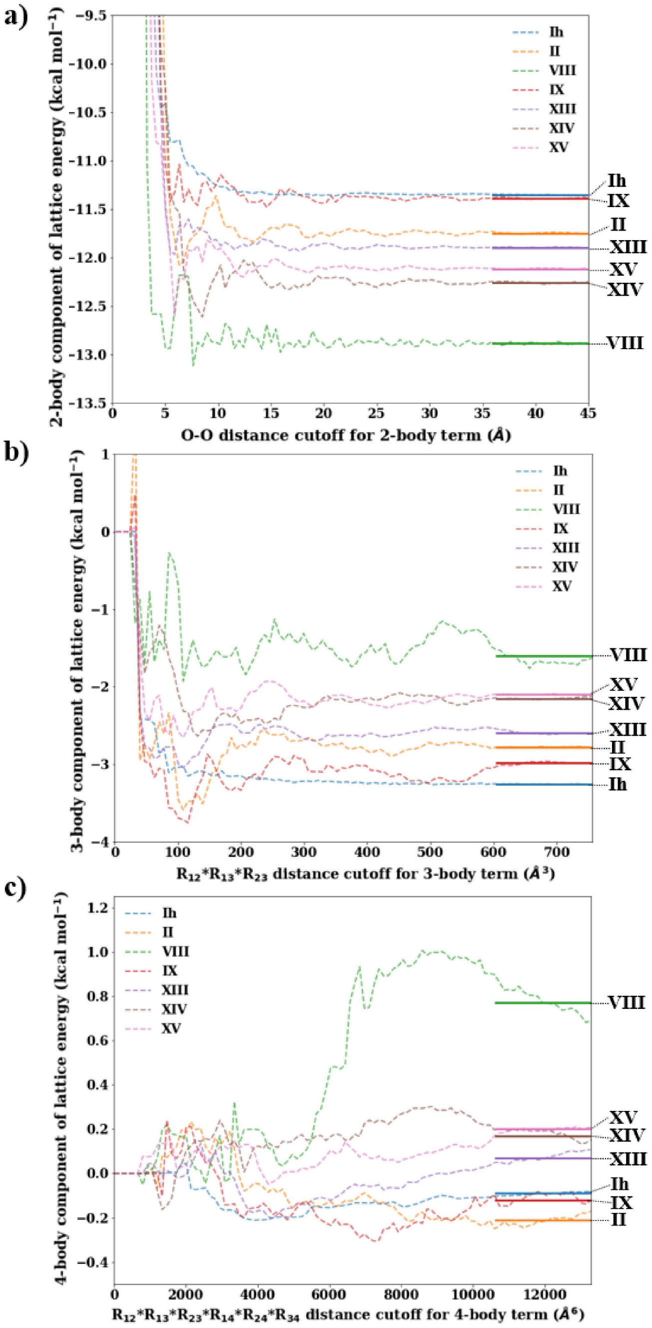


Figure 6.4: The convergence of the (a) 2-body, (b) 3-body, and (c) 4-body terms with respect to the distance cutoff for the various ice polymorphs (Ih, II, VIII, IX, XIII, XIV, XV, VIII).

	Ih	**XI**	**II**	**XIII**	**XIV**	**XV**	**VIII**
1B	0.75	0.59	0.49	0.55	0.54	0.50	0.29
2B	-11.350 (0.005)	-11.39 (0.02)	-11.75 (0.01)	-11.894 (0.007)	-12.26 (0.02)	-12.116 (0.008)	-12.88 (0.02)
3B	-3.254 (0.003)	-2.98 (0.03)	-2.78 (0.01)	-2.59 (0.03)	-2.15 (0.02)	-2.10 (0.01)	-1.6 (0.1)
4B	-0.09 (0.01)	-0.12 (0.03)	-0.21 (0.04)	0.07 (0.03)	0.17 (0.04)	0.20 (0.02)	0.77 (0.09)
E_{latt} (MBE)	-13.95 (0.02)	-13.90 (0.08)	-14.25 (0.06)	-13.86 (0.07)	-13.70 (0.08)	-13.51 (0.04)	-13.5 (0.2)
E_{latt} (PBC)	-14.09,[5] -13.96	-13.86[5]	-13.91[5]	-13.72[5]	-13.65[5]	-13.54[5]	-13.61[5]
Exp. lattice energy	-14.07[287]	-13.97[287]	-14.05[287]	-13.95[287]	-13.74[287]	-13.48[287]	-13.31[287]
Exp. density (g cm^{-3})	0.93	1.17	1.20	1.25	1.29	1.33	1.49

Table 6.3: Many-Body Terms and Lattice Energies (kcal/mol) of the Ice Phases Considered in This Study Computed with the MB-pol Potential. Uncertainties in parentheses are assigned based on the changes in the largest 20% of cutoff distances examined.

for ice Ih were performed on a 3x3x3 proton-disordered supercell (N=324), the 4-body term is calculated using a 2x2x2 proton-disordered supercell (N=96) due to the higher cost associated with evaluating the 4-body term. Even though the total 4-body contribution to the lattice energy is small ($<$ 1 kcal/mol) for all ice phases examined in this study, it requires numerous 4-body calculations to achieve a converged estimate. This is because although individual 4-body terms are small in magnitude, their number is quite large, and it may be necessary to be included in the total count to ensure accuracy.

For these reasons, the uncertainty in the total 4-body term is typically larger than for the lower order terms in the expansion, especially in ice VIII, for which we see large fluctuations with the value of the cutoff function. Despite this, the introduced MBE scheme for periodic systems yields relatively small uncertainties in the lattice energy estimates with the largest margin of uncertainty being $\sim$1.5% of the lattice energy for ice VIII and the smallest margin of uncertainty being $<$ 0.2% for ice Ih. We report the uncertainty to be half of the difference between the highest and lowest n-body term value for the 20% of the largest cutoff distances examined.

The many-body components of the lattice energies for the ice polymorphs considered here computed using the MB-pol potential are organized in Table 6.3. The lattice energies (kcal/mol) for each of the ice polymorphs are estimated as the sum of the 1-body through 4-body terms according to Eq. (6.6), averaging over the molecules in the unit cell. A

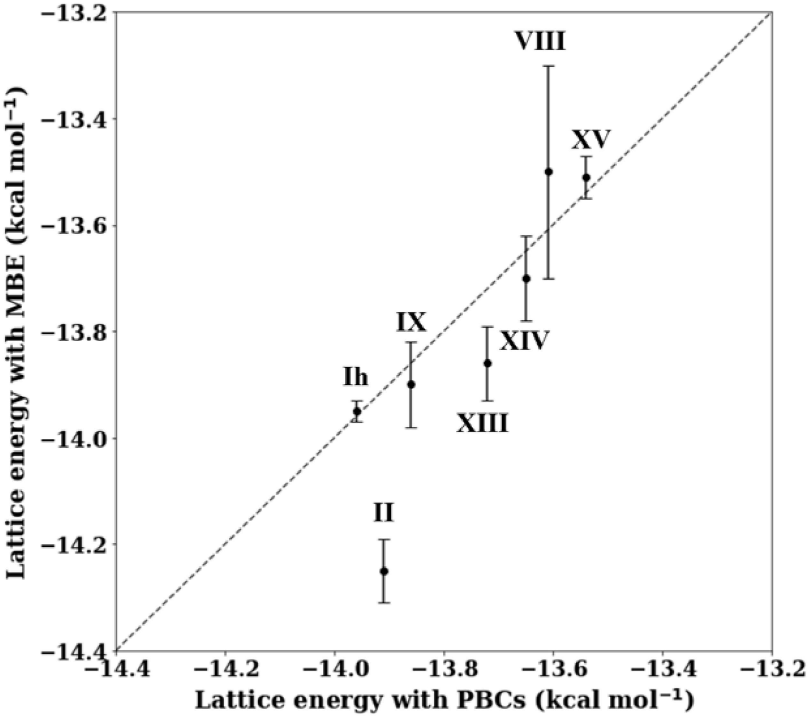


Figure 6.5: A comparison of the lattice energies obtained with the periodic MBE method (y-axis) versus the ones with a PBC implementation (x-axis) with the MB-pol potential with an Ewald summation to capture the long-range interactions outside of a 9 Å cutoff.[5]

comparison of the lattice energies produced from the MBE for periodic systems and the ones obtained using PBCs is shown in Figure 6.5. In general, we find excellent agreement between the two methods (periodic MBE vs. PBCs with Ewald sum), with the maximum error of ∼2% found for ice II. This demonstrates that this method adequately captures long-range interactions without an additional correction to the energy. More specifically, the longest cutoff necessary was 12-19 Å for the 2-body term to obtain results within 1% of the 2-body energy, computed at the maximum oxygen-oxygen cutoff distance of 45 Å. Naturally, this does not imply that the interactions beyond that point are exactly zero, but rather that the distribution of terms sums up to a value that is close to zero. The agreement between the lattice energies computed with the periodic MBE and with PBCs demonstrates the success of this method in describing a solid-state property (lattice energy) through a purely fragmentation-based approach. It should be noted that using the MBE to obtain the lattice energy with MB-pol (a classical potential) is not computationally more efficient than the conventional method with PBCs. This is because the scaling of the 3-body and 4-body terms exceeds the scaling of the MB-pol potential with increased system sizes. This idea has been discussed previously,[201] demonstrating that the scaling of the calculation method must exceed that of the many-body term in order to observe a computational speed-up. That said, the use of gold-standard electronic structure methods, which scale $\mathcal{O}(N^5) - \mathcal{O}(N^7)$, with the introduced method would significantly speed up the calculation, in comparison to the full calculation with PBCs.

The excellent agreement between the two methods (MBE vs. PBCs with Ewald summation) suggests that the MBE for ice converges at the 4-body term, like that of water clusters. The total 4-body contribution to the lattice energy estimates ranges from ∼0.5-5.7% for the 7 ice phases examined here. We did not consider a 5-body term calculation as an accurate estimate of this would likely be difficult to obtain. Because of the large number of 5-body terms and the small magnitude of the 5-body energies, an attempt to compute the total 5-body contribution would likely be thwarted by numerical instability issues, as previously discussed by Herbert and co-workers.[194, 288, 289]

The many-body decomposition of the lattice energies provides additional insight into the nature of the stabilizing forces for the various polymorphs of ice. It was noted by Pham et al. that removing the 2B dispersion and short-range 3B components of the MB-pol potential caused the relative stabilities of the polymorphs to change.[5] In this study, the many-body contributions to the lattice energy were quantified for the first time. It was found that the many-body (3B+) interactions varied significantly, contributing between 7-24% to the total lattice energies. This further informs our understanding of hydrogen bonding and differences in the nature of the stabilizing interactions for the ice phases across the phase diagram.

In addition, it is possible that this method could be used to decompose other properties of solid-state systems. For example, the total atomic gradients can be approximated from the 1-body, 2-body, 3-body, and 4-body contributions to the atomic gradients (as previously demonstrated),[201] making it possible to drive dynamics or optimize a unit cell structure.

Furthermore, using this scheme, it is, in principle, possible to obtain higher order derivatives (using finite difference) which could be used to calculate phonons of the crystal. However, this functionality has not been implemented.

6.4 Conclusions

In conclusion, the extension of the MBE from finite to infinite periodic systems was introduced for the first time. We illustrated the generalizability of this formulation in describing the MBE for systems of arbitrary size, spanning from finite supercells to infinite periodic systems. This established connection helped to validate the ability of the method to recover solid-state properties such as the lattice energy. Importantly, only a fraction of the total possible many-body terms is necessary to compute. These are determined solely through the inherent translational symmetry of the periodic systems. The absence of screening protocols in the proposed method could prove useful in applying it to optimize unit cell structures and/or drive dynamics. In addition, this method can be applied to achieve high-level lattice energy estimates using a hybrid electronic structure approach, which has already been applied to crystalline benzene using a different strategy for the MBE of periodic systems.18 The decomposition of molecular properties into the ones of smaller subsystems effectively combats the poor scaling of electronic structure methods by computing many independent calculations for smaller systems, a trivially parallelizable process that can be used to efficiently use hardware architectures with hundreds of thousands of processors. Lastly, this method opens the door for calculations that are less compatible with standard PBCs, which rely heavily on the perfect periodicity of a system. For example, since the subsystem calculations are performed independently, it is possible to compute the formation energy of a single point defect without the interactions with its periodic images. We believe that the strengths of the proposed method have the potential to fill in gaps in other approaches used for solid-state, crystalline systems.

Chapter 7

ORIGINS OF TETRAHEDRAL ORDER IN ICE

7.1 Introduction

Ice has attracted significant attention from the scientific community due to its prevalence on Earth (in ice sheets, permafrost, snowflakes), its anomalous properties,[290, 291] and the possible existence of different ice polymorphs on planets that are stable at other temperatures and pressures.[292, 293] Currently, there are 20 experimentally observed polymorphs of ice existing at different pressures and temperatures (Ih, Ic, II-XIX)[269, 271, 271, 273, 276, 279, 280, 294–299] not including numerous proposed/hypothetical polymorphs only realized in simulations.[300–306] A commonality of all polymorphs is that they obey the Bernal-Fowler ice rules,[285] meaning that they accept (A) two and donate (D) two hydrogen bonds from nearby water molecules, leading to the AADD adjacency. The natural exception to this rule is the existence of defects for which the Bernal-Fowler rules are violated (vacancies, Bjerrum defects, ionic point defects, etc.).[307–313] Furthermore, these polymorphs consist of both proton-disordered phases (with an entropic contribution from Pauling's residual entropy[314] to the Gibbs energy) and their low-temperature proton-ordered analogs. Altogether, this results in the solid portion of the H_2O phase diagram being a complex array of diverse lattice structures and hydrogen bonding configurations. It is anticipated that more stable ice configurations could be discovered through the study of isomorphic zeolite structures[302, 315] and the use of other theoretical methods (such as quantum Monte Carlo,[306, 316] molecular dynamics,[317] basin-hopping methods with classical potentials,[304, 318] and first-principles phase diagram calculations[305] to name a few), which all focus on predicting thermodynamically stable configurations of ice.[319] This detailed characterization of the ice phase diagram sparks interest in understanding the origin of the physical interactions that drive the differing stabilities of the various ice phases and are, thus, responsible for the complex layout of the ice phase diagram. A more comprehensive understanding of the underlying physics governing the fundamental interactions of the ice phases existing across different temperature and pressure regimes in the phase diagram would contribute to our foundational understanding of hydrogen bonding in ice at the molecular level and can further aid in the discovery of additional ice phases.

In this study our aim is to unravel the underlying factors responsible for the structural features, such as the tetrahedral order, in various ice polymorphs across the intricate phase diagram. We achieve this by breaking down the nature and strength of intermolecular interactions, which are the determining factors in shaping the arrangement of water molecules in the diverse ice lattices observed within the phase diagram. For this purpose, we rely on applying the many-body expansion (MBE) of the energy of a hydrogen bonded ice lattice that casts the total interaction as the sum of one- to n-body terms, where n is the size of the system, to investigate their differing magnitudes across the ice phase diagram and correlate them with the structural motifs that are characteristic of the respective structures.

The MBE of the energy of a finite cluster of water molecules is straightforward and it has been previously extensively used to show that it converges to low order, namely the 4-body term.[2,94,320] Its formal extension for a periodic system has only recently been reported,[321] allowing, for the first time, a thorough account of the non-additive effects of the lattice energies for the various ice polymorphs and their connection to the local structure. Similarly to the finite cluster case, the lattice energy, E_{latt},[321] can be cast (equation 1) as the sum of contributions from the deformation energy (1-body energy), the pairwise interactions (2-body energy), and the cooperative/nonadditive interactions (3- and 4-body), which have been found to originate almost entirely from many-body induction,[44,45] as

$$E_{latt} \simeq E_{1B} + E_{2B} + E_{3B} + E_{4B} \tag{7.1}$$

While the pairwise (2-body) interactions in water clusters constitute the largest component of the lattice energy, there are significant contributions from nonadditive interactions,[2,94] a fact that has fueled the development of a series of accurate water potentials aimed at closely reproducing the individual many-body terms.[29,45–47,49,50,88,204,322] In the following we examine the relationship between the many-body non-additive terms and various properties of the ice polymorphs such as the density, the local tetrahedral order,[323,324] and the hydrogen bonded cycles in their extended network. We considered nearly all known proton-ordered polymorphs, viz. the ice II, VIII, IX, XIII, XIV, and XV phases, along with the proton disordered ice Ih (snow, ice on Earth), which exist between 0-10 GPa in the phase diagram (see Figure 7.1). The polymorphs are grouped based on the pressure domains at which they exist, see recent depictions of the ice phase diagram,[325,326] and they were selected in order to represent a wide variety of lattice structures across the phase diagram. Ultimately, we aim to unveil the factors contributing to the different stabilities of the ice polymorphs across the phase diagram and how their local and extended hydrogen bond network influence this variation, thus shedding light on the structure-energy relationship that governs the complex phase diagram of ice. We expect these results to aid in the future discovery of stable polymorphs of ice while improving our fundamental understanding of hydrogen bonding and cooperative effects across the ice phase diagram.

7.2 Computational details

The contributions of the various terms of the MBE to the lattice energy were obtained using a method described in previous work.[321] This method computes the many-body terms in real-space (without using the minimum image convention), enumerating only the unique subsystems according to the inherent translational symmetry of the periodic system. The convergence of the MBE terms was monitored with respect to the cutoff distance with the uncertainty assigned based on the fluctuations in the estimates. We ensured that the MBE for periodic systems up to the 4-body term (equation 7.1) matched the value of the lattice energy obtained using conventional methods with periodic boundary conditions (with an Ewald summation) with a maximum error of $< 2\%$. All potential energy calculations were performed with the MB-pol potential.[46,47]

The orientational tetrahedrality order parameter (q) was computed following the method outlined by Errington and Debenedetti[324] (equation 7.2), which was modified from the original formulation introduced by Chau and Hardwick:[323]

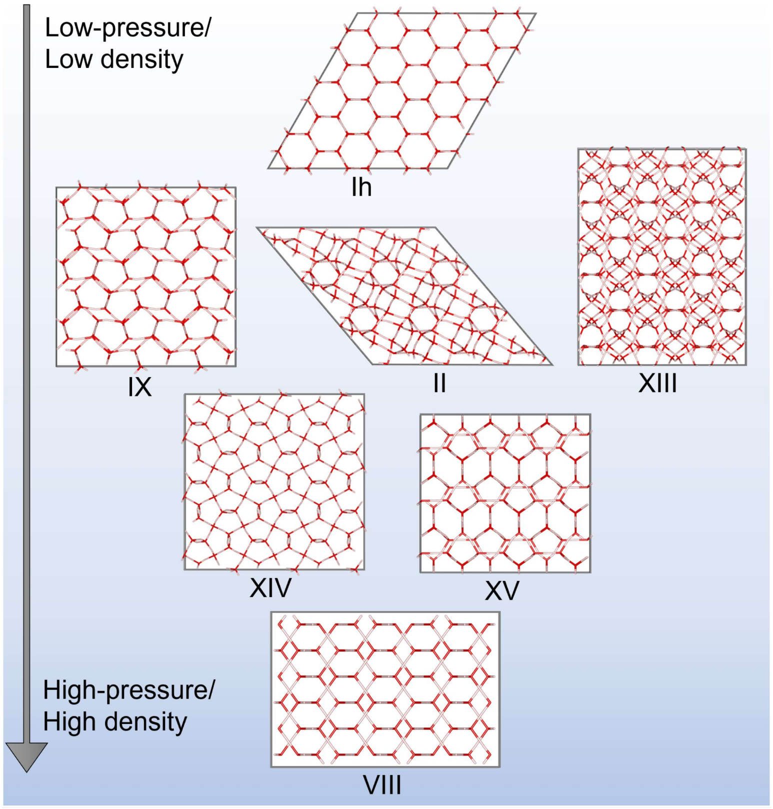


Figure 7.1: The ice polymorphs considered in this study. The polymorphs are grouped based on the pressures at which they are the thermodynamically most stable form of ice (low-pressure to high-pressure).

$$q = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}(\cos\phi_{jk} + \frac{1}{3})^2 \tag{7.2}$$

The tetrahedrality was computed for each molecule in the unit cell, where ϕ_{jk} is the O$\cdots$O$\cdots$O angle between the selected molecule and each set of its two nearest neighbors, indices j and k. The local tetrahedral order was computed using both the 4 nearest neighbor (q_{NN}) oxygen atoms (indices j and k) and also using the 4 oxygen atoms of the water molecules participating in a conventional hydrogen bond (q_{HB}), for each molecule in the unit cell. We used the definition of a hydrogen bond by Kumar et al., which utilizes r and Φ (r: intermolecular O$\cdots$H distance, Φ: angle between the intermolecular O$\cdots$H vector and the vector normal to the plane containing the molecule accepting the hydrogen bond) as descriptors to determine whether molecules are participating in a hydrogen bond.[327]

The non-short-circuited cycles were extracted from the graph structures of the ice polymorphs. The nodes of the graph represent an entire molecule and the directed edges represent the direction of the hydrogen bond. The hydrogen bonds were determined using the r-Φ method outlined by Kumar et al.[327] We note that we consider cycles to be short-circuited if 1+ nodes connect any nodes in the cycle. If more than 1 node short circuits a cycle, the minimal set of cycles are kept (to avoid counting cycles that encompass 2+ smaller cycles). The cycle sizes found are in close agreement with those reported by Salzmann et al.[326] However, we find the n = 10 cycle in ice II and the n = 10, 12 cycles in ice XIII to be short-circuited. The cycles were classified as homodromic (all molecules are acceptor-donor or AD), antidromic (one molecule is AA, one molecule is DD, and the rest are AD), heterodromic (all cycles that are not homodromic or antidromic). Examples of the connectivity of homodromic, antidromic, and heterodromic cycles of 6 molecules are shown in Figure 7.4.

7.3 Results and discussion

The MBE (1- through 4-body) components of the lattice energy, evaluated at the experimental density for the 7 ice lattice structures obtained from a previous work,[321] are listed in Table 7.1. The reference lattice energy values E_{latt} (Exp.) in Table 7.1 for ice Ih, II, VIII, and IX correspond to estimates at zero temperature and pressure and without zero-point energy corrections reported by Whalley[328] that were deduced from their experimental values, even though these ice phases are thermodynamically most stable at different temperatures and pressures. Therefore, the comparison between these values and the calculated ones is direct. The temperatures at which the densities of the ice polymorphs were obtained experimentally in previous works are noted in Table 7.1. As expected, the 1-body term is repulsive (positive) as it corresponds to the deformation energy of the water monomer from its gas phase geometry. Both the 2- and 3-body terms are attractive (negative values), whereas the 4-body term is small (0.5% – 1.5% of E_{latt}) and attractive for ice Ih, II and IX but of the same magnitude and repulsive for ice XIII, XIV, XV. Ice VIII stands out as an exception with a larger repulsive (positive) 4-body term that accounts for 5.7% of E_{latt} and will be discussed separately below. In the following we seek to identify correlations between the structure and the non-additive components of the lattice energy in an effort to identify descriptors that govern the properties of the diverse hydrogen bonded networks observed across the ice phase diagram.

Ice polymorph	Proton-ordered	Exp. density	E_{latt} (Exp.)[328]	E_{latt} (MBE)[321]	E_{1B}[321]	E_{2B}[321]	E_{3B}[321]	E_{4B}[321]	q_{NN}	q_{HB}	Cycle sizes
Ih	No	0.93[329]	-14.07	-13.95 (0.02)	0.75	-11.350 (0.005)	-3.254 (0.003)	-0.09 (0.01)	1.000	1.000	6
IX	Yes	1.17[277]	-13.97	-13.90 (0.08)	0.59	-11.39 (0.02)	-2.98 (0.03)	-0.12 (0.03)	0.859	0.859	5, 7, 8
II	Yes	1.20[328]	-14.05	-14.25 (0.06)	0.49	-11.75 (0.01)	-2.78 (0.01)	-0.21 (0.04)	0.831	0.831	6, 8
XIII	Yes	1.25[279]	–	-13.86 (0.07)	0.55	-11.894 (0.007)	-2.59 (0.03)	0.07 (0.03)	0.798	0.798	4, 5, 6, 8, 9
XIV	Yes	1.29[279]	–	-13.70 (0.08)	0.54	-12.26 (0.02)	-2.15 (0.02)	0.17 (0.04)	0.787	0.787	7, 8
XV	Yes	1.33[280]	–	-13.51 (0.04)	0.50	-12.116 (0.008)	-2.10 (0.01)	0.20 (0.02)	0.691	0.691	4, 8
VIII	Yes	1.49[328]	-13.31	-13.5 (0.2)	0.29	-12.88 (0.02)	-1.6 (0.1)	0.77 (0.09)	0.328	1.000	6

Table 7.1: Experimentally estimated and calculated properties of the ice polymorphs considered in this study. E_{nB} represents the n-body contribution to the lattice energy (kcal/mol) computed through the MBE with the value in parentheses indicating the estimated uncertainty. E_{latt} (MBE) is the lattice energy computed using the sum of the 1-body through 4-body terms in the MBE, eq. (7.1).

We first investigate the relation between non-additivity and crystal structure. Figure 7.2 traces the percentage contribution of the many-body terms (1- through 4-body) to the lattice energy versus the experimental density of each phase, which is related to the pressure at which they are stable. Note that the 1-body term is repulsive (has positive values) and exhibits the largest negative percentage (most repulsive) for the ice polymorphs having the smaller density. Overall, all MBE terms (1- to 4-body) show a remarkable, nearly monotonic variation with the density of the ice polymorph that is either positive (correlation) for the 1- and 2-body or negative (anti-correlation) for the 3- and 4-body. This finding suggests

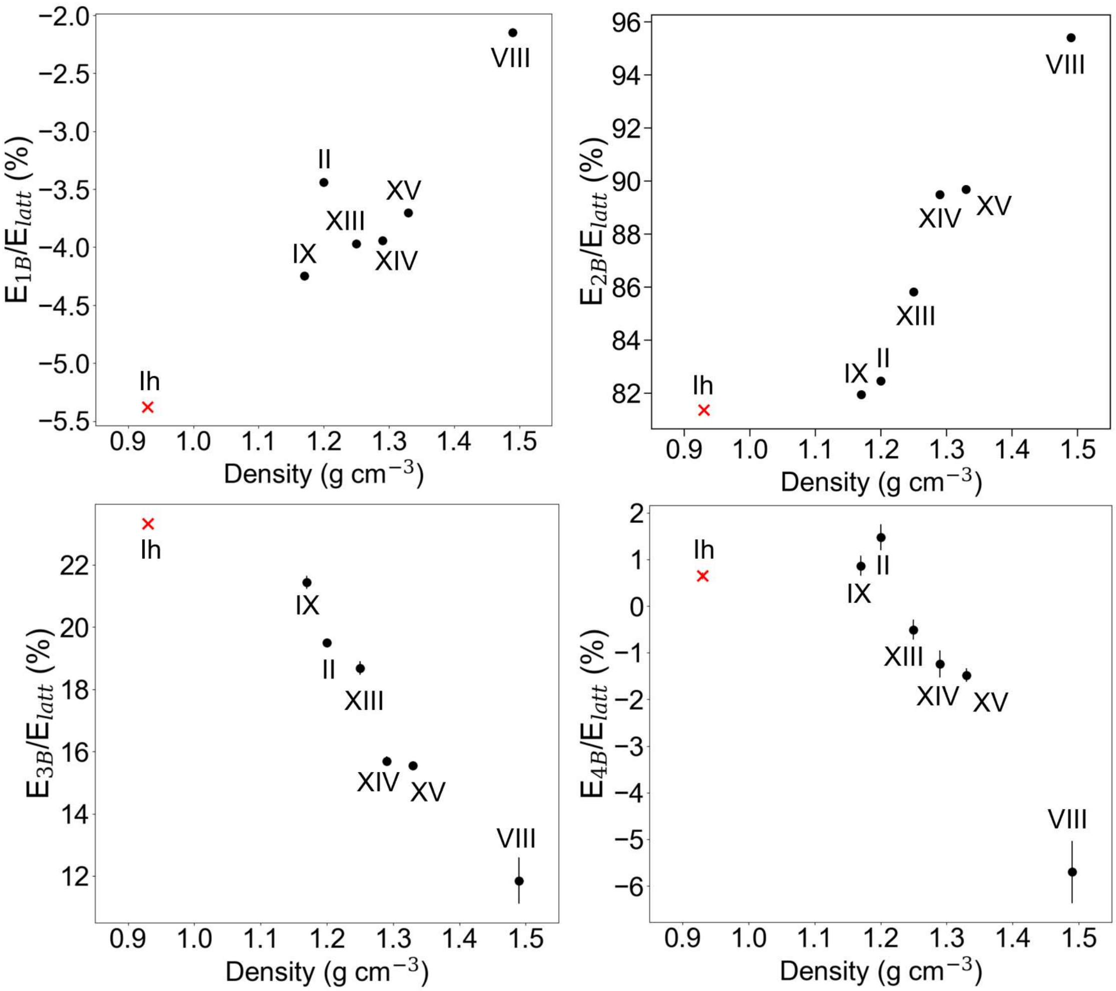


Figure 7.2: The relationship between the percentage contribution of the many-body terms (1- through 4-body) to the lattice energy and the densities of the various ice polymorphs. The results for the only proton-disordered phase considered (ice Ih) are denoted by red x's, whereas the proton-ordered phases are represented by filled black circles.

that the ice polymorphs existing at lower pressures (lower densities) are stabilized more strongly by nonadditive, cooperative interactions. Since the 3- and 4-body interactions in aqueous systems have been shown to arise almost entirely from many-body induction with very small contributions from 3-body dispersion,[44,45] we can therefore ascertain that the low-pressure polymorphs of ice (such as ice Ih, IX, II) are stabilized more strongly by many-body induction than the high-pressure ones (i.e., ice XIV, XV, VIII). These contributions from many-body induction range from 6-25% of the total lattice energy, depending on the polymorph. In addition, the polymorphs exhibiting larger cooperative effects (and smaller pairwise additive contributions) generally have larger lattice energies, a fact that demonstrates the importance of the former (non-additive contributions) in stabilizing the ice forms (cf. Table 7.1). These relationships demonstrate how the nature of the interactions changes as pressure is applied (resulting in an increase in the density and a change in the unit cell structure) and could serve as a guide for the determination of new stable ice forms.

To this end, cooperative effects can be thought of as a descriptor of the density (observable) of ice polymorphs with the two exhibiting a negative correlation (anti-correlation).

The orientational tetrahedrality order parameter[323,324] (q) is also listed in Table 7.1 for the various ice phases. The value of q ranges from 1 (orientation of nearest neighbors around a selected molecule is perfectly tetrahedral, viz. all O$\cdots$O$\cdots$O angles are 109.5°) to 0 for a stochastic distribution of the O$\cdots$O$\cdots$O angles of the nearest neighbor molecules. Note the large range of q values ($0.328 - 1.00$) across the ice phase diagram in Table 7.1. The conventional definition of the q parameter is based on the O$\cdots$O$\cdots$O angle of the oxygen atom of a given molecule with the four oxygen atoms of its nearest neighbors. However, it is also of interest to define an order parameter based on the four molecules that participate in a conventional hydrogen bond (using the r-Φ definition described by Kumar et al.[327]) with the central molecule. We will denote the originally defined tetrahedral order parameter (eq. (7.2), calculated using nearest neighbors) as q_{NN} and the modified tetrahedral order parameter (calculated using molecules participating in conventional hydrogen bonds) as q_{HB}. For nearly all the ice phases considered here q_{NN} and q_{HB} are the same with the sole exception of ice VIII, the highest-pressure phase considered in this work. The discussion regarding the different q values for ice VIII is presented in detail in the next section.

The many-body (3- and 4-body) terms comprising the lattice energies correlate nearly linearly with the respective local tetrahedral order of the ice polymorph (Figure 7.3, left panel). The phases with high tetrahedral order values (q close to 1) are associated with the strongest (in excess of 20%) nonadditive contributions to their lattice energies. While this almost linear correlation between local tetrahedral order and cooperativity is intriguing, the local structure alone is not the driving force for the strong cooperative effects. As it will be

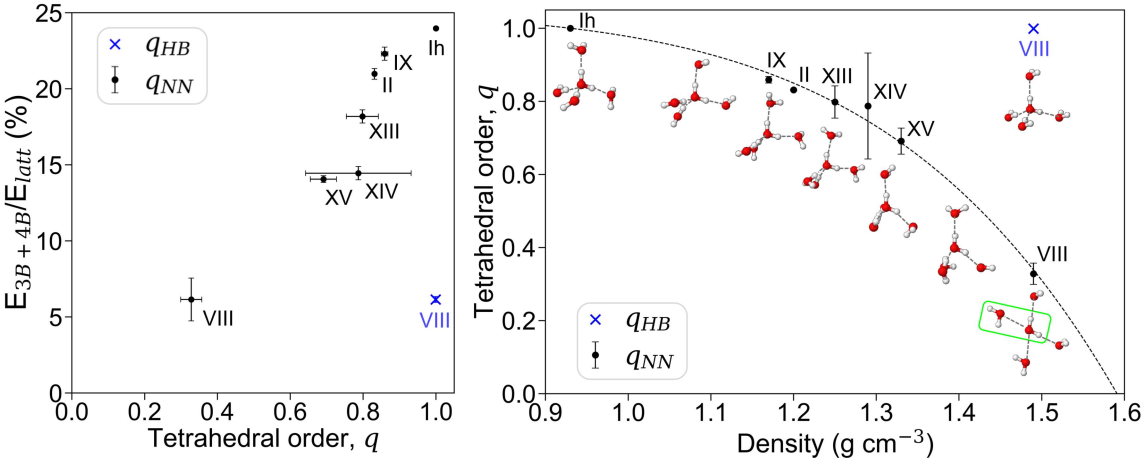


Figure 7.3: (Left) percentage contribution of the 3- and 4-body terms to the lattice energy versus the orientational tetrahedrality order parameter (q); (right) tetrahedral order q as a function of the density of the polymorph. The dimer configuration of C_i symmetry in ice VIII is highlighted with a green border. Solid circles and x's denote the two different definitions of q (see text).

shown later, it is the connectivity determined by the extended hydrogen bond network that plays a critical role in the observed effect. The analysis of the role of the extended hydrogen bond network in inducing cooperativity will be the focus of the subsequent section.

The above results suggest that cooperative effects can be thought of as a descriptor of the tetrahedral order (observable through structure determination) of ice polymorphs with the two exhibiting a positive correlation.

We have so far discussed the correlations between the many body effects with each of the density (negative correlation or anti-correlation) and local tetrahedral order (positive correlation). It stands to reason to examine the correlation between the last two, viz. the density and the local tetrahedral order. Both these two quantities have been experimentally measured, and despite the fact that the data needed to determine the relation between these two are readily available, to the best of our knowledge their correlation has not been reported previously. For supercooled liquid water, a correlation has been reported, albeit at a much smaller range of densities (approx. 0.95-1.08 g cm^{-3}) and q values (approx. 0.5-0.9).[330] Using the available experimental data (with q derived from the unit cell structure that is obtained via X-ray measurements) we construct the right panel of Figure 7.3 showing the relationship between q and the experimental density of the ice polymorphs. Because of the interlocking ring structure of ice VIII, two of the nearest neighbor oxygen atoms are participating in two highly non-linear hydrogen bonds (highlighted in green in Figure 7.3), a configuration that corresponds to a stationary point of C_i symmetry on the water dimer potential energy surface (PES) as described by Burnham and Xantheas,[32] in addition to four conventional, near linear hydrogen bonds (reminiscent of the global minimum on the water dimer PES). Because the molecules in ice VIII each participate in 6 hydrogen bonding interactions, ice VIII is a special case and will be discussed in detail. Examples of the respective tetrahedral Walrafen pentamers[331] for the different ice polymorphs are shown within the right panel in Figure 7.3. Note that some phases have molecules in the unit cell with quite different local tetrahedral orders. For example, ice XIV has some molecules in the unit cell with nearly perfect local tetrahedral geometries and others with low values of tetrahedral order. This variation in q is denoted by the error bars, which represent the standard deviation of the tetrahedral order parameters across the entire unit cell for each ice phase.

The results shown in the right panel of Figure 7.3 demonstrate that as the density of the polymorph increases, the local tetrahedral order, q_{NN} and q_{HB} (with the exception of q_{HB} for ice VIII), decreases, viz. there exists an anti-correlation between the two. That is, while the Bernal-Fowler rules impose that the molecules in ice participate in four conventional hydrogen bond interactions, the $\mathrm{O}\cdots\mathrm{O}\cdots\mathrm{O}$ angles defining the orientation of the molecules around one another can deviate significantly from 109.5°. Intuitively, it makes sense that for the water molecules to stay 4-coordinated under high pressure, the local order must distort. For example, in ice XV, which has the smallest q_{HB} value, the $\mathrm{O}\cdots\mathrm{O}\cdots\mathrm{O}$ angles have values of 129-130° and 91-92°. The only polymorph among the ones considered with a truly perfect tetrahedral local environment is ice Ih, in which all $\mathrm{O}\cdots\mathrm{O}\cdots\mathrm{O}$ angles are $\sim$109.5°. Ice VIII may also be considered perfectly tetrahedral according to the q_{HB} definition (q_{HB}=1.00, see Table 7.1 and blue "x" symbol in Figure 7.3). However, when considering the q_{NN} definition, its local structure deviates significantly from the perfect tetrahedral order (q_{NN}=0.328). The above analysis of the experimental data suggests that the local tetrahedral order in ice is anti-correlated with the density.

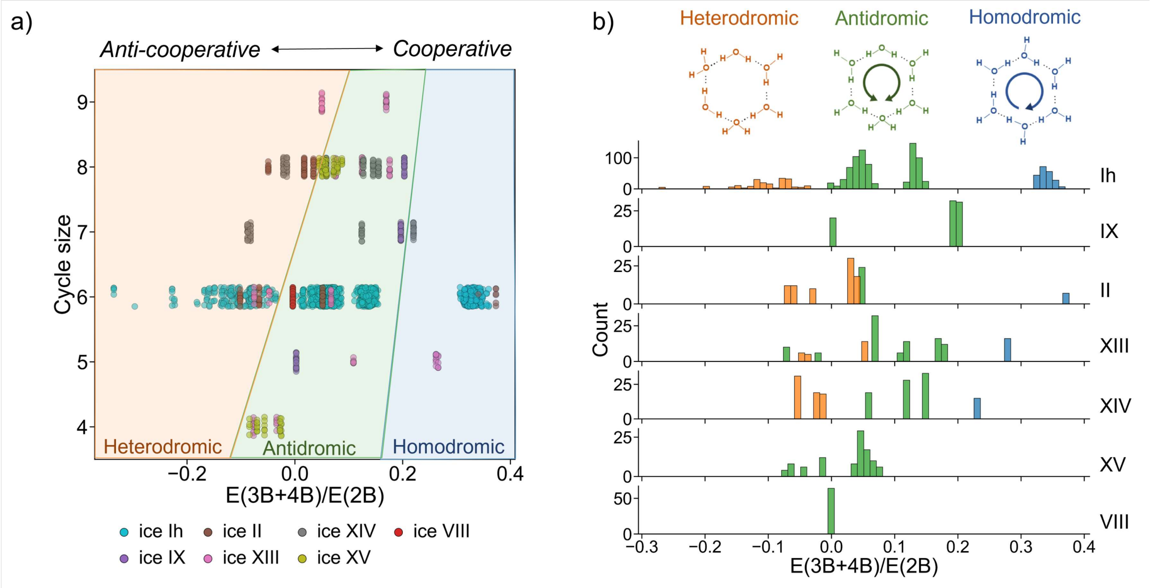


Figure 7.4: (a) The ratio of cooperative effects [E(3B+4B)] to pairwise effects [E(2B)] for heterodromic (left), antidromic (middle), and homodromic (right) hydrogen bonded cycles in the ice polymorphs colored according to the polymorph, (b) The distribution of the ratio among the 3 distinct networks for the various ice phases.

Given the previously discussed correlations / anti-correlations between structural and energetic properties that are found in different ice phases, we seek to further understand and attribute the coupling of these relationships through a common feature of the underlying hydrogen bond network. We consider the hydrogen bonded cycles as the unifying feature between structure and energy correlations due to the previously established strong cooperative effects existing in homodromic rings of gas phase clusters, in which each molecule accepts one and donates one hydrogen bond.[94] The various ice phases have differing numbers of cycle sizes and cycle counts,[326] with the exception of ice Ih and ice VIII, which both contain only 6-membered rings (cf. Table 7.1). As a consequence, the unique local tetrahedral orders (q_{HB}) of various ice polymorphs give rise to differing extended networks. We extracted the hydrogen bond cycles from the extended ice networks and computed the many-body expansion of their binding energy. It should be emphasized that the cycles in the ice lattices differ markedly from the respective gas phase ring local minima, both in terms of binding energies as well as structure / connectivity. Figure 7.4a shows the relative contribution from cooperative effects against the pairwise additive contribution for each of the cycles extracted from the ice phases against the cycle size. These networks are categorized in 3 distinct groups, namely heterodromic, antidromic and homodromic networks. The connectivity of these networks for the n=6 cycles is shown on the top of Figure 7.4b, which enumerates the distribution of the $[E(3B + 4B)]/[E(2B)]$ ratio among the 3 distinct networks for the various ice phases. Large positive values of this ratio indicate strong cooperative (favorable) effects while large negative values indicate strong anti-cooperative (unfavorable) effects for the cycles. Figure 7.4 demonstrates that the $n = 5$, 6 homodromic cycles in ice have the strongest relative cooperative effects. The ice phases that contain these cycles are ice Ih (cyan points), ice II (brown points), and ice XIII (pink points). We also note that the three naturally occurring clathrate hydrate lattices (sI, sII and sH) contain mainly 5- and 6-membered rings.[332,333] The $n = 7$, 8 antidromic cycles of ice IX (purple points) and $n = 7$ homodromic cycles of ice XIV (gray points) have the next largest relative cooperative effects.

It is found that cycles of a particular size n can exhibit a wide range of many-body effects (cf. Figure 7.4a). For example, for n=6 the ratios of cooperative effects to pairwise additive effects span the range of -0.3 to +0.4. This is largely due to the hydrogen-ordering for a particular cycle (i.e., homodromic vs. antidromic vs. heterodromic) and the interatomic O$\cdots$O distances in the cycles. The cooperativity is very sensitive to the O$\cdots$O distance with smaller O$\cdots$O distances being stabilized by stronger cooperative effects. Despite being the highest-pressure phase of ice considered in this study, ice VIII has the largest O$\cdots$O distance between molecules participating in a hydrogen bond, due to the interlocking cubic ice networks. This is why the cooperative effects in ice VIII ($R(O \cdots O) = 2.97$ Å) are much smaller than those in ice Ih ($R(O \cdots O) = 2.75$ Å) despite having the same q_{HB} and consisting of solely 6-membered hydrogen bonded cycles.

Figure 7.4b demonstrates that the homodromic cycles (blue) have the largest many-body effects, as already established in the literature,[94] followed by antidromic cycles (green) and then heterodromic cycles (orange) which tend to have weak or repulsive many-body effects (anti-cooperative). The ice polymorphs have either only antidromic cycles or a mixture of homodromic, antidromic, and heterodromic cycles. For example, for the proton-disordered supercell ($n = 324$) of ice Ih with a net zero dipole moment, the cycles consist of 18.0% homodromic, 64.7% antidromic, and 17.3% heterodromic arrangements. Ice II, XIII, and XIV also have a mixture of homodromic, antidromic, and heterodromic cycles whereas ice IX, XV, and VIII have only antidromic cycles. In addition, the low-pressure phases (top of Figure 7.4b) exhibit cycles with much stronger relative many-body effects in comparison to the high-pressure phases (bottom of Figure 7.4b), supporting the idea that the cooperativity of the hydrogen bonded cycles in an ice phase is indicative of the strength of cooperative effects contributing to the total lattice energy.

The above analysis demonstrates the connection between the local tetrahedral order and the resulting cycles in the extended hydrogen bond environment, with the highly tetrahedral local environments having many 5- and 6-membered rings. The cycles, in turn, exhibit varying contributions from cooperative effects with 5- and 6-cycles allowing for the greatest many-body contributions. As already established in the literature, homodromic rings have notably strong cooperative effects and are found to exist in ice Ih, II, XIII, and XIV. Ultimately, this demonstrates the connection between the local tetrahedral order, the cycles that it gives rise to, and the resulting impact on the magnitude of the many-body interactions.

7.4 Conclusions

The MBE analysis of the lattice energies of various ice polymorphs has established quantitative relationships between the contributions to the lattice energy from many-body induction (3-body and 4-body components of the lattice energy) and the density (pressure), at which the phase exists, as well as the local tetrahedral order. Cooperative effects are crucial in the stabilization of the ice polymorphs, contributing almost 25% of the total lattice energy for the most stable ice phases (ice Ih/XI) but only a mere 6% for the least stable ice phase considered (ice VIII), despite the latter having the strongest 2-body component of the lattice energy out of all the ice polymorphs. Nonadditive effects show a positive correlation with the local tetrahedral order (correlation) but a negative one with the density of the ice phase (anti-correlation). Additionally, the density and local tetrahedral order exhibit a negative correlation (anti-correlation). While only requiring the crystal structure to compute the tetrahedral order, to the best of our knowledge the anti-correlation between the local tetrahedral order and density has not been previously reported in the literature for ice.

The finding that cooperative effects were the largest for structures with a nearly perfect local tetrahedral order (as in ice Ih/XI), containing a large number of 6-membered hydrogen-bonded cycles, provides a better understanding of the connection between structure and stability. This lends insight into the physical origin for the nearly perfect tetrahedral structure in the most stable phase of ice (Ih/XI), which has not been previously discussed. The interlocking cubic ice networks in ice VIII result in longer $O\cdots O$ distances between near linear hydrogen bonds, which greatly reduces the nonadditive effects for these networks, despite containing only 6-membered rings in its hydrogen bond network. Because of this, the two definitions of local tetrahedral order are quite dissimilar, viz. q_{HB}=1 and q_{NN}=0.3. In addition, ice VIII could arguably be considered 6-coordinated (rather than 4-coordinated) since it participates in two additional interactions (C_i dimer[32]) with 2 nearest neighbor molecules in the opposite hydrogen bonded network. These additional hydrogen bonded interactions strengthen the 2-body component of the lattice energy for ice VIII.

Ultimately, the 3-way relationship between the density of the polymorph, the magnitude of cooperative effects, and the local tetrahedral order provides rationale for the structure-stability relationship that governs the ice phase diagram, as shown in Figure 7.5. The hydrogen bonded cycles in the ice phase demonstrate the connection between the local tetrahedral structure and the resulting cooperative effects, with 6-cycles maximizing the cooperativity (as in ice Ih/XI). The most common in everyday life, ice Ih, exhibits near perfect tetrahedral order and the largest cooperative effects that are induced by the hydrogen bonded cycles.

Although assigning causality between local structure and non-additive effects seems ambiguous, energetic stability (more precisely, thermodynamic stability) determines the stable polymorph at a particular pressure and temperature. While it is possible for additional, higher energy structures to be formed, these represent metastable states rather than the true minimum energy structure. Therefore, we reason that the strong cooperative effects (associated with the most stable lattice energies) induce the near perfect tetrahedral order in

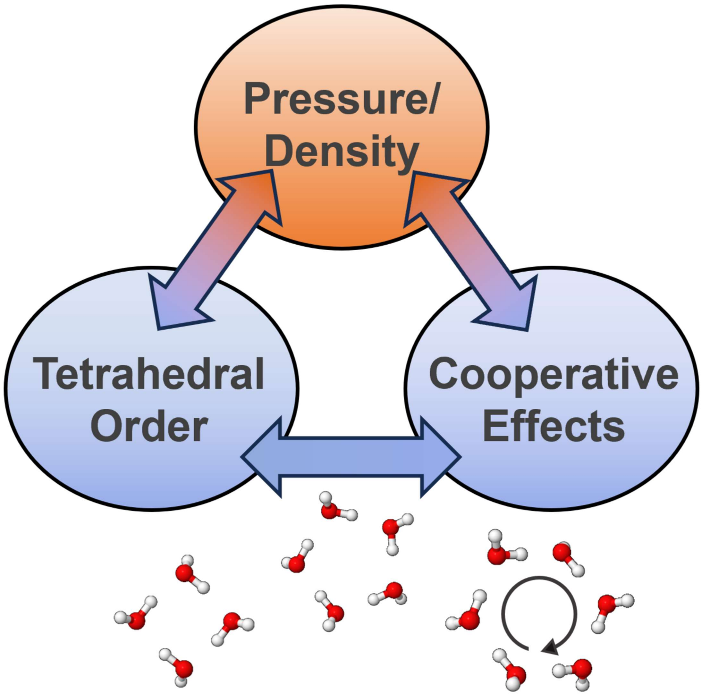


Figure 7.5: Correlation (blue arrows) and anti-correlation (orange/blue arrows) between the pressure/density of the various ice phases, the corresponding local tetrahedral order, and the cooperative (non-additive, 3- and 4-body) effects. The last two are correlated via hydrogen bonded cycle size and connectivity.

the structure of ice Ih. The strong cooperative effects also stabilize shorter $O \cdots O$ distances. Without this observed effect, ice Ih would exist at an even smaller density (larger volume). We expect that these findings will inform the discovery of additional stable polymorphs of ice while also offering a new perspective on hydrogen bonding patterns and their stabilities across the ice phase diagram.

Chapter 8

CONCLUSIONS AND OUTLOOK

The preceding chapters demonstrated the usefulness of the MBE in understanding the nature of interactions and also guiding intermolecular potential development. The vast majority of the work was applied to aqueous systems and ice, which have strong many-body interactions. Aqueous ionic clusters exhibited strong many-body interactions that were sensitive to the location and the identity of the ion. While this alone presents a challenge for interaction potential development, it was found that a classical induction and dispersion model accurately described the 3-body and higher interactions for pure water and ion-water systems. For aqueous systems, it was found that induction was the largest contributor to many-body effects (up to -8 kcal/mol) while dispersion contributed only a small amount (up to 0.2 kcal/mol). This is due to water's large molecular dipole moment, which strongly affects the charge distribution of nearby ions and molecules.

The model describing the interaction between pure water systems was improved by developing distributed multipole and distributed polarizability surfaces, which allowed the representation of the charge distribution and polarizability to change as a function of water's intramolecular degrees of freedom. While this was shown to be important for the accuracy in the case of water, it may vary for other solvents depending on the amount of their intramolecular distortion. A major benefit of the induction model is that it is transferable, only relying on parameters obtained from cheap *ab initio* calculations of the isolated fragments. Machine learning-based potentials, on the other hand, fit the many-body terms as a function of interatomic distances. That said, the higher order many-body terms pose a significant challenge for these methods due to the increasing number of degrees of freedom to sample and computational cost of obtaining the high-fidelity data. To this end, classical models based on an induction scheme can greatly simplify the development of highly accurate potential energy surfaces for molecular systems by providing accurate estimates for the

3- and higher many body terms.

While the application of the MBE to finite systems is well-defined, its extension to periodic systems is less straightforward. The MBE was formulated for periodic systems by systematically enumerating the translationally-unique subsystems. This method was applied to 7 ice polymorphs, demonstrating excellent agreement with results obtained using periodic boundary conditions (PBCs) with the minimum image convention. The polymorphs of ice presented a challenging test case due to the long range dipole-dipole interactions that, in turn, induced larger many-body effects. Furthermore, the many-body components of the lattice energy provided insight into the nature of the stabilizing interactions. It was found that the ice polymorphs were stabilized by differing amounts due to many-body induction. The polymorphs existing at low-densities (low experimental pressures) were shown to exhibit the largest 3- and 4-body interactions and, thus, were stabilized mainly by many-body induction. For ice Ih, the cooperative effects (3-body + 4-body) contributed nearly 25% to the total lattice energy. Conversely, for the the high-density (high experimental pressure) polymorphs, the cooperative effects accounted for only about 6% of the total lattice energy. Furthermore, polymorphs that exhibited a perfectly tetrahedral local order were stabilized most strongly by cooperative effects. These structures also contained more 5- and 6-membered homodromic hydrogen bond cycles, which contribute strongly to this effect. Ultimately, a 3-way relationship was established connecting the density (experimental pressure) of ice, the local and extended structure, and the cooperativity of the stabilizing forces. Importantly, this connects experimental conditions with the resulting structure of the polymorph and the nature of the stabilizing interactions, providing valuable insight into the organization of the ice phase diagram.

The extension of the MBE to periodic systems introduces a framework for accelerating the calculation of accurate lattice energies of molecular crystals using high level *ab initio* methods. Ongoing work is focused on applying this framework to numerous molecular crystals. As was shown in previous work by Heindel and Xantheas,[201] the scaling of *ab initio* methods can be circumvented to some extent by calculating the energies and gradients of small subsystems, rather than the full system. The MBE was employed in a previous study[201] to obtain the n-body components of the forces. In that work, it was shown that the

1- through 4-body contributions to the intermolecular forces sufficiently recovered dynamical and spectral features of the cluster (in comparison to the full calculation). Current work involves applying similar ideas to molecular crystals in the DMC-ICE13[334] and X23[335–337] datasets using the MBE for periodic systems. The X23 dataset is a particularly challenging application due to the large molecule sizes (i.e., anthracene with 14 carbon atoms).

The above work informs our understanding of the nature of intermolecular interactions, providing a pathway for the development of low-cost (yet highly accurate) interaction potentials. In addition, it can help to accelerate the calculation of CCSD(T)-level lattice energies directly by leveraging the MBE to expand the many-body contributions to the correlation energy. This method has the potential to achieve a speed-up of several orders of magnitude in comparison to the conventional full system calculation, which is currently unattainable. As a whole, this work helps to close the gap in system sizes that can be accessed with a high-level of accuracy, thus providing confidence in computational predictions of materials and their properties.

BIBLIOGRAPHY

[1] J.P. Heindel and S. S. Xantheas. The Many-Body Expansion for Aqueous Systems Revisited: II. Alkali Metal and Halide Ion–Water Interactions. *J. Chem. Theory Comput.*, 17(4):2200–2216, 2021.

[2] J. P. Heindel and S. S. Xantheas. The Many-Body Expansion for Aqueous Systems Revisited: I. Water–Water Interactions. *J. Chem. Theory Comput.*, 16(11):6843–6855, 2020.

[3] K. M. Herman, J. P. Heindel, and S. S. Xantheas. The many-body expansion for aqueous systems revisited: III. Hofmeister ion–water interactions. *Phys. Chem. Chem. Phys.*, 23(19):11196–11210, 2021.

[4] K. Momma and F. Izumi. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Cryst.*, 44(6):1272–1276, 2011.

[5] C. H. Pham, S. K. Reddy, K. Chen, C. Knight, and F. Paesani. Many-Body Interactions in Ice. *J. Chem. Theory Comput.*, 13(4):1778–1784, 2017.

[6] M. Riera, A. W. Götz, and F. Paesani. The i-TTM model for ab initio -based ion–water interaction potentials. II. Alkali metal ion–water potential energy functions. *Phys. Chem. Chem. Phys.*, 18(44):30334–30343, 2016.

[7] J. Řezáč, D. Nachtigallová, F. Mazzoni, M. Pasquini, G. Pietraperzia, M. Becucci, K. Müller-Dethlefs, and P. Hobza. Binding energies of the -stacked anisole dimer: New molecular beam—laser spectroscopy experiments and ccsd(t) calculations. *Chem. Eur. J.*, 21:6740–6746, 2015.

[8] K. M. Herman, E. Aprà, and S. S. Xantheas. A critical comparison of ch versus interactions in the benzene dimer: obtaining benchmarks at the ccsd(t) level and assessing the accuracy of lower scaling methods. *Phys. Chem. Chem. Phys.*, 25:4824–4838, 2023.

[9] V. S. Thimmakondu and A. Karton. Ccsd(t) rotational constants for highly challenging c5h2 isomers—a comparison between theory and experiment. *Molecules*, 28:6537, 2023.

[10] X. Liu, L. McKemmish, and J. Pérez-Ríos. The performance of ccsd(t) for the calculation of dipole moments in diatomics. *Phys. Chem. Chem. Phys.*, 25:4093–4104, 2023.

[11] C. W. Bauschlicher Jr., S. R. Langhoff, P. R. Taylor, N. C. Handy, and P. J. Knowles. Benchmark full configuration-interaction calculations on hf and nh2. *J. Chem. Phys.*, 85:1469–1474, 1986.

[12] R. J. Harrison and N. C. Handy. Full ci results for be2 and (h2)2 in large basis sets. *Chem. Phys. Lett.*, 98:97–101, 1983.

[13] J. Olsen, P. Jorgensen, H. Koch, A. Balkova, and R. J. Bartlett. Full configuration–interaction and state of the art correlation calculations on water in a valence double-zeta basis with polarization functions. *J. Chem. Phys.*, 104:8007–8015, 1996.

[14] J. P. Perdew and K. Schmidt. *Density functional theory and its application to materials: Antwerp, Belgium, 8-10 June 2000.* Number v. 577 in AIP conference proceedings. American Institute of Physics, Melville, N.Y, 2001.

[15] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka. Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. *J. Chem. Phys.*, 123(6):062201, 2005.

[16] T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90(2):1007–1023, 1989.

[17] D. E. Woon and T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.*, 98(2):1358–1371, 1993.

[18] D. E. Woon and T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. v. core-valence basis sets for boron through neon. *J. Chem. Phys.*, 103:4572–4585, 1995.

[19] T. H. Dunning, K. A. Peterson, and A. K. Wilson. Gaussian basis sets for use in correlated molecular calculations. x. the atoms aluminum through argon revisited. *J. Chem. Phys.*, 114:9244–9253, 2001.

[20] A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. IX. The atoms gallium through krypton. *J. Chem. Phys.*, 110(16):7667–7676, 1999.

[21] B. P. Prascher, D. E. Woon, K. A. Peterson, T. H. Dunning, and A. K. Wilson. Gaussian basis sets for use in correlated molecular calculations. vii. valence, core-valence, and scalar relativistic basis sets for li, be, na, and mg. *Theor. Chem. Acc.*, 128:69–82, 2011.

[22] T. Van Mourik and T. H Dunning. Gaussian basis sets for use in correlated molecular calculations. viii. standard and augmented sextuple zeta correlation consistent basis sets for aluminum through argon. *Int. J. Quantum Chem.*, 76:205–221, 2000.

[23] A. K. Wilson, T. van Mourik, and T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. vi. sextuple zeta correlation consistent basis sets for boron through neon. *J. Mol. Struct.*, 388:339–349, 1996.

[24] T. H. Dunning. A Road Map for the Calculation of Molecular Binding Energies. *J. Phys. Chem. A*, 104(40):9062–9080, 2000.

[25] K. A. Peterson, D. E. Woon, and T. H. Dunning. Benchmark calculations with correlated molecular wave functions. IV. The classical barrier height of the H+H2→H2+H reaction. *J. Chem. Phys.*, 100(10):7410–7415, 1994.

[26] E. Miliordos and S. S. Xantheas. An accurate and efficient computational protocol for obtaining the complete basis set limits of the binding energies of water clusters at the MP2 and CCSD(T) levels of theory: Application to $(H_2O)_m$, m = 2-6, 8, 11, 16, and 17. *J. Chem. Phys.*, 142(23):234303, 2015.

[27] E. Miliordos and S. S. Xantheas. On the validity of the basis set superposition error and complete basis set limit extrapolations for the binding energy of the formic acid dimer. *J. Chem. Phys.*, 142:094311, 2015.

[28] J. G. Hill, K. A. Peterson, G. Knizia, and H. Werner. Extrapolating mp2 and ccsd explicitly correlated correlation energies to the complete basis set limit with first and second row correlation consistent basis sets. *J. Chem. Phys.*, 131:194105, 2009.

[29] K. M. Herman and S. S. Xantheas. An extensive assessment of the performance of pairwise and many-body interaction potentials in reproducing ab initio benchmark binding energies for water clusters n = 2 − 25. *Phys. Chem. Chem. Phys.*, 25:7120–7143, 2023.

[30] G. S. Fanourgakis and S. S. Xantheas. The Flexible, Polarizable, Thole-Type Interaction Potential for Water (TTM2-F) Revisited. *J. Phys. Chem. A*, 110(11):4100–4106, 2006.

[31] G. S. Fanourgakis and S. S. Xantheas. Development of transferable interaction potentials for water. V. Extension of the flexible, polarizable, Thole-type model potential (TTM3-F, v. 3.0) to describe the vibrational spectra of water clusters and liquid water. *J. Chem. Phys.*, 128(7):074506, 2008.

[32] C. J. Burnham and S. S. Xantheas. Development of transferable interaction models for water. I. Prominent features of the water dimer potential energy surface. *J. Chem. Phys.*, 116:1479, 2002.

[33] C. J. Burnham, D. J. Anick, P. K. Mankoo, and G. F. Reiter. The vibrational proton potential in bulk liquid water and ice. *J. Chem. Phys.*, 128(15):154519, 2008.

[34] C. J. Burnham, J. Li, S. S. Xantheas, and M. Leslie. The parametrization of a Thole-type all-atom polarizable water model from first principles and its application to the study of water clusters (n = 2–21) and the phonon spectrum of ice Ih. *J. Chem. Phys.*, 110(9):4566–4581, 1999.

[35] S. S. Xantheas, C. J. Burnham, and R. J. Harrison. Development of transferable interaction models for water. II. Accurate energetics of the first few water clusters from first principles. *J. Chem. Phys.*, 116:1493, 2002.

[36] C. J. Burnham and S. S. Xantheas. Development of transferable interaction models for water. iii. reparametrization of an all-atom polarizable rigid model (ttm2–r) from first principles. *J. Chem. Phys.*, 116:1500–1510, 2002.

[37] C. J. Burnham and S. S. Xantheas. Development of transferable interaction models for water. iv. a flexible, all-atom polarizable potential (ttm2-f) based on geometry dependent charges derived from an ab initio monomer dipole moment surface. *J. Chem. Phys.*, 116:5115–5124, 2002.

[38] P. Ren and J. W. Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B*, 107:5933–5947, 2003.

[39] M. L. Laury, L. Wang, V. S. Pande, T. Head-Gordon, and J. W. Ponder. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B*, 119(29):9423–9437, 2015.

[40] F. Wang, M. J. Deible, and K. D. Jordan. Benchmark study of the interaction energy for an (H2O) 16 cluster: Quantum Monte Carlo and complete basis set limit MP2 results. *J. Phys. Chem. A*, 117(32):7606–7611, 2013.

[41] C. Liu, J. Piquemal, and P. Ren. AMOEBA+ Classical Potential for Modeling Molecular Interactions. *J. Chem. Theory Comput.*, 15(7):4122–4139, 2019.

[42] C. Liu, J. Piquemal, and P. Ren. Implementation of Geometry-Dependent Charge Flux into the Polarizable AMOEBA+ Potential. *J. Phys. Chem. Lett.*, 11(2):419–426, 2020.

[43] A. K Das, L. Urban, I. Leven, M. Loipersberger, A. Aldossary, M. Head-Gordon, and T. Head-Gordon. Development of an Advanced Force Field for Water using Variational Energy Decomposition Analysis. *J. Chem. Theory Comput.*, 15(9):5001–5013, 2019.

[44] K. M. Herman, A. J. Stone, and S. S. Xantheas. A classical model for three-body interactions in aqueous ionic systems. *J. Chem. Phys.*, 157(2):024101, 2022.

[45] K. M. Herman, A. J. Stone, and S. S. Xantheas. Accurate Calculation of Many-Body Energies in Water Clusters Using a Classical Geometry-Dependent Induction Model. *J. Chem. Theory Comput.*, 19(19):6805–6815, 2023.

[46] V. Babin, C. Leforestier, and F. Paesani. Development of a "First Principles" Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. *J. Chem. Theory Comput.*, 9(12):5395–403, 2013.

[47] Volodymyr Babin, Gregory R Medders, and Francesco Paesani. 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.*, 10(4):1599–1607, 2014.

[48] G. R. Medders, V. Babin, and F. Paesani. Development of a "First-Principles" Water Potential with Flexible Monomers. III. Liquid Phase Properties. *J. Chem. Theory Comput.*, 10(8):2906–10, 2014.

[49] Y. Wang, X. Huang, B. C. Shepler, B. J. Braams, and J. M. Bowman. Flexible, ab initio potential, and dipole moment surfaces for water. I. Tests and applications for clusters up to the 22-mer. *J. Chem. Phys.*, 134(9):094509, 2011.

[50] Q. Yu, C. Qu, P. L. Houston, R. Conte, A. Nandi, and J. M. Bowman. q-AQUA: A Many-Body CCSD(T) Water Potential, Including Four-Body Interactions, Demonstrates the Quantum Nature of Water from Clusters to the Liquid Phase. *J. Phys. Chem. Lett.*, 13(22):5068–5074, 2022.

[51] S. L. Bore and F. Paesani. Realistic phase diagram of water from "first principles" data-driven quantum simulations. *Nat. Commun.*, 14(1):3349, 2023.

[52] T. E. Gartner III, K. M. Hunter, E. Lambros, A. Caruso, M. Riera, G. R. Medders, A. Z. Panagiotopoulos, P. G. Debenedetti, and F. Paesani. Anomalies and local structure of liquid water from boiling to the supercooled regime as predicted by the many-body mb-pol model. *J. Phys. Chem. Lett.*, 13:3652–3658, 2022.

[53] M. C. Muniz, T. E. Gartner III, M. Riera, C. Knight, S. Yue, F. Paesani, and A. Z. Panagiotopoulos. Vapor–liquid equilibrium of water with the mb-pol many-body potential. *J. Chem. Phys.*, 154:211103, 2021.

[54] E. Lambros and F. Paesani. How good are polarizable and flexible models for water: Insights from a many-body perspective. *J. Chem. Phys.*, 153(6):060901, 2020.

[55] E. Paquet and H. L. Viktor. Molecular Dynamics, Monte Carlo Simulations, and Langevin Dynamics: A Computational Review. *Biomed Res. Int.*, 2015:e183918, 2015.

[56] K. A. Fichthorn and W. H. Weinberg. Theoretical foundations of dynamical monte carlo simulations. *J. Chem. Phys.*, 95:1090–1096, 1991.

[57] N. Metropolis and S. Ulam. The monte carlo method. *J. Am. Stat. Assoc.*, 44(247):335–341, 1949.

[58] P. S. Salmon. A neutron diffraction study on the structure of liquid germanium. *J. Phys. F: Met. Phys.*, 18:2345, 1988.

[59] C. J. Wzenzel, U. Linderstrøm-Lang, and S. A. Rice. Amorphous solid water: A neutron diffraction study. *Science*, 187:428–430, 1975.

[60] H. Hasegawa and I. Yasui. X-ray and neutron diffraction analyses of barium silicate glass. *J. Non-Cryst. Solids*, 95-96:201–208, 1987.

[61] A. K. Soper. Joint structure refinement of x-ray and neutron diffraction data on disordered materials: application to liquid water. *J. Phys.: Condens. Matter*, 19:335206, 2007.

[62] L. Verlet. Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159(1):98, 1967.

[63] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

[64] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384–2393, 1980.

[65] G. J. Martyna, M. L. Klein, and M. Tuckerman. Nosé–hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.*, 97:2635–2643, 1992.

[66] R. Kubo. A stochastic theory of line shape. *Adv. Chem. Phys.*, 15:101–127, 1962.

[67] W. Chen, M. Sharma, R. Resta, G. Galli, and R. Car. Role of dipolar correlations in the infrared spectra of water and ice. *Phys. Rev. B*, 77(24):245114, 2008.

[68] L. Buimaga-Iarinca and C. Morari. Calculation of infrared spectra for adsorbed molecules from the dipole autocorrelation function. *Theor. Chem. Acc.*, 141, 2022.

[69] Richard P. Feynman, Albert R. Hibbs, and Daniel F. Styer. *Quantum mechanics and path integrals.* Courier Corporation, 2010.

[70] M. E. Tuckerman, B. J. Berne, G. J. Martyna, and M. L. Klein. Efficient molecular dynamics and hybrid monte carlo algorithms for path integrals. *J. Chem. Phys.*, 99:2796–2808, 1993.

[71] T. E. Markland and M. Ceriotti. Nuclear quantum effects enter the mainstream. *Nat. Rev. Chem.*, 2(3):1–14, 2018.

[72] F. Uhl, D. Marx, and M. Ceriotti. Accelerated path integral methods for atomistic simulations at ultra-low temperatures. *J. Chem. Phys.*, 145:054101, 2016.

[73] J. Cao and G. A. Voth. The formulation of quantum statistical mechanics based on the feynman path centroid density. i. equilibrium properties. *J. Chem. Phys.*, 100(7):5093, 1994.

[74] J. Cao and G. A. Voth. The formulation of quantum statistical mechanics based on the feynman path centroid density. ii. dynamical properties. *J. Chem. Phys.*, 100(7):5106, 1994.

[75] S. Jang and G. A. Voth. A derivation of centroid molecular dynamics and other approximate time evolution methods for path integral centroid variables. *J. Chem. Phys.*, 111(6):2371, 1999.

[76] I. R. Craig and D. E. Manolopoulos. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *J. Chem. Phys.*, 121(8):3368–3373, 2004.

[77] B. J. Braams and D. E. Manolopoulos. On the short-time limit of ring polymer molecular dynamics. *J. Chem. Phys.*, 125(12):124105, 2006.

[78] G. S. Fanourgakis, G. K. Schenter, and S. S. Xantheas. A quantitative account of quantum effects in liquid water. *J. Chem. Phys.*, 125(14):141102, 2006.

[79] E. Mátyus, D. J. Wales, and S. C. Althorpe. Quantum tunneling splittings from path-integral molecular dynamics. *J. Chem. Phys.*, 144:114108, 2016.

[80] C. L. Vaillant, D. J. Wales, and S. C. Althorpe. Tunneling splittings in water clusters from path integral molecular dynamics. *J. Phys. Chem. Lett.*, 10(22):7300–7304, 2019.

[81] Y. Zhu, S. Yang, J. Zeng, W. Fang, L. Jiang, D. H. Zhang, and X. Li. Accurate calculation of tunneling splittings in water clusters using path-integral based methods. *J. Chem. Phys.*, 158:220901, 2023.

[82] M. Ceriotti, W. Fang, P. G. Kusalik, R. H. McKenzie, A. Michaelides, M. A. Morales, and T. E. Markland. Nuclear quantum effects in water and aqueous systems: Experiment, theory, and current challenges. *Chem. Rev.*, 116(13):7529–7550, 2016.

[83] P. Gasparotto, A. A. Hassanali, and M. Ceriotti. Probing defects and correlations in the hydrogen-bond network of ab initio water. *J. Chem. Theory Comput.*, 12(4):1953–1964, 2016.

[84] M. Ceriotti, J. Cuny, M. Parrinello, and D. E. Manolopoulos. Nuclear quantum effects and hydrogen bond fluctuations in water. *Proc. Natl. Acad. Sci. U.S.A.*, 110(39):15591–15596, 2013.

[85] J. A. Morrone and R. Car. Nuclear quantum effects in water. *Phys. Rev. Lett.*, 101:017801, 2008.

[86] F. Paesani, S. Yoo, H. J. Bakker, and S. S. Xantheas. Nuclear quantum effects in the reorientation of water. *J. Phys. Chem. Lett.*, 1(15):2316–2321, 2010.

[87] C. Swalina, Q. Wang, A. Chakraborty, and S. Hammes-Schiffer. Analysis of nuclear quantum effects on hydrogen bonding. *J. Phys. Chem. A*, 111(11):2206–2212, 2007.

[88] J. P. Heindel, K. M. Herman, and S. S. Xantheas. Many-Body Effects in Aqueous Systems: Synergies Between Interaction Analysis Techniques and Force Field Development. *Annu. Rev. Phys. Chem.*, 74(1):337–360, 2023.

[89] D. Hankins, J. W. Moskowitz, and F. H. Stillinger. Water Molecule Interactions. *J. Chem. Phys.*, 53(12):4544–4554, 1970.

[90] E. Clementi, W. Kołos, G. C. Lie, and G. Ranghino. Nonadditivity of interaction in water trimers. *Int. J. Quantum Chem.*, 17(3):377–398, 1980.

[91] S. S. Xantheas and T. H. Dunning. The structure of the water trimer from ab initio calculations. *J. Chem. Phys.*, 98(10):8037–8040, 1993.

[92] S. S. Xantheas. *Ab initio* studies of cyclic water clusters $(H_2O)_n$, n =1–6. II. Analysis of many-body interactions. *J. Chem. Phys.*, 100(10):7523–7534, 1994.

[93] S. S. Xantheas. Significance of higher-order many-body interaction energy terms in water clusters and bulk water. *Philos. Mag.*, 73(1):107–115, 1996.

[94] S. S. Xantheas. Cooperativity and hydrogen bonding network in water clusters. *Chem. Phys.*, 258(2-3):225–231, 2000.

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

[96] A. Shank, Y. Wang, A. Kaledin, B. J. Braams, and J. M. Bowman. Accurate ab initio and "hybrid" potential energy surfaces, intramolecular vibrational energies, and classical ir spectrum of the water dimer. *J. Chem. Phys.*, 130(14):144314, 2009.

[97] J. Mato, D. Tzeli, and S. S. Xantheas. The many-body expansion for metals. I. The alkaline earth metals Be, Mg, and Ca. *J. Chem. Phys.*, 157(8):084313, 2022.

[98] D. Tzeli and S. S. Xantheas. Breaking covalent bonds in the context of the many-body expansion (MBE). I. The purported “first row anomaly” in XH $_n$ (X = C, Si, Ge, Sn; n = 1–4). *J. Chem. Phys.*, 156(24):244303, 2022.

[99] T. Depastas, G. A. Souliotis, D. Tzeli, and S. S. Xantheas. Many-body expansion for light nuclear systems. *Phys. Rev. C*, 107:044004, 2023.

[100] N. Sahu, S. D. Yeole, and S. R. Gadre. Appraisal of molecular tailoring approach for large clusters. *J. Chem. Phys.*, 138:104101, 2013.

[101] Nityananda Sahu and Shridhar R. Gadre. Molecular Tailoring Approach: A Route for ab Initio Treatment of Large Clusters. *Acc. Chem. Res.*, 47(9):2739–2747, 2014.

[102] V. Ganesh, Rameshwar K. Dongare, P. Balanarayan, and Shridhar R. Gadre. Molecular tailoring approach for geometry optimization of large molecules: Energy evaluation and parallelization strategies. *J. Chem. Phys.*, 125(10):104109, 2006.

[103] N. Sahu, S. R. Gadre, A. Rakshit, P. Bandyopadhyay, E. Miliordos, and S. S. Xantheas. Low energy isomers of (H2O)25 from a hierarchical method based on Monte Carlo temperature basin paving and molecular tailoring approaches benchmarked by MP2 calculations. *J. Chem. Phys.*, 141(16):164304, 2014.

[104] M. M. Deshmukh, S. R. Gadre, and L. J. Bartolotti. Estimation of intramolecular hydrogen bond energy via molecular tailoring approach. *J. Phys. Chem. A*, 110:12519–12523, 2006.

[105] M. M. Deshmukh, L. J. Bartolotti, and S. R. Gadre. Intramolecular hydrogen bonding and cooperative interactions in carbohydrates via the molecular tailoring approach. *J. Phys. Chem. A*, 112:312–321, 2008.

[106] M. M. Deshmukh and S. R. Gadre. Molecular tailoring approach for the estimation of intramolecular hydrogen bond energy. *Molecules*, 26:2928, 2021.

[107] K. Szalewicz. Symmetry-adapted perturbation theory of intermolecular forces. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2(2):254–272, 2012.

[108] R. Bukowski, J. Sadlej, B. Jeziorski, P. Jankowski, K. Szalewicz, S. A. Kucharski, H. L. Williams, and B. M. Rice. Intermolecular potential of carbon dioxide dimer from symmetry-adapted perturbation theory. *J. Chem. Phys.*, 110:3785–3803, 1999.

[109] E. M. Mas, R. Bukowski, K. Szalewicz, G. C. Groenenboom, P. E. S. Wormer, and A. van der Avoird. Water pair potential of near spectroscopic accuracy. I. Analysis of potential surface and virial coefficients. *J. Chem. Phys.*, 113(16):6687–6701, 2000.

[110] R. Podeszwa, R. Bukowski, and K. Szalewicz. Potential Energy Surface for the Benzene Dimer and Perturbational Analysis of Interactions. *J. Phys. Chem. A*, 110(34):10345–10354, 2006.

[111] A. Jing, K. Szalewicz, and A. van der Avoird. Ammonia dimer: extremely fluxional but still hydrogen bonded. *Nat. Commun.*, 13:1470, 2022.

[112] R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. T. Bell, and M. Head-Gordon. Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. *J. Phys. Chem. A*, 111(36):8753–8765, 2007.

[113] R. Z. Khaliullin, A. T. Bell, and M. Head-Gordon. Analysis of charge transfer effects in molecular complexes based on absolutely localized molecular orbitals. *J. Chem. Phys.*, 128:184112, 2008.

[114] A. K. Das, M. Liu, and T. Head-Gordon. Development of a Many-Body Force Field for Aqueous Alkali Metal and Halogen Ions: An Energy Decomposition Analysis Guided Approach. *J. Chem. Theory Comput.*, 18(2):953–967, 2022.

[115] F. Hofmeister. Zur Lehre von der Wirkung der Salze. *Archiv f. experiment. Pathol. u. Pharmakol*, (25):1–30, 1888.

[116] W. Kunz, J. Henle, and B. W. Ninham. 'Zur Lehre von der Wirkung der Salze' (about the science of the effect of salts): Franz Hofmeister's historical papers. *Curr. Opin. Colloid Interface Sci.*, 9(1):19–37, 2004.

[117] G. Jones and M. Dole. The Viscosity of Aqueous Solutions of Strong Electrolytes with Special Reference to Barium Chloride. *J. Am. Chem. Soc.*, 51(10):2950–2964, 1929.

[118] H. D. B. Jenkins and Y. Marcus. Viscosity B-Coefficients of Ions in Solution. *Chem. Rev.*, 95(8):2695–2724, 1995.

[119] M. Boström, W. Kunz, and B. W. Ninham. Hofmeister Effects in Surface Tension of Aqueous Electrolyte Solution. *Langmuir*, 21(6):2619–2623, 2005.

[120] L. M. Pegram and M. T. Record. Hofmeister Salt Effects on Surface Tension Arise from Partitioning of Anions and Cations between Bulk Water and the AirWater Interface. *J. Phys. Chem. B*, 111(19):5411–5417, 2007.

[121] W. J. Xie and Y. Q. Gao. A Simple Theory for the Hofmeister Series. *J. Phys. Chem. Lett.*, 4(24):4247–4252, 2013.

[122] S. Y. Willow and S. S. Xantheas. Molecular-Level Insight of the Effect of Hofmeister Anions on the Interfacial Surface Tension of a Model Protein. *J. Phys. Chem. Lett.*, 8(7):1574–1577, 2017.

[123] R. L. Baldwin. How Hofmeister ion interactions affect protein stability. *Biophys. J.*, 71(4):2056–2063, 1996.

[124] H. I. Okur, J. Hladílková, K. B. Rembert, Y. Cho, J. Heyda, J. Dzubiella, P. S. Cremer, and P. Jungwirth. Beyond the Hofmeister Series: Ion-Specific Effects on Proteins and Their Biological Functions. *J. Phys. Chem. B*, 121(9):1997–2014, 2017.

[125] S. Shimizu, W. M. McLaren, and N. Matubayasi. The Hofmeister series and protein-salt interactions. *J. Chem. Phys.*, 124(23):234905, 2006.

[126] K. D. Collins. Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. *Methods*, 34(3):300–311, 2004.

[127] R. Perez-Jimenez, R. Godoy-Ruiz, B. Ibarra-Molero, and J. M. Sanchez-Ruiz. The Efficiency of Different Salts to Screen Charge Interactions in Proteins: A Hofmeister Effect? *Biophys. J.*, 86(4):2414–2429, 2004.

[128] Y. Zhang and P. S Cremer. Interactions between macromolecules and ions: the Hofmeister series. *Curr. Opin. Chem. Biol.*, 10(6):658–663, 2006.

[129] M. Lund and P. Jungwirth. Patchy proteins, anions and the Hofmeister series. *J. Phys.: Condens. Matter*, 20(49):494218, 2008.

[130] E. Sedlák, L. Stagg, and P. Wittung-Stafshede. Effect of Hofmeister ions on protein thermal stability: Roles of ion hydration and peptide groups? *Arch. Biochem. Biophys.*, 479(1):69–73, 2008.

[131] R. Piazza and M. Pierno. Protein interactions near crystallization: a microscopic approach to the Hofmeister series. *J. Phys.: Condens. Matter*, 12(8A):A443–A449, 2000.

[132] C. P. Schneider, D. Shukla, and B. L. Trout. Arginine and the Hofmeister Series: The Role of Ion–Ion Interactions in Protein Aggregation Suppression. *J. Phys. Chem. B*, 115(22):7447–7458, 2011.

[133] M. Boström, D. R. M. Williams, and B. W. Ninham. Why the properties of proteins in salt solutions follow a Hofmeister series. *Curr. Opin. Colloid & Interface Sci.*, 9(1):48–52, 2004.

[134] X. Tadeo, B. López-Méndez, D. Castaño, T. Trigueros, and O. Millet. Protein Stabilization and the Hofmeister Effect: The Role of Hydrophobic Solvation. *Biophys. J.*, 97(9):2595–2603, 2009.

[135] J. M. Fox, K. Kang, W. Sherman, A. Héroux, G. M. Sastry, M. Baghbanzadeh, M. R. Lockett, and G. M. Whitesides. Interactions between Hofmeister Anions and the Binding Pocket of a Protein. *J. Am. Chem. Soc.*, 137(11):3859–3866, 2015.

[136] L. A. Moreira, M. Boström, B. W. Ninham, E. C. Biscaia, and F. W. Tavares. Hofmeister effects: Why protein charge, pH titration and protein precipitation depend on the choice of background salt solution. *Colloids Surf. A: Physicochem. Eng. Asp.*, 282-283:457–463, 2006.

[137] Y. Chen, H. I. Okur, N. Gomopoulos, C. Macias-Romero, P. S. Cremer, P. B. Petersen, G. Tocci, D. M. Wilkins, C. Liang, and M. Ceriotti. Electrolytes induce long-range orientational order and free energy changes in the H-bond network of bulk water. *Sci. Adv.*, 2(4):e1501891, 2016.

[138] A. Shalit, S. Ahmed, J. Savolainen, and P. Hamm. Terahertz echoes reveal the inhomogeneity of aqueous salt solutions. *Nat. Chem.*, 9(3):273–278, 2017.

[139] P. B. Petersen and R. J. Saykally. On the nature of ions at the liquid water surface. *Annu. Rev. Phys. Chem.*, 57(1):333–364, 2006.

[140] C. D. Cappa, J. D. Smith, B. M. Messer, R. C. Cohen, and R. J. Saykally. Effects of Cations on the Hydrogen Bond Network of Liquid Water: New Results from X-ray Absorption Spectroscopy of Liquid Microjets. *J. Phys. Chem. B*, 110(11):5301–5309, 2006.

[141] T. P. Pollard and T. L. Beck. Toward a quantitative theory of Hofmeister phenomena: From quantum effects to thermodynamics. *Curr. Opin. Colloid Interface Sci.*, 23:110–118, 2016.

[142] J. T. O'Brien and E. R. Williams. Effects of Ions on Hydrogen-Bonding Water Networks in Large Aqueous Nanodrops. *J. Am. Chem. Soc.*, 134(24):10228–10236, 2012.

[143] J. T. O'Brien, J. S. Prell, M. F. Bush, and E. R. Williams. Sulfate Ion Patterns Water at Long Distance. *J. Am. Chem. Soc.*, 132(24):8248–8249, 2010.

[144] M. J. DiTucci, S. Heiles, and E. R. Williams. Role of Water in Stabilizing Ferricyanide Trianion and Ion-Induced Effects to the Hydrogen-Bonding Water Network at Long Distance. *J. Am. Chem. Soc.*, 137(4):1650–1657, 2015.

[145] C. Yan, Z. Xue, W. Zhao, J. Wang, and T. Mu. Surprising Hofmeister Effects on the Bending Vibration of Water. *ChemPhysChem*, 17(20):3309–3314, 2016.

[146] R. Mancinelli, A. Botti, F. Bruni, M. A. Ricci, and A. K. Soper. Hydration of Sodium, Potassium, and Chloride Ions in Solution and the Concept of Structure Maker/Breaker. *J. Phys. Chem. B*, 111(48):13570–13577, 2007.

[147] A. W. Omta. Negligible Effect of Ions on the Hydrogen-Bond Structure in Liquid Water. *Science*, 301(5631):347–349, 2003.

[148] W. Wachter, W. Kunz, R. Buchner, and G. Hefter. Is There an Anionic Hofmeister Effect on Water Dynamics? Dielectric Spectroscopy of Aqueous Solutions of NaBr, NaI, $NaNO_3$, $NaClO_4$, and NaSCN. *J. Phys. Chem. A*, 109(39):8675–8683, 2005.

[149] L. Näslund, D. C. Edwards, P. Wernet, U. Bergmann, H. Ogasawara, L. G. M. Pettersson, S. Myneni, and A. Nilsson. X-ray Absorption Spectroscopy Study of the Hydrogen Bond Network in the Bulk Water of Aqueous Solutions. *J. Phys. Chem. A*, 109(27):5995–6002, 2005.

[150] Z. Yin, I. Rajkovic, K. Kubicek, W. Quevedo, A. Pietzsch, P. Wernet, A. Föhlisch, and S. Techert. Probing the Hofmeister Effect with Ultrafast Core–Hole Spectroscopy. *J. Phys. Chem. B*, 118(31):9398–9403, 2014.

[151] P. Jungwirth and D. J. Tobias. Ions at the Air/Water Interface. *J. Phys. Chem. B*, 106(25):6361–6373, 2002.

[152] N. Galamba. Mapping Structural Perturbations of Water in Ionic Solutions. *J. Phys. Chem. B*, 116(17):5242–5250, 2012.

[153] R. D. Mountain and D. Thirumalai. Alterations in Water Structure Induced by Guanidinium and Sodium Ions. *J. Phys. Chem. B*, 108(51):19711–19716, 2004.

[154] A. P. Gaiduk and G. Galli. Local and Global Effects of Dissolved Sodium Chloride on the Structure of Water. *J. Phys. Chem. Lett.*, 8(7):1496–1502, 2017.

[155] A. Grossfield. Dependence of ion hydration on the sign of the ion's charge. *J. Chem. Phys.*, 122(2):024506, 2005.

[156] Marcel D. Baer and Christopher J. Mundy. An ab initio approach to understanding the specific ion effect. *Faraday Discuss.*, 160:89–101, 2013.

[157] C. F. Schwenk, T. S. Hofer, and B. M. Rode. "Structure Breaking" Effect of Hydrated Cs^+. *J. Phys. Chem. A*, 108(9):1509–1514, 2004.

[158] R. Leberman and A. K. Soper. Effect of high salt concentrations on water structure. *Nature*, 378(6555):364–366, 1995.

[159] F. Roberts and B. Tesman. *Applied combinatorics.* CRC Press, 2009.

[160] S. S. Xantheas. On the importance of the fragment relaxation energy terms in the estimation of the basis set superposition error correction to the intermolecular interaction energy. *J. Chem. Phys.*, 104(21):8821–8824, 1996.

[161] J. C. White and E. R. Davidson. An analysis of the hydrogen bond in ice. *J. Chem. Phys.*, 93(11):8029–8035, 1990.

[162] Y. Marcus and G. Hefter. Ion Pairing. *Chem. Rev.*, 106(11):4585–4621, 2006.

[163] C. J. Fennell, A. Bizjak, V. Vlachy, and K. A. Dill. Ion Pairing in Molecular Simulations of Aqueous Alkali Halide Solutions. *J. Phys. Chem. B*, 113(19):6782–6791, 2009.

[164] L. Götte, K. M. Parry, W. Hua, D. Verreault, H. C. Allen, and D. J. Tobias. Solvent-Shared Ion Pairs at the Air–Solution Interface of Magnesium Chloride and Sulfate Solutions Revealed by Sum Frequency Spectroscopy and Molecular Dynamics Simulations. *J. Phys. Chem. A*, 121(34):6450–6459, 2017.

[165] B. S. González, J. Hernández-Rojas, and D. J. Wales. Global minima and energetics of $Li^+(H_2O)_n$ and $Ca^{2+}(H_2O)_n$ clusters for n≤20. *Chem. Phys. Lett.*, 412(1):23–28, 2005.

[166] S. Pei, S. Jiang, Y. Liu, T. Huang, K. Xu, H. Wen, Y. Zhu, and W. Huang. Properties of Ammonium Ion–Water Clusters: Analyses of Structure Evolution, Noncovalent Interactions, and Temperature and Humidity Effects. *J. Phys. Chem. A*, 119(12):3035–3047, 2015.

[167] M. Kulichenko, N. Fedik, K. V. Bozhenko, and A. I. Boldyrev. Hydrated Sulfate Clusters ${SO_4}^{2-}(H_2O)_n$ (n = 1–40): Charge Distribution Through Solvation Shells and Stabilization. *J. Phys. Chem. B*, 123(18):4065–4069, 2019.

[168] D. S. Lambrecht, G. N. I. Clark, T. Head-Gordon, and M. Head-Gordon. Exploring the Rich Energy Landscape of Sulfate–Water Clusters ${SO_4}^{2-}(H_2O)_{n=3\text{–}7}$: An Electronic Structure Approach. *J. Phys. Chem. A*, 115(41):11438–11454, 2011.

[169] M. Kaupp, P. v. R. Schleyer, H. Stoll, and H. Preuss. Pseudopotential approaches to Ca, Sr, and Ba hydrides. Why are some alkaline earth MX $_2$ compounds bent? *J. Chem. Phys.*, 94(2):1360–1366, 1991.

[170] S. F. Boys and F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.*, 19(4):553–566, 1970.

[171] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, and W. A. de Jong. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.*, 181(9):1477–1489, 2010.

[172] K. D. Collins. Charge density-dependent strength of hydration and biological structure. *Biophys. J.*, 72(1):65–76, 1997.

[173] F. Guo and J. M. Friedman. Charge Density-Dependent Modifications of Hydration Shell Waters by Hofmeister Ions. *J. Am. Chem. Soc.*, 131(31):11010–11018, 2009.

[174] K. D. Collins. Sticky ions in biological systems. *Proc. Natl. Acad. Sci. U.S.A.*, 92(12):5553–5557, 1995.

[175] D. F. Parsons, M. Boström, P. L. Nostro, and B. W. Ninham. Hofmeister effects: interplay of hydration, nonelectrostatic potentials, and ion size. *Phys. Chem. Chem. Phys.*, 13(27):12352, 2011.

[176] B. Hribar, N. T. Southall, V. Vlachy, and K. A. Dill. How Ions Affect the Structure of Water. *J. Am. Chem. Soc.*, 124(41):12302–12311, 2002.

[177] J. C. Werhahn, D. Akase, and S. S. Xantheas. Universal scaling of potential energy functions describing intermolecular interactions. II. The halide-water and alkali metal-water interactions. *J. Chem. Phys.*, 141(6):064118, 2014.

[178] J. C. Werhahn, E. Miliordos, and S. S. Xantheas. A new variation of the Buckingham exponential-6 potential with a tunable, singularity-free short-range repulsion and an adjustable long-range attraction. *Chem. Phys. Lett.*, 619:133–138, 2015.

[179] S. S. Xantheas and J. C. Werhahn. Universal scaling of potential energy functions describing intermolecular interactions. I. Foundations and scalable forms of new generalized Mie, Lennard-Jones, Morse, and Buckingham exponential-6 potentials. *J. Chem. Phys.*, 141(6):064117, 2014.

[180] B. Sharma and A. Chandra. Ab Initio Molecular Dynamics Simulation of the Phosphate Ion in Water: Insights into Solvation Shell Structure, Dynamics, and Kosmotropic Activity. *J. Phys. Chem. B*, 121(46):10519–10529, 2017.

[181] H. J. Bakker, M. F. Kropman, A. W. Omta, and S. Woutersen. Hydrogen-Bond Dynamics of Water in Ionic Solutions. *Phys. Scr.*, 69(6):C14, 2004.

[182] D. J. Arismendi-Arrieta, M. Riera, P. Bajaj, R. Prosmiti, and F. Paesani. I-ttm model for ab initio-based ion–water interaction potentials. 1. Halide–water potential energy functions. *J. Phys. Chem. B*, 120(8):1822–1832, 2015.

[183] V. Babin, G. R. Medders, and F. Paesani. Toward a Universal Water Model: First Principles Simulations from the Dimer to the Liquid Phase. *J. Phys. Chem. Lett.*, 3(24):3765–9, 2012.

[184] K.J. Tielrooijn, N.Garcia-Araezm, , M. Bonn, and H.J. Bakker. Cooperativity in ion hydration. *Science*, 328:1006–1009, 2010.

[185] K. Stokely, M.G. Mazza, H.E. Stanley, and G. Franzese. Effect of hydrogen bond cooperativity on the behavior of water. *Proc. Natl. Acad. Sci. U.S.A.*, 107:1301–1306, 2010.

[186] J. G. Davis, B. M. Rankin, K. P. Gierszal, and D. Ben-Amotz. On the cooperative formation of non-hydrogen-bonded water at molecular hydrophobic interfaces. *Nat. Chem.*, 5:796–802, 2013.

[187] R. Cota, N. Ottosson, H. J. Bakker, and S. Woutersen. Evidence for reduced hydrogen-bond cooperativity in ionic solvation shells from isotope-dependent dielectric relaxation. *Phys. Rev. Lett.*, 120:216001, 2018.

[188] I. Bakó, A. Lábas, K. Hermansson, A. Bencsura, and J. Oláh. How can we detect hydrogen bond local cooperativity in liquid water: A simulation study. *J. Mol. Liq.*, 245:140–146, 2017.

[189] S. S. Xantheas and T. H. Dunning. Structures and energetics of f-(h2o)(n), n=1-3, clusters from ab-initio calculations. *J. Phys. Chem.*, 98:13489–13497, 1994.

[190] S. S Xantheas. Quantitative Description of Hydrogen Bonding in Chloride Water Clusters. *J. Phys. Chem.*, 100(23):9703–9713, 1996.

[191] M. P. Hodges, A. J. Stone, and S. S. Xantheas. Contribution of many-body terms to the energy for small water clusters: A comparison of ab initio calculations and accurate model potentials. *J. Phys. Chem. A*, 101:9163–9168, 1997.

[192] R. M. Richard and J. M. Herbert. A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory. *J. Chem. Phys.*, 137(6):064113, 2012.

[193] R. M. Richard, K. U. Lao, and J. M. Herbert. Approaching the complete-basis limit with a truncated many body expansion. *J. Chem. Phys.*, 139:224102, 2013.

[194] R. M. Richard, K. U. Lao, and J. M. Herbert. Understanding the many-body expansion for large systems. I. Precision considerations. *J. Chem. Phys.*, 141(1):014108, 2014.

[195] K. U. Lao, K. Liu, R. M. Richard, and J. M. Herbert. Understanding the many-body expansion for large systems. II. Accuracy considerations. *J. Chem. Phys.*, 144(16):164105, 2016.

[196] K. Y. Liu and J. M. Herbert. Understanding the many-body expansion for large systems. iii. critical role of four-body terms, counterpoise corrections, and cutoffs. *J. Chem. Phys.*, 147:161729, 2017.

[197] J. F. Ouyang, M. W. Cvitkovic, and R. P. A. Bettens. Trouble with the many-body expansion. *J. Chem. Theory Comput.*, 10(9):3699–3707, 2014.

[198] S. P. Veccham, J. Lee, and M. Head-Gordon. Making many-body interactions nearly pairwise additive: The polarized many-body expansion approach. *J. Chem. Phys.*, 194101:194101, 2019.

[199] W. Chen and M. S. Gordon. Energy Decomposition Analyses for Many-Body Interaction and Applications to Water Complexes. *J. Phys. Chem.*, 100(34):14316–14328, 1996.

[200] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko. Fragmentation methods: A route to accurate calculations on large systems. *Chem. Rev.*, 112:632–672, 2012.

[201] J. P. Heindel and S. S. Xantheas. Molecular Dynamics Driven by the Many-Body Expansion (MBE-MD). *J. Chem. Theory Comput.*, 17(12):7341–7352, 2021.

[202] C. J. Burnham and S. S. Xantheas. Development of transferable interaction models for water. iii. reparametrization of an all-atom polarizable rigid model (ttm2-r) from first principles. *J. Chem. Phys.*, 116:1500–1510, 2002.

[203] C. J. Burnham and S. S. Xantheas. Development of transferable interaction models for water. iv. a flexible, all-atom polarizable potential (ttm2-f) based on geometry dependent charges derived from an ab initio monomer dipole moment surface. *J. Chem. Phys.*, 116:5115–5124, 2002.

[204] A. Nandi, C. Qu, P. L. Houston, R. Conte, Q. Yu, and J. M. Bowman. A ccsd(t)-based 4-body potential for water. *J. Phys. Chem. Lett.*, 12(42):10318–10324, 2021.

[205] Z. Xie and J. M. Bowman. Permutationally Invariant Polynomial Basis for Molecular Energy Surface Fitting via Monomial Symmetrization. *J. Chem. Theory Comput.*, 6(1):26–34, 2010.

[206] G. A. Cisneros, K. T. Wikfeldt, L. Ojamäe, J. Lu, Y. Xu, H. Torabifard, A. P. Bartók, G. Csányi, V. Molinero, and F. Paesani. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions. *Chem. Rev.*, 116(13):7501–7528, 2016.

[207] A. Stone. *The Theory of Intermolecular Forces.* OUP Oxford, 2013.

[208] A. J. Misquitta and A. J. Stone. CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form, 2019.

[209] A. J. Stone, A. Dullweber, O. Engkvist, E. Fraschini, M. P. Hodges, A. W. Meredith, D. R. Nutt, P. L. A. Popelier, and D. J. Wales. Orient: a program for studying interactions between molecules, version 5.0. gitlab.com/anthonyjs/orient, 2018.

[210] R. A. Kendall, T. H. Dunning, and R. J. Harrison. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.*, 96:6796–6806, 1992.

[211] T. Leininger, A. Nicklass, W. Küchle, H. Stoll, M. Dolg, and A. Bergner. The accuracy of the pseudopotential approximation: Non-frozen-core effects for spectroscopic constants of alkali fluorides XF (X= K, Rb, Cs). *Chem. Phys. Lett.*, 255(4-6):274–280, 1996.

[212] K. A. Peterson, D. Figgen, E. Goll, H. Stoll, and M. Dolg. Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16–18 elements. *J. Chem. Phys.*, 119(21):11113–11123, 2003.

[213] K. A. Peterson, B. C. Shepler, D. Figgen, and H. Stoll. On the spectroscopic and thermochemical properties of clo, bro, io, and their anions. *J. Phys. Chem. A*, 110:13877–13883, 2006.

[214] E. Aprà, E. J. Bylaska, W. A. de Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. J. J. van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. W. Aquino, R. Atta-Fynn, J. Autschbach, N. P. Bauman, J. C. Becca, D. E. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauët, Y. Chen, G. N. Chuev, C. J. Cramer, J. Daily, M. J. O. Deegan, T. H. Dunning, M. Dupuis, K. G. Dyall, G. I. Fann, S. A. Fischer, A. Fonari, H. Früchtl, L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. W. Götz, J. Hammond, V. Helms, E. D. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. G. Johnson, H. Jónsson, R. A. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. D. Lins, R. J. Littlefield, A. J. Logsdail, K. Lopata, W. Ma, A. V. Marenich, J. Martin del Campo, D. Mejia-Rodriguez, J. E. Moore, J. M. Mullin, T. Nakajima, D. R. Nascimento, J. A. Nichols, P. J. Nichols,

J. Nieplocha, A. Otero-de-la Roza, B. Palmer, A. Panyala, T. Pirojsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. M. Richard, P. Sadayappan, G. C. Schatz, W. A. Shelton, D. W. Silverstein, D. M. A. Smith, T. A. Soares, D. Song, M. Swart, H. L. Taylor, G. S. Thomas, V. Tipparaju, D. G. Truhlar, K. Tsemekhman, T. Van Voorhis, Á. Vázquez-Mayagoitia, P. Verma, O. Villa, A. Vishnu, K. D. Vogiatzis, D. Wang, J. H. Weare, M. J. Williamson, T. L. Windus, K. Woliński, A. T. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao, and R. J. Harrison. NWChem: Past, present, and future. *J. Chem. Phys.*, 152(18):184102, 2020.

[215] C. Millot, J. Soetens, M. T. C. Martins Costa, M. P. Hodges, and A. J. Stone. Revised Anisotropic Site Potentials for the Water Dimer and Calculated Properties. *J. Phys. Chem. A*, 102(4):754–770, 1998.

[216] M. Li, B. Zhuang, Y. Lu, Z. Wang, and L. An. Accurate determination of ion polarizabilities in aqueous solutions. *J. Phys. Chem. B*, 121:6416–6424, 2017.

[217] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B*, 114(8):2549–2564, 2010.

[218] J. A. Rackers, R. R. Silva, Z. Wang, and J. W. Ponder. Polarizable Water Potential Derived from a Model Electron Density. *J. Chem. Theory Comput.*, 17(11):7056–7084, 2021.

[219] L. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande. Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B*, 117(34):9956–9972, 2013.

[220] G. N. Merrill and M. S. Gordon. Study of Small Water Clusters Using the Effective Fragment Potential Model. *J. Phys. Chem. A*, 102(16):2650–2657, 1998.

[221] H. M. Silva and M. S. Gordon. The effective fragment potential: Small clusters and radial distribution functions. *J. Chem. Phys.*, 121:2711, 2004.

[222] N. De Silva, M. A. Adreance, and M. S. Gordon. Application of a semi-empirical dispersion correction for modeling water clusters. *J. Comput. Chem.*, 40(2):310–315, 2019.

[223] E. B. Guidez and M. S. Gordon. Dispersion Interactions in Water Clusters. *J. Phys. Chem. A*, 121(19):3736–3745, 2017.

[224] D. J. Arismendi-Arrieta, M. Riera, P. Bajaj, R. Prosmiti, and F. Paesani. i-ttm model for ab initio-based ion–water interaction potentials. 1. halide–water potential energy functions. *J. Phys. Chem. B*, 120:1822–1832, 2016.

[225] W. L. Jorgensen. Transferable intermolecular potential functions for water, alcohols, and ethers. application to liquid water. *J. Am. Chem. Soc.*, 103(2):335–340, 1981.

[226] J. L. F. Abascal and C. Vega. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.*, 123(23):234505, 2005.

[227] M. W. Mahoney and W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, 112(20):8910–8922, 2000.

[228] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. Interaction Models for Water in Relation to Protein Hydration. In Bernard Pullman, editor, *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*, The Jerusalem Symposia on Quantum Chemistry and Biochemistry, pages 331–342. Springer Netherlands, Dordrecht, 1981.

[229] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.

[230] J. L. F. Abascal, E. Sanz, R. García Fernández, and C. Vega. A potential model for the study of ices and amorphous water: TIP4P/Ice. *J. Chem. Phys.*, 122(23):234511, 2005.

[231] S. Izadi, R. Anandakrishnan, and A. V. Onufriev. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.*, 5(21):3863–3871, 2014.

[232] M. Riera, E. Lambros, T. T. Nguyen, A. W. Götz, and F. Paesani. Low-order many-body interactions determine the local structure of liquid water. *Chem. Sci.*, 10:8211–8218, 2019.

[233] D. Zhuang, M. Riera, G. K. Schenter, J. L. Fulton, and F. Paesani. Many-Body Effects Determine the Local Hydration Structure of Cs+ in Solution. *J. Phys. Chem. Lett.*, 10(3):406–412, 2019.

[234] B. J. Braams and J. M. Bowman. Permutationally invariant potential energy surfaces in high dimensionality. *International Reviews in Physical Chemistry*, 28(4):577–606, 2009.

[235] S. K. Reddy, S. C. Straight, P. Bajaj, C. H. Pham, M. Riera, D. R. Moberg, M. A. Morales, C. Knight, A. W. Gotz, and F. Paesani. On the accuracy of the MB-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *J. Chem. Phys.*, 145(19):194504, 2016.

[236] J. M. Bowman, C. Qu, R. Conte, A. Nandi, P. L. Houston, and Q. Yu. -machine learned potential energy surfaces and force fields. *J. Chem. Theory Comput.*, 19:1–17, 2023.

[237] A. J. Stone. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.*, 83:233–239, 1981.

[238] A. J. Stone and M. Alderton. Distributed multipole analysis: methods and applications. *Mol. Phys.*, 56(5):1047–1064, 1985.

[239] A. J. Stone. Distributed Multipole Analysis: Stability for Large Basis Sets. *J. Chem. Theory Comput.*, 1(6):1128–1132, 2005.

[240] G. S. Fanourgakis and S. S. Xantheas. Development of transferable interaction potentials for water. v. extension of the flexible, polarizable, thole-type model potential (ttm3-f, v. 3.0) to describe the vibrational spectra of water clusters and liquid water. *J. Chem. Phys.*, 128:074506, 2008.

[241] U. Dinur. “flexible” water molecules in external electrostatic potentials. *J. Phys. Chem.*, 94:5669–5671, 1990.

[242] E. D. Boittier, M. Devereux, and M. Meuwly. Molecular dynamics with conformationally dependent, distributed charges. *J. Chem. Theory Comput.*, 18:7544–7554, 2022.

[243] D. R. Nutt and M. Meuwly. Theoretical investigation of infrared spectra and pocket dynamics of photodissociated carbonmonoxy myoglobin. *Biophys. J.*, 85:3612–3623, 2003.

[244] N. Plattner and M. Meuwly. The role of higher co-multipole moments in understanding the dynamics of photodissociated carbonmonoxide in myoglobin. *Biophys. J.*, 94:2505–2515, 2008.

[245] E. Ö. Jónsson, S. Rasti, M. Galynska, J. Meyer, and H. Jónsson. Transferable Potential Function for Flexible H2O Molecules Based on the Single-Center Multipole Expansion. *J. Chem. Theory Comput.*, 18(12):7528–7543, 2022.

[246] B. C. B. Symons and P. L. A. Popelier. Application of quantum chemical topology force field fflux to condensed matter simulations: Liquid water. *J. Chem. Theory Comput.*, 18:5577–5588, 2022.

[247] O. Loboda, F. Ingrosso, M. F. Ruiz-López, H. Reis, and C. Millot. Dipole and quadrupole polarizabilities of the water molecule as a function of geometry. *J. Comput. Chem.*, 37:2125–2132, 2016.

[248] B. M. Axilrod and E. Teller. Interaction of the van der Waals type between three atoms. *J. Chem. Phys.*, 11(6):299–300, 1943.

[249] Y. Muto. Force between nonpolar molecules. *J. Phys. Math. Soc. Jpn*, 17:629–631, 1943.

[250] K. T. Tang and J. Peter Toennies. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *J. Chem. Phys.*, 80(8):3726–3741, 1984.

[251] O. A. von Lilienfeld and A. Tkatchenko. Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids. *J. Chem. Phys.*, 132:234109, 2010.

[252] D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, H. F. Schaefer, A. Y. Sokolov, K. Patkowski, A. E. DePrince, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill. PSI4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.*, 152(18):184108, 2020.

[253] D. J. Margoliash, T. R. Proctor, G. D. Zeiss, and W. J. Meath. Triple-dipole energies for H, He, Li, N, O. H_2, N_2, O_2, NO, N_2O, H_2O, NH_3 and CH_4 evaluated using pseudo-spectral dipole oscillator strength distributions. *Molec. Phys.*, 35:747–757, 1978.

[254] MBX: A many-body energy and force calculator. https://paesanigroup.ucsd.edu/software/mbx.html.

[255] H. Partridge and D. W. Schwenke. The determination of an accurate isotope dependent potential energy surface for water from extensive *ab initio* calculations and experimental data. *J. Chem. Phys.*, 106(11):4618–4639, 1997.

[256] B. T. Thole. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.*, 59(3):341–350, 1981.

[257] E. R. Batista, S. S. Xantheas, and H. Jónsson. Molecular multipole moments of water molecules in ice Ih. *J. Chem. Phys.*, 109(11):4546–4551, 1998.

[258] C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.*, 110(13):6158–6170, 1999.

[259] Z. L. Glick, A. Koutsoukas, D. L. Cheney, and C. D. Sherrill. Cartesian message passing neural networks for directional properties: Fast and transferable atomic multipoles. *J. Chem. Phys.*, 154:224103, 2021.

[260] O. Demerdash and T. Head-Gordon. Convergence of the many-body expansion for energy and forces for classical polarizable models in the condensed phase. *J. Chem. Theory Comput.*, 12(8):3884–3893, 2016.

[261] J. P. Heindel, E. S. Knodel, and D. P. Schofield. Origin of Many-Body Vibrational Frequency Shifts in Water Clusters. *J. Phys. Chem. A*, 122(33):6724–6735, 2018.

[262] D. Schmitt-Monreal and C. R. Jacob. Density-Based Many-Body Expansion as an Efficient and Accurate Quantum-Chemical Fragmentation Method: Application to Water Clusters. *J. Chem. Theory Comput.*, 17(7):4144–4156, 2021.

[263] J. Hofierka and J. Klimeš. Binding energies of molecular solids from fragment and periodic approaches. *Electron. Struct.*, 3(3):034010, 2021.

[264] C. H. Borca, B. W. Bakr, L. A. Burns, and C. D. Sherrill. CrystaLattE: Automated computation of lattice energies of organic crystals exploiting the many-body expansion to achieve dual-level parallelism. *J. Chem. Phys.*, 151(14):144103, 2019.

[265] J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G. K. Chan. Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Science*, 345(6197):640–643, 2014.

[266] W. Bernd Schweizer and Jack D. Dunitz. Quantum Mechanical Calculations for Benzene Dimer Energies: Present Problems and Future Challenges. *J. Chem. Theory Comput.*, 2(2):288–291, 2006.

[267] M. R. Kennedy, A. R. McDonald, A. E. DePrince, M. S. Marshall, R. Podeszwa, and C. D. Sherrill. Communication: Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory. *J. Chem. Phys.*, 140(12):121104, 2014.

[268] A. L. P. Nguyen, T. G. Mason, B. D. Freeman, and E. I. Izgorodina. Prediction of lattice energy of benzene crystals: A robust theoretical approach. *J. Comput. Chem.*, 42(4):248–260, 2021.

[269] P. V. Hobbs. *Ice Physics.* Oxford University Press, New York, 1974.

[270] A. D. Fortes, I. G. Wood, M. Alfredsson, L. Vočadlo, and K. S. Knight. The incompressibility and thermal expansivity of D_2O ice II determined by powder neutron diffraction. *J. Appl. Crystallogr.*, 38(4):612–618, 2005.

[271] B. Kamb and B. L. Davis. ICE VII, the densest form of ice. *Proc. Natl. Acad. Sci. U.S.A.*, 52(6):1433–1439, 1964.

[272] C. Lobban, J. L. Finney, and W. F. Kuhs. The p–T dependency of the ice II crystal structure and the effect of helium inclusion. *J. Chem. Phys.*, 117(8):3928–3934, 2002.

[273] E. Whalley, D. W. Davidson, and J. B. R. Heath. Dielectric Properties of Ice VII. Ice VIII: A New Phase of Ice. *J. Chem. Phys.*, 45(11):3976–3982, 1966.

[274] Y. Yoshimura, S. T. Stewart, M. Somayazulu, H. Mao, and R. J. Hemley. High-pressure x-ray diffraction and Raman spectroscopy of ice VIII. *J. Chem. Phys.*, 124(2):024502, 2006.

[275] J. D. Jorgensen, R. A. Beyerlein, N. Watanabe, and T. G. Worlton. Structure of D_2O ice VIII from *in situ* powder neutron diffraction. *J. Chem. Phys.*, 81(7):3211–3214, 1984.

[276] E. Whalley, J. B. R. Heath, and D. W. Davidson. Ice IX: An Antiferroelectric Phase Related to Ice III. *J. Chem. Phys.*, 48(5):2362–2370, 1968.

[277] J. D. Londono, W. F. Kuhs, and J. L. Finney. Neutron diffraction studies of ices III and IX on under-pressure and recovered samples. *J. Chem. Phys.*, 98(6):4878–4888, 1993.

[278] S. J. La Placa, W. C. Hamilton, B. Kamb, and A. Prakash. On a nearly proton-ordered structure for ice IX. *J. Chem. Phys.*, 58(2):567–580, 1973.

[279] C. G. Salzmann, P. G. Radaelli, A. Hallbrucker, E. Mayer, and J. L. Finney. The Preparation and Structures of Hydrogen Ordered Phases of Ice. *Science*, 311(5768):1758–1761, 2006.

[280] C. G. Salzmann, P. G. Radaelli, E. Mayer, and J. L. Finney. Ice XV: a new thermodynamically stable phase of ice. *Phys. Rev. Lett.*, 103(10):105701, 2009.

[281] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 77(18):3865–3868, 1996.

[282] B. Santra, J. Klimeš, A. Tkatchenko, D. Alfè, B. Slater, A. Michaelides, R. Car, and M. Scheffler. On the accuracy of van der Waals inclusive density-functional theory exchange-correlation functionals for ice at ambient and high pressures. *J. Chem. Phys.*, 139(15):154702, 2013.

[283] M. Matsumoto, T. Yagasaki, and H. Tanaka. GenIce: Hydrogen-Disordered Ice Generator. *J. Comput. Chem.*, 39(1):61–64, 2018.

[284] M. Matsumoto, T. Yagasaki, and H. Tanaka. Novel Algorithm to Generate Hydrogen-Disordered Ice Structures. *J. Chem. Inf. Model.*, 61(6):2542–2546, 2021.

[285] J. D. Bernal and R. H. Fowler. A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. *J. Chem. Phys.*, 1:515, 1933.

[286] S. Nanayakkara, Y. Tao, and E. Kraka. Capturing Individual Hydrogen Bond Strengths in Ices via Periodic Local Vibrational Mode Theory: Beyond the Lattice Energy Picture. *J. Chem. Theory Comput.*, 18(1):562–579, 2022.

[287] J. G. Brandenburg, T. Maas, and S. Grimme. Benchmarking DFT and semiempirical methods on structures and lattice energies for ten ice polymorphs. *J. Chem. Phys.*, 142(12):124104, 2015.

[288] J. M. Herbert. Fantasy versus reality in fragment-based quantum chemistry. *J. Chem. Phys.*, 151(17):170901, 2019.

[289] R. M. Richard, K. U. Lao, and J. M. Herbert. Aiming for benchmark accuracy with the many-body expansion. *Acc. Chem. Res.*, 47(9):2828–2836, 2014.

[290] P. A. Giguère. On the anomalous thermodynamic properties of ice. *J. Phys. Chem. Solids*, 11(3):249–256, 1959.

[291] B. Pamuk, J. M. Soler, R. Ramírez, C. P. Herrero, P. W. Stephens, P. B. Allen, and M.-V. Fernández-Serra. Anomalous Nuclear Quantum Effects in Ice. *Phys. Rev. Lett.*, 108(19):193003, 2012.

[292] E. S. Gaffney and D. L. Matson. Water ice polymorphs and their significance on planetary surfaces. *Icarus*, 44(2):511–519, 1980.

[293] J. Klinger. Extraterrestrial Ice. A Review. *J. Phys. Chem*, 87:4209–4214, 1983.

[294] G. Tammann. Ueber die Grenzen des festen Zustandes IV. *Ann. Phys.*, 307(5):1–31, 1900.

[295] R. L. McFarlan. The Structure of Ice III. *J. Chem. Phys.*, 4(4):253–259, 1936.

[296] P. W. Bridgman. The Pressure-Volume-Temperature Relations of the Liquid, and the Phase Diagram of Heavy Water. *J. Chem. Phys.*, 3(10):597–605, 1935.

[297] B. Kamb. Structure of Ice VI. *Science*, 150(3693):205–209, 1965.

[298] R. Howe and R. W. Whitworth. A determination of the crystal structure of ice XI. *J. Chem. Phys.*, 90(8):4450–4453, 1989.