JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2018, 14, 1321−1332

pubs.acs.org/JCTC

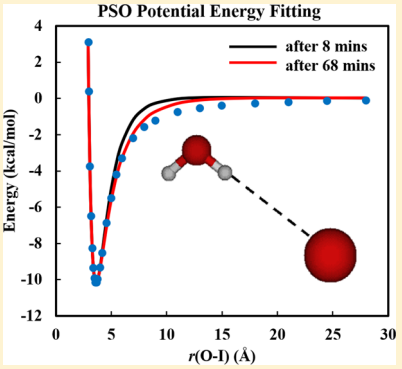
PSO Method for Fitting Analytic Potential Energy Functions. Application to $I^-(H_2O)$

H. N. Bhandari,[†] X. Ma,[‡] A. K. Paul,[§] P. Smith,[†] and W. L. Hase*[,‡]

[†]Department of Mathematics and Statistics and [‡]Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, Texas 79409, United States

[§]Department of Chemistry National Institute of Technology, Meghalaya, Shillong 793003 Meghalaya, India

ABSTRACT: In this work a particle swarm optimization (PSO) algorithm was used to fit an analytic potential energy function to $I^-(H_2O)$ intermolecular potential energy curves calculated with DFT/B97-1 theory. The analytic function is a sum of two-body terms, each written as a generalized sum of Buckingham and Lennard-Jones terms with only six parameters. Two models were used to describe the two-body terms between I^- and H_2O: a three-site model H_2O and a four-site model including a ghost atom. The fits are compared with those obtained with a genetic/nonlinear least-squares algorithm. The ghost atom model significantly improves the fitting accuracy for both algorithms. The PSO fits are significantly more accurate and much less time-consuming than those obtained with the genetic/nonlinear least-squares algorithm. Eight I^----H_2O potential energy curves, fit with the PSO algorithm for the three- and four-site models, have RMSE of 1.37 and 0.22 kcal/mol and compute times of ∼20 and ∼68 min, respectively. The PSO fit for the four-site model is quite adequate for determining densities of states and partition functions for $I^-(H_2O)_n$ clusters at high energies and temperatures, respectively. The PSO algorithm was also applied to the eight potential energy curves, with the four-site model, for a short time ∼8 min fitting. The RMSE was small, only 0.37 kcal/mol, showing the high efficiency of the PSO algorithm with retention of a good fitting accuracy. The PSO algorithm is a good choice for fitting analytic potential energy functions, and for the work presented here was able to find an adequate fit to an $I^-(H_2O)$ analytic intermolecular potential with a small number of parameters.



1. INTRODUCTION

Fitting analytic potential energy functions to electronic structure calculations, for use in chemical and molecular dynamics simulations, is an important component of theoretical and computational science and engineering.[1−3] The fit may be for an intermolecular potential between molecules,[4] including solute−solvent interactions, between different functional groups and moieties within a molecule, or between a molecule and a surface.[5,6] In addition, a complete anharmonic intramolecular potential energy function may be fit for a molecule, as well as functions describing a potential energy surface (PES) for a chemical reaction between molecules.[2,3] For the latter, experimental data has often been used to assist in fitting the PES.[7,8]

Different strategies and algorithms have been used to fit analytic potential energy functions.[1−3] The approach of the Hase research group in the past has been to use a genetic algorithm to identify possible sets of parameters for the fit and then optimize these parameters by a local nonlinear least-squares algorithm.[1] Though one is not assured that the optimum fit has been found, the fitting is assumed complete when there is an adequate fit.[6,9−12] This is often a lengthy process. This approach has been used to obtain molecule−molecule and molecule−surface intermolecular potentials.[4−6]

The work presented here is related to ongoing calculations to determine densities of states and partition functions for intermolecular modes of $I^-(H_2O)_n$ clusters at high energies and temperatures. These calculations use classical statistical mechanics and rigid water molecules and require an analytic intermolecular potential for the cluster. The density of states at energy E for smaller clusters may be determined by the energy derivative of the phase space volume, determined by a Monte Carlo calculation of the phase space integral;[13−17] e.g., for $I^-(H_2O)_2$, this is an 18-dimensional sampling problem. For larger clusters, for which direct evaluation of the phase space volume is not practical, adiabatic switching[16,18,19] may be used.

The above calculations are computationally demanding and for calculating the intermolecular density of states and partition function, at high energies and temperatures, it is sufficient to write the analytic potential for the $I^-(H_2O)_n$ cluster, with rigid water molecules, as a sum of the water intermolecular potential $V(H_2O,H_2O)$ and the I^----H_2O intermolecular potential $V(I^-,H_2O)$. Model two-body potentials, such as TIP4P,[20] are considered for $V(H_2O,H_2O)$. A number of analytic functions have been developed for the I^----H_2O intermolecular potential, with a range a complexity.[21−34] Many of the functions are a sum of two-body potentials, but some include a multibody term.[22−24,26,29,31,33] The I^----H_2O analytic functions were determined by holding H_2O rigid, but for some calculations with these

Received: November 7, 2017
Published: January 18, 2018

DOI: 10.1021/acs.jctc.7b01122
J. Chem. Theory Comput. 2018, 14, 1321−1332

potentials water is not rigid[25−27,30,33,34] and allowed to vibrate, which allows comparisons with spectroscopic studies.

For the ongoing study to calculate densities of states and partition functions for the intermolecular modes of $I^-(H_2O)_n$ clusters at high energies and temperatures, a I^----H_2O potential was developed with a small number of parameters to enhance the efficiency of the computationally demanding calculations. An accurate potential for energies up to I^----$H_2O \rightarrow I^- + H_2O$ dissociation was also important. In fitting this potential it was discovered that the particle swarm optimization (PSO) algorithm[9] is far superior to the genetic/nonlinear least-squares algorithm we have previously used.[1] An acceptable fit was obtained with the PSO algorithm, but not with the genetic/nonlinear least-squares algorithm. This finding is the focus of the work presented here and the remainder of this article is organized as follows. The next section, section 2, describes the PSO method and its implementation in a high performance computing environment. Section 3 presents the electronic structure calculations for the $I^-(H_2O)$ intermolecular potential, the analytic function used to represent this potential, and the fit to this function by the genetic/nonlinear least-squares algorithm. Section 4 describes the fit of the analytic function to the $I^-(H_2O)$ intermolecular potential by the PSO method. Section 5 compares and discusses the genetic/nonlinear least-squares and PSO fits to the analytic potential energy function and their computational efficiencies. The calculations are summarized in the concluding section 6.

2. PSO METHOD

2.1. Introduction to the PSO Algorithm. The particle swarm optimization (PSO) algorithm is a bioinspired stochastic optimization technique developed by James Kennedy and Russell Eberhart in 1995.[35] Initially, it was developed to minimize a continuous function f: $\mathbb{R}^K \rightarrow \mathbb{R}$, sometimes called a "fitness" function. However, it can also be used to minimize piecewise continuous functions, binary functions, discrete functions, and so on. It is a population-based search algorithm that mimics the behavior of bird flocking. In the algorithm, the total population is called swarm and individuals are called particles. Particles are points in the search space of the underlying optimization problem. Each particle is a candidate solution of the fitness function. Particles move in the search space stochastically seeking better fitness. Movement of each particle is influenced by its momentum, individual experience, and group experience. There are following three important heuristics in the PSO algorithm: (a) each particle has an internal memory to store its best (so far) position, called local best position; (b) particles in swarm can share information about best position (to date) of the entire group, called global best position of the swarm; (c) each particle wants to keep its current velocity, called a momentum/inertia component.

There are many PSO versions available in the literature. One of the widely accepted versions is so-called standard PSO version.[36] In this version, the inertia weight ω and velocity clamping $\mathbf{V}_{max}$ are used. In the PSO system, with M particles and K dimensions, the ith particle maintains a triple of vectors $(\mathbf{x}_i(t), \mathbf{v}_i(t), \mathbf{y}_i(t)) \in \mathbb{R}^K \otimes \mathbb{R}^K \otimes \mathbb{R}^K$ for $1 \le i \le M$, where $\mathbf{x}_i(t) \in \mathbb{R}^K$ is the current position, $\mathbf{v}_i(t) \in \mathbb{R}^K$ is the current velocity, and $\mathbf{y}_i(t) \in \mathbb{R}^K$ is the local best position. For simplicity, the dimension of the vectors is not written explicitly, which is the same as the dimension of optimization problem. The velocity and position of each particle i are updated using the following equations,[36,37]

$$\mathbf{v}_i(t+1) = \omega \mathbf{v}_i(t) + c_1 r_1(\mathbf{y}_i(t) - \mathbf{x}_i(t)) + c_2 r_2(\hat{\mathbf{y}}(t) - \mathbf{x}_i(t)) \quad (1)$$

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1) \quad (2)$$

where $\hat{\mathbf{y}}(t)$ represents the global best position and $r_1 \in U[0,1]$ and $r_2 \in U[0,1]$ are random numbers chosen from a random uniform distribution $U[0,1]$.There are three components in the velocity equation with corresponding learning parameters ω, c_1, $c_2 \in \mathbb{R}_0^+$, where $\mathbb{R}_0^+$ is the set of positive real numbers including zero. Here ω is a weight parameter for the inertia component $\omega \mathbf{v}_i(t)$, c_1 is a parameter for the cognitive component $c_1 r_1(\mathbf{y}_i(t) - \mathbf{x}_i(t))$, and c_2 is a parameter for the social component $c_2 r_2(\hat{\mathbf{y}}(t) - \mathbf{x}_i(t))$.

Furthermore, the local best and the global best positions are updated as follows:

$$\mathbf{y}_i(t+1) = \begin{cases} \mathbf{x}_i(t+1) & \text{if } f(\mathbf{x}_i(t+1)) < f(\mathbf{y}_i(t)) \\ \mathbf{y}_i(t) & \text{otherwise} \end{cases} \quad (3)$$

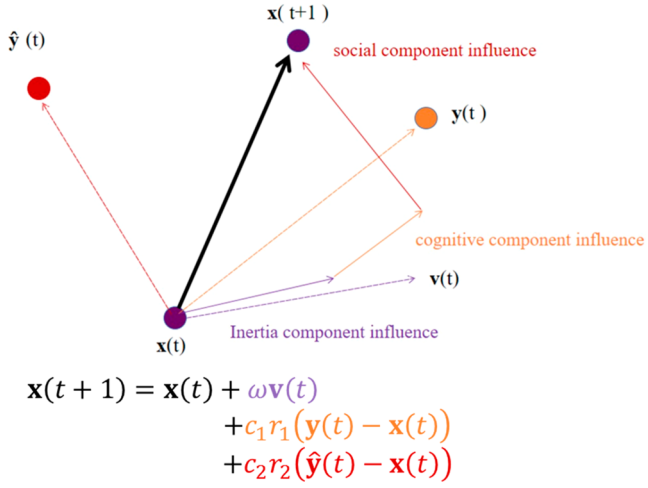


Figure 1. Illustration on how the single particle moves in the search space under the influence of inertia component, cognitive component, and social component.

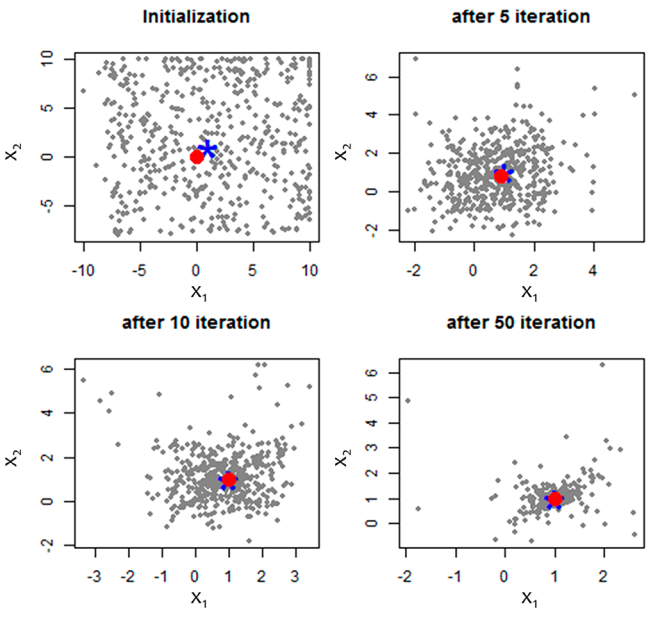


Figure 2. Illustration of the PSO system in two dimensions with the Rosenbrock function. Gray dots represent particle positions, the red dot represents the global best position, and the blue star represents the exact optimal solution. The global best position gradually moves to the exact solution.

DOI: 10.1021/acs.jctc.7b01122
J. Chem. Theory Comput. 2018, 14, 1321−1332

$$\hat{y}(t+1) = \arg\min\{f(\mathbf{y}_1(t+1)), f(\mathbf{y}_2(t+1)), ..., f(\mathbf{y}_M(t+1))\} \quad (4)$$

The PSO algorithm has become very popular in recent years and is applied widely in many fields such as function approximations, linear and nonlinear fitting problems, robotics, image/signal processing, etc.[38] The main reason for its popularity is that it is easy to implement. It does not use gradient information but searches stochastically throughout the domain. It generally can escape from being trapped by a poor local minimum. Theoretically, the PSO algorithm can locate a global minimum of even a highly multimodal function, if one applies some extra techniques such as restarting PSO multiple times.[37]

2.2. Standard PSO Pseudo Code. The following is the pseudo code of the standard PSO algorithm that was used in this work.

Step 1: Initialization:

1. Choose the PSO domain $\Omega = \{\mathbf{x} \in \mathbb{R}^K \text{ s.t } \mathbf{lb} \le \mathbf{x} \le \mathbf{ub}\}$ where **lb** and **ub** are vectors of lower and upper bounds of the vector **x.**
2. Choose the learning parameters $\omega \in U[0.4, 1]$, $c_1 \in U[1.4, 2]$, $c_2 \in U[1.4, 2]$, and an arbitrary number $a \in [0, 1]$
3. Choose the velocity restriction in the j^{th} dimension $\boldsymbol{V}_{max,j} = a(\boldsymbol{ub}_j - \boldsymbol{lb}_j)$
4. For each particle i for $1 \le i \le M$, randomly initialize the position $\mathbf{x_i} \in U[\mathbf{lb}, \mathbf{ub}]$ and velocity $\mathbf{v_i} \in U[-\boldsymbol{V}_{max}, \boldsymbol{V}_{max}]$
5. Choose the local best positions $\mathbf{y_i} = \mathbf{x_i}$ for $1 \le i \le M$
6. Choose the global best position $\hat{\mathbf{y}} = \mathbf{y}_\tau$ for some $\tau = \{1, 2, ..., M\}$

Loop:

Step 2: Update positions and velocities:

For *particle* $i = 1$ to $i = M$ **Do:**

1. Choose $r_1 = U[0, 1], r_2 = U[0, 1]$
2. Update velocity: $\mathbf{v_i} = \omega\mathbf{v_i} + c_1r_1(\mathbf{y_i} - \mathbf{x_i}) + c_2r_2(\hat{\mathbf{y}} - \mathbf{x_i})$
3. Velocity restriction:

For dimension $j = 1$ to $j = K$ **Do:**

i. **if** $\boldsymbol{v}_{ij} > \boldsymbol{V}_{max,j}$ **then** $\boldsymbol{v}_{ij} = \boldsymbol{V}_{max,j}$

ii. **if** $\boldsymbol{v}_{ij} < -\boldsymbol{V}_{max,j}$ **then** $\boldsymbol{v}_{ij} = -\boldsymbol{V}_{max,j}$

4. Update position: $\mathbf{x_i} = \mathbf{x_i} + \mathbf{v_i}$
5. Boundary check:

For dimension $j = 1$ to $j = K$ **Do:**

i. **if** $\boldsymbol{x}_{ij} > \boldsymbol{ub}_j$ **then** $\boldsymbol{x}_{ij} = \boldsymbol{ub}_j$

ii. **if** $\boldsymbol{x}_{ij} < \boldsymbol{lb}_j$ **then** $\boldsymbol{x}_{ij} = \boldsymbol{lb}_j$

End

Step 3: Update local best and global best positions:

For *particle* $i = 1$ to $i = M$ **Do:**

1. Update local best: **if** $f(\mathbf{x_i}) < f(\mathbf{y_i})$ **then** $\mathbf{y_i} = \mathbf{x_i}$
2. Update global best: **if** $f(\mathbf{y_i}) < f(\hat{\mathbf{y}})$ **then** $\hat{\mathbf{y}} = \mathbf{y_i}$

End

We used the following two stopping criteria simultaneously: (a) if predefined maximum allowed iterations is reached and (b) if there is no significant improvement[39] in the fitness value even after certain number of iterations.

2.3. PSO Movement Illustration. Figure 1 illustrates how the single particle makes the next move under the influence of three components of the velocity update rule, given by eqs 1 and 2. Here $\mathbf{x}(t)$ and $\mathbf{v}(t)$ are the current position and velocity of the particle for the current time step t. Similarly, $\mathbf{y}(t)$ and $\hat{\mathbf{y}}(t)$ are the local best and global best positions, respectively. The particle velocity for time step $t + 1$ is updated by adding three vectors toward the current velocity, local best position, and global best position weighted by parameters ω, c_1, and c_2, as shown in Figure 1. Also $r_1 \in U[0,1]$ and $r_2 \in U[0,1]$ are used so that the particle moves stochastically, maintaining exploration/exploitation trade-off.

As an example, consider a benchmark test function called the Rosenbrock function defined as

$$f(x) = \sum_{i=1}^{K-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2] \quad (5)$$

with $f(x^*) = 0$ at $x^* = (1, 1, ...)$ as an optimal solution. The operation of the PSO method, during the process of minimization of the two-dimensional Rosenbrock function, is similar to Figure 2. In each iteration, each and every particle moves in the search space and gradually converges to a common point, considered as an approximation to the optimal solution of the given optimization problem (in this case minimizing the Rosenbrock function).

2.4. Implementation of Local Optimizer. It has been proved that the standard PSO algorithm can converge to a point that might not be a local/global minimum.[37] To overcome this drawback of the standard PSO algorithm, it may be repeated multiple times to find a better solution by keeping the previous optimal solution as a global best position for the next run. This is a common technique used to find the global minimum of an optimization problem and is called the multistart approach.[37] Also, the PSO is a global search method that searches the optimal solutions in the entire domain. However, it is not as good as some existing local search methods in terms of local refinement of the solution. To address this issue, we used "Cyclic Coordinate Descent" local optimizer[40] between two PSO runs. This method uses coordinates of the global best position vector as the search directions. It refines the solution in some small domain around the global best position by searching along each coordinate of the global best position vector repeatedly. We used the Brent Algorithm[41] to perform a 1-dimensional line search in Cyclic Coordinate Descent local optimizer. The main purpose of this approach is to find a more accurate local minimum. We found that these additional strategies provided a very good combination of exploration/exploitation of the domain and produced better results in general.

3. INTERMOLECULAR POTENTIAL ENERGY FOR $I^-(H_2O)$ AND FIT WITH THE GENETIC/NONLINEAR LEAST-SQUARES ALGORITHM

3.1. $I^-(H_2O)$ Intermolecular Potential Energy. Electronic structure potential energy curves at the DFT/B97-1 level of theory[42−44] with the ECP/d basis set,[45] were calculated for different orientations of the I^----H_2O interaction, as well as for the $I^-(H_2O)$ potential energy minimum. The DFT calculations were performed with the NWChem electronic structure computer program.[46] The geometry for the potential energy minimum is depicted in Figure 3 and the depth of the minimum, without including zero-point energies, is −10.89 kcal/mol. At 298 K, the $I^-(H_2O) \rightarrow I^- + H_2O$ dissociation enthalpy predicted by DFT/B97-1 with the ECP/d basis set is 10.59 kcal/mol, which is in good agreement with the experimental value of 10.48 ± 0.14 kcal/mol.[47,48] When H_2O is treated as rigid and in its equilibrium geometry, the $I^-(H_2O)$ geometry in its potential energy minimum is almost the same, giving the depth of the minimum of

DOI: 10.1021/acs.jctc.7b01122
J. Chem. Theory Comput. 2018, 14, 1321−1332

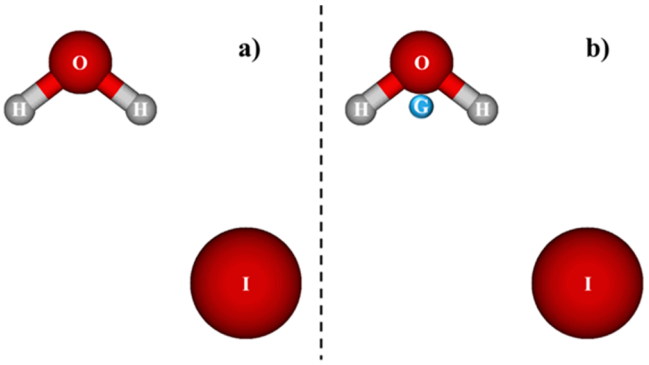


Figure 3. (a) Structure of the $I^-(H_2O)$ global potential minimum and (b) illustration of a ghost atom position.

−10.59 kcal/mol. The orientations used for calculating the potential energy curves are given in Figure 4. The B97-1/ECP/d potential energy for the different orientations are given by the graphs in Figure 5.

3.2. Analytic Intermolecular Potential Energy Function. In previous work,[4−6,49,50] it is found that intermolecular potentials may be accurately fit by a sum of two-body potentials. However, to obtain an accurate fit, it is important to use a potential more general than the Lennard-Jones function, i.e.

$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{6}$$

Different two-body potentials may be used,[4−6,49,50] and the one used here was found to be successful for previous fits.[4,6] It consists of two generalized Lennard-Jones terms and the exponential Buckingham function, i.e.

$$V(r) = Ae^{-Br} + \frac{C}{(r+c)^m} + \frac{D}{(r+d)^n} \tag{7}$$

where r is the interatomic distance, A, C, and D are linear parameters and B, c, d, m, n are nonlinear parameters. Parameter B represents the exponential range of the potential, c and d are shift parameters, and m and n are integer parameters with $n \geq m + 3$.

Two models are used for the $I^-(H_2O)$ intermolecular potential. For each, H_2O is kept rigid in its equilibrium geometry optimized at B97-1/ECP/d level of theory: $r(O-H_1) = r(O-H_2) = 0.964$ Å, and $\theta(HOH) = 104.4°$. For one, I^- interacts with the H and O atoms. For the second, a ghost atom is added, as first proposed by Clementi and co-workers[51] and used in the TIP4P potential.[20] This atom lies on the bisector of the H−O−H angle and at a distance z from the O atom.

In the first model, the water molecule without the ghost atom is treated as three atoms, H_1, H_2, and O, where H_2 refers to the hydrogen atom closest to I^-. There are three potential energies in this $I^-(H_2O)$ system, namely, the two hydrogen−iodine potentials (V_{HI}) and the oxygen−iodine potential (V_{OI}), which are dependent on the interatomic distances $r(H_1-I)$, $r(H_2-I)$, and $r(O-I)$. The total potential energy is a summation of these potentials and given by eq 8. An additional potential term, namely, ghost−iodine potential (V_{GI}), is present in the second model, and the total potential energy is given by eq 9.

$$V = V_{OI}(r_{O-I}) + \sum_{i=1}^{2} V_{HI}(r_{H_i-I}) \tag{8}$$

$$V = V_{OI}(r_{O-I}) + \sum_{i=1}^{2} V_{HI}(r_{H_i-I}) + V_{GI}(r_{G-I}) \tag{9}$$

3.3. Fitting the Intermolecular Potential. The root-mean-square error (RMSE) is used to evaluate the quality of the fit to the intermolecular potential energy function. This may be calculated as follows: Let $\{\mathbf{r}_i, t_i\}$ for $i = 1, 2, \ldots N$ be the given data for the $I^-(H_2O)$ N orientations. And the root-mean-square error (RMSE) may be calculated as

$$\phi(\lambda) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(V(\mathbf{r}_i,\boldsymbol{\lambda}) - t_i)^2} \tag{10}$$

where $\mathbf{r}_i$ is a vector of interatomic distances, t_i is the DFT/B97-1 energy, and $V(\mathbf{r}_i,\boldsymbol{\lambda})$ is the fitted potential energy. Also $\boldsymbol{\lambda}$ represents a vector of linear and nonlinear parameters. The dimension of $\boldsymbol{\lambda}$ is 16 in the first model (without the ghost atom) and 25 in the second model (with the ghost atom), including the ghost atom position parameter z. Hence, this becomes a multidimensional nonlinear least-squares fitting problem.

The goal in the fitting was to find physically acceptable optimal parameters. Therefore, the concern is not only in the quality of fitting (i.e., RMSE) but also in the correct behavior/shape of the potential functions. For example, the $I^-(H_2O)$ orientation pattern c and e shown in Figure 4 represent the repulsive intermolecular interactions and the potential energies should all be positive. Also, the fitted potential functions should give a potential minimum with correct structure and potential depth.

3.4. Genetic/Nonlinear Least-Squares Algorithm Fitting Results. The B97-1/ECP/d potential energy curves in Figure 5, for the different orientations, were fit simultaneously by the genetic/nonlinear least-squares algorithm and the fits for the models with and without the ghost atom are also shown in this figure. The fitting parameters are given in Table 1. For some of the two-body terms, zero is the best fit for some of the shift parameters. The fit with the ghost atom included is somewhat better and has a similar RMSE. It gives a potential energy minimum of −10.81 kcal/mol, in good agreement with the B97-1 value of −10.89 kcal/mol. With the H_2O moiety in its equilibrium structure optimized at the B97-1/ECP/d theory level, the fitted geometric parameters for the minimum are

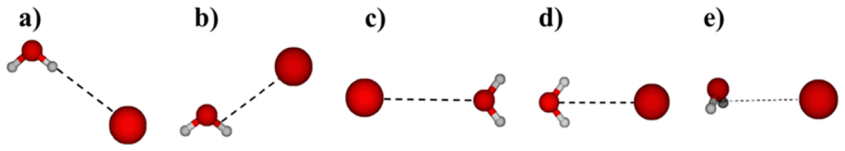


Figure 4. Orientations of $I^-(H_2O)$ used for calculating the B97-1/ECP/d potential energy curves for the genetic/nonlinear least-squares algorithm fits. The position of I^- is moved (a) in the H_2O plane and along the OH bond direction with $r(H-I)$ varied; (b) in the H_2O plane, parallel to one OH bond, and intersecting the other OH bond at its midpoint; (c) in the H_2O plane and on the oxygen side along the H−O−H angle bisector; (d) in the H_2O plane and on the hydrogen side along the HOH angle bisector; and (e) perpendicular to the H_2O plane and intersecting the plane at the H_2O geometric center. These orientations and the three additional orientations were used for the PSO algorithm fits.

DOI: 10.1021/acs.jctc.7b01122
J. Chem. Theory Comput. 2018, 14, 1321−1332

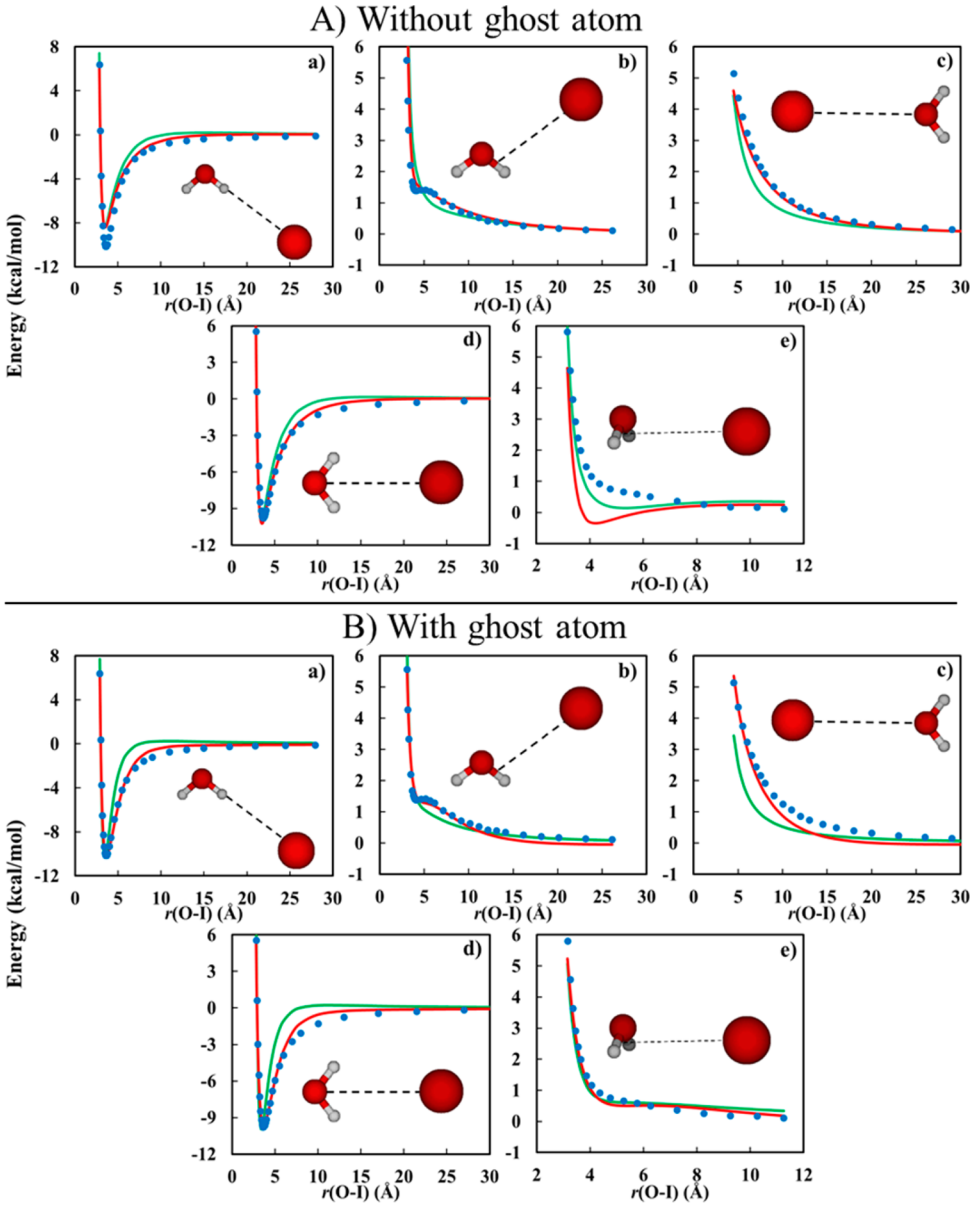


Figure 5. Blue dots: B97-1/ECP/d potential energies are represented. Green and red curves: Fitted analytic potential energy functions by genetic/nonlinear least-squares algorithm and PSO algorithm, respectively. The upper and lower figures show the fitted results without and with the ghost atom.

Table 1. Parameter Sets, without and with the Ghost Atom, from Genetic/Nonlinear Least-Squares Algorithm Fits[a,b]

Without the Ghost Atom: RMSE = 1.12 kcal/mol								
potentials	*A*	*B*	*C*	*D*	*c*	*d*	*m*	*n*
H−I	1188.2	2.6553	−55301	1956.2	4.0766	1.7372	4	15
O−I	32516×10^3	5.0927	5408.0	0	1.9971		3	
With the Ghost Atom: RMSE = 1.15 kcal/mol								
potentials	*A*	*B*	*C*	*D*	*c*	*d*	*m*	*n*	*z*
H−I	15333	2.9985	−545.99	345.13	−0.03612	0	3	10	
O−I	817.99	5.9012	761.07	0	0		3		
G−I[c]	201.90	1.7208	72.409	83359	0	0	2	9	0.3

[a]Decimal numbers are kept in 5 significant figures for display. The actual numbers are in double-precision, except for *z*. [b]If the coefficient *C* or *D* is 0, then the shift term *c* or *d* and the index m or *n* are meaningless, respectively. [c]G−I is the interaction between ghost atom and iodine.

$r(O-H_1) = r(O-H_2) = 0.964$ Å, $\theta(HOH) = 104.4°$, $r(O-I) = 3.549$ Å, and $\theta(OH_2I) = 157.8°$. These are in good agreement with the results of the B97-1 calculations: $r(O-H_1) = 0.964$ Å, $r(O-H_2) = 0.981$ Å, $\theta(HOH) = 100.8°$, $r(O-I) = 3.602$ Å, and $\theta(OH_2I) = 159.7°$, and B97-1 calculations with treating H_2O rigid: $r(O-H_1) = r(O-H_2) = 0.964$ Å, $\theta(HOH) = 104.4°$, $r(O-I) = 3.613$ Å, and $\theta(OH_2I) = 161.0°$ (Table 2).

These are the best fits that were found with the genetic/nonlinear least-squares algorithm. The constraints on the fitting were that $A, B > 0$ and $n \geq m + 3$. No constraints were placed on *c*

DOI: 10.1021/acs.jctc.7b01122
J. Chem. Theory Comput. 2018, 14, 1321−1332

Table 2. $I^-(H_2O)$ Potential Energy Minima and Corresponding Geometries Calculated at the B97-1/ECP/d Theory Level and Predicted by Analytical Potential Energy Parameter Sets[a]

	$r(H_1-I)$	$r(H_2-I)$	$r(O-I)$	$r(O-H_1)$	$r(O-H_2)$	$\theta(HOH)$	$\theta(OH_2I)$	E
B97-1[b]	3.66	2.67	3.60	0.96	0.98	100.8	159.7	−10.89
B97-1(rigid H_2O)[c]	3.75	2.69	3.61	0.96	0.96	104.4	161.0	−10.59
GA-5[d]	3.65	2.64	3.55	0.96	0.96	104.4	157.8	−10.81
PSO-5[e]	3.88	2.78	3.72	0.96	0.96	104.4	163.9	−10.27
PSO-8[f]	3.81	2.75	3.68	0.96	0.96	104.4	161.1	−10.51
PSO-8′[g]	3.78	2.75	3.67	0.96	0.96	104.4	159.6	−10.61

[a]Bond distances are in units of Å, bond angles are in units of degree, and potential minimum energies E are in units of kcal/mol. H_2O geometries in GA-5, PSO-5, PSO-8, and PSO-8′ are the same as B97-1(rigid H_2O), which is the equilibrium H_2O optimized at B97-1/ECP/d theory level. [b]Global potential minimum; i.e., none of the atoms were "frozen" during geometry optimization. [c]The H_2O is treated rigid and in its equilibrium geometry. [d]Genetic/nonlinear least-squares algorithm fits with five $I^-(H_2O)$ orientations using the ghost atom model. [e]PSO fits with five $I^-(H_2O)$ orientations using the ghost atom model. [f]PSO fits with eight $I^-(H_2O)$ orientations using the ghost atom model. [g]Lower level PSO fits with 8 $I^-(H_2O)$ orientations using ghost atom model.

Table 3. Initial Domains of the PSO System with and without the Ghost Atom

potentials	A	B	C	D	C	d	m^a	n^a	z
H---I	[0, 5000]	[0, 5]	[−5000, 0]	[0, 5000]	[0, 2]	[0, 2]	[3, 9]	[6, 14]	
O---I	[0, 5000]	[0, 5]	[0, 5000]	[0, 5000]	[0, 2]	[0, 2]	[3, 9]	[6, 14]	
G---I	[0, 5000]	[0, 5]	[0, 5000]	[0, 5000]	[0, 2]	[0, 2]	[3, 9]	[6, 14]	[0, 0.4]

[a]For integer parameters m and n, real number domains were used and rounded to the nearest integer in function evaluation.

and d, the shift parameters, or the position of the ghost atom along the z-axis. For the fit with the ghost atom, the H−I term with C is attractive, and the remaining C and D terms are repulsive. This is a physical result, because the D terms are added to enhance the accuracy of the repulsive potential and the O−I and G−I terms are expected to be repulsive. For the fit with the ghost atom, a fixed position of the ghost atom is used between the O and H atoms, with z of 0.3 Å.

In surmising the fits with the ghost atom in Figure 5, the fits for curves b and e are best, which are primarily repulsive interactions. However, curve c with a dominant O−I repulsive interaction is not fitted well. The sharp change in the shape of the potential for curve b is not well fitted by the analytic function. Somewhat surprisingly the attractive potential curves a and d are fitted better without the ghost atom. In summary, these were the best fits obtained after many searches. It is possible that there are more accurate fits that were not found with the genetic/nonlinear least-squares algorithm.

4. PSO ALGORITHM FIT OF THE $I^-(H_2O)$ ANALYTIC INTERMOLECULAR POTENTIAL

4.1. Setting up the PSO Domain and Tuning the PSO Parameters. The PSO algorithm usually works in a K-dimensional compact domain $\Omega \in \mathbb{R}^K$, which needs to be constructed before starting the PSO algorithm. Because the location of optimal solution was not known initially, the domain size may be dynamically changed throughout the optimization process. If any parameter hits its respective boundary, then the domain size of this parameter is increased by changing either the lower bound or upper bound. The initial PSO domain for the fitting parameters are given in Table 3.

The PSO velocity updating given in eq 1, has the tuning parameters ω, c_1, c_2, and $\mathbf{V}_{max}$. These parameters play a great role in the algorithm and should be chosen very carefully. Theoretical analyses have shown that the PSO system will converge to a point if learning parameters are selected, which satisfy the condition $1 > \omega > \frac{1}{2}(c_1 + c_2) - 1$.[52] However, there is no unique set of values suggested by theory. Furthermore, this limit is not guaranteed to be a local or global minimum. Empirical studies have shown that $\omega = 0.7968$, $c_1 = 1.4962$, and $c_2 = 1.4962$ produces a relatively good solution for many problems and satisfies $1 > \omega > \frac{1}{2}(c_1 + c_2) - 1$.[52] Many researchers use this set of parameters.[53−55] However, to increase flexibility in our algorithm, parameters were selected randomly from uniform distributions as follows: $\omega = U[0.4, 1.0]$, $c_1 = U[1.4, 2.0]$, and $c_2 = U[1.4, 2.0]$. Also, to avoid explosion of the velocity, velocity clamping was applied with $\mathbf{V}_{max}$. The value of $\mathbf{V}_{max}$ may be controlled by choosing parameter $a \in [0,1]$. A larger value of a means the PSO particles are allowed to take larger step sizes and vise-versa. In this study, a was chosen as 0.1, i.e., one-tenth of the domain size of each fitting parameter. Figure 6 shows that the RMSE in the PSO method initially decreases significantly and then slows down in successive runs. Application of the local optimizer slightly improves the RMSE because it helps to exploit the local domain.

It is also necessary to choose other parameters such as the number of particles, maximum allowed iterations, and the number of PSO runs. One can control the number of iterations per run and number of runs to balance the exploration−exploitation trade-off. A larger number of iterations in a single PSO run means that the algorithm is investing more time to refine the local domain, whereas a larger number of runs helps to explore the entire domain by reinitializing the PSO particles. We used 500 iterations and 100 runs in our calculations, and, additionally, $T = 100$ in the second termination criterion.[39] Applying the second termination criterion helps to stop searching the region if there is not a significant progress even after T iterations, which means that the global minimum is highly unlikely in that particular region of the domain. Selecting the number of particles is very crucial in the PSO algorithm. It should be chosen on the basis of dimension, problem difficulty, memory space availability, computing resource availability, etc. There is no common rule suggested by theory. Therefore, it should be chosen on the basis of experience. Table 4 provides the RMSE and time consumption for different numbers of particles. The RMSE is generally improved with more particles, despite the somewhat stochastic results.

DOI: 10.1021/acs.jctc.7b01122
J. Chem. Theory Comput. 2018, 14, 1321−1332

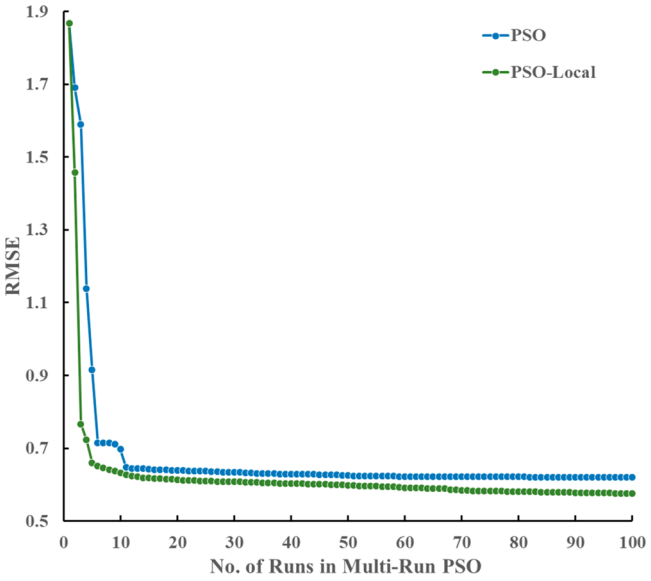


Figure 6. Dark blue curve: RMSE for a multi-run PSO without using the local optimizer. Dark green curve: RMSE for a multi-run PSO using the local optimizer. The results are for the ghost atom model with 50 particles, 100 runs, and 500 iterations per run, including the second terminating criterion[39] where T was considered as 100.

Table 4. Root Mean Square Error (RMSE) and Time Usage of PSO Algorithm Fits for the Ghost Atom Model, with Number of Particles from 20 to 100[a]

no. of particles	RMSE (kcal/mol)	time (mins)
20	0.95	7.23
40	0.65	7.80
60	0.44	8.87
80	0.55	10.03
100	0.31	10.33

[a]The PSO fits were obtained with 500 iterations per run and 50 runs.

As another aspect, the time consumption increases almost linearly but not arithmetically. This is mainly due to the application of the second termination criterion, and the overhead is almost negligible. This would allow the use of more particles in the PSO fitting without an arithmetically increased time consumption. A number of particles between 50 and 100 seems to be a good choice for the current problem for the ghost atom model and a total of 25 parameters for the potential energy function.

In our current form of the PSO algorithm, every parameter should have the real domain including m and n. Fitting the two-body analytic potential energy functions in eq 7 can be regarded as searching for the optimal parameter set in continuous search space for parameters A, B, C, D, c, and d and discrete search space for parameter m and n. Meanwhile, the evaluated function, i.e., the analytic potential energy function, is continuous to the continuous variables A, B, C, D, c, and d, and discontinuous to the discrete variables m and n. Because the PSO algorithm is intrinsically continuous, it is adapted to this problem by treating the discrete variables m and n as continuous variables, and rounding them to the nearest integers in each iteration, before the potential energy function evaluations.

4.2. PSO Results. The PSO algorithm followed by local optimizer was performed with 500 particles, 500 iterations per run, and 100 runs for the two intermolecular potential models, i.e., without and with a ghost atom. Because the algorithm is stochastic in nature, it was performed 10 times in both models and best fits were selected out of these 10 different tries, which provided the smallest RMSE. The same constraints on the fitting parameters described in section 3.4 were used here for the PSO fitting. Additionally, the shift terms, i.e., c and d, were considered only positive, to avoid the denominators becoming zeros for the second and third terms in eq 7. In the PSO algorithm, the position parameter of the ghost atom, z, is allowed to be optimized simultaneously along with the other parameters.

The PSO algorithm gave better fits to the five potential energy curves than those found with the genetic/nonlinear least-squares algorithm. The fitting parameters are shown in Table 5, and the fitting results are shown in Figure 5. The ghost atom model improves the fitting results significantly by reducing the RMSE from 0.87 to 0.27 kcal/mol, and gives a potential energy minimum of −10.27 kcal/mol. The fitted geometric parameters for this minimum are r(O−I) = 3.72 Å and θ(OH$_2$I) = 163.9°. The running time for executing PSO fitting without and with the ghost atom are 7.85 and 16.63 min, respectively.

The PSO algorithm was also performed on an expanded data set to provide additional constraints on the PSO fitting. Three additional I$^-$(H$_2$O) orientations and their potential energy curves were used, to give a total of eight potential energy curves for the PSO fitting. These three additional orientations are identified as f, g, and h and shown in Figure 7. Orientation f includes the global potential energy minimum and the region around it. Orientations g and h are introduced for additional descriptions of interactions between I$^-$ and H$_2$O when I$^-$ is positioned out of and in the H$_2$O plane, respectively. The potential energy curves for orientations f, g, and h are included in Figure 8. The fitting parameters are shown in Table 6, and the fitting results are shown in Figure 8. The ghost atom model improves the fitting results significantly, by reducing the RMSE to 0.22 from 1.37 kcal/mol without the ghost atom. This is so far the best fitted parameter set, and the two-body potential energy functions are plotted in Figure 9. The PSO fitting costs 19.45 min without the ghost atom, and 67.87 min with the ghost atom. The potential energy minimum predicted by the ghost atom model

Table 5. Parameter Sets from PSO Algorithm Fits of Five I$^-$(H$_2$O) Orientations Shown in Figure 5

Without the Ghost Atom: Elapsed Time = 7.85 min, RMSE = 0.87 kcal/mol								
potentials	A	B	C	D	c	d	m	n
H−I	41734	4.9692	−16264	0	5.3255	1.0854	3	
O−I	11093×10^2	3.8748	36984	0	5.8998	2.0000	3	
With the Ghost Atom: Elapsed Time = 16.63 min, RMSE = 0.27 kcal/mol								
potentials	A	B	C	D	c	d	m	n	z
H−I	432.72	1.4834	−5500.0	14059	1.8387	0.38892	3	8	
O−I	17261	3.0905	45713×10^1	0	8.2118		4		
G−I	54142	3.9691	79311	9258.6	3.1317	0.091315	4	9	0.33998

DOI: 10.1021/acs.jctc.7b01122
J. Chem. Theory Comput. 2018, 14, 1321−1332

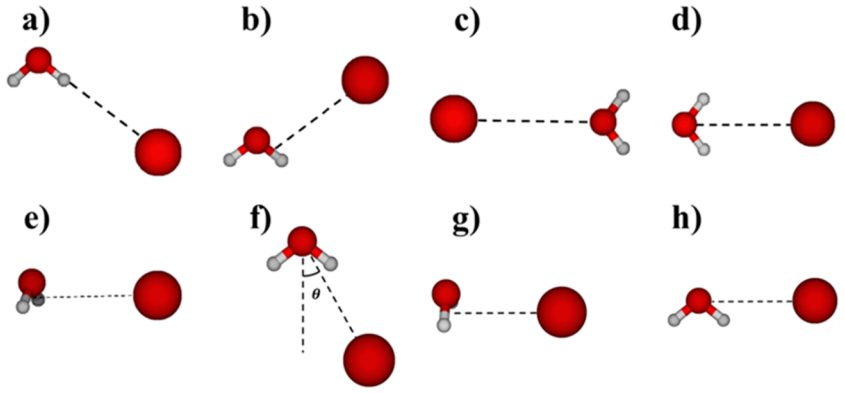


Figure 7. $I^-(H_2O)$ orientations used to calculate three additional potential energy curves for the PSO fitting. The position of I^- is moved (f) in the H_2O plane with a fixed oxygen−iodine distance r(O−I) = 3.6 Å and varied G−O−I angle, (g) perpendicular to the H_2O plane and intersecting the plane at the midpoint of a H−O bond, and (h) in the H_2O plane, perpendicular to the H−O−H angle bisector, and directed toward the O atom.

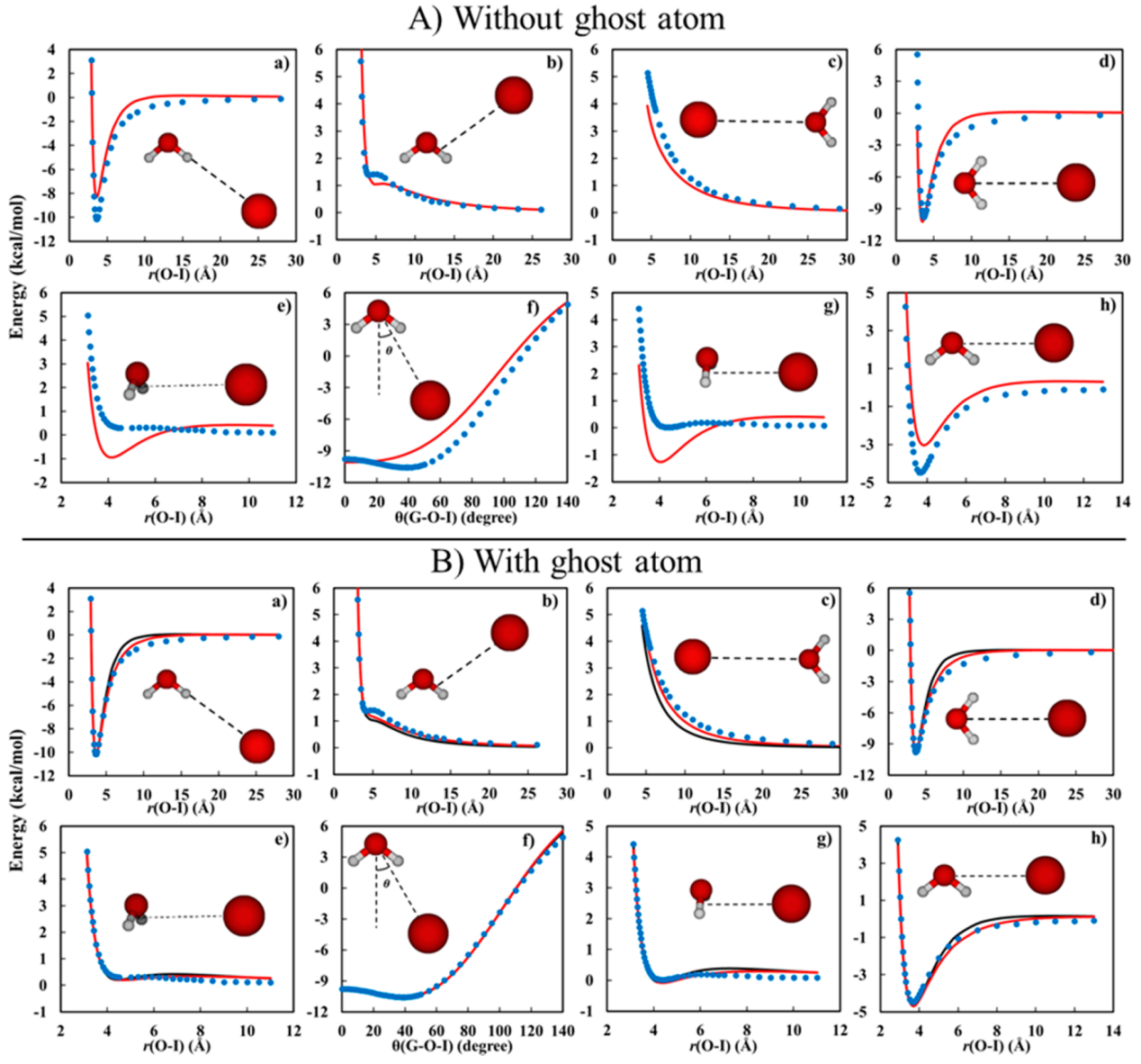


Figure 8. Blue dots: B97-1/ECP/d potential energies. Red curves: PSO algorithm fitted results. The upper and lower figures show the fitted results without and with the ghost atom, respectively. The black curves in the lower figure for the ghost atom model represent PSO fitting in short time, i.e., 50 particles, 500 iterations per run, and 50 runs.

fitting parameter set is −10.51 kcal/mol, with the geometric parameters r(O−I) = 3.68 Å and θ(OH_2I) = 161.1°.

The PSO algorithm was also tested for a short time optimization to illustrate its performance. Using 50 particles,

DOI: 10.1021/acs.jctc.7b01122
J. Chem. Theory Comput. 2018, 14, 1321−1332

Table 6. Parameter Sets from PSO Algorithm Fits of Eight $I^-(H_2O)$ Orientations Shown in Figure 8

Without the Ghost Atom: Elapsed Time = 19.45 min, RMSE = 1.37 kcal/mol								
potentials	*A*	*B*	*C*	*D*	*c*	*d*	*m*	*n*
H–I	12430.74	3.5838	−5312.91	0.0	3.1074		3	
O–I	4677.26	2.1062	14468.46	8965.55	4.1427	0.0	3	8

With the Ghost Atom: Elapsed Time = 67.87 min, RMSE = 0.22 kcal/mol									
potentials	*A*	*B*	*C*	*D*	*c*	*d*	*m*	*n*	*z*
H–I	4674.5	2.9220	−15086.8	0.0	3.6184		3		
O–I	26810	2.9712	0	1.9300×10^{-4}		2.1999	9	14	
G–I	7990.2	2.7378	32643.1	0.0	3.8974		3	8	0.28422

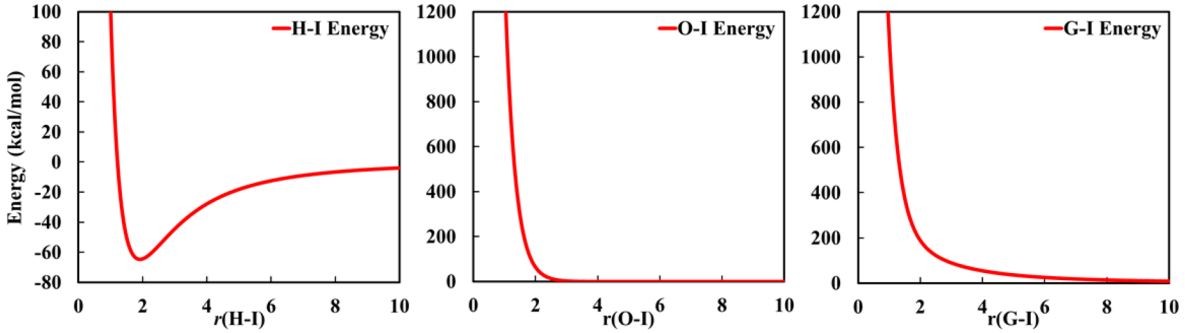


Figure 9. Two-body potentials between H–I, O–I, and G–I as a function of individual interatomic distances, for the refined PSO fitted parameters in Table 6.

Table 7. Parameter Set from Short Time PSO Algorithm Fit with the Ghost Atom, i.e., 50 Particles, 500 Iterations per Run, and 50 Runs

Elapsed Time = 8.03 min, RMSE = 0.37 kcal/mol									
potentials	*A*	*B*	*C*	*D*	*c*	*d*	*m*	*n*	*z*
H–I	1908.7	2.1256	−3593.3	4825.1	1.4326	0.67597	3	7	
O–I	22975	3.1460	8857.8	10626	4.4901	0.088909	4	9	
G–I	14266	3.0477	8072.6	0	1.8325		3		0.24992

500 iterations per run, and 50 runs, a relatively lower level calculation that costs much shorter time of 8.03 min, the fits to the potential energy curves are actually quite good, as shown in Figure 8B with RMSE of 0.37 kcal/mol. The potential energy minimum of −10.61 kcal/mol and its geometric parameter r(O–I) = 3.67 Å and $\theta(OH_2I)$ = 159.6° are also in good agreement with B97-1 calculation results, as shown in Table 2. However, the fitted parameters, given in Table 7, are substantially different from those for the longer time fitting in Table 6.

5. COMPARISON OF THE GENETIC/NONLINEAR LEAST-SQUARES AND PSO ALGORITHM FITS

5.1. Fitting Quality. The PSO algorithm has not been widely used in fitting analytic potential energy functions to electronic structure calculations.[56,57] González et al. used the PSO algorithm to fit an Ar–Ar potential with both a Lennard-Jones and a six-parameter generic potential, and a Cu–Cu potential with the Sutton–Chen potential of the embedded atom model (EAM).[56] Stinson et al.[57] fitted a two-state empirical valence bond (EVB) potential model to sulfuric acid/water DFT calculations with the PSO algorithm. The potential was then used to study proton transfer between sulfuric acid and water molecules.[57] The PSO fitting error was small for each of these studies.

Comparison of the genetic/nonlinear least-squares and PSO algorithm fits to five potential energy curves for $I^-(H_2O)$ are plotted in Figure 5. Comparison of their RMSEs shows the PSO fits are better. The RMSEs given by the PSO fits are 0.87 and 0.27 kcal/mol, respectively, without and with the ghost atom, and those given by genetic/nonlinear least-squares fit are 1.12 and 1.15 kcal/mol. When the shape of the fitted analytic potential energy curves is examined, both fitting algorithms give curves similar to those obtained with the electronic structure calculations. For the I^----H_2O attractive interactions in curves a and d, both algorithms describe the asymptote behavior correctly and predict the potential wells correctly, except in Figure 5A-a where the PSO fitted well depth is 1.9 kcal/mol higher than the electronic structure value. In these curves, the genetic/nonlinear least-squares fits are less attractive and the analytic potential wells narrower.

The I^----H_2O repulsive interactions are represented in curves b, c, and e. The fitted analytic potential energy curves for both algorithms retain the correct shape for these repulsive interactions, except that curve e is fitted poorly without the ghost atom. As the I^----H_2O separation increases in curve b, there is a "kink" in the electronic structure potential energy curve, which is reproduced by the PSO fit, but not by that of the genetic/nonlinear least-squares algorithm. Curve c is only well fit by the PSO algorithm without the ghost atom. The PSO fit worsens with the ghost atom, but as described next, the fit for curve e is dramatically improved. Without the ghost atom, curve e is fitted poorly by PSO, with the fit including a potential energy minimum for the purely repulsive potential energy curve. This is corrected by including the ghost atom. The PSO algorithm improves the fitting, for the ghost atom model, by lowering the

DOI: 10.1021/acs.jctc.7b01122
J. Chem. Theory Comput. 2018, 14, 1321−1332

RMSE to 0.27 kcal/mol from 1.15 kcal/mol given by the genetic/nonlinear least-squares algorithm.

To provide additional constraints on the fitting, the three additional curves f, g, and h in Figure 8 were calculated, with g representing an out-of-plane I^- + H_2O interaction. Only PSO fittings were performed for these curves, with a short time fitting included (black curve) for the model with the ghost atom. The fits without the ghost atom are of poor quality, but with the ghost atom included the fits for all of the curves are quite good, including the short time fitting. With the three additional curves included and for the ghost atom model, RMSE is lowered from 0.27 kcal/mol for the five curves in Figure 5 to 0.22 kcal/mol. The improved fitting is illustrated by comparing curve e in Figures 5 and 8, where the energies for short I^----H_2O separations are fit better in Figure 8 than in Figure 5. The constraints provided by the three additional curves enhanced the fitting.

5.2. Improved Fitting with the Ghost Atom Model. Including the ghost atom is critical for obtaining quality fits to the potential energy curves. For the genetic/nonlinear least-squares algorithm fits in Figure 5, the fits for curves a and e are poor, with the well depth for curve a underestimated by ∼1.9 kcal/mol. These are corrected with the ghost atom model, where the fitted potential minimum curve a has an error less than 0.01 kcal/mol. Such improvements are also seen in the PSO algorithm fitting results in Figures 5 and 8. The PSO RMSEs are 0.27 and 0.22 kcal/mol with the ghost atom for the curves in Figures 5 and 8, respectively, but 0.87 and 1.37 kcal/mol without the ghost atom. Similarly, in curves a, f, and h, the I^----H_2O attractions are underestimated without the ghost atom and corrected with the ghost atom. The I^----H_2O repulsive interaction dominates in configurations e and g for I^- perpendicular to the H_2O plane. Fitting without the ghost atom mistakenly gives an attractive potential minimum in both curves, which is corrected with the ghost atom model.

5.3. Comparison of Fitting Efficiency. Previous works have compared the PSO and genetic algorithms in benchmark fitting problems.[58−60] Ab Wahab et al. showed that the PSO algorithm achieves 18 optimal results among 30 benchmark continuous functions, whereas the genetic algorithm only achieves 14 optimal results with a computational time averaging 1.7 times longer.[58] Hassan et al. examined the effectiveness and efficiency of the two algorithms by the *t*-test. Both algorithms achieved high quality results, but overall the PSO algorithm was less time-consuming.[59] Kachitvichyanukul compared multiple aspects of these two algorithms and showed that (1) updating/evolving populations is simple with the PSO algorithm but computationally costly with the genetic algorithm and (2) the population size has a linear and exponential influence on the PSO and genetic algorithm fittings, respectively.[60]

Performing the genetic/nonlinear least-squares fitting of large search spaces is straightforward, however, time-consuming to obtain an accurate parameter set for an analytic potential. One may perform the fitting in a shorter time, but with less accuracy, by using smaller search spaces. The search space is divided into subspaces, each with a faster fitting to obtain an accurate parameter set. With this strategy, the genetic/nonlinear least-squares fits to the five potential energy curves in Figure 5, without the ghost atom, have a total cost of ∼10 h on an Intel Xeon E5430 @ 2.66 GHz processor. Fitting with the ghost atom consumes an even longer time, because more parameters are included in the optimization. The main reason for the long computation time is that the genetic algorithm needs numerous generations to obtain a converged result and this number increases when a larger search space is used. However, the PSO algorithm proposed here for accurate fitting of potential energy curves is much faster. The fitting was performed on an Intel Xeon X5650 @ 2.67 GHz processor. The CPU time is ∼8 min for fitting the five curves in Figure 5 without the ghost atom and ∼17 min with the ghost atom. For fitting the eight curves in Figure 8, the CPU time is ∼19 and 68 min without and with the ghost atom, respectively. The PSO algorithm fitting improves the computational efficiency by reducing the computation time by nearly a factor of 10 compared to the fit with the genetic/nonlinear least-squares algorithm.

Furthermore, the lower level PSO fitting with 50 particles, 500 iterations per run, and 50 runs, for the ghost atom fit in Figure 8, costs only ∼8 min, and as shown in Table 7 and Figure 8 the quality of the fit is retained. This is very encouraging for using the PSO approach to fit complex analytic potential energy functions. Another advantage of the PSO algorithm is that it is easily parallelized. As seen in the pseudo code of the algorithm, the most expensive part of computation are the function evaluations. However, function evaluations can be carried out in a parallel fashion because they are independent of each other. Parallel implementation of the PSO algorithm is expected to have a near linear reduction in the wall-clock time as the number of cores employed is increased. In the future one may prefer to perform the above lower level PSO fitting with parallel implementation on (1) a larger set of training data and/or (2) an improved H_2O model such as TIP5P[61] for a more accurate fit. In addition, the high efficiency of the PSO algorithm will allow accurate fittings possible of large and complex molecular systems.

6. CONCLUSION

In the work presented here, electronic structure potential energy curves for $I^-(H_2O)$ are fit with the PSO and genetic/nonlinear least-squares algorithms. The fittings are performed for the five curves in Figure 5 and including the three additional curves in Figure 8. The ghost atom model significantly improves the quality of the fit for both algorithms. Using this model, a superior fit to the potential energy curves including their minima is only obtained with the PSO algorithm. Though the PSO fit is more accurate, the PSO algorithm is ∼10 times more computationally efficient than the genetic/nonlinear least-squares algorithm. As shown in Figure 8, the PSO fit to the I^----H_2O intermolecular potential is quite acceptable for calculating densities of states and partition functions of $I^-(H_2O)_n$ clusters at high energies and temperatures, respectively. A quite encouraging aspect of the PSO fittings reported here is that the PSO algorithm should be quite beneficial for future fittings of analytic potential energy functions for large and complex molecular system.

In concluding, fitting potential energy points to an analytic potential energy function is a root searching problem, with a very large number of possible solutions. There is no way to determine if the best possible fit has been found, and the fitting is ended when an acceptable fit has been obtained.[6,9−12] The fitting is often a quite lengthy process. For the current fitting, a sum of two-body potentials is used to represent the I^----H_2O intermolecular potential. Each two-body potential is a generalized sum of Buckingham and Lennard-Jones terms with six parameters. A quite good fit is obtained with the PSO algorithm, but not with the genetic/nonlinear least-squares algorithm we have previously used, even though the latter fitting required a factor of 10 more in time. As for any fitting problem, better fits might be obtained with the algorithms if the fitting was carried out for a longer period of time. Also, better fits could be obtained with the

DOI: 10.1021/acs.jctc.7b01122
J. Chem. Theory Comput. 2018, 14, 1321−1332

genetic/nonlinear least-squares algorithm if a more complex analytic potential was used with more parameters, e.g., including multibody terms. However, with the PSO algorithm a quite good fit was obtained with a relatively simple two-body potential. This work illustrates the importance of the fitting algorithm in fitting analytic potential energy functions.

AUTHOR INFORMATION

ORCID

X. Ma: 0000-0002-0923-8758
A. K. Paul: 0000-0001-7074-1232
W. L. Hase: 0000-0002-0560-5100

Funding

The research reported here was supported by the National Science Foundation under Grant No. CHE-1416428, the Robert A. Welch Foundation under Grant No. D-0005, and the Air Force Office of Scientific Research under AFOSR Award No. FA9550-16-1-0133.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The DFT calculation and genetic/nonlinear least-squares fitting were performed on the Chemdynm computer cluster of the Hase research group. The PSO fitting was performed at the High Performance Computing Center (HPCC) at Texas Tech University (TTU). The authors acknowledge the lecture given by Wolfgang Domcke at the International Bunsen Discussion Meeting on "Nonadiabatic Effects in Chemical Kinetics, Dynamics, and Spectroscopy" held at the Max-Planck-Institut für Biophysikalische Chemie, Göttingen, Germany, November 1−4, 2015. The code used for the PSO algorithm fitting in this work is available on GitHub (https://github.com/humnath5/PSO-Method).

REFERENCES

(1) Marques, J. M. C.; Prudente, F. V.; Pereira, F. B.; Almeida, M. M.; Maniero, A. M.; Fellows, C. E. A New Genetic Algorithm to Be Used in the Direct Fit of Potential Energy Curves to ab Initio and Spectroscopic Data. *J. Phys. B: At., Mol. Opt. Phys.* **2008**, *41*, 085103.

(2) Braams, B. J.; Bowman, J. M. Permutationally Invariant Potential Energy Surface in High Dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577−606.

(3) Bowman, J. M.; Czako, G.; Fu, B. High-dimensional ab initio potential energy surfaces for reaction dynamics calculations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 8094−8111.

(4) Paul, A. K.; Kohale, S. C.; Pratihar, S.; Sun, R.; North, S. W.; Hase, W. L. A Unified Model for Simulating Liquid and Gas Phase Intermolecular Energy Transfer: N_2 + C_6F_6 Collisions. *J. Chem. Phys.* **2014**, *140*, 194103.

(5) Meroueh, O.; Hase, W. L. Dynamics of Energy Transfer in Peptide-Surface Collisions. *J. Am. Chem. Soc.* **2002**, *124*, 1524−1531.

(6) Pratihar, S.; Kohale, S.; Vázquez, S. A.; Hase, W. L. An Intermolecular Potential for Binding of Protonated Peptide Ions with Perfluorinated Hydrocarbon Surfaces. *J. Phys. Chem. B* **2014**, *118*, 5577−5588.

(7) Hase, W. L.; Mrowka, G.; Brudzynski, R. J.; Sloane, C. S. An Analytic Function Describing the H + C_2H_4 → C_2H_5 Potential Energy Surface. *J. Chem. Phys.* **1978**, *69*, 3548−3562. Hase, W. L.; Mrowka, G.; Brudzynski, R. J.; Sloane, C. S. Erratum: An Analytic Function Describing the H + C_2H_4 → C_2H_5 Potential Energy Surface. *J. Chem. Phys.* **1980**, *72*, 6321.

(8) Vande Linde, S. R.; Hase, W. L. A Complete Multi-Dimensional Analytic Potential Energy Surface for Cl^- + CH_3Cl S_N2 Nucleophilic Substitution. *J. Phys. Chem.* **1990**, *94*, 2778−2788.

(9) Opalka, D.; Domcke, W. High-order Expansion of $T_2 \times t_2$ Jahn−Teller Potential-Energy Surfaces in Tetrahedral Molecules. *J. Chem. Phys.* **2010**, *132*, 154108.

(10) Bhattacharyya, S.; Opalka, D.; Poluyanov, L. V.; Domcke, W. The $(E + A) \times (e + a)$ Jahn−Teller and Pseudo-Jahn−Teller Hamiltonian Including Spin−Orbit Coupling for Trigonal Systems. *J. Phys. Chem. A* **2014**, *118*, 11962−11970.

(11) Janke, S. M.; Auerbach, D. J.; Wodtke, A. M.; Kandratsenka, A. An Accurate Full-Dimensional Potential Energy Surface for H-Au(111): Importance of Nonadiabatic Electronic Excitation in Energy Transfer and Adsorption. *J. Chem. Phys.* **2015**, *143*, 124708.

(12) Kammler, M.; Janke, S. M.; Kandratsenka, A.; Wodtke, A. M. Genetic Algorithm Approach to Global Optimization of the Full-Dimensional Potential Energy Surface for Hydrogen Atom at fcc-Metal Surfaces. *Chem. Phys. Lett.* **2017**, *683*, 286−290.

(13) Farantos, S. C.; Murrell, J. N.; Hajduk, J. C. Monte Carlo Calculations of Classical Density of States for Non-Separable Polyatomic Potential Energy Surfaces. *Chem. Phys.* **1982**, *68*, 109−117.

(14) Hase, W. L. Classical Number and Density of States. *J. Chem. Educ.* **1983**, *60*, 379−381.

(15) Weerasinghe, S.; Amar, F. G. Absolute Classical Densities of States for Very Anharmonic Systems and Applications to the Evaporation of Rare Gas Clusters. *J. Chem. Phys.* **1993**, *98*, 4967−4982.

(16) Peslherbe, G. H.; Hase, W. L. Statistical Anharmonic Unimolecular Rate Constants for the Dissociation of Fluxional Molecules: Application to Aluminum Clusters. *J. Chem. Phys.* **1996**, *105*, 7432−7447.

(17) Kolakkandy, S.; Paul, A. K.; Pratihar, S.; Kohale, S. C.; Barnes, G. L.; Wang, H.; Hase, W. L. Energy and Temperature Dependent Dissociation of $Na^+(Benzene)_{1,2}$ Clusters. Importance of Anharmonicity. *J. Chem. Phys.* **2015**, *142*, 044306.

(18) Hertz, P. Über die Mechanischen Grundlagen der Thermodynamik. *Ann. Phys. (Berlin, Ger.)* **1910**, *338*, 537−552.

(19) Reinhardt, W. P. Calculation of the VIB-Rotational State Density of Polyatomic Molecules by Adiabatic Switching. *J. Mol. Struct.* **1990**, *223*, 157−170.

(20) 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.

(21) Sheu, W. S.; Rossky, P. J. Charge-Transfer-to-Solvent Spectra of an Aqueous Halide Revisited via Computer Simulation. *J. Am. Chem. Soc.* **1993**, *115*, 7729−7735.

(22) Dang, L. X.; Garrett, B. C. Photoelectron Spectra of the Hydrated Iodine Anion from Molecular Dynamics Simulations. *J. Chem. Phys.* **1993**, *99*, 2972−2977.

(23) Peslherbe, G. H.; Ladanyi, B. M.; Hynes, J. T. Free Energetics of NaI Contact and Solvent-Separated Ion Pairs in Water Clusters. *J. Phys. Chem. A* **2000**, *104*, 4533−4548.

(24) Dang, L. X. Computational Study of Ion Binding to the Liquid Interface of Water. *J. Phys. Chem. B* **2002**, *106*, 10388−10394.

(25) Ayala, R.; Martınez, J. M.; Pappalardo, R. R.; Sánchez Marcos, E. On the Halide Hydration Study: Development of First-Principles Halide Ion-Water Interaction Potential Based on a Polarizable Model. *J. Chem. Phys.* **2003**, *119*, 9538−9548.

(26) Lamoureux, G.; Roux, B. Absolute Hydration Free Energy Scale for Alkali and Halide Ions Established from Simulations with a Polarizable Force Field. *J. Phys. Chem. B* **2006**, *110*, 3308−3322.

(27) Jensen, K. P.; Jorgensen, W. L. Halide, Ammonium, and Alkali Metal Ion Parameters for Modeling Aqueous Solutions. *J. Chem. Theory Comput.* **2006**, *2*, 1499−1509.

(28) Wick, C. D.; Kuo, I. F. W.; Mundy, C. J.; Dang, L. X. The Effect of Polarizability for Understanding the Molecular Structure of Aqueous Interfaces. *J. Chem. Theory Comput.* **2007**, *3*, 2002−2010.

(29) Wick, C. D.; Xantheas, S. S. Computational Investigation of the First Solvation Shell Structure of Interfacial and Bulk Aqueous Chloride and Iodide Ions. *J. Phys. Chem. B* **2009**, *113*, 4141−4146.

DOI: 10.1021/acs.jctc.7b01122
J. Chem. Theory Comput. 2018, 14, 1321−1332

(30) Wick, C. D. Electrostatic Dampening Dampens the Anion Propensity for the Air-Water Interface. *J. Chem. Phys.* **2009**, *131*, 084715.
(31) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D., Jr; Roux, B. Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 774−786.
(32) Werhahn, J. C.; Akase, D.; Xantheas, S. S. Universal Scaling of Potential Energy Functions Describing Intermolecular Interactions. II. The Halide-Water and Alkali Metal-Water Interactions. *J. Chem. Phys.* **2014**, *141*, 064118.
(33) Kiss, P. T.; Baranyai, A. A New Polarizable Force Field for Alkali and Halide Ions. *J. Chem. Phys.* **2014**, *141*, 114501.
(34) Arismendi-Arrieta, D. J.; Riera, M.; Bajaj, P.; Prosmiti, R.; Paesani, F. i-TTM Model for Ab Initio-Based Ion−Water Interaction Potentials. 1. Halide−Water Potential Energy Functions. *J. Phys. Chem. B* **2016**, *120*, 1822−1832.
(35) Kennedy, J.; Eberhart, R. Particle Swarm Optimization. *Proceedings of IEEE International Conference on Neural Networks, Vol. 4*, Perth, Australia, November 27−December 1, 1995; IEEE Press: Piscataway, NJ, 1995; pp1942−1948.
(36) Shi, Y.; Eberhart, R. C. A Modified Particle Swarm Optimizer. *Proceedings of the IEEE International Conference on Evolutionary Computation*, Anchorage, AK, USA, May 4−9, 1998; IEEE Press: Piscataway, NJ, 1998; pp 69−73.
(37) van den Bergh, F. An Analysis of Particle Swarm Optimizers. *Ph.D. Dissertation*, University of Pretoria, South Africa, 2002.
(38) Poli, R.; Kennedy, J.; Blackwell, T. Particle Swarm Optimization: An Overview. *Swarm Intell* **2007**, *1*, 33−57.
(39) Consider $f(\hat{\mathbf{y}}(t))$ and $f(\hat{\mathbf{y}}(t+T))$ as fitness values for global best positions at time t and $t + T$, respectively, for some integer $T > 0$. Then the algorithm is terminated from criterion (b) if $\left|\frac{f(\hat{\mathbf{y}}(t)) - f(\hat{\mathbf{y}}(t+T))}{f(\hat{\mathbf{y}}(t)) + \epsilon}\right| < \epsilon$ for some small $\epsilon > 0$.
(40) Bazarra, M. S.; Shetty, C. M. *Nonlinear Programming, Theory and Algorithms*; John Wiley & Sons: New York, 1979.
(41) Brent, R. P. *Algorithms for Minimization without Derivatives*; Prentice-Hall: Englewood Cliffs, NJ, 1973.
(42) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.
(43) Becke, A. D. Density-Functional Thermochemistry. V. Systematic Optimization of Exchange-Correlation Functionals. *J. Chem. Phys.* **1997**, *107*, 8554−8560.
(44) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. Development and Assessment of New Exchange-Correlation Functionals. *J. Chem. Phys.* **1998**, *109*, 6264−6271.
(45) Zhang, J.; Hase, W. L. Electronic Structure Theory Study of the $F^- + CH_3I \rightarrow FCH_3 + I^-$ Potential Energy Surface. *J. Phys. Chem. A* **2010**, *114*, 9635−9643.
(46) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. NWChem: A Comprehensive and Scalable Open-Source Solution for Large Scale Molecular Simulations. *Comput. Phys. Commun.* **2010**, *181*, 1477−1489.
(47) Bässmann, C.; Boesl, U.; Yang, D.; Drechsler, G.; Schlag, E. W. Mass Selective Anion-ZEKE Spectroscopy of the Iodine-Water Cluster. *Int. J. Mass Spectrom. Ion Processes* **1996**, *159*, 153−167.
(48) Schlicht, F.; Entfellner, M.; Boesl, U. Anion ZEKE-Spectroscopy of the Weakly Bound Iodine Water Complex. *J. Phys. Chem. A* **2010**, *114*, 11125−11132.
(49) Alberti, M. Rare Gas-Benzene-Rare Gas Interactions: Structural Properties and Dynamics Behavior. *J. Phys. Chem. A* **2010**, *114*, 2266−2274.
(50) Ulusoy, I. S.; Andrienko, D. A.; Boyd, E. D.; Hernandez, R. Quantum and Quasi-Classical Collisional Dynamics for O_2-Ar at High Temperatures. *J. Chem. Phys.* **2016**, *144*, 234311.
(51) Kistenmacher, H.; Lie, G. C.; Popkie, H.; Clementi, E. Study of the Structure of Molecular Complexes. VI. Dimers and Small Clusters of Water Molecules in the Hartree-Fock Approximation. *J. Chem. Phys.* **1974**, *61*, 546−561.
(52) Eberhart, R. C.; Shi, Y. Comparing Inertia Weights and Constriction Factors in Particle Swarm Optimization. *Proceedings of the IEEE Congress on Evolutionary, Computation, Vol. 1*, July 16−19, 2000, La Jolla, CA, USA; IEEE Press: Piscataway, NJ, 2000; pp 84−88.
(53) van den Bergh, F.; Engelbrecht, A. P. A Study of Particle Swarm Optimization Particle Trajectories. *Inf. Sci.* **2006**, *176*, 937−971.
(54) Poli, R. Mean and Variance of the Sampling Distribution of Particle Swarm Optimizers During Stagnation. *IEEE Trans. on Evol. Comput.* **2009**, *13*, 712−721.
(55) Trelea, I. C. The Particles Swarm optimization algorithm: Convergence analysis and parameter selection. *Inform. Proc. Lett.* **2003**, *85*, 317−325.
(56) González, D.; Davis, S. Fitting of Interatomic Potentials without Forces: A Parallel Particle Swarm Optimization Algorithm. *Comput. Phys. Commun.* **2014**, *185*, 3090−3093.
(57) Stinson, J. L.; Kathmann, S. M.; Ford, I. J. A Classical Reactive Potential for Molecular Clusters of Sulphuric Acid and Water. *Mol. Phys.* **2016**, *114*, 172−185.
(58) Ab Wahab, M. N.; Nefti-Meziani, S.; Atyabi, A. A Comprehensive Review of Swarm Optimization Algorithms. *PLoS One* **2015**, *10*, e0122827.
(59) Hassan, R.; Cohanim, B.; de Weck, O.; Venter, G. A Comparison of Particle Swarm Optimization and the Genetic Algorithm. *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Structures, Structural Dynamics, and Materials and Co-located Conferences*, Austin, Texas, USA, April 18−21, 2005; AIAA: Reston, VA, 2015.
(60) Kachitvichyanukul, V. Comparison of Three Evolutionary Algorithms: GA, PSO, and DE. *Ind. Eng. Manag. Syst.* **2012**, *11*, 215.
(61) Mahoney, M. W.; Jorgensen, W. L. A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions. *J. Chem. Phys.* **2000**, *112*, 8910−8922.

DOI: 10.1021/acs.jctc.7b01122
J. Chem. Theory Comput. 2018, 14, 1321−1332