

Article

A Benchmark Protocol for DFT Approaches and Data-Driven Models for Halide-Water Clusters

Raúl Rodríguez-Segundo ^{1,2,†}, Daniel J. Arismendi-Arrieta ³  and Rita Prosimi ^{1,*} 

¹ Institute of Fundamental Physics (IFF-CSIC), Consejo Superior de Investigaciones Científicas, Serrano 123, 28006 Madrid, Spain; raul@iff.csic.es

² Atelgraphics S.L., Mota de Cuervo 42, 28043 Madrid, Spain

³ Department of Chemistry, Ångström Laboratory, Uppsala University, P.O. Box 538, 75121 Uppsala, Sweden; daniel.arismendi@kemi.uu.se

* Correspondence: rita@iff.csic.es; Tel.: +34-91-5616800 (ext. 442292)

† Doctoral Programme in Theoretical Chemistry and Computational Modelling, Doctoral School, Universidad Autónoma de Madrid, 28049 Madrid, Spain.

Abstract: Dissolved ions in aqueous media are ubiquitous in many physicochemical processes, with a direct impact on research fields, such as chemistry, climate, biology, and industry. Ions play a crucial role in the structure of the surrounding network of water molecules as they can either weaken or strengthen it. Gaining a thorough understanding of the underlying forces from small clusters to bulk solutions is still challenging, which motivates further investigations. Through a systematic analysis of the interaction energies obtained from high-level electronic structure methodologies, we assessed various dispersion-corrected density functional approaches, as well as ab initio-based data-driven potential models for halide ion–water clusters. We introduced an active learning scheme to automate the generation of optimally weighted datasets, required for the development of efficient bottom-up anion–water models. Using an evolutionary programming procedure, we determined optimized and reference configurations for such polarizable and first-principles-based representation of the potentials, and we analyzed their structural characteristics and energetics in comparison with estimates from DF-MP2 and DFT+D quantum chemistry computations. Moreover, we presented new benchmark datasets, considering both equilibrium and non-equilibrium configurations of higher-order species with an increasing number of water molecules up to 54 for each F, Cl, Br, and I anions, and we proposed a validation protocol to cross-check methods and approaches. In this way, we aim to improve the predictive ability of future molecular computer simulations for determining the ongoing conflicting distribution of different ions in aqueous environments, as well as the transition from nanoscale clusters to macroscopic condensed phases.

Keywords: electronic structure calculations; molecular interactions; benchmark protocol; data-driven modeling; genetic algorithm; ions hydration



Citation: Rodríguez-Segundo, R.; Arismendi-Arrieta, D.J.; Prosimi, R. A Benchmark Protocol for DFT Approaches and Data-Driven Models for Halide-Water Clusters. *Molecules* **2022**, *27*, 1654. <https://doi.org/10.3390/molecules27051654>

Academic Editor: Franck Rabilloud

Received: 4 February 2022

Accepted: 26 February 2022

Published: 2 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The study of water's properties is one of the most intense research topics [1–7]. In practice, you can rarely find water in its pure state, and the presence of other molecules is common, which can produce changes in the observed behavior of the water network. This is the case of ions, and depending on their charge and size could be strongly hydrated or disrupt the structure of the surrounding water hydrogen bonding [8–19]. Ion hydration is a key point in understanding a variety of physicochemical processes; thus both experimental and theoretical studies are seeking to give a definitive picture of how ions influence the water molecules around them [10,12,14–16,18–37]. On the one hand, it has been possible to obtain a clearer approximation of specific structure effects thanks to neutron and X-ray diffraction experiments [4,6,16,38,39]. However, on the other hand, the interpretation of these data strongly depends on the models used to analyze them. Thus, developing

potential models of high quality implies the parameterization of rather sophisticated functions of atomic coordinates to properly describe/represent the underlying short- and long-range many-body interactions.

Interaction potentials can be parameterized either from experimental data or first-principles values, with the latter representing a more adequate way to reproduce energies, as they are not biased by the subsequent reinterpretation that must be made of the laboratory data. Indeed, very valuable information has been obtained from numerous quantum chemistry calculations of water and ion–water clusters [30–37,40–46], at different levels of theory; however, they are usually limited to a few water molecules systems. A systematic analysis of the cluster properties as a function of its size provides a protocol for the selection of the best computational method for modeling ion–water interactions when going from systems in gas to condensed phase. In the theoretical studies, the majority of the limitations are highly dependent on the computational power available, and some of the most advanced *ab initio* and modern first-principles-based approaches could combine computational efficiency with model accuracy for investigating molecular properties and ion-hydration effects.

Traditionally, the improvement of model potentials is sought by adding physical terms that take into account the many-body influence, normally neglected since the two-body terms comprise most of the interaction energy [19,21,24,25,27,28,30,35,47–54]. This makes the potential much more flexible, presenting less error in conflict zones, although when more parameters are taken into account, part of the meaning assigned to each term is lost and the potential increasingly resembles a black box that makes it more difficult to interpret. More recently, the revolution of machine-learning techniques in the field makes such tools necessary components in improving decisions to be made during the model building process, such as choices of input training set configurations, level of electronic structure theory, etc., that influence the quality of results, and hence should be carefully considered [33,55–59]. The associated design choices typically require the expertise and intuition of experienced scientists; nevertheless, machine-learning approaches could automate and accelerate analysis in a purely data-driven fashion [60–62]. Of course, the level of model accuracy/precision is always dependent on the type of application or interest you have, and is traditionally restricted by the limited computational power available.

A reliable potential model must perform well in all ranges from clusters with a few water molecules to nano-droplets. With the intention to find the best approach, investigations are carried out in halide-water systems of increasing size both at the equilibrium points and outside of them, which will tell us which of all methodologies, approaches and models under study is the most reliable for further molecular dynamics simulations. Thus, this work presents a comparison of four polarizable data-driven model potentials (namely, *i*-TTM2, *i*-TTM3, *i*-TTM4, and *i*-MBpol) together with interaction energies obtained from electronic structure calculations using different type/rung density functionals, such as TPSS, M06L, M062X, B3LYP, LC- ω PBE, CAMB3LYP, and ω -B97XD, as implemented in the Gaussian package [63], considering dispersion corrections [64–66] and results from reference wave-function-based calculations at DF-MP2 level of theory [67,68]. All these computations were performed for F[−], Cl[−], Br[−], and I[−] halide ions in selected size water clusters, containing up to 54 water molecules.

The present work has the following distribution: In the first section, we present a brief description of the computational details of the methods and models used in the study. In the next section, the results obtained are presented and discussed starting with the reference data obtained from DF-MP2 calculations, the different data-driven interaction models, and the DFT/DFT+D energies. Finally, given the importance of knowing the behavior out of equilibrium, an analysis of the best-performing approaches and models is carried out by scanning the position of the halide ion in the water clusters. Some conclusive remarks are reported in the last section.

2. Computational Details, Results and Discussion

2.1. Building Up Data-Driven Interaction Models

Through the many-body expansion formalism [69], the total energy of the n -bodies/monomers system, E_n , can be expressed as a sum over higher n -body (nB) energies, $E^{(nB)}$, $E_n = \sum_n E_i^{(1B)} + \sum_n E_{ij}^{(2B)} + \sum_n E_{ijk}^{(3B)} + \dots + E^{(nB)}$.

In our case, for a halide anion and n -water molecules system, the potential energy form reads

$$V_n = \sum_n V^{(1B)} + \sum_n V^{(2B)} + \sum_n V^{(3B)} + \dots + \sum_n V^{(nB)} = \sum_n (V_W^{intra} + V_{W-W}^{inter} + V_{i-W}^{inter}),$$

where $V^{(1B)}$ is the intramolecular distortions of each water molecule (V_W^{intra}), the $V^{(2B)}$ contains all 2B halide anion–water and water–water pairwise intermolecular interactions, while $V^{(3B)}, \dots, V^{(nB)}$ are the 3B and higher-order nonadditive interaction terms, respectively.

Here, the V_W^{intra} term is represented by the Partridge Schwenke potential [70], while the intermolecular water–water interactions (V_{W-W}^{inter}) with any of the Thole Type Model (TTM) such as potentials (TTM2, TTM3, TTM4, and MB-pol) [48,50,51,71,72]. In turn, the ion–water intermolecular interaction (V_{i-W}^{inter}) is expressed as a sum of 2B short- and long-range contributions, and many-body polarization terms as described by the TTM model [71], $V_{i-W}^{inter} = V_{rep}^{2B} + V_{disp}^{2B} + V_{TTM,elec}^{2B} + V_{TTM,ind}^{nB}$. Briefly, the extended TTM model [71] utilizes for the water three geometry-dependent points charges calculated from the Partridge–Schwenke ab initio dipole moment surface (DMS) [70], two positive smeared charges at the hydrogen atoms and a negative smeared charge shifted from the oxygen atom to the M-site (located near the O atom and along the bisector of the HOH bond angle) [71], and also three polarizable sites at each atom with and an increased damping between the intramolecular dipoles on the H-atom, while a negative smeared charge and a polarizable site are placed on the halide anion. The $V_{TTM,elec}^{2B}$ represents the 2B permanent charge–charge electrostatic contributions, while the $V_{TTM,ind}^{nB}$ includes higher-order induced dipoles interactions, which are all described by a classical many-body polarization model [35,71,73]. The V_{rep}^{2B} term is the repulsive energy between pairs of anion and oxygen or hydrogen atoms, expressed as $\sum_{i,j=(O/H),X^-} A_{ij} e^{\beta_{ij} R_{ij}}$, while V_{disp}^{2B} describes the dispersion energy as a sum of pairwise interactions between OX^- and HX^- , $\sum_{i,j=(O/H),X^-} \frac{C_6^{(ij)} f(R_{ij}, \delta_{ij})}{R_{ij}^6}$, with $f(R, \delta)$ being the Tang–Toennies damping functions [74], and $C_6^{OX^-}$ and $C_6^{HX^-}$ are the dispersion coefficients. The δ_{ij} parameters determine the shape to the $f(R, \delta)$ damping functions, and were set equal to the β_{ij} of the corresponding Born–Mayer functions [75].

All C_6 , A , and β parameters have been determined by their fit to energies from DFT+D and CCSD(T)-F12 level of theories. In particular, the parameterization procedure, shown in Figure 1, consists of a two-stage process. The C_6 coefficients impose the correct asymptotic behavior, and were evaluated first by the fit to energy data obtained from LC- ω PBE/AVTZ calculations [63], employing the exchange-hole dipole moment (XDM) model [64,76]. For the ion TTM contributions to the interaction, the atomic dipole polarizability of each halide ion was obtained from CCSD(T)/t-aug-cc-pV5Z/t-aug-cc-pV5Z-PP calculations [35], while those for the O and H atoms were obtained from ref. [71]. The remaining model's parameters were adjusted to the interaction energies at CCSD(T)-F12b/VQZ-F12(-PP) level of theory calculations, including basis set superposition error (BSSE) corrections via the counterpoise (CP) method [77]. For each X^-H_2O cluster, interaction energies were computed in a grid of near 500 points distributed equidistantly along the R (anion–water (oxygen) distance), θ (polar angle), and ϕ (azimuthal angle) coordinates.

Specifically, during the fitting procedure the linear and nonlinear parameters were optimized using linear regression through singular value decomposition, and the simplex algorithm, respectively. The weighted sum of the squared residuals, given as

$$\chi^2 = \sum_{\mathcal{N} \in \mathcal{T}} w_{\mathcal{N}} [(V_{i-W}^{\text{inter}}(\mathcal{N}) - \Delta E^{\text{ref}}(\mathcal{N}))^2],$$

were computed and minimized, with ΔE^{ref} being the reference energies, and \mathcal{N} the number of configurations in the training set \mathcal{T} . The weight-factors, such as $w_{\mathcal{N}}(E) = (\frac{1}{E - E_{\text{min}} + DE})^2$, where E_{min} and DE correspond to the minimum energy in the \mathcal{T} , and the range of favorably weighted energies, respectively, are traditionally introduced to overcome difficulties in the fit when some space regions need a larger number of points to have a good representation. Usually, this can be also solved by adding new points in such regions of interest; however, this repetitive (forward-backward) procedure turns out to be a time-consuming and computationally expensive task.

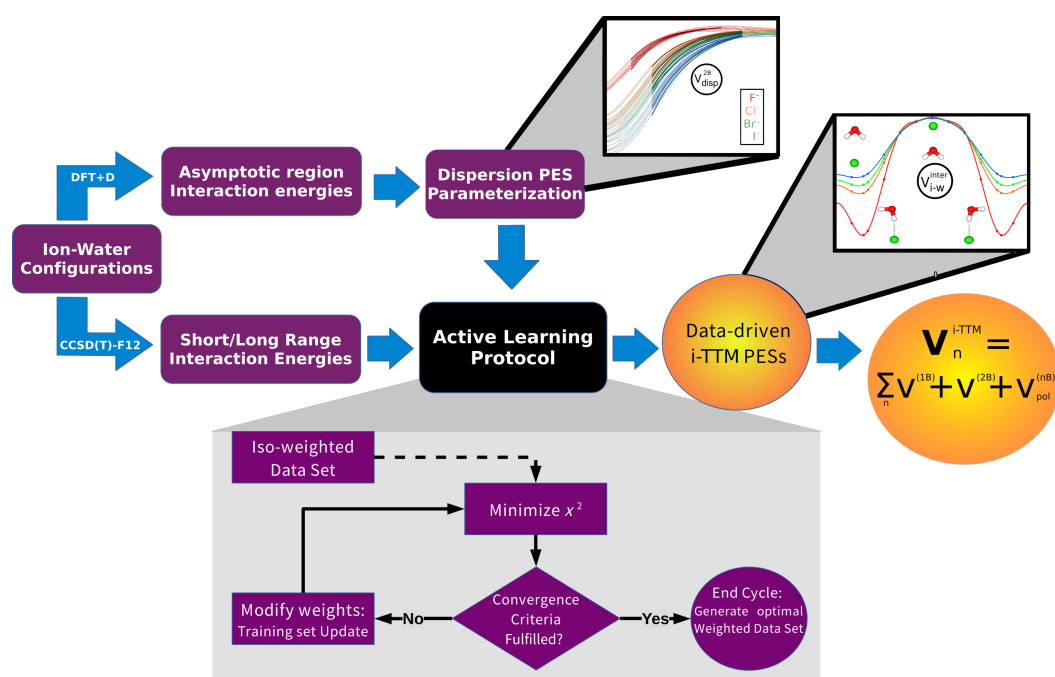


Figure 1. Full automated flowchart of the PES's parameterization process.

Nowadays, machine-learning schemes and data processing tools offer the possibility to also automate the most repetitive tasks in conventional procedures, accelerating or even improving the prediction of parameters [61,62]. Thus, in this work, we explore the application of a fully automated approach for the generation of weighted datasets based on the concept of active learning. Our active learning scheme improves the results obtained from an arbitrary pool of initial data without the calculation of new points, allowing us to weight in an efficient and smart way the potential configurations, reducing the training set size, and thus minimizing the computational cost. The flowchart is displayed in the inset plot of Figure 1, which consists of a pool of equidistant configurations points in R , θ , and ϕ coordinates of each X^-H_2O dimer, summarized below:

- We start from a pool of fixed points, in our case those points are equidistant in all R , θ , and ϕ coordinates.
- For each iteration, a minimization of the parameters for the target function χ^2 is performed, using a simplex procedure.
- Once the minimization has ended, we check the fulfillment of the convergence criteria, such as the maximum number of steps, minimum average error, and certain error value, along the whole curve to see if it reached convergence.
- If none of the convergence criteria have been reached, the error at each point is obtained.
- For the N points with an error larger than a predefined percentage of the maximum error value, we increase their weights in a gradual descent manner throughout the iterations.

- Minimization is performed with the new training set of weights.
- This procedure is repeated until convergence is achieved.

In this way, we generated four families of data-driven V_n PESs (see data given in supplementary material), namely i-TTM2, i-TTM3, i-TTM4, and i-MBpol, depending on which TTM water model we chose, to represent in an efficient manner from on a bottom-up approach the halide–water interactions in larger $X^-(H_2O)_n$ systems, with X being F, Cl, Br, and I anions.

2.2. Evolutionary Programming Procedure and Selected Optimal Reference Structures

Given the availability of the i-TTM potential models, halide $(H_2O)_n$ clusters containing an increasing number of water molecules are considered, and their structuring is investigated through optimal potential configurations. Looking for systems with an increasing number of water molecules, we decided to pick some concrete numbers corresponding to n equal to 1, 2, 4, 8, 16, 20, 24, 36, 48, and 54 water molecules, which allows having a large distribution of water nanoclusters to explore trends or patterns in the growth of the clusters during the anion hydration/solvation process.

Searching for energy minima in high-dimensional potential surfaces is a challenging task. As the number of atoms in the system increases, its degrees of freedom increase too, which leads to high computational cost if we want to explore the whole configurational space. To deal with this, an evolutionary algorithm (EA) [78–81] is used in this work. The genetic algorithm is based on natural selection, and begins with an initial guess, and applies modifications, creating a population of possible solutions; these solutions “compete” against a fitness function that we want to optimize, and we retain the best candidates that are the parents for the next generation. In the second step, we create a new population by mixing the values of the previous best solutions (reproduction) and giving some random values (mutation), and then we repeat the previous step, keeping the best individuals of that generation. This process continues until we reach the optimization of the function. For structure optimization, the fitness function is the potential energy surface we want to explore, and the variables we mix in every generation are the coordinates of the atoms of the system.

Such algorithms present an efficient way to solve numerical optimization problems in multi-dimensional space, as gradients of the potential are not needed, while the search space-narrowing is automatically realized by the self-adaptation of mutation [78–80]. Briefly, we start with an initial generation of \mathcal{M} individuals (for each $X^-(H_2O)_n$ cluster, with X = F, Cl, Br, and I in our case). Each individual is characterized by a pair of real vectors $(\chi_i, \eta_i)_{i=1-\mathcal{M}}$, containing the Cartesian coordinates, χ_i , of all cluster atoms, and the strategy parameters, η_i for all individuals, which controls the evolution of the dispersion of the population in time. The initial coordinates χ_i are chosen randomly in the interval $(0, \Delta)$, with $\eta_i = 1$ and Δ being a displacement factor that increases the revolution power. Each parent set (χ_i, η_i) evolves to generate a new population according to the evolution law: $\chi'_i(j) = \chi_i(j) + \eta_i(j)\mathcal{N}_j(0, 1)$ and $\eta'_i(j) = \eta_i(j) \exp(\tau'\mathcal{N}(0, 1) + \tau''\mathcal{N}_j(0, 1))$, where $\mathcal{N}(0, 1)$ and $\mathcal{N}_j(0, 1)$ are Gaussian random numbers of average zero and standard deviation 1 generated for each j coordinate, while τ'/τ'' are adjustable parameters. For each individual of the joint parent–child group ($2\mathcal{M}$ individuals) q (tournament size), opponents are randomly chosen from the $2\mathcal{M}-1$ individuals, to compare each other, and the individual in each encounter with the lower potential energy wins. The best individuals then become parents for the next generation and so on. Convergence is achieved when the potential energy difference between two consecutive generations is below a threshold value.

As a starting point, we used the minimum energy structures found in the Cambridge energy landscape database [82] for pure water clusters, and generated initial anion–water configurations by replacing one water molecule in all positions with the halide ion using the DENE software package [83]. We used generations with 100 individuals and a threshold for the energy of 0.01 kcal/mol. The displacement of the atoms' coordinates is governed by a Gaussian function and is multiplied by a factor of $\delta_0 = 0.15$ that controls the size of

the movement and $\delta_j = 0.85$ that updates the value in every generation, which means that every following generation moves 85% less than the previous one.

In Figure 2, the optimal structures obtained from the EA using the i-TTM4 interaction potentials for all $X^-(H_2O)_n$ ($n=16,20,24,36,48,54$) clusters under study are shown, together with their corresponding energies, while their geometry coordinates are given in the supplementary material. For the small halide–water systems of up to eight water molecules, the preference of anions to be on the water cluster surface has been proven by previous studies [35,41,84,85]. As can be seen in Figure 2, for all cases the ion presents a tendency to stay at the surface of the cluster. In Figure 3, we defined a plane perpendicular to the line defined by the center of mass of the cluster and the ion position, containing the center of mass point (see upper panel), and we displayed (lower panels) the distances, r , from this plane of each oxygen atom of the water molecules or halide ion of the indicated cluster. One can see that the r values for the ions are larger than those for oxygens, with the only exception to this rule being found in the $I^-(H_2O)_{54}$ cluster, where although it is found near the limits of the cluster, already allows it to be partially surrounded by a few water molecules. However, as the number of water molecules in the cluster increases, the identification of one global minimum becomes less significant, as multiple structures will likely be important for a physical description of the system.

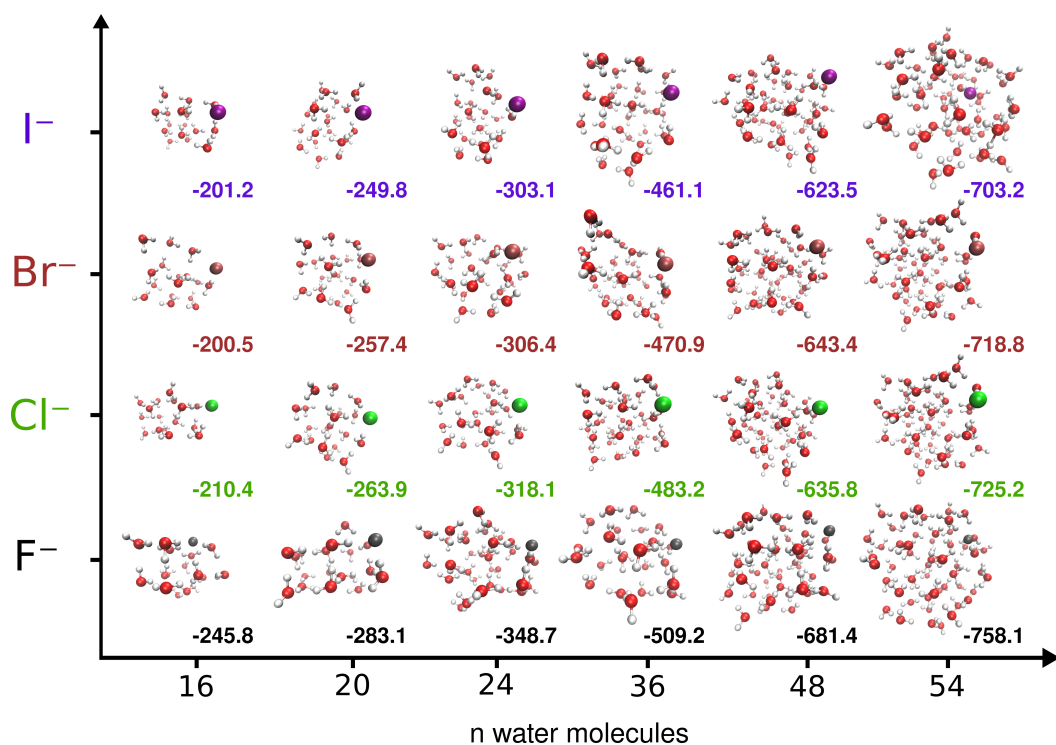


Figure 2. Optimal configurations and energies (in kcal/mol) for the indicated clusters formed by a halide ion and 16, 20, 24, 36, 48, and 54 water molecules.

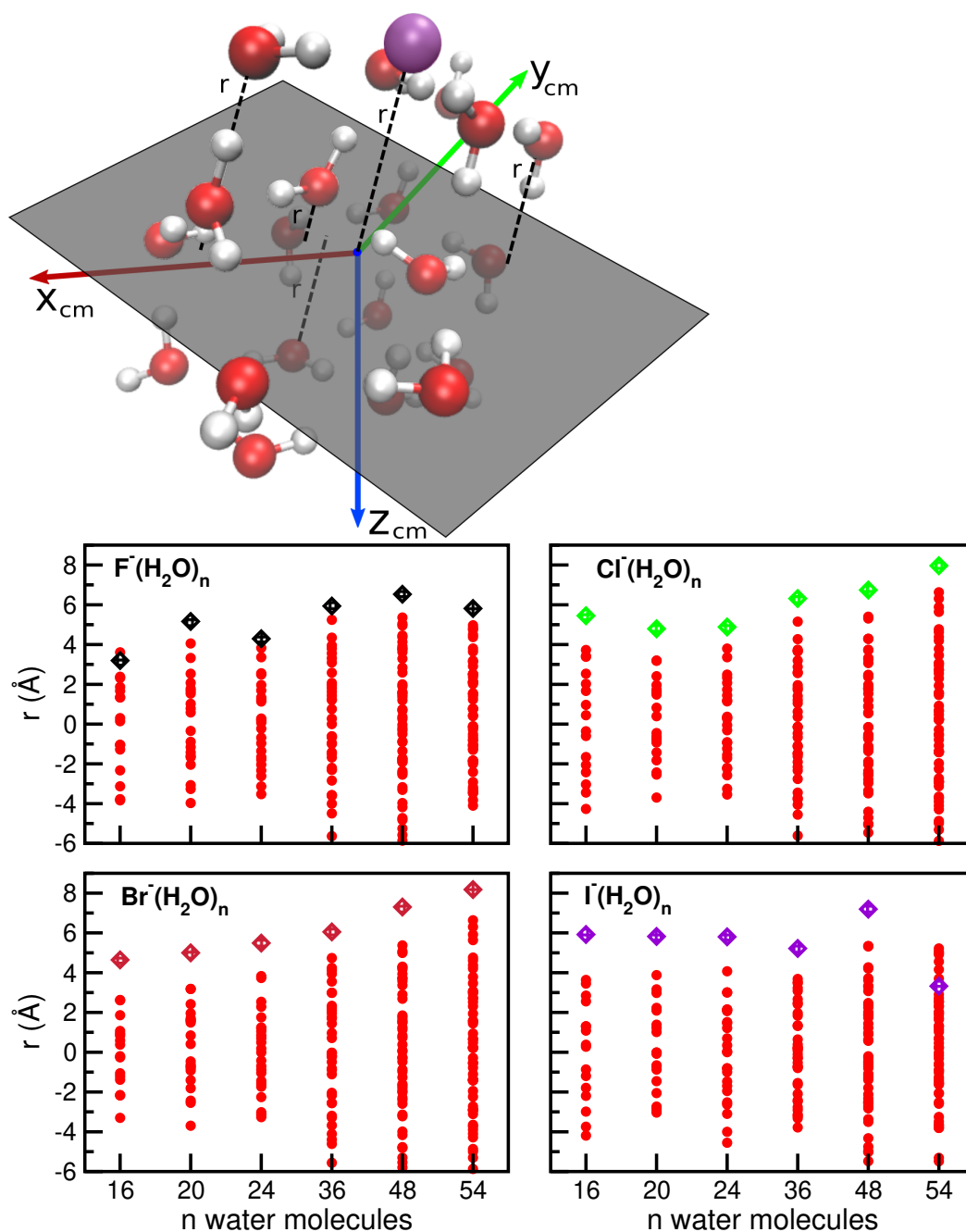


Figure 3. Distances, r , of all oxygen (water) atoms (red circles) and halide ion (color diamonds) to the indicated plane (see text) for each n water molecules clusters.

2.3. Electronic Structure Calculations and Reference Energy Data

Looking for a computationally accessible wave-function based method to obtain the reference energies for large halide–water clusters with up to 54 water molecules, we chose the density fitting DF-MP2 computations, as implemented in the MOLPRO package [68,86]. Following our previous studies on small-size $X^-(H_2O)_{n=1-8}$ clusters [35,85], where high-level theories, such as CCSD(T) or explicitly correlated CCSD(T)-F12 calculations can be carried out, we found that DF-MP2 performs reasonable well, and could be a computationally efficient method for larger systems under study. Thus, for the F, Cl, O, and H atoms, we employed all electron aug-cc-pVTZ (AVTZ) basis sets [68,87,88], while for Br and I we used the ECP10MDF and ECP28MDF small-core effective core potentials in conjunction with the aug-cc-pVTZ-PP(AVTZ-PP) valence basis sets [68,89–91], and the standard mp2fit auxiliary basis sets for density fitting in the DF-MP2 calculations [67,68].

The interaction energies for each system were calculated as the difference between the total energies of the $X^-(H_2O)_n$ cluster and the sum of the energies of the water cluster and X^- ion, $\Delta E = E_{X^-(H_2O)_{n=16-54}} - E_{(H_2O)_{n=16-54}} - E_{X^-}$.

In order to assess comparable alternative first-principles methodologies, we also carried out DFT-based calculations using the Gaussian16 package [63] with the additional dispersion corrections computed by the DFT-D3 and DFT-D4 programs [92,93]. We chose to employ different density functionals from various groups/rungs (according to the Jacob's ladder classification), such as TPSS [94] and M06L [95] (meta-GGAs/3rd rung), B3LYP [96], and M06-2X [97] (hybrid/meta-GGAs/4th rung), and CAMB3LYP [98], LC- ω PBE [99] and ω B97XD [100] (range-separated hybrid). All DFT computations were carried out with the AVQZ basis set and for the numerical integration, the ultrafine grid was used. In turn, the DFT approaches were corrected, including different semi-classical dispersion schemes, such as the DFT-D3 and its variants regarding the damping functions available in the literature, the original zero-damping function D3(0) [101] and its modified D3M(0) version [102], the most popular Becke–Johnson damping function D3BJ [65] and a modified version D3M(BJ) [102] with an emphasis on non-equilibrium particularly compressed geometries, as well as the most recently developed DFT-D4 [66] model, including many-body dispersion interactions beyond the pairwise terms. In all electronic structure computations (see data given in the supplementary material) and data analyses, the DENEb toolkit utilities [83] were employed.

2.4. Comparative Analysis of the Different Halide–Water Potentials: Equilibrium Structures

Correlation plots with interaction energies from the DF-MP2 calculations and potential values from the i-TTM2, i-TTM3, i-TTM4, and i-MBpol halide–water models are presented in Figure 4. Energies for the different F^- to I^- ions clusters, with $n = 1, 2, 4, 8, 16, 20, 24, 36, 48,$ and 54 water molecules, were obtained at their optimal energy structures (see Figure 2).

One can see that all model potentials show similar behaviors, with i-TTM3 having larger deviations than the i-TTM2, i-MBpol, and i-TTM4 ones. We should note that i-TTM2 energies are very close to the i-MBpol for all $Cl^-(H_2O)_n$, $Br^-(H_2O)_n$, and $I^-(H_2O)_n$ clusters studied, while this is not the case for the $F^-(H_2O)_n$ ones. In all type and size cluster cases, the i-TTM4 provides energy values that are in closer agreement with the DF-MP2 energies. Energy differences (δ in %) between the i-TTM4 and i-MBpol potential values of each $X^-(H_2O)_n$ cluster structure from the corresponding DF-MP2 reference energies are shown in the lower panel of Figure 4. The calculated errors are found to be less than 3.2, 4.3, 3.6, and 3.7% for the $F^-(H_2O)_n$, $Cl^-(H_2O)_n$, $Br^-(H_2O)_n$, and $I^-(H_2O)_n$ i-TTM4 potentials, respectively, while somehow larger deviations, up to 16%, were obtained for the i-MBpol values. The main difference between the i-TTM4 and i-MBpol model potentials is that the latter incorporates the MB-pol water–water model [72], which includes explicitly short-range three-body terms, while the TTM4-F water model [71] contains only pair (two-body) short-range components, with the higher-body contributions described in both models by induction terms.

2.5. Comparative Analysis of DFT/DFT+D Approaches: Equilibrium Structures

In the search for a computationally affordable method for studies in which the speed of calculation is important, in addition to the model potentials, the DFTs could present an alternative, which is fast enough to be used in direct (on-the-fly) computational simulations. Dozens of DFT functionals exist in the literature along with a variety of dispersion corrections and can be classified according to the approximations they use. On the basis of previous studies on smaller $X^-(H_2O)_n$ systems [35,85], in this work we decided to consider two meta-GGA functionals, TPSS and M06L; one hybrid-GGA, B3LYP; one hybrid-meta-GGA, M062X; and three range separated hybrids, CAMB3LYP, LC- ω PBE and ω B97XD. Figure 5 shows the correlation plots of the DF-MP2 and DFT interaction energies of these functionals. One can see that all functionals show the same behavior for all $X^-(H_2O)_n$ systems. Among the functionals with the “best-performance” for the four ion–water clusters

is ω B97XD, which intrinsically includes the D2 dispersion [100], followed by the M062X, M06L and CAMB3LYP, while TPSS, LC- ω PBE and B3LYP show significant differences.

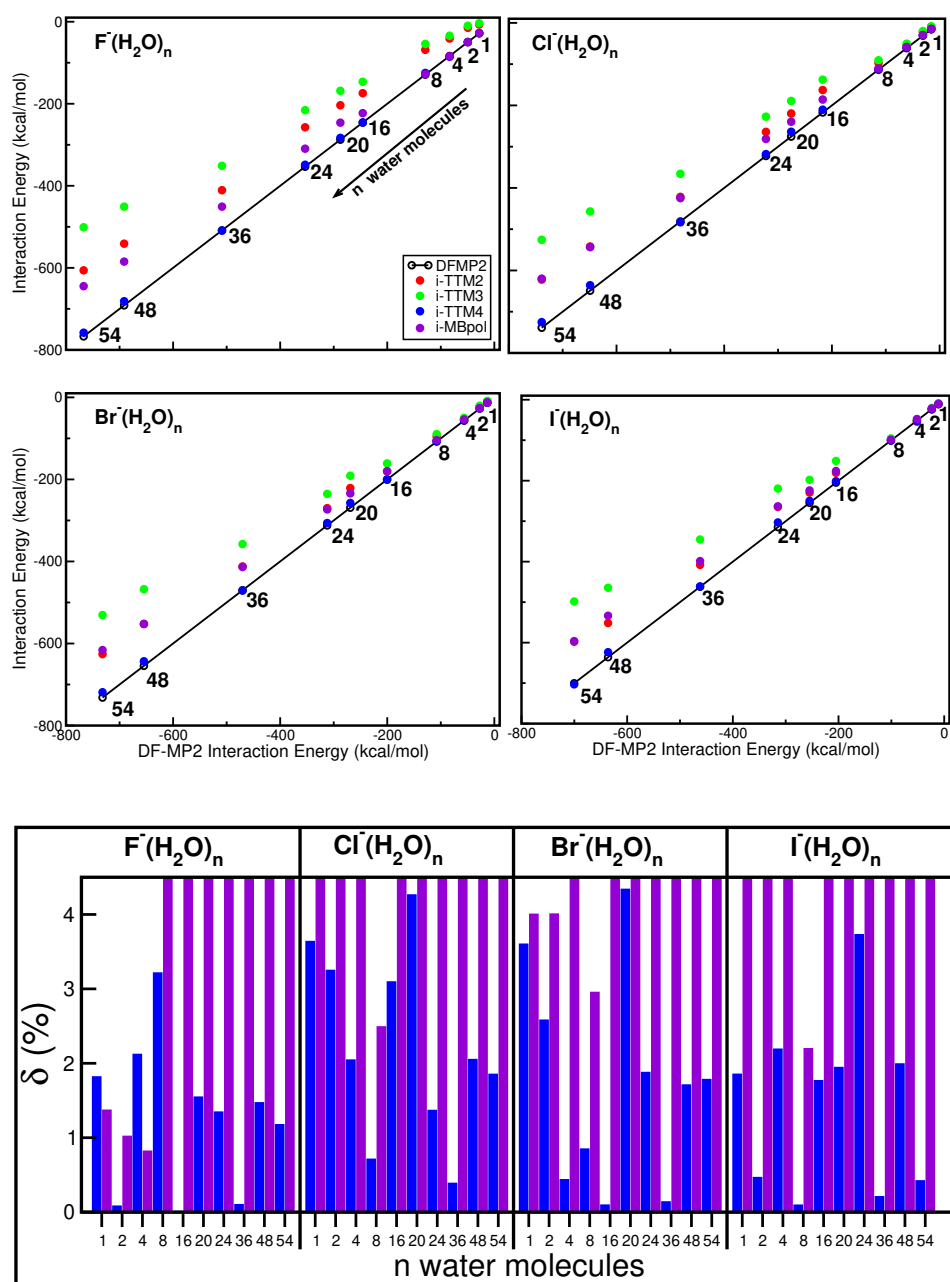


Figure 4. Correlation plots (upper and middle panels) and percentage (%) difference between the ion–water potential (i-TTM2, i-TTM3, i-TTM4, i-MBpol) values and the DF-MP2 energies at the corresponding optimal geometry structure of each indicated $X^-(H_2O)_n$ cluster.

Dispersion plays an important role in these systems, since long-distance interactions are important both in hydrogen bonds between water molecules and in interactions between water and ions, especially in those ions whose electron cloud is very large, e.g., Br^- and I^- . Therefore, we carried out calculations for all remaining functionals by adding different semi-classical dispersion corrections, such as D3 with its variants D3(0) and D3(BJ) [65,101], as well as the modified versions D3M(0) and D3M(BJ) [102], and the recent D4 schemes (Damped Axilrod-Teller-Muto D4(ATM) and many-body dispersion D4(MBD)) [66].

In the left-side panels of Figure S1, we present the percentage (%) error, δ , of the indicated selected best-performing DFT/DFT+D energies and the i-TTM4 model values from the corresponding DF-MP2 energies for each halide–water cluster of up to 54 water molecules, while in the right-side panels of the figure a graphical representation of the errors (in %) of the four best-performing DFT/DFT+D functionals or the i-TTM4 model from the DF-MP2 energies for each cluster with $n = 16$ –54 are displayed.

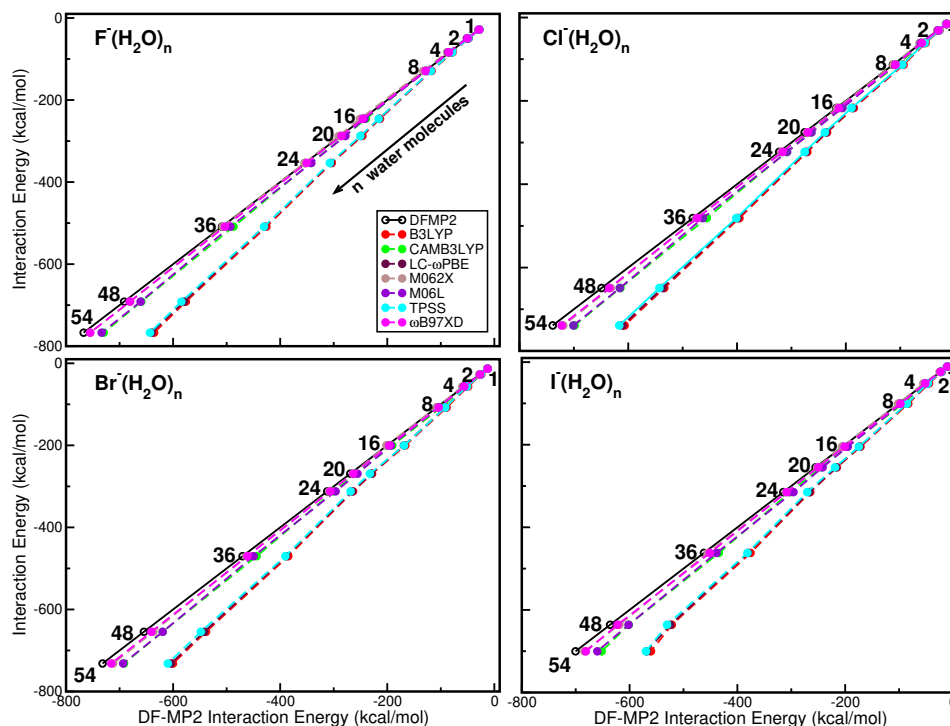


Figure 5. Correlation plots for the indicated DFT values and the DF-MP2 interaction energies at the corresponding optimal geometry structure of each indicated $X^{-}(H_2O)_n = 1 - 54$ cluster.

The importance of adding dispersion corrections in the DFT approaches for these systems can be seen. We found that the error increases as the number of water molecules increases, indicating the inability of the DFT functionals to deal with additive long-range interactions, while by adding dispersion such errors have been balanced or even eliminated. In general, we obtain better results when adding the corresponding corrections, except in the case of CAM-B3LYP, for which dispersion worsens their energies, as well as the M062X+D3(0) for the smaller size clusters. In particular, among the TPSS approaches, D3 and D4 show better results for $F^{-}/Cl^{-}/Br^{-}$ and I^{-} water clusters, respectively, while the M06L dispersion corrections yield in general rather high deviations. Next, dispersion schemes seem to work well for the B3LYP, with the D4 notably improving results, showing a balanced behavior for all sizes and halide–water systems studied, with errors of less than 1%, with LC- ω PBE+D3M following with errors of around 1%, and then TPSS+D3(D4) with slightly higher than 1%.

2.6. Overall Comparative Analysis: Non-Equilibrium Configurations

In Figure S1, one should also notice the performance of the i-TTM4, especially for the $F^{-}(H_2O)_n$ clusters. However, we should mention that the good behavior of the i-TTM4 potential compared to the other potentials could be due to the fact that the minimum energy geometries have been optimized using that potential. Thus, to further test the DFT+D and model potentials, a study was carried out by moving the ions out of their lowest energy configuration. We started by placing the ion into the center of the mass of the cluster structure, $R_{ion-water} = 0$, and moved it with variable size steps along the direction defined by the line joins it to the ion's position at the corresponding minimum energy

configuration. In Figures S2–S5 (see Supplementary Materials), we display such scans for all F, Cl, Br, and I halide anions and 16, 20, 24, 36, 48, and 54 water molecules clusters. In addition to the four i-TTM potentials, the energies are presented at the center of mass and at intermediate points for DF-MP2, B3LYP and B3LYP+D4(MBD); as such, DFT+D approaches were found to yield the best and balanced performance in all halide ion–water clusters. One can see that the chosen non-equilibrium structures sample various repulsive, attractive and dissociative areas of the potential surfaces, and for all those sizes of halide ion–water clusters, configurations the i-TTM4 and B3LYP+D4 were found closer to the reference DF-MP2 energies. In Figure 6, we summarize all the results obtained for each halide ion for the indicated size water clusters. By analyzing averaged errors along the scans, as in the case of equilibrium structures, we found that the B3LYP+D4 DFT approach yields the best results with errors of less than 0.1 kcal/mol in most cases, indicating clearly the improvement of including dispersion correction in the B3LYP functional. Next, the i-TTM4 potential showed errors of 0.2, 0.4, 0.05, 0.2 kcal/mol for $F^-(H_2O)_n$, $Cl^-(H_2O)_n$, $Br^-(H_2O)_n$ and $I^-(H_2O)_n$ clusters, respectively.

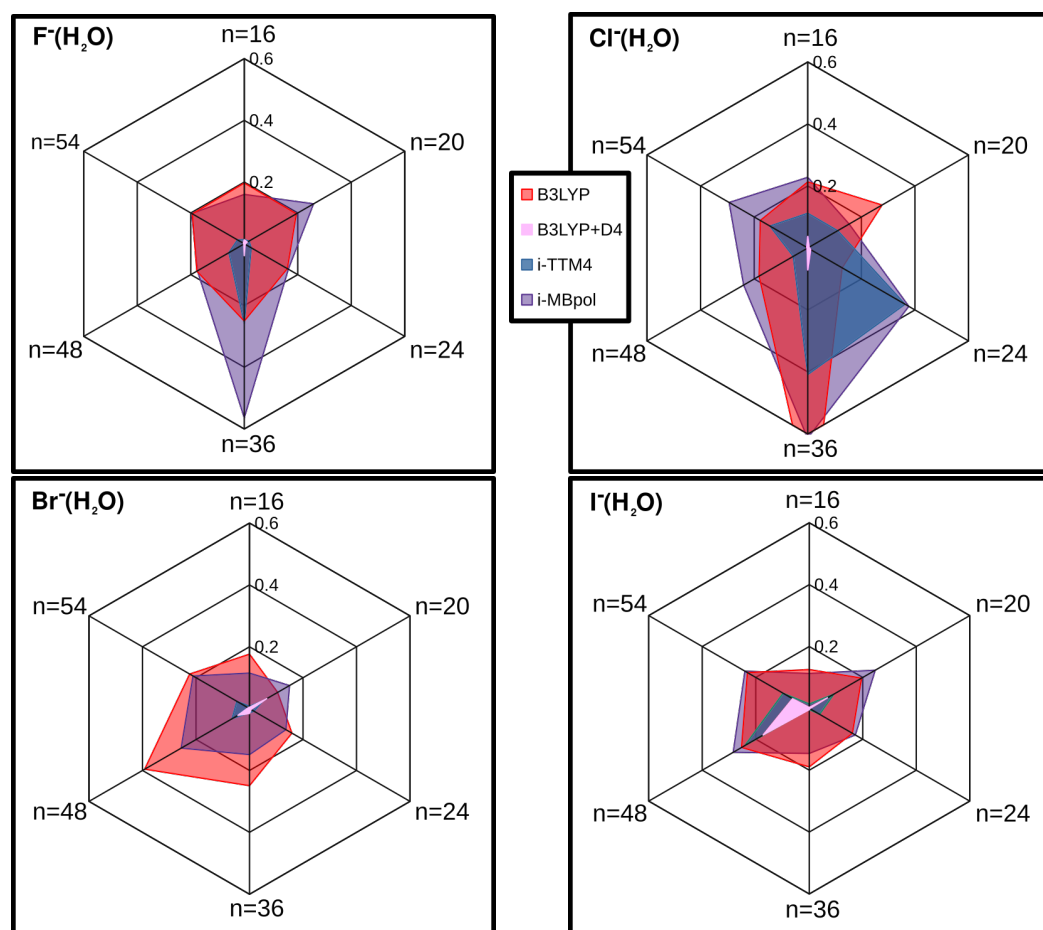


Figure 6. Average errors (normalized) for the indicated DFT+D approaches and model potentials obtained from the scans of the different halide–water clusters (see Figures S2–S5 in SI).

3. Conclusions

We have adopted a general protocol scheme to access promising model potentials and DFT/DFT+D approaches against wave-function-based reference data. First of all, we built up data-driven bottom-up interaction potentials for various halide ion–water systems, such as F^- , Cl^- , Br^- , and I^- . We explored the application of a fully automated approach for the generation of optimally weighted datasets, accelerating the prediction of adjusted parameters without adding new energies to the dataset. In general, during the fitting processes, difficulties are commonly present due to the poor grid point representation

of the potential surface at some space regions. Our active learning scheme allows us to weight in an efficient and smart way the potential configurations, increasing their values in those areas, where the prediction presents higher errors, reducing the training set size, and minimizing the computational cost.

In this way, we presented four i-TTM families of parameterized model PESs trained on X^-H_2O CCSD(T)-F12 data, including implicitly many-body polarization contributions terms. Our intention is to employ such model interaction potentials in computationally high-demanding dynamics simulations for reliable nanoclusters' and bulk property investigations. Thus, we have reported new benchmark datasets by increasing the size of the clusters, up to 54 water molecules. We span a broad range of different representative configurations for each benchmark structure, from equilibrium, moderate binding to nonbinding ones to ensure an overall description of the interactions at different levels of theory.

By analyzing the energy differences between electronic structure wave-function-based methodologies, such as DF-MP2, and various DFT+D approaches, we evaluated the performance of the DFT functionals in the presence/absence of dispersion corrections, as well as the new i-TTM model potentials. We highlighted the importance of adding dispersion to the DFT approaches, providing improved data. We found that the B3LYP+D4 functional and the ab initio-based i-TTM4 model yields the best overall results in comparison to the DF-MP2 reference data for all clusters sizes and structures studied. Such an outcome reveals that less demanding computational methods could present alternatives to adequately describe halide–water interactions in finite-size clusters. As such systems are large enough, we should underline the importance of considering well-converged reference energy data, as well as a variety of representative configurations for cross-checking computational approaches. Investigations in this direction could further benefit from the application of active learning techniques to generate the automatic sampling of more objective representative datasets.

Moreover, finite-size clusters can serve as model microsolutions, so the reported stable potential structures for each cluster, through evolutionary programming algorithm energy optimizations, should be expected to influence halide ions' hydration. All their structures show an on-surface preference of the halide ion as clusters grow, indicating a competition between the ion–water and water–water interactions. However, such a classical description of energetics and structuring of halide–water clusters should be further extended to include nuclear quantum and thermal effects. Such additional information could be useful for investigating the transition from molecular nanoscale aggregates to dissolved anions. Currently, path-integral molecular dynamics (PIMD) simulations are in progress for these halide–water clusters, employing the current data-driven i-TTM4 model potentials to explore the anions' distribution in water clusters of increasing size.

Supplementary Materials: The following supporting information can be downloaded online. Figure S1: Errors δ (in %) for the indicated DFT/DFT+D functionals and i-TTM4 model. Figure S2: Scan of fluorine ion with clusters of 16, 20, 24, 36, 48, and 54 water molecules. Figure S3: Scan of chlorine ion with clusters of 16, 20, 24, 36, 48, and 54 water molecules. Figure S4: Scan of bromine ion with clusters of 16, 20, 24, 36, 48, and 54 water molecules. Figure S5: Scan of iodine ion with clusters of 16, 20, 24, 36, 48, and 54 water molecules.

Author Contributions: Conceptualization, R.P. and D.J.A.-A.; computations and software, R.R.-S. and D.J.A.-A.; formal analysis & validation, R.R.-S., D.J.A.-A. and R.P.; writing—original draft preparation, R.P.; writing—review and editing, R.R.-S., D.J.A.-A. and R.P.; visualization, R.R.-S. and R.P.; supervision, R.P.; funding acquisition, R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Comunidad de Madrid grant No. IND2017/AMB-7696, MINECO grant No. FIS2017-83157-P, MICINN grant No, PID2020-114654GB-I00 and COST Action CA18212(MD-GAS).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting reported results are available from the corresponding author upon request. The parameters of potential models, electronic structure DFMP2 and DFT/DFT-D calculations and i-TTM models energies, as well as optimized geometry coordinates, can be seen in the Supplementary Materials.

Acknowledgments: The authors thank to Centro de Cálculo del IFF, SGAI (CSIC) and CESGA for allocation of computer time. All authors acknowledge José A. Torres Alonso for technical support with DENEb-Atelgraphics software.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Ball, P. *Life's Matrix: A Biography of Water*; University of California Press: Berkeley, CA, USA, 2001.
2. Sharp, K.A. *Water: Structure and Properties*. In *eLS*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2001.
3. Brini, E.; Fennell, C.J.; Fernandez-Serra, M.; Hribar-Lee, B.; Lukšič, M.; Dill, K.A. How Water's Properties Are Encoded in Its Molecular Structure and Energies. *Chem. Rev.* **2017**, *117*, 12385–12414. [[CrossRef](#)]
4. Perakis, F.; Camisasca, G.; Lane, T.J.; Späh, A.; Wikfeldt, K.T.; Sellberg, J.A.; Lehmkuhler, F.; Pathak, H.; Kim, K.H.; Amann-Winkel, K.; et al. Coherent X-rays reveal the influence of cage effects on ultrafast water dynamics. *Nat. Commun.* **2018**, *9*, 1917. [[CrossRef](#)]
5. Markland, T.E.; Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nat. Rev. Chem.* **2018**, *2*, 0109. [[CrossRef](#)]
6. Shi, R.; Tanaka, H. Direct Evidence in the Scattering Function for the Coexistence of Two Types of Local Structures in Liquid Water. *J. Am. Chem. Soc.* **2020**, *142*, 2868–2875. [[CrossRef](#)]
7. Yang, J.; Dettori, R.; Nunes, J.P.F.; List, N.H.; Biasin, E.; Centurion, M.; Chen, Z.; Cordones, A.A.; Deponte, D.P.; Heinz, T.F.; et al. Direct observation of ultrafast hydrogen bond strengthening in liquid water. *Nature* **2021**, *596*, 531–535. [[CrossRef](#)]
8. Kebarle, P.; Arshadi, M.; Scarborough, J. Hydration of Negative Ions in the Gas Phase. *J. Chem. Phys.* **1968**, *49*, 817–822. [[CrossRef](#)]
9. Marcus, Y. Effect of Ions on the Structure of Water: Structure Making and Breaking. *Chem. Rev.* **2009**, *109*, 1346–1370. [[CrossRef](#)]
10. Marcus, Y. *Ions in Solution and Their Solvation*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2015.
11. Robertson, W.H.; Johnson, M.A. Molecular Aspects of Halide Ion Hydration: The Cluster Approach. *Ann. Rev. Phys. Chem.* **2003**, *54*, 173–213. [[CrossRef](#)]
12. Smith, J.D.; Saykally, R.J.; Geissler, P.L. The Effects of Dissolved Halide Anions on Hydrogen Bonding in Liquid Water. *J. Am. Chem. Soc.* **2007**, *129*, 13847–13856. [[CrossRef](#)]
13. Tielrooij, K.J.; Garcia-Araez, N.; Bonn, M.; Bakker, H.J. Cooperativity in Ion Hydration. *Science* **2010**, *328*, 1006–1009. [[CrossRef](#)] [[PubMed](#)]
14. Funkner, S.; Niehues, G.; Schmidt, D.A.; Heyden, M.; Schwaab, G.; Callahan, K.M.; Tobias, D.J.; Havenith, M. Watching the Low-Frequency Motions in Aqueous Salt Solutions: The Terahertz Vibrational Signatures of Hydrated Ions. *J. Am. Chem. Soc.* **2012**, *134*, 1030–1035. [[CrossRef](#)] [[PubMed](#)]
15. Tobias, D.J.; Stern, A.C.; Baer, M.D.; Levin, Y.; Mundy, C.J. Simulation and Theory of Ions at Atmospherically Relevant Aqueous Liquid-Air Interfaces. *Ann. Rev. Phys. Chem.* **2013**, *64*, 339–359. [[CrossRef](#)] [[PubMed](#)]
16. Migliorati, V.; Sessa, F.; Aquilanti, G.; D'Angelo, P. Unraveling halide hydration: A high dilution approach. *J. Chem. Phys.* **2014**, *141*, 044509. [[CrossRef](#)] [[PubMed](#)]
17. Piatkowski, L.; Zhang, Z.; Backus, E.H.G.; Bakker, H.J.; Bonn, M. Extreme surface propensity of halide ions in water. *Nat. Commun.* **2014**, *5*, 1–7. [[CrossRef](#)]
18. Waluyo, I.; Nordlund, D.; Bergmann, U.; Schlesinger, D.; Pettersson, L.G.M.; Nilsson, A. A different view of structure-making and structure-breaking in alkali halide aqueous solutions through X-ray absorption spectroscopy. *J. Chem. Phys.* **2014**, *140*, 244506. [[CrossRef](#)] [[PubMed](#)]
19. Wang, L.P.; McKiernan, K.A.; Gomes, J.; Beauchamp, K.A.; Head-Gordon, T.; Rice, J.E.; Swope, W.C.; Martínez, T.J.; Pande, V.S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121*, 4023–4039. [[CrossRef](#)] [[PubMed](#)]
20. Kistenmacher, H.; Popkie, H.; Clementi, E. Study of the structure of molecular complexes. II. Energy surfaces for a water molecule in the field of a sodium or potassium cation. *J. Chem. Phys.* **1973**, *58*, 1689–1699. [[CrossRef](#)]
21. Perera, L.; Berkowitz, M.L. Structures of $\text{Cl}^-(\text{H}_2\text{O})_n$ and $\text{F}^-(\text{H}_2\text{O})_n$ ($n = 2, 3, \dots, 15$) clusters. Molecular dynamics computer simulations. *J. Chem. Phys.* **1994**, *100*, 3085–3093. [[CrossRef](#)]
22. Ayotte, P.; Nielsen, S.B.; Weddle, G.H.; Johnson, M.A.; Xantheas, S.S. Spectroscopic Observation of Ion-Induced Water Dimer Dissociation in the $\text{X}-(\text{H}_2\text{O})_2$ ($\text{X} = \text{F}, \text{Cl}, \text{Br}, \text{I}$) Clusters. *J. Phys. Chem. A* **1999**, *103*, 10665–10669. [[CrossRef](#)]
23. Cabarcos, O.M.; Weinheimer, C.J.; Lisy, J.M.; Xantheas, S.S. Microscopic hydration of the fluoride anion. *J. Chem. Phys.* **1999**, *110*, 5–8. [[CrossRef](#)]

24. Guàrdia, E.; Skarmoutsos, I.; Masia, M. On Ion and Molecular Polarization of Halides in Water. *J. Chem. Theory Comput.* **2009**, *5*, 1449–1453. [[CrossRef](#)] [[PubMed](#)]
25. Sala, J.; Guàrdia, E.; Masia, M. The polarizable point dipoles method with electrostatic damping: Implementation on a model system. *J. Chem. Phys.* **2010**, *133*, 234101. [[CrossRef](#)] [[PubMed](#)]
26. Horvath, S.; McCoy, A.B.; Elliott, B.M.; Weddle, G.H.; Roscioli, J.R.; Johnson, M.A. Anharmonicities and Isotopic Effects in the Vibrational Spectra of X-H₂O, HDO, and D₂O [X = Cl, Br, and I] Binary Complexes. *J. Phys. Chem. A* **2010**, *114*, 1556–1568. [[CrossRef](#)]
27. Vega, C.; Abascal, J.L.F. Simulating water with rigid non-polarizable models: A general perspective. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19663–19688. [[CrossRef](#)]
28. Trumm, M.; Martínez, Y.O.G.; Réal, F.; Masella, M.; Vallet, V.; Schimmelpfennig, B. Modeling the hydration of mono-atomic anions from the gas phase to the bulk phase: The case of the halide ions F⁻, Cl⁻, and Br⁻. *J. Chem. Phys.* **2012**, *136*, 044509. [[CrossRef](#)]
29. Wolke, C.T.; Menges, F.S.; Tötsch, N.; Gorlova, O.; Fournier, J.A.; Weddle, G.H.; Johnson, M.A.; Heine, N.; Esser, T.K.; Knorke, H.; et al. Thermodynamics of Water Dimer Dissociation in the Primary Hydration Shell of the Iodide Ion with Temperature-Dependent Vibrational Predissociation Spectroscopy. *J. Phys. Chem. A* **2015**, *119*, 1859–1866. [[CrossRef](#)] [[PubMed](#)]
30. Wick, C.D.; Xantheas, S.S. Computational Investigation of the First Solvation Shell Structure of Interfacial and Bulk Aqueous Chloride and Iodide Ions. *J. Phys. Chem. B* **2009**, *113*, 4141–4146. [[CrossRef](#)] [[PubMed](#)]
31. Baer, M.D.; Mundy, C.J. Toward an Understanding of the Specific Ion Effect Using Density Functional Theory. *J. Phys. Chem. Lett.* **2011**, *2*, 1088–1093. [[CrossRef](#)]
32. Werhahn, J.C.; Akase, D.; Xantheas, S.S. Universal scaling of potential energy functions describing intermolecular interactions. II. The halide-water and alkali metal-water interactions. *J. Chem. Phys.* **2014**, *141*, 064118. [[CrossRef](#)]
33. Cisneros, G.A.; Wikfeldt, K.T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A.P.; Csányi, G.; Molinero, V.; Paesani, F. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions. *Chem. Rev.* **2016**, *116*, 7501–7528. [[CrossRef](#)]
34. Gillan, M.J.; Alfé, D.; Michaelides, A. Perspective: How good is DFT for water? *J. Chem. Phys.* **2016**, *144*, 130901. [[CrossRef](#)] [[PubMed](#)]
35. Arismendi-Arrieta, D.J.; Riera, M.; Bajaj, P.; Prosmi, R.; Paesani, F. i-TTM Model for Ab Initio-Based Ion–Water Interaction Potentials. 1. Halide–Water Potential Energy Functions. *J. Phys. Chem. B* **2016**, *120*, 1822–1832. [[CrossRef](#)] [[PubMed](#)]
36. Dasgupta, S.; Lambros, E.; Perdew, J.P.; Paesani, F. Elevating density functional theory to chemical accuracy for water simulations through a density-corrected many-body. *Nat. Commun.* **2021**, *12*, 6359. [[CrossRef](#)]
37. Joseph P. Heindel, J.P.; Xantheas, S.S. The Many-Body Expansion for Aqueous Systems Revisited: II. Alkali Metal and Halide Ion–Water Interactions. *J. Chem. Theory Comput.* **2021**, *17*, 2200–2216.
38. Amann-Winkel, K.; Bellissent-Funel, M.C.; Bove, L.E.; Loerting, T.; Nilsson, A.; Paciaroni, A.; Schlesinger, D.; Skinner, L. X-ray and Neutron Scattering of Water. *Chem. Rev.* **2016**, *116*, 7570–7589. [[CrossRef](#)]
39. Antalek, M.; Pace, E.; Hedman, B.; Hodgson, K.O.; Chillemi, G.; Benfatto, M.; Sarangi, R.; Frank, P. Solvation structure of the halides from x-ray absorption spectroscopy. *J. Chem. Phys.* **2016**, *145*, 044318. [[CrossRef](#)] [[PubMed](#)]
40. Dorsett, H.E.; Watts, R.O.; Xantheas, S.S. Probing Temperature Effects on the Hydrogen Bonding Network of the Cl⁻(H₂O)₂ Cluster. *J. Phys. Chem. A* **1999**, *103*, 3351–3355. [[CrossRef](#)]
41. Kim, J.; Lee, H.M.; Suh, S.B.; Majumdar, D.; Kim, K.S. Comparative ab initio study of the structures, energetics and spectra of X⁻(H₂O)_{n=1-4} [X = F, Cl, Br, I] clusters. *J. Chem. Phys.* **2000**, *113*, 5259–5272. [[CrossRef](#)]
42. Prakash, M.; Gopalsamy, K.; Subramanian, V. Studies on the structure, stability, and spectral signatures of hydride ion-water clusters. *J. Chem. Phys.* **2011**, *135*, 214308. [[CrossRef](#)]
43. Neogi, S.G.; Chaudhury, P. Structure and spectroscopic aspects of water-halide ion clusters: A study based on a conjunction of stochastic and quantum chemical methods. *J. Comput. Chem.* **2013**, *34*, 471–491. [[CrossRef](#)]
44. Kamarchik, E.; Toffoli, D.; Christiansen, O.; Bowman, J.M. Ab initio potential energy and dipole moment surfaces of the F-(H₂O) complex. *Spectrochim. Acta A* **2014**, *119*, 59–62. [[CrossRef](#)] [[PubMed](#)]
45. León-Merino, I.; Rodríguez-Segundo, R.; Arismendi-Arrieta, D.J.; Prosmi, R. Assessing Intermolecular Interactions in Guest-Free Clathrate Hydrate Systems. *J. Phys. Chem. A* **2018**, *122*, 1479–1487. [[CrossRef](#)] [[PubMed](#)]
46. Bizzarro, B.B.; Egan, C.K.; Paesani, F. Nature of Halide–Water Interactions: Insights from Many-Body Representations and Density Functional Theory. *J. Chem. Theory Comput.* **2019**, *15*, 2983–2995. [[CrossRef](#)] [[PubMed](#)]
47. Perera, L.; Berkowitz, M.L. Many-body effects in molecular dynamics simulations of Na⁺(H₂O)_n and Cl⁻(H₂O)_n clusters. *J. Chem. Phys.* **1991**, *95*, 1954–1963. [[CrossRef](#)]
48. Burnham, C.J.; Xantheas, S.S. 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.* **2002**, *116*, 5115–5124. [[CrossRef](#)]
49. Rheinecker, J.; Bowman, J.M. The calculated infrared spectrum of Cl-H₂O using a new full dimensional ab initio potential surface and dipole moment surface. *J. Chem. Phys.* **2006**, *125*, 133206. [[CrossRef](#)]
50. Fanourgakis, G.S.; Xantheas, S.S. The Flexible, Polarizable, Thole-Type Interaction Potential for Water (TTM2-F) Revisited. *J. Phys. Chem. A* **2006**, *110*, 4100–4106. [[CrossRef](#)]

51. Fanourgakis, G.S.; Xantheas, S.S. 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.* **2008**, *128*, 074506. [CrossRef]
52. Babin, V.; Leforestier, C.; Paesani, F. Development of a “First Principles” Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. *J. Chem. Theory Comput.* **2013**, *9*, 5395–5403. [CrossRef]
53. Kamarchik, E.; Bowman, J.M. Coupling of Low- and High-Frequency Vibrational Modes: Broadening in the Infrared Spectrum of F-(H₂O)₂. *J. Phys. Chem. Lett.* **2013**, *4*, 2964–2969. [CrossRef]
54. Werhahn, J.C.; Miliordos, E.; Xantheas, S.S. 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.* **2015**, *619*, 133–138. [CrossRef]
55. Metz, M.P.; Piszczatowski, K.; Szalewicz, K. Automatic Generation of Intermolecular Potential Energy Surfaces. *J. Chem. Theory Comput.* **2016**, *12*, 5895–5919. [CrossRef] [PubMed]
56. Zhai, Y.; Caruso, A.; Gao, S.; Paesani, F. Active learning of many-body configuration space: Application to the Cs⁺—water MB-nrg potential energy function as a case study. *J. Chem. Phys.* **2020**, *152*, 144103. [CrossRef] [PubMed]
57. Bilbrey, J.A.; Heindel, J.P.; Schram, M.; Bandyopadhyay, P.; Xantheas, S.S.; Choudhury, S. A look inside the black box: Using graph-theoretical descriptors to interpret a Continuous-Filter Convolutional Neural Network (CF-CNN) trained on the global and local minimum energy structures of neutral water clusters. *J. Chem. Phys.* **2020**, *153*, 024302. [CrossRef]
58. Nandi, A.; Qu, C.; Houston, P.L.; Conte, R.; Yu, Q.; Bowman, J.M. A CCSD(T)-Based 4-Body Potential for Water. *J. Phys. Chem. Lett.* **2021**, *12*, 10318–10324. [CrossRef]
59. Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121*, 10037–10072. [CrossRef]
60. Smith, J.S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A.E. Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148*, 241733. [CrossRef]
61. Westermayr, J.; Gastegger, M.; Schütt, K.T.; Maurer, R.J. Perspective on integrating machine learning into computational chemistry and materials science. *J. Chem. Phys.* **2021**, *154*, 230903. [CrossRef]
62. Ceriotti, M.; Clementi, C.; Anatole von Lilienfeld, O. Machine learning meets chemical physics. *J. Chem. Phys.* **2021**, *154*, 160401. [CrossRef]
63. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. *Gaussian16 Revision C.01*; Gaussian Inc.: Wallingford, CT, USA, 2016.
64. Becke, A.D.; Johnson, E.R. Exchange-hole dipole moment and the dispersion interaction revisited. *J. Chem. Phys.* **2007**, *127*, 154108. [CrossRef]
65. Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465. [CrossRef] [PubMed]
66. Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **2019**, *150*, 154122. [CrossRef] [PubMed]
67. Baudin, P.; Ettenhuber, P.; Reine, S.; Kristensen, K.; Kjærgaard, T. Efficient linear-scaling second-order Møller-Plesset perturbation theory: The divide—expand—consolidate RI-MP2 model. *J. Chem. Phys.* **2016**, *144*, 054102. [CrossRef]
68. Werner, H.J.; Knowles, P.; Knizia, G.; Manby, R.; Schütz, M.; Celani, P.; Györfy, W.; Kats, D.; Korona, T.; Lindh, R.; et al. MOLPRO, Version 2012.1, a Package of ab Initio Programs. 2012. Available online: <http://www.molpro.net> (accessed on 28 January 2022).
69. Murrell, J.N.; Carter, S.; Farantos, S.C.; Huxley, P.; Varandas, A. *Molecular Potential Energy Function*; Wiley: Hoboken, NJ, USA, 1984.
70. Partridge, H.; Schwenke, D.W. The determination of an accurate isotope dependent potential energy surface for water from extensive ab initio calculations and experimental data. *J. Chem. Phys.* **1997**, *106*, 4618–4639. [CrossRef]
71. Burnham, C.J.; Anick, D.J.; Mankoo, P.K.; Reiter, G.F. The vibrational proton potential in bulk liquid water and ice. *J. Chem. Phys.* **2008**, *128*, 154519. [CrossRef]
72. Medders, G.R.; Babin, V.; Paesani, F. Development of a “First-Principles” Water Potential with Flexible Monomers. III. Liquid Phase Properties. *J. Chem. Theory Comput.* **2014**, *10*, 2906–2910. [CrossRef]
73. Medders, G.R.; Paesani, F. Infrared and Raman Spectroscopy of Liquid Water through “First-Principles” Many-Body Molecular Dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1145–1154. [CrossRef]
74. Tang, K.T.; Toennies, J.P. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *J. Chem. Phys.* **1984**, *80*, 3726–3741. [CrossRef]
75. Born, M.; Mayer, J.E. Zur Gittertheorie der Ionenkristalle. *Z. Phys.* **1932**, *75*, 1–18. [CrossRef]
76. Otero-de-la Roza, A.; Johnson, E.R. Non-covalent interactions and thermochemistry using XDM-corrected hybrid and range-separated hybrid density functionals. *J. Chem. Phys.* **2013**, *138*, 204109. [CrossRef]
77. Boys, S.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566. [CrossRef]
78. Iwamatsu, M. Applying evolutionary programming to structural optimization of atomic clusters. *Comput. Phys. Commun.* **2001**, *142*, 214–218. [CrossRef]
79. Fogel, D. *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*; IEEE Press-Wiley: Hoboken, NJ, USA, 2006.
80. Pérez de Tudela, R.; Barragán, P.; Valdés, A.; Prosimiti, R. Energetics and Solvation Structure of a Dihalogen Dopant (I₂) in ⁴He Clusters. *J. Phys. Chem. A* **2014**, *118*, 6492–6500. [CrossRef]

81. Alharzali, N.; Rodríguez-Segundo, R.; Prosimiti, R. Modelling interactions of cationic dimers in He droplets: Microsolvation trends in He_nK_2^+ clusters. *Phys. Chem. Chem. Phys.* **2021**, *23*, 7849–7859. [CrossRef]
82. The Cambridge Cluster Database. 2021. Available online: <http://www.wales.ch.cam.ac.uk/CCD.html> (accessed on 30 December 2021).
83. DENEb 1.30 Beta: The Nanotechnology Software by Atelgraphics. 2020. Available online: <https://www.atelgraphics.com> (accessed on 28 January 2022).
84. Bajaj, P.; Riera, M.; Lin, J.K.; Mendoza Montijo, Y.E.; Gazca, J.; Paesani, F. Halide Ion Microhydration: Structure, Energetics, and Spectroscopy of Small Halide–Water Clusters. *J. Phys. Chem. A* **2019**, *123*, 2843–2852. [CrossRef] [PubMed]
85. Rodríguez-Segundo, R.; Arismendi-Arrieta, D.; Prosimiti, R. A bottom-up approach for ion-water interactions: From clusters to bulk. In *State-of-the-Art Reviews on Energetic Ion-Atom and Ion-Molecule Collisions*; World Scientific: Singapore, 2019; Chapter 8, pp. 205–225.
86. Werner, H.J.; Knowles, P.J.; Knizia, G.; Manby, F.R.; Schütz, M. Molpro: A general-purpose quantum chemistry program package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 242–253. [CrossRef]
87. Dunning, T.H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023. [CrossRef]
88. Woon, D.E.; Dunning, T.H. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *J. Chem. Phys.* **1994**, *100*, 2975–2988. [CrossRef]
89. Peterson, K.A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. 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.* **2003**, *119*, 11113–11123. [CrossRef]
90. Peterson, K.A.; Shepler, B.C.; Figgen, D.; Stoll, H. On the Spectroscopic and Thermochemical Properties of ClO, BrO, IO, and Their Anions. *J. Phys. Chem. A* **2006**, *110*, 13877–13883. [CrossRef]
91. Hill, J.G.; Peterson, K.A. Correlation consistent basis sets for explicitly correlated wavefunctions: Pseudopotential-based basis sets for the post-d main group elements Ga–Rn. *J. Chem. Phys.* **2014**, *141*, 094106. [CrossRef] [PubMed]
92. Universität-Bonn. A Dispersion Correction for Density Functionals, Hartree-Fock and Semi-Empirical Quantum Chemical Methods. Available online: <https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/dft-d3/dft-d3> (accessed on 28 January 2022).
93. Universität-Bonn. D4—A Generally Applicable Atomic-Charge Dependent London Dispersion Correction. Available online: <https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/dftd4> (accessed on 28 January 2022).
94. Tao, J.; Perdew, J.P.; Staroverov, V.N.; Scuseria, G.E. Climbing the Density Functional Ladder: Non empirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401. [CrossRef] [PubMed]
95. Zhao, Y. and Truhlar, D.G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006**, *125*, 194101. [CrossRef] [PubMed]
96. Becke, A.D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652. [CrossRef]
97. Zhao, Y.; Truhlar, D.G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, non-covalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215.
98. Yanai, T.; Tew, D.P.; Handy, N.C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51. [CrossRef]
99. Vydrov, O.A.; Scuseria, G.E. Assessment of a long-range corrected hybrid functional. *J. Chem. Phys.* **2006**, *125*, 234109. [CrossRef]
100. Chai, J.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 084106. [CrossRef]
101. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **2010**, *132*, 154104. [CrossRef]
102. Smith, D.; Burns, L.; Patkowski, K.; Sherrill, C. Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory. *J. Phys. Chem. Lett.* **2016**, *7*, 2197–2203. [CrossRef]