Chemical Accuracy in Modeling Halide Ion Hydration from Many-Body Representations

Francesco Paesani,*,†,#,+ Pushp Bajaj,† Marc Riera†

†Department of Chemistry and Biochemistry, #Materials Science and Engineering, and + San Diego Supercomputer Center, University of California San Diego
La Jolla, CA 92093, United States

* Corresponding author: fpaesani@ucsd.edu

Abstract

Despite the key role that ionic solutions play in several natural and industrial processes, a unified, molecular-level understanding of how ions affect the structure and dynamics of water across different phases remains elusive. In this context, computer simulations can provide new insights that are difficult, if not impossible, to obtain by other means. However, the predictive power of a computer simulation directly depends on the level of “realism” that is used to represent the underlying molecular interactions. Here, we report a systematic analysis of many-body effects in halide-water clusters and demonstrate that the recently developed MB-nrg full-dimensional many-body potential energy functions achieve high accuracy by quantitatively reproducing the individual terms of the many-body expansion of the interaction energy, thus opening the door to realistic computer simulations of ionic solutions.
I. Introduction

A quantitative characterization of the driving forces and molecular mechanisms that regulate ion hydration is key to the microscopic understanding of fundamental processes taking place in aqueous clusters, solutions, and interfaces, which, in turn, have major implications for different fields of science and engineering. For example, charged species are often found as intermediates in chemical reactions and catalytic processes [1, 2]. In biochemistry, ions play a central role in the stabilization of biomolecules [3-5] as well as in mediating protein-protein interactions [6, 7], intracellular signal transduction [8, 9], and enzyme and nucleic acid catalysis [10-13]. In the atmosphere, ionic clusters carry electric currents and are involved in the formation and evolution of aerosol particles [14, 15]. In materials science, ionic solutions are central to many devices such as electrolytic cells, capacitors, and batteries [16].

Due to the long-range nature of Coulomb interactions, ions do not exist as isolated species under equilibrium at ambient conditions in the gas phase, but form neutral ionic aggregates (e.g., ionic salts or, in special cases, ionic liquids). Completely different behavior is observed when ions are dissolved in water. In this case, ion-ion interactions are mediated by the presence of water molecules that allow ions to exist as individual solvated species, up to saturation. The stabilization of individual hydrated ions results from the interplay of energetic contributions associated with ion-water interactions, which counteract the direct Coulomb attraction between ions of opposite charges, and entropic contributions associated with the reorganization of the water hydrogen-bond (H-bond) network required to accommodate the ionic species. Given the central role played by water in mediating many fundamental processes, it is not surprising that much experimental and theoretical work has been and continues to be devoted to the development of a microscopic understanding of ion hydration [17-20].
The first attempt to rationalize specific ion effects on the properties of aqueous solutions dates back to Hofmeister who classified different salts according to their ability to precipitate egg-white proteins and, subsequently, explained his observations in terms of the corresponding strength of hydration [21, 22]. Over the years, Hofmeister’s original explanation has led to the development of the concept of “structure makers” and “structure breakers” to indicate ions that are strongly and weakly hydrated, respectively [23, 24]. Nearly equivalent terms “kosmotropes” and “chaotropes” were later introduced in biophysical contexts to identify “structure makers” and “structure breakers”, respectively [25]. Within this picture, small and highly charged ions are considered as “structure makers”, strengthening the water H-bond network and inducing tetrahedral order. In contrast, large, monovalent ions correspond to “structure breakers”, disrupting the water H-bond network and leading to a more disordered liquid structure. This classification, which finds support in measurements of macroscopic properties (e.g., viscosity and ion mobility), implies that ions that act as “structure makers” can affect the structural properties of aqueous solutions at long range, effectively inducing long-range ordering.

Measurements carried out using different vibrational spectroscopic techniques, including terahertz [26, 27], infrared [18], and Raman [28] spectroscopies as well as X-ray absorption spectroscopy [29, 30], have challenged this picture, suggesting that dissolved ions have no influence on the water H-bond network at long range. However, this interpretation is apparently at odds with infrared photodissociation spectra measured for sulfate ions in water clusters, which indicate that sulfate ions induce H-bond ordering beyond the first hydration shell [31]. Femtosecond elastic second harmonic scattering (fs-ESHS) measurements carried out for highly dilute (from 10 µM to few mM concentration) salt solutions have provided further support for long-range
ion-induced structural ordering by detecting orientational correlations between water molecules at distances up to 10 nm from the ions, which were originally explained invoking ion-dipole and collective H-bond interactions [32]. Alternative interpretations of these apparent long-range effects have recently been proposed from molecular dynamics (MD) simulations [33, 34].

One of the most striking examples of present uncertainties associated with our understanding of specific ion effects is perhaps represented by the debate around the distribution of ions at the air/water interface [17, 20]. Measurements carried out by Heydweiller more than 100 years ago showed that the surface tension of electrolyte solutions is larger than that of pure water, and displays a strong dependence on the nature of the anions but varies little as a function of the cations [35]. This trend, which follows a reverse Hofmeister series, was independently explained by Wagner [36], and Onsager and Samaras [37] by considering the effects of image charges which cause a depletion of ions in the interfacial region and thus account for the observed increase in surface tension. Subsequent experiments, however, found that the electrostatic potential difference across the air/water interface measured for solutions of halide salts (with the exception of fluoride salts) is more negative than that for pure water, thus suggesting that larger halide ions have higher propensity for the interface than cations, in contradiction with predictions derived from Wagner, and Onsager and Samaras theories [38].

Renewed interest in specific ion effects at aqueous interfaces was generated by the development of both surface sensitive experimental techniques. In particular, vibrational sum-frequency generation (vSFG) spectra measured for NaF and NaCl solutions were found to be very similar to those of neat water while appreciable differences were found in the analogous spectra of NaBr and NaI solutions [39]. Since
the vSFG spectra are sensitive to changes in the water H-bond network, the results of Ref. [39] were interpreted as an indication of the presence of Br- and I- ions in the interfacial region. Opposite conclusions were, however, reached from analogous vSFG studies in Ref. [40] which indicated “a significantly diminished population of the anions at the uppermost layer of the surface region.” Subsequent analyses of second harmonic generation (SHG) spectra in terms of Langmuir adsorption isotherms as well as photoelectron (PE) measurements of anion/cation ratios as a function of photoelectron kinetic energy provided further support for the adsorption of larger halide ions at the air/water interface [41].

On the theoretical side, pioneering MD simulations of halide-water clusters, carried out with polarizable force fields (FFs), found that all halide ions except fluoride had a preference for the surface of the clusters [42-45]. Analogous simulations carried out with nonpolarizable FFs predicted anions to always be in the interior of the clusters, thus emphasizing the importance of many-body effects in ion-water interactions [46-48]. Subsequent MD simulations of concentrated electrolyte solutions as well as calculations of single-ion potentials of mean force (PMFs) carried out in slab geometries using polarizable FFs found a significant adsorption of the larger halide ions at the interface, with the surface propensity increasing from Cl- to I- while F- as well as Na+ ions were repelled from the interface [49, 50]. Qualitatively similar predictions were made from analogous MD simulations carried out with different polarizable models [51-53]. In an attempt to reconcile Wagner, and Onsager and Samaras theory with recent surface-sensitive measurements and computer simulations, an extension of the dielectric continuum (DC) theory has been proposed which takes into account both the dimension and the polarizability of the ions [54]. Within the extended DC theory, it is predicted that highly polarizable ions are indeed adsorbed at the water surface,
although to a much less extent than found in MD simulations with polarizable FFs. An alternative explanation, emphasizing the impact that ions may have on surface fluctuations, has also been proposed [55]. More recently, MD simulations based on density functional theory (DFT) reported a much lower propensity of iodide for the water surface compared to predictions obtained with polarizable FFs [56]. However, no further simulations for other ions have so far been reported to provide support for similar differences between polarizable FFs and DFT models.

While much progress has been made over the last two decades in identifying different factors that affect ion hydration in different environments, it has become increasingly apparent that any further step towards a unified and quantitative theory of ion hydration requires an accurate description of the underlying molecular interactions. In this article we systematically investigate many-body effects in halide-water interactions and provide a quantitative assessment of the accuracy of recently developed many-body (MB) potential energy functions (PEFs) in comparison with polarizable FFs and various DFT models that are commonly employed in computer simulations of electrolyte solutions. With this objective in mind, in Section 2 we introduce the many-body expansion (MBE) of the interaction energy and provide a brief historical perspective on its application to aqueous systems. Section 3 briefly overviews recently developed MB PEFs. A systematic analysis of many-body effects in halide hydration is presented in Section 4 along with detailed comparisons between MB PEFs, polarizable FFs, and various DFT models. Conclusions and outlook are given in Section 5.

2. Many-body expansion of the interaction energy

The energy of a system containing N (atomic or molecular) monomers can be formally expressed through the MBE as [57]
\[ V_N(r_1, ..., r_N) = \sum_{i=1}^{N} V^{1B}(r_i) + \sum_{i<j}^{N} V^{2B}(r_i, r_j) + \sum_{i<j<k}^{N} V^{3B}(r_i, r_j, r_k) + \ldots + V^{NB}(r_1, r_2, r_3, ..., r_N) \]  

(1)

where \( r_i \) collectively denotes the coordinates of all atoms within the \( i^{th} \) monomer. In Eq. (1), \( V^{1B} \) is the one-body (1B) potential representing the energy required to deform an individual monomer from its equilibrium geometry. Therefore, by construction, \( V^{1B}(i) = 0 \) for atomic monomers, and \( V^{1B}(i) = E(i) - E_{eq}(i) \) for molecular monomers, where \( E(i) \) and \( E_{eq}(i) \) are the monomer energies for distorted and equilibrium configurations, respectively. All higher n-body (nB) interaction terms \( (V^{nB}) \) in Eq. (1) are defined recursively through

\[ V^{nB}(r_1, ..., r_n) = V_n(r_1, ..., r_n) - \sum_{i=1}^{N} V^{1B}(r_i) - \sum_{i<j}^{N} V^{2B}(r_i, r_j) - \ldots - \sum_{i<j<k}^{N} V^{(n-1)B}(r_i, r_j, r_k) \]  

(2)

The importance of many-body effects in water-water and ion-water interactions was already recognized in the 1950s by Frank and Wen who introduced a molecular picture of water consisting of “flickering clusters of hydrogen-bonded molecules”, emphasizing the “cooperative nature” of hydrogen bonding [58]. Pioneering electronic structure calculations carried out on small aqueous clusters demonstrated that nonadditive contributions to the interaction energy beyond the pairwise additive 2B term are not negligible [59, 60]. In particular, it was found that the 3B term can contribute up to \(~15 - 20\%\) of the interaction energy of cyclic water structures, while 4B contributions represent, on average, \(~1\%\) of the interaction energy in generic water clusters [61-71].

Due to the rapid convergence of the MBE for non-metallic systems, Eq. 1 also provides a rigorous and efficient framework for the development of full-dimensional PEFs in which individual terms of the MBE can be determined using different
electronic structure methods. Most empirical FFs describing water-water and ion-water interactions are pairwise additive (i.e., the MBE is truncated at the 2B term) and attempt to account for many-body contributions in an average fashion by using an effective $V^{2B}$ [72-83]. Although in the early times of MD simulations this simplification was a necessity dictated by computational efficiency, it was soon recognized that “pair potentials do not realistically reproduce both gas and condensed phase water properties” [84]. The first attempts to derive PEFs for aqueous systems which could rigorously represent the individual terms of the MBE were made in the late 1970s and 1980s [85-90]. In particular, Clementi and coworkers developed a series of analytical PEFs for water which were fitted to ab initio reference data obtained at the fourth-order Møller-Plesset (MP4) and Hartree-Fock levels of theory for the 2B and 3B terms, respectively, and represented many-body effects implicitly through a classical polarization term [89, 90]. Stillinger and David developed a polarizable model for water in which treats $\text{H}^+$ and $\text{O}^{2-}$ were considered as the basic dynamical and structural elements [91]. Building upon these pioneering studies, several polarizable FFs for water have been developed over the years (see Ref. [92] for a recent review).

Nonadditive effects in ion-water interactions were first investigated by Kollman and co-workers who developed a series of polarizable FFs specifically parameterized to reproduce several experimental data [84]. These polarizable FFs were subsequently used in MD simulations of small $\text{Cl}^-(\text{H}_2\text{O})_n$ and $\text{Na}^+(\text{H}_2\text{O})_n$ clusters, which provided the first indication of the propensity of polarizable halide ions for the water surface [42-45]. Several empirical polarizable FFs with different degrees of sophistication were then developed and shown to reproduce, at least to some extent, several properties of ion-water systems, from small clusters to bulk solutions and interfaces [46-48, 93]. Although these studies emphasized the importance of polarization in correctly
representing ion-water interactions, especially in the case of large anions, it was shown that predictions of MD simulations carried out with polarizable FFs depend sensitively on the type of electrostatic dampening adopted in the models, which can often lead to overpolarization of the ions at short ion-water distances [94-96].

The development of efficient algorithms for correlated electronic structure methods along with continued improvements in computer performance has recently made possible to determine, with chemical accuracy, individual (low-order) terms of Eq. 1. In parallel, tremendous progress has been made in constructing multidimensional mathematical functions that are capable to reproduce interaction energies in generic N-molecule systems, with high fidelity [97-99]. By combining these three factors, it has been realized that the MBE effectively provides a rigorous and efficient framework for the development of full-dimensional PEFs entirely from “first principles”, in which the low-order terms are accurately determined from correlated electronic structure data (e.g., using coupled cluster theory with single, double, and perturbative triple excitations, CCSD(T), in the complete basis set, CBS, limit, the current “gold standard” for chemical accuracy) and the higher-order terms are represented implicitly by classical many-body induction. Along these lines, several MB PEFs for water have been proposed in the last decade, the most notable of which are CC-pol [100], WHBB [101], HBB2-pol [102], and MB-pol [103-105]. When employed in computer simulations that allow for explicit treatment of nuclear quantum effects, these many-body PEFs have been shown to correctly predict structural, thermodynamic, dynamical, and spectroscopic properties of water, from the dimer in the gas phase to liquid water and ice (see Ref. [92] for a recent review). Among existing many-body PEFs for water, MB-pol has been shown to correctly predict the vibration-rotation tunneling spectrum of the water dimer [103], the energetics, quantum equilibria, and infrared spectra of small
clusters [104, 106-108], the structural, thermodynamic, and dynamical properties of liquid water [105, 109], the energetics of different ice phases [110], infrared and Raman spectra of liquid water [111-113], the vibrational sum-frequency generation spectrum of the air/water interface [114, 115], the infrared and Raman spectra of ice Ih [116]. More recently, molecular configurations extracted from classical (MD) and quantum path-integral molecular dynamics (PIMD) simulations with MB-pol have been used to determine the electronic band gap of liquid water, both in the bulk and at the air/water interface, through many-body perturbation theory electronic structure calculations [117].

3. Many-body potential energy functions for halide-water interactions

Building upon the accuracy and computational efficiency demonstrated by MB-pol in predicting the properties of water across different phases [118], the MBE in Eq. 1 has recently been applied to the development of two families of MB PEFs (TTM-nrg for “Thole-type model energy” [119] and MB-nrg for “many-body energy” [120]) describing halide-water interactions. Both TTM-nrg and MB-nrg PEFs adopt MB-pol to describe all water properties (i.e., water monomer distortion, dipole moment, and polarizability, as well as water-water interactions) [118] and have been derived from CCSD(T)-F12b/CBS data for halide-water dimers and halide-water-water trimers but differ in the functional forms employed to represent halide-water many-body effects.

In the TTM-nrg PEFs, halide-water interactions are described by four terms,

\[ V_{TTM-nrg} = V_{TTM} + V_{rep} + V_{disp} \]  

(3)

where \( V_{TTM} \) represents permanent and induced electrostatic interactions, with the induction term being derived from the extended Thole-type model originally introduced in Ref. [121]. The repulsive term, \( V_{rep} = \sum_i V_{rep}^i \), is represented by a sum of pairwise Born-Mayer functions \( (V_{rep}^i) \) between the halide ion (X) and the oxygen (O) and
hydrogen (H₁ and H₂) atoms of the \(i^{th}\) water molecule in the system,

\[
V_{rep}^i = A_{XO}e^{-b_{XO}R_{XO}} + A_{XH}e^{-b_{XH}R_{XH}} + A_{XH}e^{-b_{XH}R_{XH}}
\]  

(4)

where \(R_{XY}\) are interatomic distances, and \(A_{XY}\) and \(b_{XY}\) are fitting parameters, with \(Y = O, H_1, \) and \(H_2\). Similarly, the dispersion energy, \(V_{disp}\), is also represented through a sum of pairwise additive contributions, \(V_{disp} = \sum_i V_{disp}^i\),

\[
V_{disp}^i = - f_6(\delta_{XO}) \frac{C_{6,XO}}{R_{XO}^6} - f_6(\delta_{XH}) \frac{C_{6,XH}}{R_{XH_1}^6} - f_6(\delta_{ion-H2}) \frac{C_{6,XH}}{R_{XH_2}^6}
\]  

(5)

where \(f_6(\delta_{XY})\) are Tang-Toennies damping functions [122] with coefficients \(\delta_{XY}\), and \(C_{6,XY}\) are dispersion coefficients calculated using the exchange-hole dipole model (XDM) [123].

In the MB-nrg PEFs, the interactions between individual halide ions and water molecules are described through the MBE of Eq. 1 and include explicit 2B (ion-water) and 3B (ion-water-water) terms, with all higher-order interactions being implicitly taken into account through many-body induction as in the TTM-nrg model. The 2B term of the MB-nrg PEFs includes three contributions,

\[
V^{2B} = V^{2B}_{short} + V^{2B}_{TTM} + V^{2B}_{disp}
\]  

(6)

where the electrostatic term \((V^{2B}_{TTM})\) corresponds to the 2B component of \(V_{TTM}\) in Eq. 3 and \(V^{2B}_{disp}\) is the same sum of pairwise additive contributions as in the TTM-nrg model (Eq. 5). \(V^{2B}_{short}\) is expressed by a 4th-degree permutationally invariant polynomial [97] in variables that are functions of the distances between the ion and the six sites of an MB-pol water molecule. Similarly, the 3B term of the MB-nrg PEFs includes two contributions,

\[
V^{3B} = V^{3B}_{short} + V^{3B}_{TTM}
\]  

(7)

where the electrostatic term \((V^{3B}_{TTM})\) corresponds to the 3B component of \(V_{TTM}\) in Eq. 3 and \(V^{3B}_{short}\) is expressed by a 4th-degree permutationally invariant polynomial [97] in
variables that are functions of the distances between the ion and the six sites of an MB-pol water molecule. The coefficients of both 2B and 3B permutationally invariant polynomials are optimized using Tikhonov regression [124] (also known as ridge regression) and supervised learning [125] to reproduce the interaction energies calculated at the CCSD(T)-F12b level of theory in the CBS limit [120, 126]. It can be demonstrated that 2B and 3B MB-nrg permutationally invariant polynomials effectively recovers quantum-mechanical effects that cannot be described by purely classical expressions (e.g., charge transfer and penetration, and Pauli repulsion) employed in classical polarizable FFs.

4. Many-body effects in halide ion hydration

To provide a systematic analysis of many-body effects in halide ion hydration at the fundamental level, the accuracy of both TTM-nrg and MB-nrg PEFs is assessed through extensive comparisons with reference data for individual terms of the MBE calculated at the CCSD(T)/CBS level as well as with corresponding values calculated with various DFT models that are commonly used in MD simulations of ionic solutions. Also included in these comparisons are results obtained with the AMOeba polarizable FF [127-130].

4.1 Molecular models and computational details

All reference geometries of the halide-water clusters used in the comparisons shown in the following sections were optimized at the DF-MP2/aug-cc-pVTZ(-PP) level of theory using a strict gradient convergence threshold of $10^{-6}$ a.u. Effective core potentials (PP) with 10 and 28 electrons were used for bromide and iodide ions, respectively. For each cluster, all energies were calculated using the stratified approximation many-body (SAMBA) approach [71], using a hierarchy of theory levels
for the individual terms in the MBE. Specifically, the MBE terms were calculated using explicitly correlated coupled cluster with single, double and perturbative triple excitations, CCSD(T)-F12b [131, 132], method with successively smaller basis sets for higher-body interactions [133-136]. The CBS limit for the two-body term was achieved by extrapolating the values obtained with the aug-cc-pVTZ(-PP) and aug-cc-pVQZ(-PP) basis sets using a two-point extrapolation [71, 137] as in Ref. [120]. The three-body interaction energies were calculated using the aug-cc-pVTZ(-PP) basis sets and corrected for basis set superposition error (BSSE) using the counterpoise method [138]. The four-body and higher-order terms were calculated using the aug-cc-pVTZ(-PP) basis sets without correcting for BSSE. All geometry optimizations and single point energy calculations were performed using the MOLPRO quantum chemistry software package [139].

For comparison with TTM-nrg and MB-nrg, various DFT models, with and without dispersion contributions [140], are considered in this analysis. GGA (Generalized Gradient Approximation) functionals: BLYP(-D3) [141, 142], PBE(-D3) [143], and revPBE(-D3) [144]. Meta-GGA functionals: TPSS(-D3) [145] and SCAN(-D3) [146]. Hybrid GGA functionals: PBE0(-D3) [147], revPBE0(-D3) [144, 147], B3LYP(-D3) [148]. Meta-hybrid functionals: M06-2X(-D3) [149]. Range-separated, meta-hybrid functionals: ωB97X [150], ωB97X-D [150], and ωB97M-V [151]. All DFT calculations were carried out with the aug-cc-pVQZ basis using Gaussian 09 [152], except those for revPBE(-D3), SCAN(-D3), revPBE0(-D3), and ωB97M-V that were carried out with Q-Chem [153].

Finally, results obtained with AMOEBA, a generic polarizable FF for biomolecular simulations [127-130], are also included in the comparisons. AMOEBA adopts an induction model with distributed atomic polarizabilities up to the quadrupole
moment based on a Thole’s damping scheme and was originally parameterized using both experimental and ab initio data. AMOEBA is under continued development and new versions have recently been proposed [129, 130]. All AMOEBA calculations presented were performed with the TINKER package [154], using the AMOEBA 2009 force field [128], which includes the 2003 parameterizations for both water and halide ions [127].

4.2 $X\,(H_2O)$: Fundamental 2B halide-water interactions

The minimum energy structures predicted by the MB-nrg PEFs for the four $X\,(H_2O)$ dimers are shown in Figure 1. Although in all cases the halide ions are found to preferentially form one H-bond with the water molecule, the structural parameters vary appreciably going from $F\,(H_2O)$ to $I\,(H_2O)$. In particular, the H-bond increases in length from 1.39 Å to 2.68 Å and becomes less linear, while the HOH angle of the water molecule becomes progressively narrower. These differences can be explained by considering the relatively more pronounced molecular character of HF (or higher basicity of $F^-\,$) compared to that of other hydrogen halides, which results in more directional and covalent-like interactions between $F^-\,$ and $H_2O$. On the other hand, due to lower and more diffuse charge density, the interactions of iodide with water are primarily electrostatic in nature and dominated by polarization effects.

Figure 1. Minimum energy geometries for $X\,(H_2O)$ dimers, with $X = F$, Cl, Br, and I.
The binding energies (E_{bind}) and harmonic frequencies for the bending (ν_{bend}),
and H-bonded (ν_{H-bond}) and free OH (ν_{free}) stretching modes of the water molecule
computed with both TTM-nrg and MB-nrg PEFs for the four X(H_2O) dimers of Fig. 1
are compared in Table 1 with the corresponding CCSD(T)-F12/aug-cc-pVQZ(-PP)
values. In all cases, the values predicted by the MB-nrg PEFs are in quantitative
agreement with the reference data. On the other hand, noticeable deviations are found
with the TTM-nrg PEFs, which are more pronounced in F(H_2O) and progressively
decrease as ion’s size and polarizability increase (and charge density decreases). These
differences can be attributed to intrinsic deficiencies associated with the TTM-nrg PEFs
which, as all common polarizable FFs, are unable to correctly describe charge transfer
and penetration effects through purely classical expressions.

Table 1. Comparison of MB-nrg and TTM-nrg binding energies (in kcal/mol) and
harmonic frequencies (in cm\(^{-1}\)) calculated for X(H_2O) dimers, with X = F, Cl, Br,
and I, with reference data obtained at the CCSD(T)-F12/aug-cc-pVQZ(-PP) level of
theory.

<table>
<thead>
<tr>
<th></th>
<th>F(H_2O)</th>
<th>Cl(\Gamma)(H_2O)</th>
<th>Br(H_2O)</th>
<th>I(H_2O)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CCSD(T)-F12 / aug-cc-pVQZ(-PP)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E_{bind})</td>
<td>-27.5429</td>
<td>-14.9352</td>
<td>-12.8941</td>
<td>-10.7151</td>
</tr>
<tr>
<td>(ν_{bend})</td>
<td>1723.04</td>
<td>1692.21</td>
<td>1685.52</td>
<td>1679.73</td>
</tr>
<tr>
<td>(ν_{H-bond})</td>
<td>2263.58</td>
<td>3423.01</td>
<td>3528.03</td>
<td>3629.75</td>
</tr>
<tr>
<td>(ν_{free})</td>
<td>3883.40</td>
<td>3892.49</td>
<td>3891.43</td>
<td>3888.11</td>
</tr>
<tr>
<td><strong>MB-nrg</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E_{bind})</td>
<td>-27.2548</td>
<td>-14.7367</td>
<td>-12.7575</td>
<td>-10.6048</td>
</tr>
<tr>
<td>(ν_{bend})</td>
<td>1716.30</td>
<td>1688.96</td>
<td>1683.96</td>
<td>1679.00</td>
</tr>
<tr>
<td>(ν_{H-bond})</td>
<td>2264.61</td>
<td>3416.98</td>
<td>3532.19</td>
<td>3641.66</td>
</tr>
<tr>
<td>(ν_{free})</td>
<td>3893.35</td>
<td>3893.84</td>
<td>3894.06</td>
<td>3888.71</td>
</tr>
<tr>
<td><strong>TTM-nrg</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E_{bind})</td>
<td>-27.8798</td>
<td>-14.2776</td>
<td>-12.1804</td>
<td>-10.0466</td>
</tr>
<tr>
<td>(ν_{bend})</td>
<td>1703.23</td>
<td>1699.26</td>
<td>1697.31</td>
<td>1695.54</td>
</tr>
<tr>
<td>(ν_{H-bond})</td>
<td>3100.39</td>
<td>3546.31</td>
<td>3641.33</td>
<td>3716.10</td>
</tr>
<tr>
<td>(ν_{free})</td>
<td>4001.25</td>
<td>3950.34</td>
<td>3939.26</td>
<td>3927.66</td>
</tr>
</tbody>
</table>
It should be noted that fully full-dimensional vibrational spectra calculated for the four \( \text{X}(\text{H}_2\text{O}) \) dimers and the corresponding isotopologues, \( \text{X}(\text{D}_2\text{O}) \), at the quantum-mechanical level with both TTM-nrg and MB-nrg PEFs have recently been reported [155]. In all cases, the MB-nrg vibrational spectra are in close agreement the the available experimental data, correctly reproducing frequency shifts and tunneling splittings, while significant deviations from the experimental values are found with the TTM-nrg PEFs, which again emphasize the difficulties of purely classical representations in correctly describing halide-water interactions.

Figure 2: Comparisons between RMSDs associated with various DFT models, TTM-nrg and MB-nrg PEFs, and AMOEBA polarizable FF relative to CCSD(T)-F12b/CBS interaction energies \( (E_{\text{ref}}) \) calculated for \( \text{X}^-(\text{H}_2\text{O}) \) dimers, with \( \text{X} = \text{F}, \text{Cl}, \text{Br}, \text{and I} \), in the DF-MP2/aug-cc-pVTZ(-PP) optimized geometries shown on the left.
The accuracy of both TTM-nrg and MB-nrg PEFs in reproducing 2B halide-water interactions is further investigated through comparisons with various DFT models as well as with the AMOEBA polarizable FF. Specifically, Figure 2 shows comparisons of deviations relative to CCSD(T)-F12b/CBS interaction energies of the four X−(H2O) dimers calculated for the DF-MP2/aug-cc-pVTZ(-PP) optimized geometries. DFT models, from GGA to range-separated hybrid functionals, are on the left side, while the AMOEBA polarizable FF is on the right side, with TTM-nrg and MB-nrg PEFs in between. As a reference also shown as a dashed line is the ±1 kcal/mol threshold that is generally used to define “chemical accuracy”. It should be noted that, in the case of the X−(H2O) dimers, the interaction energies exactly correspond to the 2B term in Eq. 1. As discussed in Ref. [120], a large variability is observed among different DFT models, with GGA and long-range separated hybrid functionals respectively providing the poorest and the best agreement with the CCSD(T)-F12b/CBS reference data.

From the analysis of Fig. 2, it is possible to derive general trends that provide guidance in the assessment of different molecular models currently available for halide-water interactions. Overall, the MB-nrg PEF provides the most accurate description of the interaction energies of all X−(H2O) dimers, with the largest deviation being 0.18 kcal/mol for F−(H2O). BLYP and revPBE underestimate the interaction energies of all dimers and the deviations are always larger than 1 kcal/mol, with the exception of I−(H2O) for BLYP. Inclusion of the empirical D3 dispersion correction improves the accuracy of both functionals although the description of F−(H2O) still remains outside the boundaries of “chemical accuracy”. Opposite trend is found with the PBE functional that predicts interaction energies within “chemical accuracy” for all dimers and becomes less accurate after inclusion of the empirical D3 dispersion correction, which pushes the deviation for I−(H2O) beyond the 1 kcal/mol threshold. Among the meta-
GGA functionals, TPSS predictions are always within 1 kcal/mol while SCAN overestimates the interaction energies of all dimers. In both cases, inclusion of the D3 dispersion correction appears to have deleterious effects on the ability of these functionals to reproduce 2B interactions. B3LYP, with and without the D3 dispersion correction, is the best performer among all hybrid functionals considered in Fig. 2, while PBE0 is unable to correctly predict the interaction energy of F(H₂O). The D3 dispersion correction has a significant effect on the accuracy of revPBE0, bringing the interaction energy within 1 kcal/mol of the CCSD(T)/CBS reference data for all dimers. Similar to PBE0, the meta-GGA and hybrid M062X functional is not able to correctly reproduce the interaction energy in the F(H₂O) dimer. Interestingly, the D3 dispersion correction has little effect on the performance of M062X. All meta-GGA, hybrid, and long-range separated functionals accurately reproduce the CCSD(T)/CBS reference data, with ωB97M-V consistently providing similar deviations independently of the halide ion.

On the other hand, interaction energies calculated with the TTM-nrg PEF and the AMOEBA polarizable FF are within “chemical accuracy” for all dimers except for F(H₂O), for which both models significantly underestimate the interaction strength. These deviations can be attributed, at least qualitatively, to the inability of purely classical MB induction schemes adopted by polarizable models to correctly account for quantum-mechanical effects such as charge transfer and penetration, and Pauli repulsion, which are emphasized in F(H₂O) by the covalent-like nature of the interactions at short and medium ranges.

4.3 X(H₂O)₂: Importance of 3B contributions
Halide dihydrates, X(H₂O)₂, are the smallest halide-water systems in which two water molecules can, in principle, be simultaneously H-bonded to each other and to the ion,
thus representing ideal systems for characterizing the interplay between ion-water and water-water many-body contributions to the total interaction energies. As shown in Figure 3, however, the two water molecules are H-bonded to each other in the minimum energy structures of the X−(H2O)2 complexes only in the presence of chloride, bromide, and iodide. This can be explained by considering the different size, charge density, and polarizability of each ion, which, directly affecting the relative strengths of halide-water and water-water H-bonds, progressively stabilize the cyclic configuration, going from F− to I− [120, 155].

Deviations from CCSD(T)/CBS interaction energies calculated for the minimum energy structures of the X−(H2O)2 complexes also shown in Figure 3 follow similar trends seen for the dimers in Section 4.2. Specifically, the MB-nrg PEFs consistently provide the smallest deviations, while GGA and range-separated, hybrid functionals, including dispersion contributions, respectively provide the least and most accurate descriptions of the energetics among all DFT models considered in this analysis. From a quantitative point of view, a larger number of DFT models are unable to achieve chemical accuracy than in the case of the X−(H2O) dimers, indicating that halide-water and water-water interactions are effectively described with different levels of accuracy by the same functionals. Finally, both the TTM-nrg PEF and the AMOEBA polarizable FF exhibit larger deviations, on the order of ∼1 kcal/mol, for all X−(H2O) complexes except F−(H2O)2, for which TTM-nrg predicts an interaction energy in nearly quantitative agreement with the CCSD(T)/CBS reference data while AMOEBA underestimates the interaction strength by ∼6 kcal/mol. The different performance of TTM-nrg and AMOEBA in describing F−(H2O)2 is certainly unexpected since both models represent fluoride-water interactions through similar classical expressions based on many-body induction schemes. These differences thus highlight both difficulties and
range of variability associated with polarizable models when applied to molecular interactions that are more quantum-mechanical in nature as $\text{F}^-\text{(H}_2\text{O})_2$.

Figure 3: Comparisons between RMSDs associated with various DFT models, TTM-nrg and MB-nrg PEFs, and AMOEBA polarizable FF relative to CCSD(T)-F12b/CBS interaction energies ($E_{\text{ref}}$) calculated for X (H$_2$O)$_2$ complexes, with X = F, Cl, Br, and I, in the DF-MP2/aug-cc-pVTZ(-PP) optimized geometries shown on the left.
To provide fundamental insights into the performance of different models in representing halide-water systems, the interaction energies of the $X^-(\text{H}_2\text{O})_2$ complexes are further decomposed into individual 2B and 3B contributions associated with ion-water and water-water interactions. The comparisons shown in Figure 4 demonstrate that all models describe 2B water-water interactions relatively more accurately than 2B ion-water interactions, with the latter thus being responsible for the largest fractions of the total deviations from the CCSD(T)-F12b/CBS reference data. Noticeable exceptions are the MB-nrg PEFs and, to some extent, the ωB97M-V functional that represent both 2B ion-water and water-water interactions with similar accuracy. Importantly, some functionals (e.g., BLYP-D3, revPBE-D3, PBE0, revPBE0-D3, and ωB97X-D) appear to benefit from fortuitous error cancellation between 2B ion-water and water-water interactions.

Figure 4 also shows that all models represent 3B ion-water-water interaction within chemical accuracy, with AMOEBA and TTM-nrg effectively providing deviations comparable to those associated with the DFT models, except for $F^-(\text{H}_2\text{O})_2$. In this case, the comparison between TTM-nrg and MB-nrg predictions is particularly insightful. As discussed in Section 3, both models share the same Thole-type induction model with the difference being that this classical term is supplemented in the MB-nrg PEFs by an explicit short-range 3B term expressed as a multidimensional permutationally invariant polynomial (Eq. 7). As shown in Figure 4, the inclusion of this explicit 3B term allows MB-nrg PEFs to reproduce, nearly quantitatively, the CCSD(T)/CBS reference data, independently of the halide ion. It is also important to note that the 3B ion-water-water deviations associated with both AMOEBA and TTM-nrg systematically decrease from $F^-(\text{H}_2\text{O})_2$, for which they are significantly larger than 1 kcal/mol, to $I^-(\text{H}_2\text{O})_2$. This trend indicates that 3B contributions due to the quantum-
mechanical nature of halide-water interactions (e.g., charge transfer and penetration, and Pauli repulsion), which cannot be correctly described by purely classical many-body induction schemes adopted by polarizable FFs, such as AMOEBA and TTM-nrg, play an important role in $F(H_2O)_2$.

4.4 $X(H_2O)_{n=3,4}$ clusters: Assessment of higher-body contributions

After assessing 2B and 3B effects, in this section we investigate the role played by 4B
and higher-order terms of the MBE in Eq. 1. A comparison between the performance of different models in reproducing the interaction energies of $X^-(\text{H}_2\text{O})_3$ clusters is first analyzed in Figure 5a. As for the smallest systems, the MB-nrg results are again in close agreement with the CCSD(T)-F12b/CBS reference data, which provides further support for the accuracy of the MB-nrg PEFs. On the other hand, the performance of various DFT models varies significantly depending on the type of functional and nature of the ion. In particular, both positive and negative deviations on the order of several kcal/mol are observed for BLYP, revPBE, SCAN, SCAN-D3, B3LYP, and revPBE0. With the exception of ωB97M-V, whose performance is close to that of the MB-nrg PEFs, all other functionals are associated with deviations larger than 1 kcal/mol at least for one cluster type, with the accuracy of most functionals improving going from $F^-$ to $\Gamma$.

Defined trends can be established from the analysis of individual contributions to the total interaction energies shown in Figs 5b-d. Several functionals benefit from fortuitous error cancellation between halide-water and water-water contributions. Hybrid functionals overall perform consistently better than GGA functionals. As expected from the analyses of smaller complexes reported in Sections 4.2 and 4.3, the inclusion of the empirical D3 dispersion correction has unpredictable effects on the performance of different functionals in reproducing 2B ion-water and water-water interactions. Deviations on the order of ~1 kcal/mol are in 3B contributions are associated with all functionals, except revPBE0, revPBE0-D3, M06-2X, M06-2X-D3, ωB97X, ωB97X-D, and ωB97M-V. Relatively larger deviations are associated with AMOEBA at all orders.
Figure 5: Comparisons between RMSDs associated with various DFT models, TTM-nrg and MB-nrg PEFs, and AMOEBA polarizable FF relative to interaction energies (a), and individual 2B (b), 3B (c) and 4B (d) contributions calculated at the CCSD(T)-F12b/CBS 2B level of theory for X\(\cdot\)(H\(_2\)O)\(_3\) clusters, with X = F, Cl, Br, and I, in the DF-MP2/aug-cc-pVTZ(-PP) optimized geometries shown in panel (a).
The comparisons shown in Figs. 5b-c provides further evidence for the important role played by the 2B and 3B permutationally invariant polynomials in the MB-nrg PEFs which effectively allow the MB-nrg PEFs to achieve quantitative agreement with the CCSD(T)-F12b/CBS reference data. Importantly, with the exception of AMOEBA results for \( F^- (H_2O)_3 \), Fig. 5d demonstrates that all models considered in this analysis reproduce 4B effects within “chemical accuracy”, with both TTM-nrg and MB-nrg PEFs exhibiting same accuracy as DFT models. This analysis thus indicates that 4B contributions to halide-water interactions are essentially classical-like in nature and can accurately be described through MB induction schemes. Figure 6, which shows the analysis of 5B contributions to the interaction energies in \( X (H_2O)_4 \) clusters, provide further evidence for the classical-like nature of higher-body halide-water interactions, demonstrating that both TTM-nrg and MB-nrg PEFs are able to quantitatively reproduce these effects within a classical Thole-type scheme for MB induction.

5. Conclusions and outlook
The last decade has witnessed tremendous progress in the development of accurate molecular models for computer simulations of aqueous systems across different phases. These efforts have been supported by continued growth in computing power as well as by the development of efficient algorithms for correlated electronic structure methods that nowadays enable efficient calculations of interaction energies of small molecular systems, with chemical accuracy. This progress has been accompanied by the development of several approaches to the analytical representation of multidimensional potential energy surfaces [97-99].

Building upon pioneering studies by Stillinger and coworkers [60], and Clementi and coworkers [85, 90], and recent advances in the modeling of water from the gas to the condensed phase, in this study we have shown that the many-body expansion of the
interaction energy provide a rigorous and efficient approach to the development of analytical potential energy functions for computer simulations of halide-water systems, with chemical accuracy. In particular, through a systematic analysis of the energetics of halide-water clusters, we have demonstrated that the MB-nrg PEFs introduced in Ref. [120] achieve high accuracy by quantitatively reproducing the individual terms of the many-body expansion of the interaction energy by combining explicit short-range representations of 2B and 3B interactions expressed through permutationally invariant polynomials with a physically correct description of many-body electrostatic interactions at all ion-water separations.

Figure 6: Comparisons between RMSDs associated with various DFT models, TTM-nrg and MB-nrg PEFs, and AMOEBA polarizable FF relative to 5B contributions to the interaction energies calculated at the CCSD(T)-F12b/CBS 2B level of theory for X⁻(H₂O)₄ complexes, with X = F, Cl, Br, and I, in the DF-MP2/aug-cc-pVTZ(-PP) optimized geometries shown on the left.
Direct comparisons with various DFT models, ranging from GGA functionals to range-separated, meta and hybrid functionals with explicit inclusion of dispersion contributions, demonstrate that most functionals exhibit significant halide-dependent errors in all terms of the many-body expansion of the interaction energy. Among different DFT models, range-separated, meta and hybrid functionals provide the closest agreement with CCSD(T)-F12b reference data. Importantly, the addition of the empirical D3 dispersion correction appears to not necessarily improve the accuracy of the “parent” functionals. On the other hand, our analysis also demonstrate that classical polarizable force fields suffer from intrinsic deficiencies in their functional forms which prevent them from correctly describing short-range contributions to the interaction energies that are quantum mechanical in nature (e.g., charge transfer and penetration, and Pauli repulsion).

Despite representing a step forward toward realistic computer simulations of ionic solutions, it should always be kept in mind that the MB-nrg PEFs are still “models” for ion-water interactions, with their own approximations and limitations. In particular, the MB-nrg PEFs always assume that each halide ion maintain its own chemical identity throughout the simulation, which implies that associated acid-base equilibrium is completed neglected (i.e., for each halide ion X, the corresponding HX acid is always assumed to be a strong acid). While this is a reasonable approximation for Cl\(^-\), Br\(^-\), and I\(^-\) in pure water, the situation drastically changes for F\(^-\) as well as in complex aqueous solutions with pH values lower or higher than 7. Finally, the application of MB-nrg PEFs in computer simulations of complex ionic solutions will require further theoretical and computational developments. In this context, given the unconventional functional form adopted by the MB-nrg PEFs, which is not supported by popular software for computer simulations, synergistic efforts between
theoretical/computational chemists/physicists and software engineers will be key to the development of specialized and more efficient software for many-body molecular dynamics simulations which can take full advantage of modern hardware.

Acknowledgements

This work was supported by the National Science Foundation through grants Nos. CHE-1453204, CHE-1305427, and CHE-ACI-1642336, and used resources of the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation through grant No. ACI-1053575 under allocation TG-CHE110009 on Comet at the San Diego Supercomputer Center and Stampede2 at the Texas Advanced Computing Center. M.R. was supported by a Software Fellowship from the Molecular Sciences Software Institute, which is funded by the National Science Foundation through grant No. ACI-1547580.

References


[38] A. Frumkin, Zeitschrift Fur Physikalische Chemie--Stochiometrie Und Verwandtschaftslehre 109 (1924) p. 34.


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

A.W. Lange, A. Behn, J. Deng, X.T. Feng, D. Ghosh, M. Goldey, P.R. Horn,
Proynov, Y.M. Rhee, R.M. Richard, M.A. Rohrdanz, R.P. Steele, E.J.
Sundstrom, H.L. Woodcock, P.M. Zimmerman, D. Zuev, B. Albrecht, E.
Alguire, B. Austin, G.J.O. Beran, Y.A. Bernard, E. Berquist, K. Brandhorst,
K.B. Bravaya, S.T. Brown, D. Casanova, C.M. Chang, Y.Q. Chen, S.H. Chien,
Dutoi, R.G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-
Zadorozhnaya, J. Gomes, M.W.D. Hanson-Heine, P.H.P. Harbach, A.W.
Khistyaev, J. Kim, J. Kim, R.A. King, P. Klunzinger, D. Kosenkov, T.
Kowalczyk, C.M. Krauter, K.U. Lao, A.D. Laurent, K.V. Lawler, S.V.
S.F. Manzer, S.P. Mao, N. Mardirossian, A.V. Marenich, S.A. Maurer, N.J.
Mayhall, E. Neuscamman, C.M. Oana, R. Olivares-Amaya, D.P. O'Neill, J.A.
Parkhill, T.M. Perrine, R. Peverati, A. Prociuk, D.R. Rehn, E. Rosta, N.J. Russ,
S.M. Sharada, S. Sharma, D.W. Small, A. Sodt, T. Stein, D. Stuck, Y.C. Su,
A.J.W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang,
M.A. Watson, J. Wenzel, A. White, C.F. Williams, J. Yang, S. Yeganeh, S.R.
Yost, Z.Q. You, I.Y. Zhang, X. Zhang, Y. Zhao, B.R. Brooks, G.K.L. Chan,
D.M. Chipman, C.J. Cramer, W.A. Goddard, M.S. Gordon, W.J. Hehre, A.
Klamt, H.F. Schaefer, M.W. Schmidt, C.D. Sherrill, D.G. Truhlar, A. Warshel,
X. Xu, A. Aspuru-Guzik, R. Baer, A.T. Bell, N.A. Besley, J.D. Chai, A. Drewu,
B.D. Dunietz, T.R. Furlani, S.R. Gwaltney, C.P. Hsu, Y.S. Jung, J. Kong, D.S.
Lambrecht, W.Z. Liang, C. Ochsenfeld, V.A. Rassolov, L.V. Slipchenko, J.E.
Subotnik, T. Van Voorhis, J.M. Herbert, A.I. Krylov, P.M.W. Gill, M. Head-
Gordon, Advances in molecular quantum chemistry contained in the Q-Chem