Novel Computational Chemistry Infrastructure for Simulating Astatide in Water: From Basis Sets to Force Fields Using Particle Swarm Optimization

Kennet J. Rueda Espinosa,[†] Alexei A. Kananenka,[*,†] and Alexander A. Rusakov[*,‡]

[†]*Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716, USA*

[‡]*Department of Chemistry, Oakland University, Rochester, Michigan, 48309 USA*

E-mail: akanane@udel.edu; rusakov@oakland.edu

Abstract

Using the example of astatine, the heaviest naturally occurring halogen whose isotope At-211 has promising medical applications, we propose a new infrastructure for large-scale computational models of heavy elements with strong relativistic effects. In particular, we focus on developing an accurate force field for At^- in water based on reliable relativistic DFT calculations. To ensure such calculations' reliability, we design novel basis sets for relativistic DFT, via the particle swarm optimization algorithm to optimize the coefficients of the new basis sets and the polarization-consistent basis set idea's extension to heavy elements, to eliminate the basis-set error from DFT calculations. The resulting basis sets enable the well-grounded evaluation of relativistic DFT against "gold-standard" CCSD(T) results. Accounting for strong relativistic effects, including spin-orbit interaction, via our redesigned infrastructure, we elucidate a noticeable dissimilarity between At^- and I^- in

halide–water force field parameters, radial distribution functions, diffusion coefficients, and hydration energies. This work establishes the framework for the systematic development of polarization-consistent basis sets for relativistic DFT and accurate force fields for molecular dynamics simulations to be used in large-scale models of complex molecular systems with the elements from the bottom of the periodic table, including actinides and even superheavy elements.

I Introduction

The establishment of The World Astatine Community[1,2] at the 12th International Symposium for Targeted Alpha Therapy[3] champions the joint effort of scientists from the United States, European Union, and Japan to advance the isotope Astatine-211 as a uniquely suitable agent for treating blood, ovarian, brain, as well as disseminated malignancies. The transition from ^{211}At's cyclotron production to a systematic development of radiopharmaceuticals and subsequent clinical trials is predicated on the element's thorough physical and chemical characterization. The recent progress in that direction,[4,5] while undeniably significant, remains insufficient to realize the full clinical potential of the ^{211}At isotope.[6,7] Therefore, rational design of ^{211}At-based pharmaceuticals resistant to *in vivo* deastatination requires reliable theoretical predictions of astatine's chemistry involving, in particular, complex molecules in solutions.

Modeling small molecules containing astatine is possible with sophisticated *ab initio* methods of relativistic quantum chemistry, although it requires significant computing resources even for diatomic molecules.[8] Large-scale molecular simulations are only feasible within cheaper density functional theory (DFT) or molecular dynamics (MD). The former requires the approximations of the exchange-correlation functional (XCF) and basis sets capable of accurately describing significant relativistic effects in many-electron systems. The latter should quantitatively account for such effects on intra- and intermolecular interactions through force fields (FFs), thus also depending on the access to high-level electronic-structure calculations. Consequently, developing suitable molecular FFs for complex systems with heavy elements, such as astatine, in conjunction

with establishing reliable DFT-based approaches to such systems is critical for the task at hand.

The development of accurate force fields to model ions in water has received considerable attention in the past few decades.[9–13] Halides X^- are of particular interest due to their aqueous solutions' universal presence in biological systems, which explains substantial progress in developing FFs for X^-–$(H_2O)_n$ models.[9,14–16] However, the heaviest halide At^- has hitherto remained in relative obscurity with only one FF[17] and two microsolvation[18,19] models developed for it, and researchers resorting to standard ones when studying biological systems with astatine.[20] In ref 17, Réal *et al.* developed a FF for At^--water systems by reparametrizing a polarizable TPCE/2013[21] water model. They found At^--water and I^--water radial distribution functions (RDFs) virtually indistinguishable. Furthermore, they established the diffusion coefficient of At^- is only 6% smaller and the hydration free energy 1.4 kcal/mol less negative than the respective quantities for I^-. These findings allow the authors of ref 17 to conclude that the basic structural and dynamical properties of I^-– and At^-–water systems are quite similar. Chamorro *et al.*[18] studied finite-size halide-water clusters using second-order Møller–Plesset perturbation theory (MP2) and reported small but noticeable differences in energetic properties of At^-–$(H_2O)_n$ and I^-–$(H_2O)_n$ clusters. However, several experimental studies[22–26] indicate that I^- and At^-, which are the products of ^{131}I- and ^{211}At-radiopharmaceuticals' *in vivo* dehalogenation, have different distribution and retention times in various organs. While this phenomenon is likely much more complex than halide–water interactions and remains to be explained, it suggests a potential discrepancy between the available theoretical model of I^-/At^- structural and dynamical properties in water and the known empirical data. Given that water is the major component of the human body,[27] revisiting the theory of I^-/At^-–water interaction is the necessary first step towards quantifying the At^- behavior *in vivo*.

In this article, we focus on the systematic development of FFs for accurate and efficient large-scale X^-–$(H_2O)_n$ MD simulations of heavy halides, primarily the astatide anion At^-. In X^-–$(H_2O)_n$ systems, the water molecule is described by a single Lennard-Jones (LJ) site centered on the oxygen atom and three charge sites, each centered on hydrogen atoms and oxygen atom or

dummy atom. The halide is described by the LJ site and a charge. The latter is typically taken to be that of the isolated gas-phase halide ions -1. Such simple non-polarizable FFs remain the method of choice for MD simulations of large biological systems due to their affordable computational cost. Polarizable models are more computationally expensive but not necessarily better.[28] The effects of polarizability can be approximately added to the framework on non-polarizable fixed charge FFs. Leontyev and Stuchebrukov have combined a non-polarizable fixed-charge FF with a phenomenological electronic continuum model for electronic polarization.[29–34] This method has been referred to as molecular dynamics in a dielectric continuum (MDEC).[29–34] It includes the effects of electronic polarization and screening approximately via a simple scaling of partial charges. MDEC generally provides good results[35–39] sometimes outperforming polarizable models.[28]

Successful FF development for X^-–$(H_2O)_n$ systems [9,14–17] relies on DFT data carefully evaluated against highly accurate yet still computationally tractable *ab initio* approaches such as MP2 or CCSD(T) with the complete basis set (CBS) extrapolation. Descending from light halides to I^- and beyond calls for using relativistic versions of the wave function and DFT methods. Relativistic wave function calculations are as rigorous as their established non-relativistic counterparts, albeit much more demanding computationally. Relativistic density functional theory (RDFT), while derived formally[40–44] and implemented efficiently as two-component (2c)[45–50] and four-component (4c)[51–54] extensions of generalized Kohn–Sham DFT, faces several practical complications. There are hardly any relativistic XCF approximations relevant to chemical modeling, which makes it standard to resort to the plethora of XCF approximations developed in the non-relativistic quantum-chemical community. With very few exceptions of (almost) non-empirical XCFs,[55–58] such approximations have parameters fitted to reproduce mostly light-element data. Hence, these XCFs' transferability to modeling heavy-element chemistry with significant relativistic effects cannot be guaranteed and requires careful evaluation.[59–61] Furthermore, Aebersold and Wilson demonstrate[62] that deficiencies in XCFs' performance are heavily entangled with those of the basis sets. Surprisingly, there are very few classes of basis sets developed

explicitly for DFT,[63–69] with even fewer[69–73] available for heavy elements: e.g., polarization-consistent basis sets[63–66] cover elements H through Kr only. Significant errors from using common DFT–non-specific basis sets in RDFT, as explored in Ref. 62, and a limited scope of DFT-specific ones to choose from make basis-set development for RDFT as necessary as the XCFs' revision.

The accuracy of the X^-–$(H_2O)_n$ RDFT electronic-structure data for FF parametrization depends on both the adequacy of the XCF approximation and the basis set's quality, of which we address the latter. With At^-–$(H_2O)_n$ interactions as this work's focal point, we construct bespoke RDFT basis sets for At. To this end, we extend the idea of a polarization-consistent[63–66,74] basis set and combine it with the particle swarm optimization (PSO),[75] a metaheuristic biologically inspired algorithm for a simple and efficient global minimum search of a multivariable function. This algorithm has been successfully applied to several problems in physical chemistry, such as minimum-energy structure search[76–78] and kinetic mechanisms.[79] To the best of our knowledge, this article presents the first PSO application for basis set optimization.

To put our primary At^-–$(H_2O)_n$ case into perspective, we also develop FFs for the lighter halides F^- through I^-. Additionally, for an exhaustive assessment of the existing DFT functionals' behavior in X^-–water systems, we include some results for the superheavy halide tennesside, Ts^-. Our basis-set design for F^- – I^- and Ts^- follows a less scrupulous protocol than for At^-. A detailed study of the Ts case, where the magnitude of relativistic effects is extreme,[80,81] will be a subject of our future work.

II Theory

A Force field definition

In this work, a simple non-polarizable force field for the halide ions in water was optimized. The nonbonded water-water and water-ion interactions are modeled by the sum of the standard 12-6

Lennard-Jones (LJ) and Coulomb potentials

$$U(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{1}$$

where r_{ij} is the separation between two interacting sites, q_i and q_j are the partial atomic charges of sites i and j correspondingly, ε_0 is the permittivity of vacuum, σ_{ij} is the distance at which the potential energy $U(r_{ij})$ is zero and ε_{ij} is the depth of the potential well.

B Basis set

Arguably, the single most fruitful idea in quantum chemistry is the representation of molecular one-electron wave functions, or molecular orbitals (MOs), as finite linear combinations of fixed pre-selected functions centered on atomic nuclei, or atomic orbitals (AOs).[82] These latter functions, referred to collectively as a basis set, ultimately allow for casting the problem of evaluating an approximate molecular many-electron wave function as a finite problem of computational linear algebra. Choosing the basis set for a given electronic-structure simulation is a balancing act between capturing the system's "physics" (or "chemistry") to the best of the underlying quantum-chemical method's ability and keeping the respective calculation numerically stable and computationally affordable.[74,83]

In this article, we work exclusively within the most popular framework of Gaussian basis sets. In this formalism, an AO, typically resembling a hydrogen-like wave function, is approximated using a linear combination of the type[83,84]

$$\Phi^{\mathrm{cGTO}}(x,y,z) = A\sum_{i=1}^{N} c_i \phi_i^{\mathrm{GTO}}(x,y,z), \tag{2}$$

known as a contracted Gaussian-type orbital (cGTO), where A is the normalization factor, N is the contraction length, c_i are contraction coefficients, and $\phi_i^{\mathrm{GTO}}(x,y,z)$ are primitive Gaussian-type

orbitals (GTOs). These primitive GTOs have the form

$$\phi_i^{\mathrm{GTO}}(x,y,z) = \mathcal{N} x^a y^b z^c e^{-\alpha_i(x^2+y^2+z^2)}, \tag{3}$$

where $a+b+c=l$ is the angular momentum quantum number ($l=0$ for s-functions, $l=1$ for p-functions, etc.), α_i is the exponent of the ith GTO, and $\mathcal{N}$ is a normalization constant.

The exponents of primitive GTOs α_i and contraction coefficients c_i are adjustable parameters subject to optimization. The starting point is typically a primitive GTO basis set whose exponents $\{\alpha_i\}$ are optimized variationally in an atomic calculation. Then, optimized exponents $\{\alpha_i\}$ for nl-shells are contracted to form more compact basis sets.[83] A combination of exponents' optimization in atomic configuration interaction singles and doubles (CISD) calculations with the contraction schemes based on atomic natural orbitals (ANO) or Hartree–Fock (HF) solutions for atoms give rise to, respectively, ANO and correlation-consistent (cc) basis sets. In contrast, HF exponents' optimization with segmented contractions from atomic HF solutions results in the def2 basis-set class. Final basis sets have a well-established hierarchy, allowing for systematic error reduction and extrapolation to the complete basis set (CBS) limit, which makes their presence universal in quantum-chemical applications. However, by their construction, ANO and cc sets (partially def2 ones, too, as they take polarization functions from cc sets) are inherently tailored for correlated wavefunction calculations. The inferior performance of cc and related basis sets in DFT calculations has been long observed, leading Jensen to the development of an alternative hierarchy of polarization-consistent (pc) basis sets, arguably the only family designed for rapid systematic convergence of DFT energies and other properties with the basis-set size.[63] Unlike cc and similarly constructed sets that use high angular momentum, or correlation, functions for an accurate representation of the r_{ij}^{-1} two-electron repulsion operator in an atom, pc basis sets utilize such functions for a systematic description of the electron density's deformation, or polarization, in response to atoms forming molecules. Revising the high angular momentum functions' role, including molecules in exponents' optimization, and applying a segmented contraction scheme

result in compact basis sets with a massively superior DFT performance.

Because of the pc basis sets' unavailability beyond Kr, we extend Jensen's approach to elements with strong relativistic effects and benchmark it on the astatine example. We adopt a two-step optimization process. First, we optimize exponents with the PSO algorithm using a representative set of At-containing species. Then, we apply a segmented contraction scheme based on the atomic HF solutions while controlling all species' total energies.

C RECP and SO

To explore electron-correlation and relativistic effects on chemical properties of heavy elements economically, we rely on the accurate small-core shape-consistent relativistic effective core potential (RECP) model.[85,86] These RECPs allow for the explicit treatment of valence ns- and np- and subvalence (n-1)s-, (n-1)p-electrons, as well as (n-1)d ones for Br, I, At, and Ts. Thus, our models consider 15 electrons of Cl and 25 of heavier halogens. For consistency, we also use a zero-electron RECP on the F atom, thus adding finite-nucleus, scalar-relativistic, and spin-orbit effects to the Coulomb Hamiltonian of fluorine's 9 electrons.

D Particle Swarm Optimization

The PSO algorithm[75] was inspired by the observation of the motion of swarms of birds and insects, where each swarm member is guided not only by the best solution for itself but also by the best solution seen by the entire population. An object of a swarm that moves around in the search space is called a particle. The particle in the swarm includes variables (parameters or coefficients to be optimized) that are updated each iteration during the optimization based on the information about the previous best states of the particle and the swarm as a whole. Each particle is characterized by its velocity $\mathbf{v}(t)$, position $\mathbf{x}(t)$, and its best position $\mathbf{x}_p(t)$, the position that has generated the smallest error along the particle's trajectory. Additionally, the swarm's best global position $\mathbf{x}_g(t)$ is being recorded. In each iteration, the position and velocity of ith particle are updated according

to

$$\begin{aligned} \mathbf{v}_i(t+1) &= w\mathbf{v}_i(t) + c_p\left(\mathbf{x}_{p_i}(t) - \mathbf{x}_i(t)\right) + c_g\left(\mathbf{x}_g(t) - \mathbf{x}_i(t)\right) \\ \mathbf{x}_i(t+1) &= \mathbf{x}_i(t) + \mathbf{v}_i(t), \end{aligned} \tag{4}$$

where c_p, c_g, and w are the so-called cognitive, social, and inertia coefficients. These coefficients quantify how much the particle is directed toward the best solution seen by itself, by the swarm, and in the previous direction. The PSO code used in this work is available at https://github.com/kananenka-group/PSO_Astatine.

III Computational details

A Basis set optimization for At

Following a two-step procedure, we developed three uncontracted and a series of contracted basis sets. In the first step for a given uncontracted input basis set, PSO was used to optimize all the primitive exponents $\{\alpha\}$. The resulting optimized primitive Gaussians were further contracted to produce more compact basis sets. The details of these steps are given below.

A.1 Reference data for basis set generation

We develop astatine basis sets for two-component relativistic density functional theory (2c-RDFT)[45,47,48] calculations within the RECP model[85,86] described in Section C. We rely on accurate data for the neutral At atom, its ions, and small molecules to attain sufficient basis sets' flexibility for various chemical "surroundings," or oxidation states.

Reference data for the At basis set contain total ground-state energies of an At atom, At^{2+}, At^{+}, At^{-} ions, and also At_2 and HAt molecules for equilibrium internuclear separations R_e and for four nonequilibrium configurations at $R_e \pm 0.1$ and $R_e \pm 0.2$ Å. For each atom, ion, and molecular configuration, we calculate Hartree–Fock and PBE0[56] total energies for a sequence of triple-, quadruple-, and quintuple-ζ (respectively, TZ, QZ, and 5Z) correlation-consistent basis sets developed in ref 8 for spin-orbit calculations with the chosen RECP. The final energy values E_{CBS}

result from a complete basis set (CBS) extrapolation using the exponential formula[87]

$$E(X) = E_{CBS} + Be^{-\alpha X} \tag{5}$$

where $E(X)$ are energies for $X = 3,\ 4,\ 5$ corresponding to TZ, QZ, and 5Z sets, and $E_{CBS},\ B,\ \alpha$ are fitting parameters.

A.2 Optimization of exponents

We develop three uncontracted basis sets of varying sizes: PSO-L, PSO-M, and PSO-S. To initiate the PSO optimization of exponents for the largest one, PSO-L, we create an even-tempered basis of primitive exponents $\alpha_n = \alpha_1/q^{n-1}$ with the parameters $q = 1.958149706358805$ and α_1 taken from the universal Gaussian basis set (UGBS).[88] To adapt the basis sets for the RECP model, we choose the α_1 values of 243.715119463182, 243.715119463182, 124.461944187287, and 4.32321786011102 for s-, p-, d-, and f-functions, respectively. The resulting [15s, 15p, 13d, 6f] basis set, which we label "L," is the optimization procedure's input for PSO-L.

In addition to the PSO-L basis with the highest optimization flexibility and accuracy, we design smaller PSO-M [15s, 15p, 13d] and PSO-S [11s, 11p, 9d, 4f] sets to lower the computational cost. To initialize the PSO-M basis, we remove all f-functions from the L one to create a set "M." For PSO-S, we reduce the number of primitive functions (15s→11s, 15p→11p, 13d→9d, 6f→4f) by fixing the smallest and the largest exponents for each angular momentum at the respective initial L values and assigning the remaining exponents to form an even-tempered sequence, or an "S" basis.

Primitive exponents in PSO-L, PSO-M, and PSO-S basis sets are optimized using the PSO method. We kept the search within the space of physically allowable values of exponents through two restrictions. First, each exponent was varied in the range limited by the values of the neighboring ones. If, during the optimization, the PSO algorithm takes an exponent outside its allowable range, a new value is assigned to the exponent to put it back within the range. This value is chosen as follows. If an exponent α_n becomes larger than the largest of the two neighboring

exponents (α_{n+1}), the value is set below α_{n+1} at $\alpha \rightarrow R\alpha_{n+1}$ value where $R > 1$. Similarly, if in a PSO step, the exponent α_n becomes smaller than the smallest of the two neighboring exponents (α_{n-1}), the value is set slightly above α_n to $R\alpha_{n-1}$. The R value was determined in preliminary PSO calculations. We found that $R = 1.1$ allows us to avoid linear dependencies in the basis sets. A small fraction of intermediate basis sets in the PSO calculations still suffered from the linear dependency issue and were removed from the optimization process. Second, the values of the smallest exponents were restricted to be positive at each step. If an exponent takes a negative value, it is replaced by a random number in the interval of values initially set for this exponent. In all calculations involving the HAt molecule, the H basis set was kept fixed at aug-cc-pVTZ.

A swarm of 15 particles was used in PSO calculations. The following values of the coefficients are recommended in the literature:[89] $w = 0.729$, $c_p = 1.49445$, $c_g = 1.49445$. After preliminary tests, we found that a larger inertia coefficient $w = 0.829$ allowed for a more efficient optimization process and used this value in all subsequent calculations.

A.3 Contraction coefficients

Following the exponents' optimization, we produce a more compact basis set by applying an appropriate contraction scheme to the innermost primitive Gaussians. To this end, we adopt a segmented contraction pattern (see ref 73 and references therein) based on the radial parts of $5s_{1/2}$, $5p_{1/2,\,3/2}$, $5d_{3/2,\,5/2}$, $6s_{1/2}$, and $6p_{1/2,\,3/2}$ 2c-Hartree–Fock spinors of At^-. The following principles guide the choice of segments. First, segments include only the inner exponents, usually 2.0 or larger, for a given lj combination. Thus, we guarantee sufficient flexibility of the basis set for various situations of chemical binding. Second, a segment covers consecutive exponents with approximately proportionate contributions to the lj spinors. Thus, for example, we retain only one segment for $5s_{1/2}$ and $6s_{1/2}$ spinors where their radial parts have similar functional forms. Third, we admix the generalized contraction idea and create two segments over the same span of exponents when spin-dependent relativistic effects (spin-orbit interaction) are responsible for a substantial difference in the spatial parts of the respective spinors, as is the case with $5d_{3/2}$ and

$5d_{5/2}$. We also find this approach useful to properly describe the difference in the radial parts of $5p_{1/2}$ and $6p_{1/2}$ spinors past the few innermost exponents with relatively small coefficients. Fourth, we monitor the total atomic 2c-Hartree–Fock and PBE0 energies relative to their values in the uncontracted basis, striving at the deviation below 1 kcal/mol. We choose the one with the smallest energy deviation when deciding between multiple plausible schemes. Each stage of the contraction process ascends from s- to d-functions. In each of the functions' groups, we propose a plausible segment or, when warranted, generalized-style segments of the thus far uncontracted innermost exponents and accept or reject it based on the atomic total energy. Contracted PSO-C, PSO-M, and PSO-L basis sets are denoted as PSO-Sc, PSO-Mc, and PSO-Lc, respectively.

For other halides, we construct similar basis sets, albeit without optimization, using the exponents from def2-QZVP (F, Cl, Br, I)[67,68] and Dyall's quadruple-ζ (Ts)[90,91] basis sets and creating contractions schemes for the chosen RECP model.

B Optimization of force field parameters

LJ parameters ε and σ were optimized by minimizing a sum of square differences between the gradients as well as energies obtained directly from eq 1 and the reference gradients and energies for a data set of halide-water clusters. The data set comprised 800 $X^-\cdot(H_2O)_3$, 400 $X^-\cdot(H_2O)_4$, 100 $X^-\cdot(H_2O)_5$, and 50 $X^-\cdot(H_2O)_6$ clusters for each halide (X = F, Cl, Br, I, or At). To generate the structures for the data set for each halide, except for At^-, an MD trajectory was generated using Optimized Potentials for Liquid Simulations All-atom (OPLS/AA)[92] FF with scaled ion charges (see below) and a TIP4P water model.[93] Then, clusters with a halide and a given number of water molecules were extracted randomly from the MD trajectory. Because OPLS/AA parameters do not exist for At^-, At^-–$(H_2O)_n$ clusters were generated from the corresponding I^-–$(H_2O)_n$ clusters by substituting the halides. The assumption that I^-–$(H_2O)_n$ and At^-–$(H_2O)_n$ clusters are structurally very similar is supported by the results of ref 17 which show that At^-–O and At^-–H radial distribution functions are indistinguishable from their I^-–O and I^-–H counterparts.

For each X^-–$(H_2O)_n$ cluster, 2c-RDFT/PBE0 calculations with RECPs[85,86] on halogen atoms

were performed to determine total energies and gradients. As shown in Sec. A, PBE0 accurately describes X^-–H_2O dissociation energies. For At, we use the optimized uncontracted basis set PSO-L. A modified def2-QZVP basis set was used for O and H atoms. The modification amounts to removing the highest angular momentum functions: f for H and g for O atoms. Most of the calculations were performed with NWCHEM software.[94] However, we used TURBOMOLE[48,95] for the ωB97X-D[96] XCF approximation, which is currently unavailable in NWCHEM, and all Ts calculations because the strength of spin-dependent relativistic effects requires the non-collinear[47,48] 2c-RDFT approach.

The optimization of LJ parameters was carried out with the FORCEBALANCE[97] program version 1.8 coupled to GROMACS package version 2018.1.[98–102] Newton–Raphson method with a convergence threshold of 0.1 was used in the FORCEBALANCE optimization. Initial values of LJ parameters for all halides except for At^- were taken from the OPLS/AA force field. Initial values for At^- were taken to be the same as for I^-. LJ parameters and the charges of the TIP4P water were held fixed. The charges of the halides are scaled appropriately for the TIP4P water model, as discussed below. The geometric mean combining rules $\sigma_{ij} = \left(\sigma_{ii}\sigma_{jj}\right)^{1/2}$ and $\varepsilon_{ij} = \left(\varepsilon_{ii}\varepsilon_{jj}\right)^{1/2}$ were used for interactions between unlike atoms as is done in the OPLS/AA force field[92] and implemented in GROMACS package.

C Molecular dynamics simulations

Classical MD simulations are performed using GROMACS package. Water was described by a rigid water model TIP4P. Initial configurations were generated by taking snapshots from a pure water simulation with 500 water molecules in a cubic box with an edge length of 24.6 Å and randomly replacing one water molecule with a halide. Therefore, each simulation box contained one ion and 499 TIP4P water molecules. Initial velocities were generated randomly from a Maxwellian distribution. In all calculations, three-dimensional periodic boundary conditions are applied, and electrostatic interactions are calculated using the particle-mesh Ewald summation.[103,104] A switch function was used for the LJ interactions to shift the LJ force starting from 10 Å to make it zero

at a cut-off distance of 12 Å. All simulations described below were performed with a time step of 1 fs. The geometries of water molecules for the rigid TIP4P water models were constrained using the SETTLE algorithm.[105]

The MDEC method[29–34] considers point charges moving in an electronic continuum of known dielectric constant ε_{el}. In MD simulations, the effects of electronic dielectric screening are taken into account implicitly by using scaled partial charges $q_{eff} = q/\sqrt{\varepsilon_{el}}$. In their work Leontyev *et al.*[29–34] used the experimentally determined optical dielectric constant of water $\varepsilon_{\infty} = 1.78$ to approximate ε_{el} giving the scaling factor of $\varepsilon_{el}^{-1/2} = 0.75$. For MD simulations with a given water simulation model, the true dielectric constant is given by $\varepsilon_{sim}\varepsilon_{\infty}$, where ε_{sim} is the dielectric constant of the water model.[38] In this work we follow Kann *et al.*[39] and use dielectric constant of TIP4P water model $\varepsilon_{el} = \varepsilon_0/\varepsilon_{sim}$ where $\varepsilon_0 = 78.4$ is the experimental value of the dielectric constant of water.[106] This guarantees that the true dielectric constant for any simulation model is given by experimental value. This approach reproduces the concentration dependence of the density and diffusion coefficients of solutions of alkali halides outperforming even some polarizable models.[39] For TIP4P water model $\varepsilon_{sim} \approx 50$[39,107] which yields the scaling factor of 0.80. Because the unscaled charges of all anions are taken to be those of gas-phase anions, -1 a.u, the charges of all halides in MD simulations are -0.8 a.u.

Radial distribution functions (RDFs) are calculated as follows. Firstly, energy minimization was performed for 20,000 steps using the steepest descent method. Then an isothermal-isobaric (NPT) ensemble simulation was conducted for 1 ns at 298.15 K and 1 atm, followed by a 1.0 ns NVT ensemble simulation at 298.15 K. The Berendsen barostat with the relaxation time of 0.5 ps was used to control the pressure in NPT ensembles.[108] Finally, the RDFs were calculated for the production run trajectory generated by a 10 ns NVT ensemble simulation at 298.15 K. In all NVT simulations, the temperature was controlled by velocity rescaling with the time constant of 0.5 ps.[109]

Diffusion coefficients were calculated from ions' mean squared displacements (MSD) using the approach adopted from ref 110. For each halide-water system, 20 independent MD simulations

were performed, as described below. Firstly, energy minimization was performed for 10,000 steps using the steepest descent method. Then, a 360 ps simulation in the NVT ensemble was performed to heat the system to 298.15 K. Langevin thermostat with the collision frequency of 2.0 ps^{-1} was used to control the temperature.[111] The NVT simulation is followed by 2 ns equilibration in the NPT ensemble at 298.15 K and 1 atm. The Berendsen barostat with the relaxation time of 0.5 ps was used to control the pressure. A second NVT simulation is performed for 1 ns at 298.15 K to equilibrate the system further. The resulting equilibrated system was further propagated in 80 successive steps consisting of NPT and NVE simulations. Each NPT simulation was performed for 5 ps to equilibrate the system. The following production NVE calculation was 20 ps in length, and the snapshots were saved every 0.2 ps. These steps are required to guarantee MSD vs. time linearity.[110] In each of the 80 cycles, the diffusion coefficients were calculated using

$$D = \lim_{t\to\infty} \frac{\langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle}{6t}, \tag{6}$$

where $\mathbf{r}(t) - \mathbf{r}(0)$ is the displacement of the particle at time t.

After the diffusion coefficients for each of the 20 ps NVE production runs are obtained, they are averaged

$$D_R = \frac{\sum_{i=1}^{80} D_i}{80}. \tag{7}$$

These values are then averaged over 20 independent runs, which gave the diffusion coefficients

$$D_{sim} = \frac{\sum_{i=1}^{20} D_R^i}{20}. \tag{8}$$

The error bars were determined from the standard deviation calculated in these 20 independent runs.

All diffusion coefficients reported herein include two corrections. The first correction is due to finite-size effects[112]

$$D_{sim}^c = D_{sim} + \frac{\zeta k_B T}{6\pi\eta L}, \tag{9}$$

where D is the diffusion coefficient in the infinite size limit, $\zeta = 2.83729$ for a cubic simulation box, η is the viscosity, k_B is the Boltzmann's constant, and L is the box length. Here, we used the experimental viscosity of pure water 8.93×10^{-4} kg/m s.[113] Finally, each of D^c_{sim} values were scaled by the ratio of the experimental diffusion coefficient of water and that of the TIP4P water $D = D^c_{sim} D^{wat}_{exp}/D^{wat}_{\mathrm{TIP4P}}$, where $D^{wat}_{exp} = 2.3 \times 10^{-5}$ cm^2/s.

Gibbs hydration free energies are calculated using Bennett's acceptance ratio (BAR)[114] method as implemented in GROMACS. The hydration free energy of an ion (ΔG_{hyd}) was calculated along a path that decouples a single solvated ion from the bulk solvent. Since ΔG_{hyd} is a path-dependent state function, we performed a two-step path that first decouples the ion electrostatic interactions along the path coordinate λ_{Coul} and then decouples the ion-water LJ terms along λ_{LJ} path. This two-step path is computationally convenient because it avoids simulation instabilities resulting from bare Coulombic contacts in the absence of LJ repulsion. The total hydration free energy is the sum of the two contributions $\Delta G_{hyd} = \Delta G^{LJ}_{hyd} + \Delta G^{Coul}_{hyd}$. We used 20 equally spaced steps each for λ_{LJ} and λ_{Coul} coordinates from 0 to 1. For each value of λ_{LJ} (λ_{Coul}) the following simulations were carried out. Energy minimization was performed for a simulation box containing one ion and 499 TIP4P water molecules for 20,000 steps using the steepest descent method. Then, a canonical ensemble (NVT) calculation was performed for 500 ps and at 298 K. Initial velocities were generated from a Maxwellian distribution. An NVT simulation was followed by 500 ps of simulation in the NPT ensemble to equilibrate the system. Finally, the data for hydration energy calculations were generated in 5 ns NPT ensemble calculation at a temperature of 298.15 K and a pressure of 101.3 kPa. Long-range dispersion corrections were applied for energy and pressure in all MD simulations. Equations of motion were integrated using Langevin stochastic dynamics (SD) algorithm.[115] Temperature coupling was done implicitly while using the SD algorithm, and the Parrinello–Rahman method was used to control pressure.

In MD simulations with periodic boundary conditions, the calculated absolute values of hydration free energies need to be corrected to account for artifacts arising from the interaction between ions and their periodic replicas and periodic ion-solvent artifacts in the Ewald summa-

tion.[116–119] Furthermore, for simulations at constant pressure corrections due to the difference in the ideal gas standard state between simulation and experiment,[116,119] and the correction due to the surface potential of the vacuum/water interface.[116] More detailed explanations for these corrections can be found, *e.g.*, in ref 116. Even after applying these standard corrections, the agreement between simulations and experiments is often still not satisfactory, and more corrections have been suggested. In this work, the absolute hydration free energies calculated for each ion were not corrected. Instead, we choose to report the differences between the hydration energies of halides.

IV Results and Discussion

A Choice of the exchange-correlation functional

Computationally feasible electronic-structure calculations of large X^-–$(H_2O)_n$ clusters to generate reference data for force field optimization rely on accurate DFT models. We use the ability of an exchange-correlation functional (XCF) approximation to reproduce CCSD(T) structures and binding energies of X^-–H_2O systems as the XCF's suitability criterion.[9,14,15,120,121] Direct CCSD(T) geometry optimization within the scalar-relativistic approximation for lighter halides should be accurate and computationally affordable. However, to remain at the same level of accuracy for heavier ones, it is necessary to account for spin-dependent relativistic effects using relativistic coupled-cluster 2c-CCSD(T) while doing so renders calculations unreasonably demanding. We have developed the following scheme for a consistent and computationally affordable evaluation of XCFs' suitability.

For the initial optimization of X^-–H_2O structures, we rely on the long-range (LR) corrected version of the hybrid functional PBE0,[56] LRC-ωPBEh.[122] The rationale for this choice stems from the non-empirical nature of PBE0, its overall reliability for systems with significant relativistic effects, and the proper LR behavior of the electron densities necessary for the accurate description of X^-–H_2O electrostatic interactions towards dissociation. It turns out that for F^-–H_2O and Cl^-–

H_2O, where spin-orbit effects are negligible, further structure optimization at the CCSD(T)/aug-cc-pVTZ level lowers the total energies by less than 0.01 kcal/mol relative to the optimal geometries from LRC-ωPBEh calculations with the basis sets developed in this paper. Thus, LRC-ωPBEh proves remarkably accurate to yield optimal X^-–H_2O geometries and allows us to rely on it to generate optimized X^-–H_2O structures for heavy halides with significant spin-orbit effects.

To build X^-–H_2O binding energy profiles, we create a 0.1 Å-step grid of $R(X^-–O)$ distances and optimize the $\angle(X^-–O–H)$ angle for each R value. The water molecule's configuration is kept fixed at the TIP4P parameters in these optimizations. It is worth noting that the resulting X^-–H_2O complexes for X^- from F^- through At^- monodentate except for the $R(X^-–O)$ shorter than the equilibrium value by *ca.* 0.3 Å. These findings agree with the results in ref 17. Interestingly, unlike its lighter counterparts, the superheavy halide Ts^- demonstrates a bidentate binding pattern.

For all subsequent calculations, we take the X^-– H_2O geometries optimized, as explained above, at the spin-orbit two-component relativistic DFT (2c-RDFT) level with the halogen basis sets developed in this work (PSO-Lc for astatine and similar sets for others) and def2-TZVPP basis set on H and O atoms. On these geometry profiles, we generate reference CCSD(T) data via a series of calculations in triple- and quadruple-ζ correlation-consistent basis sets followed by the CBS extrapolation with the two-point formula[123] $E_n = E_{CBS} + B\,(n+1/2)^{-4}$. For H, O, and F atoms, we use standard aug-cc-pVXZ ($X = 3,\ 4$) basis sets, while for Cl and heavier halogens, we adapt them to the RECP model. Furthermore, from Br on, we incorporate additional outer-core basis functions to properly describe electronic correlations involving the sub-valence d-shells, creating aug-cc-pwCVXZ sets.

To estimate spin-orbit effects on the X^-–H_2O binding energy, $\Delta SO(E_b)$, we rely on spin-orbit, or two-component (2c), and scalar, or one-component (1c), versions of the same method. This way, we find that for the At^-–H_2O system, the spin-orbit contribution evaluated at the MP2/CBS and LRC-ωPBEh levels are practically identical and below 0.5 kcal/mol, as illustrated in Figure 1. Thus, we can justify the assessment of DFT results against 1c-CCSD(T) for lighter halogens: in the case of I^-, the effect is already an order of magnitude smaller. The situation with Ts is markedly

different. The spin-orbit splitting of the valence 7p- and sub-valence 6d-shells is of the order of 10 eV, which renders the scalar relativistic approximation meaningless even as a starting point. Thus, we resort to using the 2c-CCSD(T) data derived from the CBS extrapolation of aug-cc-pwCVDZ and aug-cc-pwCVTZ results, with aug-cc-pwCVQZ becoming prohibitively expensive.

It is worth noting that our *ab initio* results at the CCSD(T)/CBS level of theory with the inclusion of spin-orbit effects elucidate a more pronounced difference between I^- and At^- than found in ref 17. While our At^-–H_2O *vs.* I^-–H_2O binding energy difference of *ca.* 1.25 kcal/mol supports *qualitative* similarity between these two halides, we demonstrate in Section C that it translates to a noticeable *quantitative* difference in force field parameters and radial distribution functions.

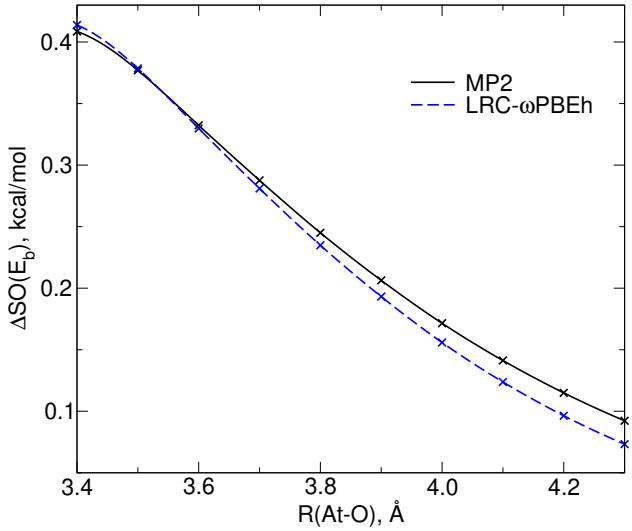

Figure 1. Spin-orbit contribution to the At^-–H_2O binding energy evaluated at the MP2/CBS and LRC-ωPBEh level.

In Figure 2, we illustrate the behavior of several common XCFs (PBE0,[56] LRC-ωPBEh,[122] B3LYP,[124] B3LYP D3,[125] and ωB97X-D[96]) relative to the CCSD(T) binding energy curves for X^-–H_2O systems. As discussed above, LRC-ωPBEh gives accurate molecular geometries; however, PBE0 is superior to other XCFs in reproducing CCSD(T) binding energy curves. For the At^-–H_2O case, the quality of the results is remarkable; we attribute it to the success of our At basis set optimization procedure. In Figure S1 of Supplemental Information, we also demonstrate

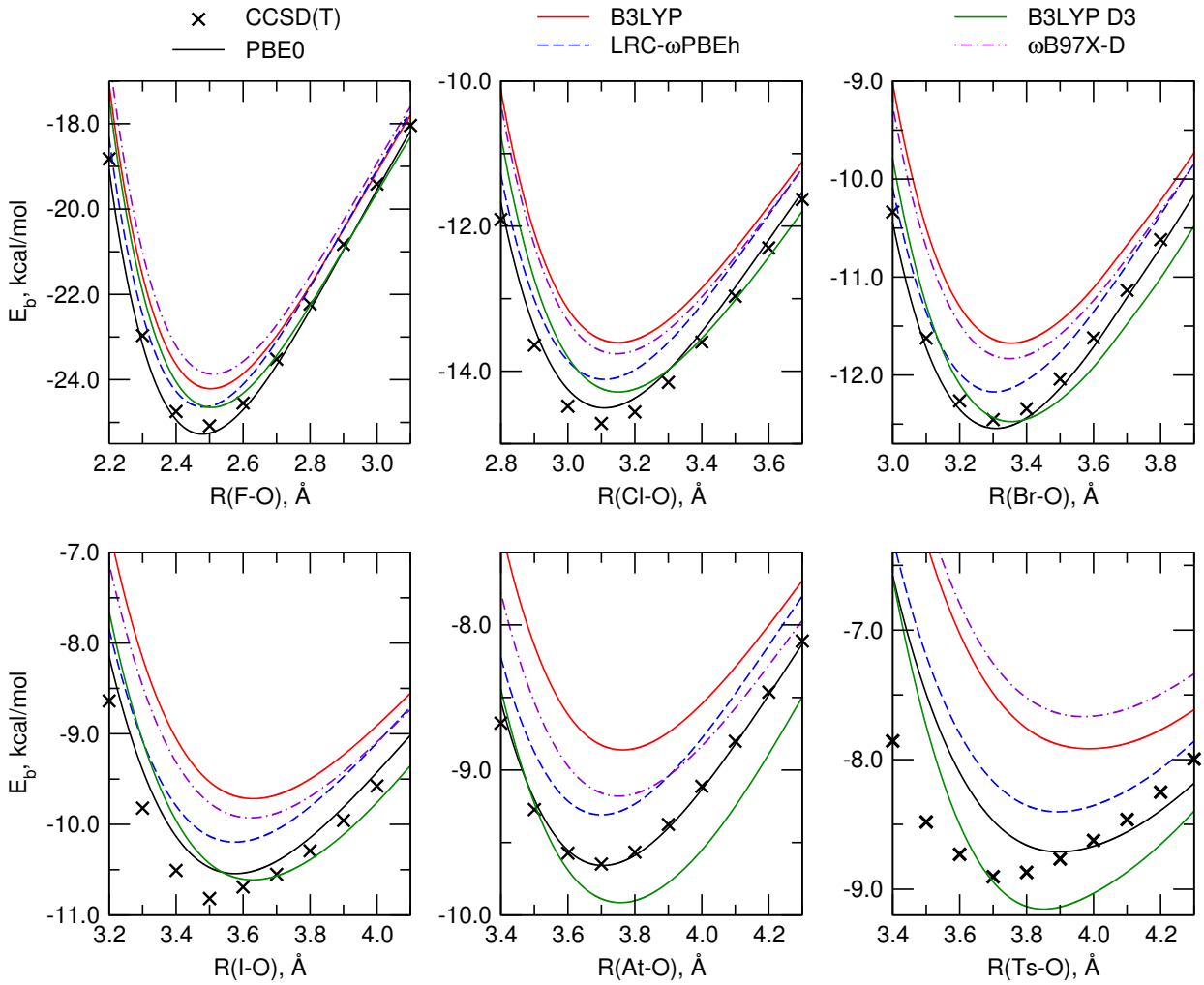

Figure 2. Halide-oxygen potential energies for several functionals compared to CCSD(T). For F^- through I^-, coupled-cluster are at the 1c-CCSD(T) level; for At^-, the spin-orbit effects are added to 1c-CCSD(T) *a posteriori*; Ts^- calculations are at the 2c-CCSD(T) level. All DFT calculations are at the 2c-RDFT level with this work's PSO-Lc basis for At and similar ones for other halogens.

that these 2c-RDFT/PBE0 binding energy profiles with our PSO-Lc basis are more accurate than with several other DFT-specific basis sets[69,72,73] for all-electron relativistic Hamiltonians — Douglas–Kroll–Hess (DKH2)[126–128] and exact two-component (X2C).[129–131] Additionally, the use of bespoke DFT-adapted basis sets allows us to revisit earlier computational recipes for modeling X^-–H_2O interactions (*cf.* ref 132).

We see the most pronounced divergence across the binding-energy profiles for the superheavy element Ts. Given the relativistic effects' magnitude, it is natural to expect irregularities in the behavior of common nonrelativistic XCFs. We will examine the Ts case in more detail in our future work.

B Basis set optimization

Figure 3 shows the decay of the mean absolute error (MAE) during the PSO optimization of the three uncontracted basis sets developed in this work. The breakdown of these errors into contributions from At^-, At, At^+, At^{2+}, HAt, and At_2 systems for the initial, optimized uncontracted, and optimized contracted basis sets is given in Table 1. For HAt and At_2, only the average error over five internuclear distances is shown in Table 1 for brevity.

Table 1: Total energy errors w.r.t to complete basis set limit $(E_{basis} - E_{CBS})$ in kcal/mol for each basis set optimized in this work. Shown are the errors for the initial uncontracted basis set (modified UGBS), optimized uncontracted (PSO-L, PSO-M, PSO-S) basis sets, and optimized contracted basis sets (PSO-Lc, PSO-Mc, PSO-Sc). For At_2 and HAt "*" denotes the average error over five internuclear distances. m.a.e. is the mean absolute error over all fourteen reference systems.

System	Large			Medium			Small		
	L	PSO-L	PSO-Lc	M	PSO-M	PSO-Mc	S	PSO-S	PSO-Sc
At	13.0	0.61	1.03	13.4	0.66	1.11	39.7	2.68	4.19
At^+	13.0	0.52	1.02	13.2	0.44	0.89	39.9	2.85	4.28
At^{2+}	13.1	0.52	1.08	13.1	0.39	0.83	40.3	3.11	4.51
At^-	13.0	0.48	0.98	13.0	0.29	0.74	39.6	2.54	4.05
HAt*	13.3	0.72	1.21	14.7	2.03	2.51	39.5	2.90	4.34
At_2*	26.3	1.10	2.09	30.0	4.63	5.56	79.6	5.87	8.86
m.a.e.	17.9	0.80	1.47	20.8	2.51	3.14	53.9	3.93	5.93

We will begin by analyzing the initial basis sets labeled L, M, and S. As Figure 3 and Table 1 demonstrate, the initial modified UGBS basis sets have significant MAEs exceeding chemical accuracy of 1 kcal/mol. Errors of this magnitude for heavy elements with standard DFT functionals have already been reported in refs 62,133. Expectedly, larger basis sets give smaller errors, but even the largest one still has a significant average error of 17.9 kcal/mol. The largest errors are for At_2 molecule, while the errors for At and its charged species are similar to those for the HAt molecule and approximately half of the errors for At_2 molecule. The basis set errors for At add up for each At atom in the system, while basis sets for H atom used in HAt calculations are close to the complete basis set limit and do not produce further errors. This result reassures that basis set optimization that also uses HAt molecules in the data set employs an already optimal basis set for the H atom.

The difference between L and M basis sets is in the absence of f-functions in the latter. As seen in Table 1, this difference results in larger errors for all the systems but especially for HAt and At_2 molecules where using more diffuse functions improves the description of At-At and H-At interactions. Overall, however, the differences of $\sim$3 kcal/mol seem small compared to the absolute values of errors for such systems reaching 30 kcal/mol.

Naturally, the smallest basis set S generates the largest errors exceeding 50 kcal/mol. However, its sole purpose is to further optimize the smallest basis set in the hierarchy.

Figure 3 showcases PSO as a highly efficient method for optimizing basis sets. The errors are reduced by half with only about 20 iterations in the case of the PSO-S basis set and much faster for PSO-L and PSO-M basis sets. In those cases, the MAEs drop by half within the first few, $<$5, iterations. Further illustrating the efficiency of the PSO, the MAEs are decreasing almost every single iteration until they are reduced significantly enough such that the search for a better set of primitive exponents slows down. To continue efficient optimization beyond this level, PSO would probably require a different set of coefficients c_p, c_g, and w, which was not done here because it was not deemed necessary. As most clearly seen from Table 1, in the case of PSO-L basis sets, the MAE eventually reached chemical accuracy of 1 kcal/mol, an improvement of over 22 times

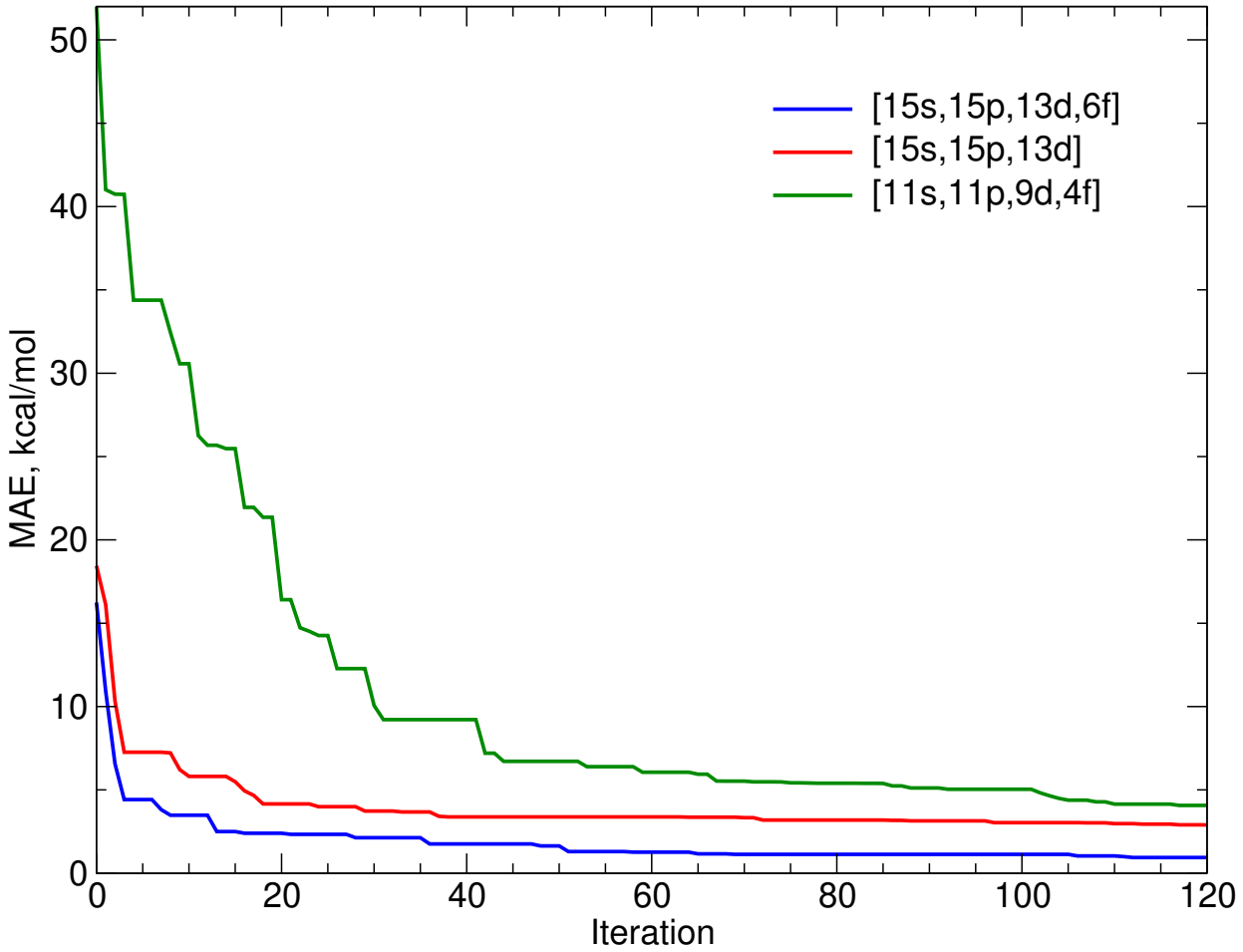

Figure 3. Average error decay during the PSO optimization for three basis sets generated in this work.

compared to the initial basis set. The largest errors are, again, expectedly for At_2 molecule, but even in this case, the errors are close to the chemical accuracy. This optimized uncontracted PSO-L basis set was further contracted as described above, resulting in a final basis set PSO-Lc. This final basis is only slightly less accurate than PSO-L and still offers an overall significant improvement compared to the initial basis set L.

Basis sets PSO-M and PSO-S were also optimized significantly. In the case of the PSO-M basis set, the MAE was reduced by a factor of 8, while for the PSO-S basis set, by a factor of $\sim$14. Notably, the resulting errors are now close to chemical accuracy even for the smallest basis set, PSO-S. Both basis sets were contracted to produce PSO-Mc and PSO-Sc, respectively. As seen from Table 1, the errors for these contracted basis sets are 25% to 50% larger than those of the corresponding uncontracted basis sets. Still, the improvement over the initial basis sets is significant. Remarkably, even the least accurate PSO-Sc basis set outperforms the large initial basis set L by a factor of 3. Furthermore, an At_2 calculation with the PSO-Sc basis set runs on average 4.5 times faster. All optimized basis sets are available in ref 134 and Supporting Information.

Given that this work is the first PSO application to basis-set optimization, we find it helpful to comment on the computational efficiency of our procedure. Our PSO calculations used a swarm of 15 particles. At each iteration, 14 2c-RDFT calculations were performed for each particle. These calculations are independent and were performed in parallel, so the total computational cost for each particle is only the cost of the most expensive calculation, which is an At_2 one. Such a calculation for the largest basis set takes about 6 minutes on a 2 x 20C Intel Xeon Gold 6230 (192 GB DDR4 RAM) processor. Therefore, in the fastest scenario, when all 210 calculations required for each iteration can be performed in parallel, the basis set optimization with 120 PSO iterations takes 12 hours. Thus, our basis-set optimization method has a highly desirable error reduction/time ratio of $\sim$2 hour^{-1}, as measured for the PSO-L basis set.

C Force field

C.1 Final force field parameters

The initial (OPLS/AA) and optimized LJ parameters for the ions are compiled in Table 2. Note that, as discussed above, the initial At^- parameters were set to the I^- OPLS/AA values because the At^- LJ parameters have not been optimized before.

Table 2: Initial (OPLS/AA) and optimized Lennard-Jones parameters. σ and ε parameters are in Å and kJ/mol, respectively.

Ion	Initial		Optimized	
	σ	ε	σ	ε
F^-	2.73	3.012	2.61	3.006
Cl^-	4.42	0.493	4.31	0.183
Br^-	4.62	0.377	4.51	0.388
I^-	5.40	0.293	5.32	0.215
At^-	5.40	0.293	5.23	0.553

The σ parameter of the LJ potential shows the same trend of a modest increase from F^- to I^- in both OPLS/AA and our optimized FFs, while the ε parameter decreases in these models. However, the continuation from I^- to At^- reverses this trend. Finding the reason for this reversal requires further investigation, partially because the ForceBalance optimization can be sensitive to the initial parameters. Additionally, the inclusion of Ts^- will be necessary for this trend's thorough evaluation. In subsequent subsections, we demonstrate that despite this apparent trend reversal in the LJ parameters, the calculated properties of At^- follow the halides' trends.

C.2 Radial distribution functions

Figure 4 presents the RDFs calculated with our FFs. The O–X^- and H–X^- RDFs all present a narrow first peak corresponding to a well-defined first hydration shell. As expected, peak positions increase along the halide series. Furthermore, as the anions get larger, both O–X^- and H–X^- first peaks broaden. This observation is consistent with a reduction in dipole alignment as ions get bigger.

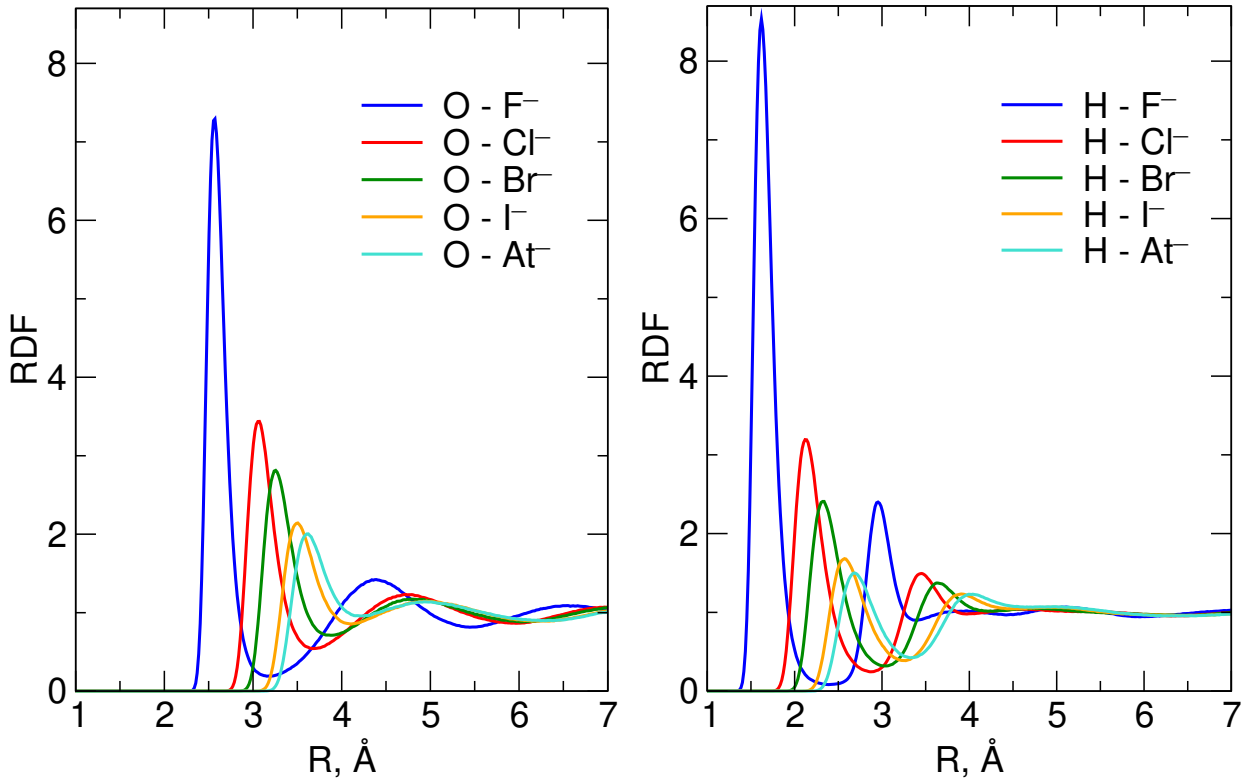

Figure 4. Radial distributions function between indicated halides and water oxygen (left) and water hydrogen (right) calculated using optimized force field parameters.

Table 3: The first peak of the radial distribution functions, Å. [a] Neutron diffraction data from ref 135. [b] Extended X-ray absorption fine structure (EXAFS) data from ref 136.

Ion	O–X$^-$			H–X$^-$	
	this work	ref 17	experiment	this work	ref 17
F^-	2.56	2.78	2.54[a]	1.62	1.90
Cl^-	3.07	3.14	3.14[a]/3.2[b]	2.12	2.27
Br^-	3.26	3.36	3.32[a]	2.32	2.45
I^-	3.50	3.71	3.63[a]/3.6[b]	2.58	2.80
At^-	3.62	3.75	-	2.68	2.82

In Table 3, we summarize O–X^- and H–X^- RDF first peaks from our and ref 17 simulations and available experimental data from the extended X-ray absorption fine structure (EXAFS)[136] method and neutron diffraction.[135] Overall, our O–X^- first peaks convincingly agree with the experimental results. For F^-, our model recovers the experiment noticeably better than the earlier simulation in ref 17. Currently, there are no experimental data for At^-. However, our model predicts a meaningful dissimilarity between I^- and At^-: astatide's first peak is shifted by 0.12 Å relative to iodide and predicts a more diffuse first hydration sphere of the former, which is in agreement with the halides' trend. We stress that this result contrasts the findings in ref 17, where O–I^- and O–At^- first peaks are virtually indistinguishable.

Similar behavior is found for the H–X^- RDFs, also shown in Figure 4 and summarized in Table 3. The first peak positions increase along the halide series. Here, we also report a noticeable difference of 0.1 Å between the positions of the first peaks in H–At^- and H–I^- RDFs with the former located at 2.58 Å and the latter one at 2.68 Å. This finding is in contrast with the corresponding RDFs obtained with a polarizable FF by Réal *et al.*[17] Similarly to the ion-oxygen RDFs, they report virtually no difference between H–At^- and H–I^- RDFs.

The width of the first peak and the relative depth of the first minimum with respect to that of the first maximum characterize how structured the first hydration shell is and the rate of exchange between the first and the second hydration shells. According to our simulations, I^- and At^- expectedly have a more flexible first hydration shell than the other halides.

We also determined the halide-water coordination (hydration) numbers by integrating O-X^- RDFs. The calculated hydration numbers are 5.7 for F^-, 6.0 for Cl^-, 6.4 for Br^-, 6.6 for I^-, and 7.1 for At^-, see also Table 4. These coordination numbers are in good agreement with the values measured by neutron diffraction, which are 5.2–6.9 for F^-, 6.4–7.1 for Cl^-, 6.4–6.7 for Br^-, and 6.7–7.1 for I^-. Since the experimental value of the hydration number for At^- is unknown, we compare our result to Réal *et al.*[17] Surprisingly, they report a noticeably larger coordination number of 9.1. For I^-, they also obtained a larger than experimental coordination number 9.2. We attribute this to issues with their FFs because a similar polarizable force field for lighter halides

from Trumm *et al.*[137] also consistently overestimates ion hydration numbers. Given the good agreement between calculated and experimental coordination numbers for F^-–I^-, we believe that our coordination number for At^- is more accurate. However, one should also be mindful that locating the first solvation shell for I^- and At^- is problematic because of the diffuseness of the first peak of the corresponding RDF.

Table 4: Hydration numbers.

Ion	this work	ref 17	experiment
F^-	5.7	6.3	5.2–6.9
Cl^-	6.0	8.4	6.4–7.1
Br^-	6.4	9.0	6.4–6.7
I^-	6.6	9.2	6.7–7.1
At^-	7.1	9.1	-

C.3 Diffusion coefficients

The ability of the force field to reproduce the dynamic transport properties of ions in aqueous solutions is an essential test of its accuracy. To test this ability, we calculated the diffusion coefficients for halide ions in aqueous solutions. In Table 5, diffusion coefficients obtained in this work are compared to the diffusion coefficients reported by Réal *et al.*[17] and to one most recently reported diffusion coefficients for F^-–Br^- halides (in TIP4P-FB water[97,138]) as well as to experimental values. TIP4P-FB is a reparametrization of a non-polarizable TIP4P water model done with the ForceBalance method. This model accurately describes many physical properties of water.

Table 5: Diffusion coefficients in 10^{-5} cm^2/s.

Ion	this work	Réal *et al.*[17]	Li *et al.*[138]	Experiment[139]
F^-	1.26±0.115	1.77	0.93-1.14	1.48
Cl^-	1.88±0.156	1.97	1.63-1.87	2.03
Br^-	1.90±0.154	2.00	1.73-1.86	2.08
I^-	1.94±0.169	1.99	-	2.05
At^-	1.91±0.140	1.88	-	-

Our simulated diffusion coefficients are in a better agreement with the experimental values than those of Li *et al.*,[138] but slightly worse, in general, than the diffusion coefficients reported of Réal *et al.*[17] Overall, however, our diffusion coefficients agree with the experiment within 10%.[139] The only exception is F^- whose diffusion coefficient is underestimated by 15%. Interestingly, there is a significant scatter in the simulated diffusion coefficients for F^-. For example, Koneshan *et al.* reports the diffusion coefficient that is too small ($1.04 \cdot 10^{-5}$ cm^2/s compared to experiment ($1.48 \cdot 10^{-5}$ cm^2/s), while the polarizable TCPE FF of Réal *et al.*[17] predicts that F^- ions diffuse nearly 20% faster ($1.77 \cdot 10^{-5}$ cm^2/s) than in the experiment. In light of this scatter, it is interesting to note that Li and Wang used a similar charge scaling approach and obtained the diffusion coefficient for F^- of $1.27 \pm 0.06 \cdot 10^{-5}$ cm^2/s[140] which is in a good agreement with ours. Their and our diffusion coefficients provide the best agreement with the experiment.

Our FF predicts the diffusion coefficient for I^- that is larger than that of Br^-, while experiment and simulated diffusion coefficients reported by others predict that the maximum of the diffusion coefficient in the halide series is reached by Br^-.[140–142] The focus of this work is on At, for which our force field predicts the diffusion coefficient to be slower than that of I^-. Not only is this trend in agreement with the TPCE model of Réal *et al.*,[17] the absolute value of the At diffusion coefficient is in excellent agreement with their value.

C.4 Hydration free energies

Single ion hydration free energies are not measurable in a condensed phase experiment. It is also not straightforward to calculate them from MD simulations. Proper calculation of ion hydration free energies requires several corrections.[116,140] Because even the sign of some of the correction terms is being debated, we decided to report the differences in anion hydration free energies instead.

Table 6 compares our results with the hydration free energies reported by Réal *et al.*[17] as well as those obtained from OPLS/AA force field that provided initial LJ parameters for the FF optimization done in this work. The differences in hydration free energies predicted by our FF

Table 6: Differencies in hydration free energies of halides in water (in kcal/mol).

Ion	this work	OPLS/AA	Réal *et al.*[17]	Experiment[116]
$F^- \rightarrow Cl^-$	28.1	28.3	27.1	29.3-30.6
$Cl^- \rightarrow Br^-$	5.1	1.98	6.6	5.3-8.9
$Br^- \rightarrow I^-$	7.3	6.9	9.3	8.3-9.0
$I^- \rightarrow At^-$	0.8	4.2	1.4	-

consistently underestimate the experimental values, but the overall agreement is reasonable. There is no experimental value for At^-. In agreement with a polarizable TCPE model of Réal *et al.*,[17] the difference in hydration free energies between I^- and At^- is minimal, less than 1.5 kcal/mol. Our FF predicts the correct trend that the hydration free energy of At^- is still less negative than that of I^-, yet our difference is almost twice smaller. Even smaller hydration free energy difference of 0.5 kcal/mol calculated using MP2 for finite-size I^-–$(H_2O)_n$ and At^-–$(H_2O)_n$ clusters was reported in ref 18. Given that our FF underestimates the hydration free energy differences compared to the experiment, the value by Réal *et al.*[17] is likely more accurate. However, as of now, no experiments have been done to validate this assessment.

We also include the difference in hydration free energies evaluated using the OPLS/AA FF. It predicts the $Cl^- \rightarrow Br^-$ hydration free energy difference noticeably smaller than the experimental values in ref 116. However, in another experimental hydration free energies' estimate by Tissandier *et al.*,[143] the hydration free energy difference $Cl^- \rightarrow Br^-$ is 3.15 kcal/mol. This range of empirical data illustrates the challenges in the experimental determination of hydration free energies. In light of this, it is interesting that our optimized FF parameters predict $Cl^- \rightarrow Br^-$ hydration free energies that are in better agreement with other experiments, although the optimization was initialized with the OPLS/AA parameters. Therefore, the ForceBalance approach can indeed be helpful in reparametrizing the existing FFs, as was done for water.[97]

V Conclusions

How similar are the structural, dynamical, and energetic properties of At^- in water to those of I^-? Motivated by this specific question and the increasing interest in astatine chemistry, we have revisited the computational infrastructure for accurate MD simulations of At^-–water systems, which are yet to be studied experimentally. In particular, we parameterized a non-polarizable force field for At^-–water systems based on the scaled ion charge approach. With no empirical data available, we optimized the FF based on the calculated reference data for At^-–$(H_2O)_n$ clusters. A feasible approach toward such a data set relies on accurate 2c-RDFT calculations. To attain the desired accuracy, we assessed the ability of several XCF models to reproduce CBS-extrapolated CCSD(T) results with an added spin-orbit contribution to the binding energy for the At^-–H_2O system.

To disentangle the errors of XCF and basis set that afflict RDFT applications to heavy elements,[62] we developed accurate basis sets for At by combining the polarization consistent approach with the inherent inclusion of scalar and spin-dependent relativistic effects and employing a remarkably efficient particle swarm optimization algorithm. For lighter halogens and Ts, we developed similar yet not as rigorously optimized basis sets. With these reliable RDFT-tailored basis sets at hand, we single out PBE0 as the XCF universally reproducing X^-–H_2O CCSD(T) data with chemical accuracy across all halides.

Based on the carefully assessed RDFT/PBE0 approach with the newly optimized basis sets, we develop non-polarizable FFs for halides in water that allow us to put astatine's behavior in perspective. Using MD simulations, we determined that the first peak of O–At^- and H–At^- RDFs extends to noticeably larger distances compared to I^- counterparts, although the difference is smaller compared to the $Br^- \rightarrow I^-$ ones. Therefore, structurally, At^-–water systems are still different than I^-–water systems. At^- follows the halide trend. This finding contrasts previously published results for the same systems, where O–X^- and H–X^- RDFs are identical for $X^- = I^-$ or At^-.[17] The diffusion coefficient for At^- found in this work is slightly (1.5%) smaller than that

of I^-, which agrees with the previously reported value.[17] The hydration free energy difference between At^- and I^- is found to be smaller than 1 kcal/mol, which is in good agreement with previously reported values, albeit somewhat too small.[17]

We reiterate the most significant results here. Firstly, At^-'s behavior in water follows the halides' trend but is not identical to I^-. This observation and the experimental difficulty inherent to radioactive materials such as At call for more development of reliable computational methods for accurately predicting the properties of At-containing molecules. Such systems will continue to play an increasingly important role in emerging radiopharmaceuticals.

Secondly, the properties of At^- in water obtained here show that At continues to follow the trend among halides. The calculated properties are either in good agreement with previously reported values obtained with a polarizable FF or, arguably, more realistic, as is the case for RDFs. Our results favor a simple non-polarizable approach with scaled ion charges and show that explicit treatment of ion polarizability might not be necessary.

Thirdly, there is a necessity for a considerable improvement to basis sets and exchange-correlation functionals for RDFT calculations of heavy elements. RDFT methods offer a compromise between accuracy and computational cost. As such, RDFT is a method of choice to model complex heavy-element systems with numerous national security, defense, energy, and medicine applications. However, insufficient infrastructure development impedes its successful use in routine calculations. In this paper, we have established a promising basis-set development protocol that showcases the efficiency of global optimization techniques such as PSO. Extending our approach beyond main-group heavy elements will substantially improve the RDFT reliability.

VI Supporting Information

Basis sets developed in this work and relevant RECPs. Evaluation of basis sets from this work against SARC-DKH2 and x2c-TZVPPall-2c/x2c-QZVPPall-2c ones.

Acknowledgement

A.A.K. is grateful to Prof. Sandeep Patel for helpful discussions. The research reported in this article was supported by the National Institutes of Health under Award Number P20GM104316 (A.A.K.). A.A.R. thanks the Biomedical Research Center at Oakland University for support through the Research Excellence Program. K.J.R.E. acknowledges support from the Internationalization Program of the Faculty of Sciences of Universidad Nacional de Colombia. High-performance computing facilities were provided by the University of Delaware and via collaboration between the Oakland University Research Office and University Technology Services.

VII Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

(1) Nations join together to announce the World Astatine Community. https://www.isotopes.gov/WAC, (accessed June 26, 2023).

(2) World Astatine Community expands access to cancer-fighting At-211. https://www.ans.org/news/article-5099/world-astatine-community-expands-access-to-cancerfighting-at211/, (accessed June 26, 2023).

(3) The 12th International Symposium on Targeted Alpha Therapy (TAT 12). https://tat-12.com, (accessed June 26, 2023).

(4) Li, F.; Yang, Y.; Liao, J.; Liu, N. Recent progress of astatine-211 in endoradiotherapy: Great

advances from fundamental properties to targeted radiopharmaceuticals. *Chin. Chem. Lett.* **2022**, *33*, 3325–3338.

(5) Guérard, F.; Maingueneau, C.; Liu, L.; Eychenne, R.; Gestin, J.-F.; Montavon, G.; Galland, N. Advances in the Chemistry of Astatine and Implications for the Development of Radiopharmaceuticals. *Acc. Chem. Res.* **2021**, *54*, 3264–3275.

(6) Wilbur, D. S. Enigmatic astatine. *Nature Chem.* **2013**, *5*, 246.

(7) Lemonick, S. Astatine is a chemistry puzzle that shows anticancer promise. https://cen.acs.org/physical-chemistry/Astatine-chemistry-puzzle-shows-anticancer/98/i31, (accessed June 26, 2023).

(8) Casetti, V. T.; MacLean, J.; Ayoub, A. D.; Fredericks, R. J.; Adamski, J. A.; Rusakov, A. A. Investigating the Heaviest Halogen: Lessons Learned from Modeling the Electronic Structure of Astatine's Small Molecules. *J. Phys. Chem. A* **2023**, *127*, 46–56.

(9) Paesani, F.; Bajaj, P.; Riera, M. Chemical accuracy in modeling halide ion hydration from many-body representations. *Adv. Phys.: X* **2019**, *4*, 1631212.

(10) Zeron, I. M.; Abascal, J. L. F.; Vega, C. A force field of Li^+, Na^+, K^+, Mg^{2+}, Ca^{2+}, Cl^-, and SO_4^{2-} in aqueous solution based on the TIP4P/2005 water model and scaled charges for the ions. *J. Chem. Phys.* **2019**, *151*, 134504.

(11) Li, P.; Merz, K. M. Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* **2017**, *117*, 1564–1686.

(12) Blazquez, S.; Conde, M. M.; Abascal, J. L. F.; Vega, C. The Madrid-2019 force field for electrolytes in water using TIP4P/2005 and scaled charges: Extension to the ions F^-, Br^-, I^-, Rb^+, and Cs^+. *J. Chem. Phys.* **2022**, *156*, 044505.

(13) Heindel, J. P.; Herman, K. M.; Xantheas, S. S. Many-Body Effects in Aqueous Systems:

Synergies Between Interaction Analysis Techniques and Force Field Development. *Annu. Rev. Phys. Chem.* **2023**, *74*, 337–360.

(14) Bajaj, P.; Götz, A. W.; Paesani, F. Toward Chemical Accuracy in the Description of Ion-Water Interactions through Many-Body Representations. I. Halide-Water Dimer Potential Energy Surfaces. *J. Chem. Theory Comput.* **2016**, *12*, 2698–2705.

(15) Caruso, A.; Zhu, X.; Fulton, J. L.; Paesani, F. Accurate Modeling of Bromide and Iodide Hydration with Data-Driven Many-Body Potentials. *J. Phys. Chem. B* **2022**, *126*, 8266–8278.

(16) Caruso, A.; Paesani, F. Data-driven many-body models enable a quantitative description of chloride hydration from clusters to bulk. *J. Chem. Phys.* **2021**, *155*, 064502.

(17) Réal, F.; Gomes, A. S. P.; Martínez, Y. O. G.; Ayed, T.; Galland, N.; Masella, M.; Vallet, V. Structural, dynamical, and transport properties of the hydrated halides: How do At -bulk properties compare with those of the other halides, from F^- to I^-? *J. Chem. Phys.* **2016**, *144*, 124513.

(18) Chamorro, Y.; Flórez, E.; Maldonado, A.; Aucar, G.; Restrepo, A. Microsolvation of heavy halides. *Int. J. Quantum Chem.* **2021**, *121*, e26571.

(19) Gómez, S.; Flórez, E.; Acelas, N.; Hadad, C.; Restrepo, A. Encapsulation of Astatide by a water cage. *Phys. Chem. Chem. Phys.* **2023**, *25*, 12284–12289.

(20) Gu, H.; Liu, L. Molecular modeling and rational design of noncovalent halogenoxygenhydrogen motif at the complex interface of EGFR kinase domain with RALT peptide. *Chem. Phys.* **2021**, *550*, 111309.

(21) Réal, F.; Vallet, V.; Flament, J.-P.; Masella, M. Revisiting a many-body model for water based on a single polarizable site: From gas phase clusters to liquid and air/liquid water systems. *J. Chem. Phys.* **2013**, *139*, 114502.

(22) Larsen, R. H.; Slade, S.; Zalutsky, M. R. Blocking [211At]Astatide Accumulation in Normal Tissues: Preliminary Evaluation of Seven Potential Compounds. *Nucl. Med. Biol.* **1998**, *25*, 351–357.

(23) Carlin, S.; Mairs, R. J.; Welsh, P.; Zalutsky, M. R. Sodium-iodide symporter (NIS)-mediated accumulation of [211At]astatide in NIS-transfected human cancer cells. *Nucl. Med. Biol.* **2002**, *29*, 729–739.

(24) Carlin, S.; Akabani, G.; Zalutsky, M. R. In vitro cytotoxicity of ^{211}At-astatide and ^{131}I-iodide to glioma tumor cells expressing the sodium/iodide symporter. *J. Nucl. Med.* **2003**, *44*, 1827–38.

(25) Watabe, T.; Kaneda-Nakashima, K.; Liu, Y.; Shirakami, Y.; Ooe, K.; Toyoshima, A.; Shimosegawa, E.; Fukuda, M.; Shinohara, A.; Hatazawa, J. Enhancement of 211At Uptake via the Sodium Iodide Symporter by the Addition of Ascorbic Acid in Targeted -Therapy of Thyroid Cancer. *J. Nucl. Med.* **2019**, *60*, 1301–1307.

(26) Leggett, R.; Samuels, C. Basis for the ICRP's updated biokinetic model for systemic astatine. *J. Radiol. Prot.* **2022**, *42*, 021502.

(27) Mitchell, H. H.; Hamilton, T. S.; Steggerda, F. R.; Bean, H. W. The chemical composition of the adult human body and its bearing on the biochemistry of growth. *J. Biol. Chem.* **1945**, *158*, 625–637.

(28) Kim, J. S.; Wu, Z.; Morrow, A. R.; Yethiraj, A.; Yethiraj, A. Self-Diffusion and Viscosity in Electrolyte Solutions. *J. Phys. Chem. B* **2012**, *116*, 12007–12013.

(29) Leontyev, I. V.; Stuchebrukhov, A. A. Electronic continuum model for molecular dynamics simulations. *J. Chem. Phys.* **2009**, *130*, 085102.

(30) Leontyev, I. V.; Stuchebrukhov, A. A. Polarizable molecular interactions in condensed phase and their equivalent nonpolarizable models. *J. Chem. Phys.* **2014**, *141*, 014103.

(31) Leontyev, I. V.; Stuchebrukhov, A. A. Electronic Polarizability and the Effective Pair Potentials of Water. *J. Chem. Theory Comput.* **2010**, *6*, 3153–3161.

(32) Leontyev, I. V.; Stuchebrukhov, A. A. Electronic Continuum Model for Molecular Dynamics Simulations of Biological Molecules. *J. Chem. Theory Comput.* **2010**, *6*, 1498–1508.

(33) Leontyev, I. V.; Stuchebrukhov, A. A. Polarizable Mean-Field Model of Water for Biological Simulations with AMBER and CHARMM Force Fields. *J. Chem. Theory Comput.* **2012**, *8*, 3207–3216.

(34) Leontyev, I.; Stuchebrukhov, A. Accounting for electronic polarization in non-polarizable force fields. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2613–2626.

(35) Pluhařová, E.; Mason, P. E.; Jungwirth, P. Ion Pairing in Aqueous Lithium Salt Solutions with Monovalent and Divalent Counter-Anions. *J. Phys. Chem. A* **2013**, *117*, 11766–11773.

(36) Mason, P. E.; Wernersson, E.; Jungwirth, P. Accurate Description of Aqueous Carbonate Ions: An Effective Polarization Model Verified by Neutron Scattering. *J. Phys. Chem. B* **2012**, *116*, 8145–8153.

(37) Vazdar, M.; Pluhařová, E.; Mason, P. E.; Vácha, R.; Jungwirth, P. Ions at Hydrophobic Aqueous Interfaces: Molecular Dynamics with Effective Polarization. *J. Phys. Chem. Lett.* **2012**, *3*, 2087–2091.

(38) Pegado, L.; Marsalek, O.; Jungwirth, P.; Wernersson, E. Solvation and ion-pairing properties of the aqueous sulfate anion: explicit versus effective electronic polarization. *Phys. Chem. Chem. Phys.* **2012**, *14*, 10248–10257.

(39) Kann, Z. R.; Skinner, J. L. A scaled-ionic-charge simulation model that reproduces enhanced and suppressed water diffusion in aqueous salt solutions. *J. Chem. Phys.* **2014**, *141*, 104507.

(40) Rajagopal, A. K.; Callaway, J. Inhomogeneous Electron Gas. *Phys. Rev. B* **1973**, *7*, 1912–1919.

(41) MacDonald, A. H.; Vosko, S. H. A relativistic density functional formalism. *J. Phys. C: Solid State Phys.* **1979**, *12*, 2977.

(42) Ramana, M. V.; Rajagopal, A. K. Theory of spin polarised inhomogeneous relativistic electron gas. *J. Phys. C: Solid State Phys.* **1981**, *14*, 4291.

(43) Wüllen, C. v. Relativistic Methods for Chemists. *Challenges and Advances in Computational Chemistry and Physics* **2010**, 191–214.

(44) Engel, E.; Dreizler, R. M. *Density Functional Theory, An Advanced Course*; Springer Berlin, Heidelberg, 2011.

(45) Choi, Y. J.; Lee, Y. S. Spin-orbit density functional theory calculations for heavy metal monohydrides. *J. Chem. Phys.* **2003**, *119*, 2014–2019.

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

Richard, R. M.; Sadayappan, P.; Schatz, G. C.; Shelton, W. A.; Silverstein, D. W.; Smith, D. M. A.; Soares, T. A.; Song, D.; Swart, M.; Taylor, H. L.; Thomas, G. S.; Tipparaju, V.; Truhlar, D. G.; Tsemekhman, K.; Voorhis, T. V.; Vázquez-Mayagoitia, ; Verma, P.; Villa, O.; Vishnu, A.; Vogiatzis, K. D.; Wang, D.; Weare, J. H.; Williamson, M. J.; Windus, T. L.; Woliński, K.; Wong, A. T.; Wu, Q.; Yang, C.; Yu, Q.; Zacharias, M.; Zhang, Z.; Zhao, Y.; Harrison, R. J. NWChem: Past, present, and future. *J. Chem. Phys.* **2020**, *152*, 184102.

(47) Wüllen, C. v. A Quasirelativistic Two-component Density Functional and Hartree-Fock Program. *Z. Phys. Chem.* **2010**, *224*, 413–426.

(48) Baldes, A.; Weigend, F. Efficient two-component self-consistent field procedures and gradients: implementation in TURBOMOLE and application to Au_{20}^-. *Mol. Phys.* **2013**, *111*, 2617–2624.

(49) Williams-Young, D. B.; Petrone, A.; Sun, S.; Stetina, T. F.; Lestrange, P.; Hoyer, C. E.; Nascimento, D. R.; Koulias, L.; Wildman, A.; Kasper, J.; Goings, J. J.; Ding, F.; DePrince III, A. E.; Valeev, E. F.; Li, X. The Chronus Quantum Software Package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2019**, e1436.

(50) Li, X.; Williams-Young, D.; Valeev, E.; III, E. D.; Hammes-Schiffer, S.; Sun, Q.; Petrone, A.; Wildman, A.; Hu, H.; Zhang, T.; Stetina, T.; Grofe, A.; Cooper, B.; Hoyer, C.; Liu, H.; Goings, J.; Koulias, L.; Lu, L.; Zhao, L.; Sun, S.; Liu, X. Chronus Quantum, Beta Version. 2020; http://www.chronusquantum.org.

(51) Liu, W.; Wang, F.; Li, L. The Beijing Density Functional (BDF) Program Package: Methodologies and Applications. *J. Theor. Comput. Chem.* **2003**, *2*, 257–272.

(52) Shiozaki, T. BAGEL: Brilliantly Advanced General Electronic-structure Library. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*.

(53) Repisky, M.; Komorovsky, S.; Kadek, M.; Konecny, L.; Ekström, U.; Malkin, E.;

Kaupp, M.; Ruud, K.; Malkina, O. L.; Malkin, V. G. ReSpect: Relativistic spectroscopy DFT program package. *J. Chem. Phys.* **2020**, *152*, 184101.

(54) Belpassi, L.; Santis, M. D.; Quiney, H. M.; Tarantelli, F.; Storchi, L. BERTHA: Implementation of a four-component Dirac–Kohn–Sham relativistic framework. *J. Chem. Phys.* **2020**, *152*, 164118.

(55) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

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

(57) Perdew, J. P.; Tao, J.; Staroverov, V. N.; Scuseria, G. E. Meta-generalized gradient approximation: Explanation of a realistic nonempirical density functional. *J. Chem. Phys.* **2004**, *120*, 6898–6911.

(58) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.

(59) Wang, F.; Li, L. Numerical examination of performance of some exchange-correlation functionals for molecules containing heavy elements. *J. Comput. Chem.* **2004**, *25*, 669–677.

(60) Trombach, L.; Ehlert, S.; Grimme, S.; Schwerdtfeger, P.; Mewes, J.-M. Exploring the chemical nature of super-heavy main-group elements by means of efficient plane-wave density-functional theory. *Phys. Chem. Chem. Phys.* **2019**, *21*, 18048–18058.

(61) Laub, J. A.; Vogiatzis, K. D. Theoretical Investigation of Actinide Ligation in Aqueous and Organic Phase for Nuclear Waste Treatment. *J. Phys. Chem. A* **2023**, *127*, 5523–5533.

(62) Aebersold, L. E.; Wilson, A. K. Considering Density Functional Approaches for Actinide Species: The An66 Molecule Set. *J. Phys. Chem. A* **2021**, *125*, 7029–7037.

(63) Jensen, F. Polarization consistent basis sets: Principles. *J. Chem. Phys.* **2001**, *115*, 9113–9125.

(64) Jensen, F. Unifying General and Segmented Contracted Basis Sets. Segmented Polarization Consistent Basis Sets. *J. Chem. Theory Comput.* **2014**, *10*, 1074–1085.

(65) Jensen, F. How Large is the Elephant in the Density Functional Theory Room? *J. Phys. Chem. A* **2017**, *121*, 6104–6107.

(66) Jensen, F. Computational Chemistry: The Exciting Opportunities and the Boring Details. *Isr. J. Chem* **2021**, *61*, 1–15.

(67) Weigend, F.; Furche, F.; Ahlrichs, R. Gaussian basis sets of quadruple zeta valence quality for atoms H-Kr. *J. Chem. Phys.* **2003**, *119*, 12753–12762.

(68) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(69) Pantazis, D. A.; Neese, F. All-electron scalar relativistic basis sets for the 6p elements. *Theoretical Chemistry Accounts* **2012**, *131*, 1292.

(70) Armbruster, M. K.; Klopper, W.; Weigend, F. Basis-set extensions for two-component spin–orbit treatments of heavy elements. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4862–4865.

(71) Weigend, F.; Baldes, A. Segmented contracted basis sets for one- and two-component Dirac–Fock effective core potentials. *J. Chem. Phys.* **2010**, *133*, 174102.

(72) Pollak, P.; Weigend, F. Segmented Contracted Error-Consistent Basis Sets of Double- and Triple- Valence Quality for One- and Two-Component Relativistic All-Electron Calculations. *Journal of Chemical Theory and Computation* **2017**, *13*, 3696–3705.

(73) Franzke, Y. J.; Spiske, L.; Pollak, P.; Weigend, F. Segmented Contracted Error-Consistent Basis Sets of Quadruple-ζ Valence Quality for One- and Two-Component Relativistic All-Electron Calculations. *J. Chem. Theory Comput.* **2020**, *16*, 5658–5674.

(74) Nagy, B.; Jensen, F. Reviews in Computational Chemistry. *Rev. Comput. Chem.* **2018**, 93–149.

(75) Kennedy, J.; Eberhart, R. Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks* **1995**, *4*, 1942–1948 vol.4.

(76) Call, S. T.; Zubarev, D. Y.; Boldyrev, A. I. Global minimum structure searches via particle swarm optimization. *J. Comput. Chem.* **2007**, *28*, 1177–1186.

(77) Tkachenko, N. V.; Tkachenko, A. A.; Kulyukin, V. A.; Boldyrev, A. I. DFT Study of Microsolvated $[NO_3 \cdot (H_2O)_n]^-$ (n = 1–12) Clusters and Molecular Dynamics Simulation of Nitrate Solution. *J. Phys. Chem. A* **2021**, *125*, 8899–8906.

(78) Niu, Z.; Tang, M.; Ge, N. Structure, stability, infrared spectra, and bonding of $OH^m(H_2O)_7$ (m = 0, ±1) clusters: ab initio study combining the particle swarm optimization algorithm. *Phys. Chem. Chem. Phys.* **2020**, *22*, 26487–26501.

(79) El Rassy, E.; Delaroque, A.; Sambou, P.; Chakravarty, H. K.; Matynia, A. On the Potential of the Particle Swarm Algorithm for the Optimization of Detailed Kinetic Mechanisms. Comparison with the Genetic Algorithm. *J. Phys. Chem. A* **2021**, *125*, 5180–5189.

(80) Pyper, N. C. Relativity and the periodic table. *Philos. Trans. R. Soc., A* **2020**, *378*, 20190305.

(81) Malli, G. L.; Siegert, M.; Macedo, L. G. M. d.; Loveland, W. Relativistic effects for the superheavy reaction Og + 2Ts2 $\rightarrow$ OgTs4 (Td or D4h): dramatic relativistic effects for atomization energy of superheavy Oganesson tetratennesside OgTs4 and prediction of the existence of tetrahedral OgTs4. *Theor. Chem. Acc.* **2021**, *140*, 75.

(82) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69–89.

(83) Hill, J. G. Gaussian basis sets for molecular applications. *Int. J. Quantum Chem.* **2013**, *113*, 21–34.

(84) Helgaker, T.; Jorgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; Wiley, 2014.

(85) Mosyagin, N.; Zaitsevskii, A.; Titov, A. Shape-consistent relativistic effective potentials of small atomic cores. *International Review of Atomic and Molecular Physics* **2010**, *1*, 63–72.

(86) Petersburg Nuclear Physics Institute, Quantum Chemistry Laboratory, Effective potentials and basis sets. http://www.qchem.pnpi.spb.ru/recp, (accessed November 4, 2022).

(87) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. Basis-set convergence of the energy in molecular Hartree-Fock calculations. *Chem. Phys. Lett.* **1999**, *302*, 437–446.

(88) Castro, E. V. R. d.; Jorge, F. E. Accurate universal Gaussian basis set for all atoms of the Periodic Table. *J. Chem. Phys.* **1998**, *108*, 5225–5229.

(89) Mercangöz, B. *Applying Particle Swarm Optimization: New Solutions and Cases for Optimized Portfolios*; International Series in Operations Research & Management Science; Springer International Publishing, 2021.

(90) Dyall, K. G. Relativistic double-zeta, triple-zeta, and quadruple-zeta basis sets for the 7p elements, with atomic and molecular applications. *Theor. Chem. Acc.* **2012**, *131*, 1172.

(91) 7p Block Quadruple-Zeta Basis Set Archive. http://dirac.chem.sdu.dk/basisarchives/dyall/7p_qz_archive.txt, (accessed May 11, 2023).

(92) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(93) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(94) Aprà, E.; Bylaska, E. J.; de Jong, W. A.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; van Dam, H. J. J.; Alexeev, Y.; Anchell, J.; Anisimov, V.; Aquino, F. W.; Atta-Fynn, R.; Autschbach, J.; Bauman, N. P.; Becca, J. C.; Bernholdt, D. E.; Bhaskaran-Nair, K.; Bogatko, S.; Borowski, P.; Boschen, J.; Brabec, J.; Bruner, A.; Cauët, E.; Chen, Y.; Chuev, G. N.; Cramer, C. J.; Daily, J.; Deegan, M. J. O.; Dunning, T. H.; Dupuis, M.; Dyall, K. G.; Fann, G. I.; Fischer, S. A.; Fonari, A.; Früchtl, H.; Gagliardi, L.; Garza, J.; Gawande, N.; Ghosh, S.; Glaesemann, K.; Götz, A. W.; Hammond, J.; Helms, V.; Hermes, E. D.; Hirao, K.; Hirata, S.; Jacquelin, M.; Jensen, L.; Johnson, B. G.; Jónsson, H.; Kendall, R. A.; Klemm, M.; Kobayashi, R.; Konkov, V.; Krishnamoorthy, S.; Krishnan, M.; Lin, Z.; Lins, R. D.; Littlefield, R. J.; Logsdail, A. J.; Lopata, K.; Ma, W.; Marenich, A. V.; Martin del Campo, J.; Mejia-Rodriguez, D.; Moore, J. E.; Mullin, J. M.; Nakajima, T.; Nascimento, D. R.; Nichols, J. A.; Nichols, P. J.; Nieplocha, J.; Otero-de-la Roza, A.; Palmer, B.; Panyala, A.; Pirojsirikul, T.; Peng, B.; Peverati, R.; Pittner, J.; Pollack, L.; Richard, R. M.; Sadayappan, P.; Schatz, G. C.; Shelton, W. A.; Silverstein, D. W.; Smith, D. M. A.; Soares, T. A.; Song, D.; Swart, M.; Taylor, H. L.; Thomas, G. S.; Tipparaju, V.; Truhlar, D. G.; Tsemekhman, K.; Van Voorhis, T.; Vázquez-Mayagoitia, ; Verma, P.; Villa, O.; Vishnu, A.; Vogiatzis, K. D.; Wang, D.; Weare, J. H.; Williamson, M. J.; Windus, T. L.; Woliński, K.; Wong, A. T.; Wu, Q.; Yang, C.; Yu, Q.; Zacharias, M.; Zhang, Z.; Zhao, Y.; Harrison, R. J. NWChem: Past, present, and future. *J. Chem. Phys.* **2020**, *152*, 184102.

(95) Balasubramani, S. G.; Chen, G. P.; Coriani, S.; Diedenhofen, M.; Frank, M. S.; Franzke, Y. J.; Furche, F.; Grotjahn, R.; Harding, M. E.; Hättig, C.; Hellweg, A.; Helmich-Paris, B.; Holzer, C.; Huniar, U.; Kaupp, M.; Khah, A. M.; Khani, S. K.; Müller, T.; Mack, F.; Nguyen, B. D.; Parker, S. M.; Perlt, E.; Rappoport, D.; Reiter, K.; Roy, S.;

Rückert, M.; Schmitz, G.; Sierka, M.; Tapavicza, E.; Tew, D. P.; Wüllen, C. v.; Voora, V. K.; Weigend, F.; Wodyński, A.; Yu, J. M. TURBOMOLE: Modular program suite for ab initio quantum-chemical and condensed-matter simulations. *J. Chem. Phys.* **2020**, *152*, 184107.

(96) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(97) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885–1891.

(98) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(99) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(100) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(101) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(102) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(103) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(104) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(105) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(106) Barthel, J.; Bachhuber, K.; Buchner, R.; Hetzenauer, H. Dielectric spectra of some common solvents in the microwave region. Water and lower alcohols. *Chem. Phys. Lett.* **1990**, *165*, 369–373.

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

(108) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(109) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(110) Li, Z.; Merz, K. M. Systematic Evaluation of Ion Diffusion and Water Exchange. *J. Chem. Theory Comput.* **2021**,

(111) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N-methylamide. *Biopolymers* **1992**, *32*, 523–535.

(112) Yeh, I.-C.; Hummer, G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J. Phys. Chem. B* **2004**, *108*, 15873–15879.

(113) Harris, K. R.; Woolf, L. A. Temperature and Volume Dependence of the Viscosity of Water and Heavy Water at Low Temperatures. *J. Chem. Eng. Data* **2004**, *49*, 1064–1069.

(114) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268.

(115) van Gunsteren, W.; Berendsen, H. Algorithms for brownian dynamics. *Mol. Phys.* **1982**, *45*, 637–647.

(116) Lee Warren, G.; Patel, S. Hydration free energies of monovalent ions in transferable intermolecular potential four point fluctuating charge water: An assessment of simulation methodology and force field performance and transferability. *J. Chem. Phys.* **2007**, *127*, 064509.

(117) Darden, T.; Pearlman, D.; Pedersen, L. G. Ionic charging free energies: Spherical versus periodic boundary conditions. *J. Chem. Phys.* **1998**, *109*, 10921–10935.

(118) Hummer, G.; Pratt, L. R.; García, A. E. Ion sizes and finite-size corrections for ionic-solvation free energies. *J. Chem. Phys.* **1997**, *107*, 9275–9277.

(119) Kastenholz, M. A.; Hünenberger, P. H. Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation. *J. Chem. Phys.* **2006**, *124*, 224501.

(120) Egan, C. K.; Paesani, F. Assessing Many-Body Effects of Water Self-Ions. I: $OH^{-}(H_2O)_n$ Clusters. *J. Chem. Theory Comput.* **2018**, *14*, 1982–1997.

(121) Dasgupta, S.; Shahi, C.; Bhetwal, P.; Perdew, J. P.; Paesani, F. How Good Is the Density-Corrected SCAN Functional for Neutral and Ionic Aqueous Systems, and What Is So Right about the Hartree-Fock Density? *J. Chem. Theory Comput.* **2022**, *18*, 4745–4761.

(122) Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M. A long-range-corrected density functional that performs well for both ground-state properties and time-dependent density functional theory excitation energies, including charge-transfer excited states. *J. Chem. Phys.* **2009**, *130*, 054112.

(123) Vasiliu, M.; Peterson, K. A.; Dixon, D. A. Bond Dissociation Energies in Heavy Element Chalcogen and Halogen Small Molecules. *J. Phys. Chem. A* **2021**, *125*, 1892–1902.

(124) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(125) 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.

(126) Douglas, M.; Kroll, N. M. Quantum electrodynamical corrections to the fine structure of helium. *Annals of Physics* **1974**, *82*, 89–155.

(127) Hess, B. A. Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators. *Physical Review A* **1986**, *33*, 3742–3748.

(128) Jansen, G.; Hess, B. A. Revision of the Douglas-Kroll transformation. *Physical Review A* **1989**, *39*, 6016–6017.

(129) Kutzelnigg, W.; Liu, W. Quasirelativistic theory equivalent to fully relativistic theory. *The Journal of Chemical Physics* **2005**, *123*, 241102.

(130) Liu, W.; Peng, D. Infinite-order quasirelativistic density functional method based on the exact matrix quasirelativistic theory. *The Journal of Chemical Physics* **2006**, *125*, 044102.

(131) Iliaš, M.; Saue, T. An infinite-order two-component relativistic Hamiltonian by a simple one-step transformation. *The Journal of Chemical Physics* **2007**, *126*, 064102.

(132) 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.

(133) Aebersold, L. E.; Yuwono, S. H.; Schoendorff, G.; Wilson, A. K. Efficacy of Density Functionals and Relativistic Effective Core Potentials for Lanthanide-Containing Species: The Ln54 Molecule Set. *J. Chem. Theory Comput.* **2017**, *13*, 2831–2839.

(134) In-house basis sets for some heavy elements. https://sites.google.com/oakland.edu/rusakov/resources/in-house-basis-sets-for-some-heavy-elements, (accessed May 11, 2023).

(135) Soper, A. K.; Weckström, K. Ion solvation and water structure in potassium halide aqueous solutions. *Biophys. Chem.* **2006**, *124*, 180–191, Ion Hydration Special Issue.

(136) Ozutsumi, K.; Ohtaki, H. Application of a compact synchrotron radiation facility to studies on the structure of solvated chloride and iodide ions in various solvents. *Pure and Applied Chemistry* **2004**, *76*, 91–96.

(137) 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.

(138) Li, Z.; Merz, K. M. J. Systematic Evaluation of Ion Diffusion and Water Exchange. *J. Chem. Theory Comput.* **2022**, *18*, 3017–3026.

(139) Haynes, W. *CRC Handbook of Chemistry and Physics, 96th Edition*; 100 Key Points; CRC Press, 2015.

(140) Li, J.; Wang, F. Accurate Prediction of the Hydration Free Energies of 20 Salts through Adaptive Force Matching and the Proper Comparison with Experimental References. *J. Phys. Chem. B* **2017**, *121*, 6637–6645.

(141) Koneshan, S.; Rasaiah, J. C.; Lynden-Bell, R. M.; Lee, S. H. Solvent Structure, Dynamics, and Ion Mobility in Aqueous Solutions at 25 °C. *J. Phys. Chem. B* **1998**, *102*, 4193–4204.

(142) Lee, S. H.; Rasaiah, J. C. Molecular Dynamics Simulation of Ion Mobility. 2. Alkali Metal and Halide Ions Using the SPC/E Model for Water at 25 °C. *J. Phys. Chem.* **1996**, *100*, 1420–1425.

(143) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. The Proton's Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from Cluster-Ion Solvation Data. *J. Phys. Chem. A* **1998**, *102*, 7787–7794.

TOC Graphic