

# Hydration Shell Structure and Dynamics of Curium(III) in Aqueous Solution: First Principles and Empirical Studies

## Supporting Information

Raymond Atta-Fynn<sup>1</sup>, Eric J. Bylaska<sup>1</sup>, Gregory K. Schenter<sup>2</sup>, and Wibe A. DeJong<sup>1</sup>

<sup>1</sup> W.R. Wiley Environmental Molecular Sciences Laboratory and <sup>2</sup> Chemical and Material Sciences Division, Pacific Northwest National Laboratory, P. O. Box 999, Richland, WA 99352

### 1. Computational Methodology: Analytic model potentials for classical simulations

#### 1.1 Curium(III)-water potential derived from *ab initio* ROHF data (CMD-3B simulation)

As a precursor to the development of pair  $\text{Cm}^{3+}\text{-H}_2\text{O}$  and three-body  $\text{H}_2\text{O-Cm}^{3+}\text{-H}_2\text{O}$  analytic potentials from *ab initio* restricted open-shell Hartree-Fock (ROHF) data, we studied the dissociation potential energy curves for  $[\text{Cm}(\text{OH}_2)_n]^{3+}(\text{g}) \rightarrow [\text{Cm}(\text{OH}_2)_{n-1}]^{3+}(\text{g}) + \text{OH}_2(\text{g})$ . Plane-wave DFT calculations and two Gaussian-type basis set calculations were carried out using the NWChem computation chemistry package <sup>1</sup>.

The pseudopotential plane wave DFT equations were solved in a large *aperiodic* box using the gradient corrected PBE functional<sup>2</sup> and the pseudopotentials and energy cut-off described in the computational methodology section of manuscript. Gaussian-type basis set DFT and ROHF calculations were carried out with aug-cc-pVTZ<sup>3</sup> basis set for O and H and the Stuttgart RSC 1997 relativistic effective core potential (ECP) and the associated (12s11p10d8f)/(8s7p6d4f) basis set <sup>4</sup> was used for Cm. In the ECP, 60 electrons were treated as core states and the remaining 36 electrons were treated as valence states. The hybrid B3LYP functional<sup>5</sup> was employed for the DFT calculations. Localized basis set calculations were corrected for the basis set superposition error using the counterpoise method by Boys and Bernadi.<sup>6</sup>

In Figure S1, we depict the potential energy curves for  $n=1-3$  as a function of the  $\text{Cm}^{3+}\text{-O}$  bond distance. The curves were obtained by radially detaching a water molecule from an optimized  $[\text{Cm}(\text{OH}_2)_n]^{3+}$  along the  $\text{Cm}^{3+}\text{-O}$  bond. Clearly from the Figure, it can be observed that the DFT curves dissociate to a lower energy state whereas the ROHF dissociates to a high energy state. For  $n=1$  in the DFT calculations, an electron is transferred from water to  $\text{Cm}^{3+}$  and the resulting fully dissociated species are  $\text{Cm}^{2+}$  and  $\text{H}_2\text{O}^+$ . According to our studies, as the

number of water molecules around Cm,  $n$ , increases, the charge transfer from the detached water to  $\text{Cm}^{3+}$  slowly decreases until there is almost no charge transfer when  $n=6$ . The charge transfer destroys the spin state of Cm due partial population of the half-empty  $f$  orbitals. On the other hand, the Figure indicates that the dissociated species in ROHF calculations for each  $n$  is  $\text{Cm}^{3+}$  and  $\text{H}_2\text{O}$ . The implications of these results in relation to the empirical development of the  $\text{Cm}^{3+}$ - $\text{H}_2\text{O}$  interaction can be summarized as follows: (i) Using ROHF, pure pair potentials can be constructed from the potential energy surface for  $[\text{Cm}(\text{OH}_2)]^3$  and approximate three-body  $\text{H}_2\text{O}$ - $\text{Cm}^{3+}$ - $\text{H}_2\text{O}$  interaction can also be constructed from the potential energy surface of  $[\text{Cm}(\text{OH}_2)_2]^{3+}$  (ii) Using DFT, effective pair and three-body potentials can be constructed from potential energy surfaces of  $[\text{Cm}(\text{OH}_2)_n]^{3+}$  clusters, where  $n \geq 6$ . We adopted the ROHF approach to construct the interaction of  $\text{Cm}^{3+}$  with water because the degrees of freedom involved in the fitting procedure are considerably less than the DFT case and therefore it provides a balance between computational cost and efficiency. We must mention that the gas phase dissociation of  $\text{Cm}(\text{OH}_2)^{3+}$  to  $\text{Cm}^{2+}$  and  $\text{H}_2\text{O}^+$  in the DFT calculations is correct in principle but not inconvenient for the construction of analytic potentials. We must also mention that the DFT gas phase dissociations is common in highly charge metal ions (with charges of +3 or higher). Specifically, this has been observed for  $\text{Cm}^{3+}$ <sup>7</sup> at the CASPT2 level of theory and for  $\text{Al}^{3+}$  at the CASSCF<sup>8</sup> and the Hartree-Fock levels of theory.<sup>9</sup> According to Hagberg *et al.*<sup>7</sup> the problem of  $\text{Cm}^{3+}$ -water dissociation disappears for  $n=2$  onward but pointed out earlier, our work indicates that the problem disappears from  $n=6$  onward. We wish to point out that the ROHF dissociation energies may be inaccurate compared to correlated methods, however our focus is not on accurate energetics but rather on obtaining a potential energy surface that approximately captures the long range dissociation behavior. In that respect, our use of ROHF theory is justified.

Following the scheme developed by Inada and co-workers.<sup>10</sup>, the pair potential energy surface for interaction  $\text{Cm}^{3+}$  with a single water molecule was constructed using the ROHF *ab initio* data. Throughout the calculations, water molecule fixed at the SPC/E geometry<sup>17</sup> [ $r(\text{O}-\text{H})=1 \text{ \AA}$ ,  $\theta(\text{H}-\text{O}-\text{H})=109.47^\circ$ ]; see section 1.2 for a description of the SPC/E model for water]. The approach is outlined as follows.

(i) The O part of the water molecule was fixed at the origin with the the H atoms pointing down in the  $xz$  plane and the  $C_{2v}$  axis of the molecule point down along the negative  $z$ -axis.

(ii) The spherical coordinates  $(r_{\text{CmO}}, \varphi, \theta)$  of  $\text{Cm}^{3+}$  relative to the origin, where  $r_{\text{CmO}}$  is the  $\text{Cm}^{3+}$ –O distance,  $\varphi$  is the angle the between the projection of  $\mathbf{r}_{\text{CmO}}$  onto the  $xy$  plane and the positive  $x$ -axis, and  $\theta$  is the angle between  $\mathbf{r}_{\text{CmO}}$  and the positive  $z$ -axis were generated by varying  $r_{\text{CmO}}$  in the interval  $[1.8 \text{ \AA}, 8.0 \text{ \AA}]$ ,  $\theta$  in the interval  $[0^\circ, 180^\circ]$ , and  $\varphi$  in the interval  $[0^\circ, 180^\circ]$  (it should be noted that because symmetry the potential energy surface in the regions  $[-180^\circ, 0^\circ]$  and  $[0^\circ, 180^\circ]$  are equivalent). In all 57 radial points and 112 angular grid points—resulting in a total of 6384 points—were used to construct the pair potential.

(ii) For each point, the two-body interaction,  $E^{(2)}$  with counterpoise BSSE corrections to the *ab initio* energies using the full basis set, was computed:

$$E^{(2)} = E(\text{Cm}^{3+}\text{--H}_2\text{O}) - E(\text{Cm}^{3+}) - E(\text{H}_2\text{O}) \quad (1)$$

$E^{(2)}$  was then expressed as a sum of long range Coulomb interaction between the  $\text{Cm}^{3+}$  and  $\text{H}_2\text{O}$  and a Buckingham-type short range dispersion and repulsion . Specifically

$$E^{(2)} = \sum_{\alpha=\text{O,H1,H2}} \frac{q_{\text{Cm}}q_{\alpha}}{r_{\text{Cm}\alpha}} + \sum_{\alpha=\text{O,H1,H2}} \left[ A_{\text{Cm}\alpha} \exp(-B_{\text{Cm}\alpha} r_{\text{Cm}\alpha}) + \frac{C_{\text{Cm}\alpha}}{r_{\text{Cm}\alpha}^4} + \frac{D_{\text{Cm}\alpha}}{r_{\text{Cm}\alpha}^6} \right] \quad (2)$$

where  $r_{\text{CmO}}, r_{\text{CmH1}}, r_{\text{CmH2}}$  are the  $\text{Cm}^{3+}$ –O,  $\text{Cm}^{3+}$ –H1, and  $\text{Cm}^{3+}$ –H2 separations respectively and  $q_{\text{Cm}} = +3$  is the charge on Cm and the atomic  $q_{\text{O}}$  and  $q_{\text{H}}$  on O and H respectively are the SPC/E charges (Table S1). A nonlinear least square fit was performed to determine constants  $A_{\text{Cm}\alpha}$ ,  $B_{\text{Cm}\alpha}$ ,  $C_{\text{Cm}\alpha}$ , and  $D_{\text{Cm}\alpha}$ . The values of the constants obtained from the fit are reported in Table II in the main text.

Purely pair-wise additive *ab initio*-derived potentials for highly charged hydrated ions are usually too attractive and therefore overestimates the shell hydrations numbers and underestimates the first shell ion-ligand distance because of the absence of non-additive repulsive terms.<sup>11</sup> To obtain a realistic  $\text{Cm}^{3+}$ – $\text{H}_2\text{O}$  interaction, at least a three-body term should be included. We thus constructed the three body term as follows<sup>10</sup>.

(i) Starting with an optimized  $[\text{Cm}(\text{OH}_2)_2]^{3+}$  structure, the geometry of each water molecule was converted into the SPC/E geometry by changing the equilibrium H–O–H angle symmetrically along the  $\text{Cm}^{3+}$ –O bond and radially stretching the O–H bond. The water molecules are labeled  $\text{H}_2\text{O}(\text{I})$  and  $\text{H}_2\text{O}(\text{II})$  for convenience

(ii) The O(I)–Cm<sup>3+</sup>–O(II) angle was uniformly varied from 60° to 180° and for each angular variation, the Cm<sup>3+</sup>–O(I) and Cm<sup>3+</sup>–O(II) distances are varied between 2.2 Å and 7.0 Å. For a given [Cm(OH<sub>2</sub>)<sub>2</sub>]<sup>3+</sup> configuration, six fragments are required to compute the non-additive three-body energy: [Cm(OH<sub>2</sub>)(I)]<sup>3+</sup>, [Cm(OH<sub>2</sub>)(II)]<sup>3+</sup>, Cm<sup>3+</sup>, H<sub>2</sub>O(I), H<sub>2</sub>O(II), and H<sub>2</sub>O(I)–H<sub>2</sub>O(II), where I and II are used to label the two water molecules. The total energy of [Cm(OH<sub>2</sub>)<sub>2</sub>]<sup>3+</sup> and the first five fragments were computed at the *ab initio* ROHF level and the energy of the last fragment, namely, H<sub>2</sub>O(I)–H<sub>2</sub>O(II), was computed using the SPC/E intermolecular potential (see Equation 8). Each *ab initio* total energy calculation was BSSE-corrected using the full basis set. The non-additive three-body energy, defined as the total interaction energy minus the sum of the pairwise additive interaction energies, was computed as:

$$E^{(3)} = E_{\text{int}}^{\text{tot}} - E_{\text{Cm}^{3+}-\text{H}_2\text{O(I)}}^{\text{tot}} - E_{\text{Cm}^{3+}-\text{H}_2\text{O(II)}}^{\text{tot}} - E_{\text{H}_2\text{O(I)}-\text{H}_2\text{O(II)}}^{\text{LJ}} \quad (3)$$

where the total *ab initio* interaction energy is

$$E_{\text{int}}^{\text{tot}} = E([\text{Cm}(\text{OH}_2)_2]^{3+}) - E(\text{Cm}^{3+}) - E(\text{H}_2\text{O(I)}) - E(\text{H}_2\text{O(II)}) \quad (4)$$

and the *ab initio* pair interaction energies are

$$E_{\text{Cm}^{3+}-\text{H}_2\text{O(I)}}^{\text{tot}} = E([\text{Cm}(\text{OH}_2)(\text{I})]^{3+}) - E(\text{Cm}^{3+}) - E(\text{H}_2\text{O(I)}) \quad (5)$$

$$E_{\text{Cm}^{3+}-\text{H}_2\text{O(II)}}^{\text{tot}} = E([\text{Cm}(\text{OH}_2)(\text{II})]^{3+}) - E(\text{Cm}^{3+}) - E(\text{H}_2\text{O(II)}) \quad (6)$$

and the last term  $E_{\text{H}_2\text{O(I)}-\text{H}_2\text{O(II)}}^{\text{LJ}}$  is the SPC/E intermolecular water interaction energy.  $E^{(3)}$  was computed using over 6400 [Cm(OH<sub>2</sub>)<sub>2</sub>]<sup>3+</sup> configurations. With the exception of 31 energies which were slightly attractive, all the energies were repulsive. Approximating the H<sub>2</sub>O–Cm<sup>3+</sup>–OH<sub>2</sub> interaction by an O–Cm<sup>3+</sup>–O interaction, the three-body repulsive energies were fitted according to<sup>12 13</sup>

$$E^{(3)} = \alpha \exp(-\beta r_1 - \beta r_2 - \gamma r_3) \quad (7)$$

where  $r_1$ ,  $r_2$ , and  $r_3$  are the Cm<sup>3+</sup>–O(I), Cm<sup>3+</sup>–O(II), and O(I)–O(II) respectively and  $\alpha$ ,  $\beta$ , and  $\gamma$  are fitting parameters. Ideally, the three-body should be short-ranged, that is, it should decay rapidly to zero beyond the first shell. However, we observed that the fitted repulsion persists in the second shell. This stems mainly from the use of the Lennard-Jones energy for the H<sub>2</sub>O(I)–H<sub>2</sub>O(II) interaction as opposed to using the *ab initio* SCF energies. If such a repulsive potential is used in a simulation in aqueous phase, the effects of thousands of such long-ranged three-body

repulsions cumulate into a snowball effect resulting in significant repulsion in the first and second solvation shell.<sup>13</sup> To alleviate this problem, the leading term  $\alpha$  was further adjusted to ensure that the interaction energy vanishes midway to the end of the second shell. The values parameters are reported in Table S1.

## **1.2 Empirical Lennard-Jones Potential for $\text{Cm}^{3+}$ -water interaction (CMD-LJ simulation)**

Recently, an approach has been developed whereby the Lennard-Jones parameters  $\epsilon$  and  $\sigma$  of metals ions are continually adjusted with respected model potentials for water (e.g. SPC/E and TIP3P) to yield realistic structural and dynamical properties in aqueous solution. This approach has been used to successfully simulated  $\text{Al}^{3+}$ <sup>14</sup>, and  $\text{Cd}^{2+}$  and  $\text{Pb}^{2+}$ <sup>15</sup> in aqueous solution. To mimick this approach, we thus adjusted Lennard Jones parameters  $\epsilon$  and  $\sigma$  for  $\text{Cm}^{3+}$ , starting with Unified Force Field (UFF) parameters due to Rappe and co-workers<sup>16</sup>. The adjustment was carried out with respect to the SPC/E intermolecular potentials to approximately yield the (i) experimental average first shell  $\text{Cm}^{3+}$ -O distance, (ii) experimental first shell coordination number, and (iii) average experimental second shell  $\text{Cm}^{3+}$ -O distance. The modification was deemed necessary because when the values of  $\epsilon$  and  $\sigma$  for  $\text{Cm}^{3+}$  by Rappe and co-workers<sup>16</sup> were used in conjunction with the variants of the SPC intermolecular in classical MD simulations, the average  $\text{Cm}^{3+}$ -O distances in the first shell is and second shell are shorter than *ab initio* and experimental values. Also, first shell coordination number differs from the experimental value. An important point worth mentioning is that in principle, the values of  $\epsilon$  and  $\sigma$  must be continually adjusted (e.g. via an inverse Monte Carlo scheme) until all simulated data and the corresponding experimental data (structural and thermodynamic) are consistent with each other. The caveat here in our work is that only *a priori* structural experimental data was used to adjust  $\epsilon$  and  $\sigma$ . The lack of inclusion of thermodynamic data obviously puts the reliability of the LJ potential in question. This is due to the fact that no thermodynamic data for  $\text{Cm}^{3+}$  in aqueous solution was found in literature. Therefore the quality/reliability of the potential will be based on how well it predicts properties like internal structural arrangement of the first shell and the experimental self-diffusion coefficient of  $\text{Cm}^{3+}$ . The modified values  $\epsilon$  and  $\sigma$  for  $\text{Cm}^{3+}$  are presented in Table S1 in the manuscript.

### 1.3 Water-water interaction potential for classical molecular dynamics simulations

The non-polarizable rigid body *extended simple point charge* (SPC/E) potential model<sup>17</sup> was used to describe the intermolecular water interaction in all classical molecular dynamics (CMD) simulations. The SPC/E model is one of the most widely used potentials to describe liquid water and it accurately predicts various properties water over wide range of temperatures and pressure quite well. The explicit analytic form for the intermolecular pair potential  $U_{\text{inter}}$  for the SPC/E potential is simply a long-ranged Coulomb electrostatic potential  $U_{\text{ele}}$  plus a Lennard-Jones<sup>18</sup> term  $U_{\text{LJ}}$  given by

$$U_{\text{inter}} = \sum_{i=1}^{M-1} \sum_{j=i+1}^M \sum_{\mathbf{L}} \sum_{a_i=\text{O,H}_1,\text{H}_2} \sum_{b_j=\text{O,H}_1,\text{H}_2} \left( U_{a_i b_j}^{\text{LJ}} + U_{a_i b_j}^{\text{ele}} \right) \quad (8)$$

where  $U_{a_i b_j}^{\text{ele}} = \frac{q_{a_i} q_{b_j}}{r_{a_i b_j}}$  and  $U_{a_i b_j}^{\text{LJ}} = 4\epsilon_{a_i b_j} \left( \left( \frac{\sigma_{a_i b_j}}{r_{a_i b_j}} \right)^{12} - \left( \frac{\sigma_{a_i b_j}}{r_{a_i b_j}} \right)^6 \right)$ .  $M$  is the total number of water molecules,  $q_{a_i}$  is the partial charge on atom  $a$  in molecule  $i$ ,  $\mathbf{L}$  is the lattice vectors, and  $r_{a_i b_j} = |\mathbf{r}_{a_i} - \mathbf{r}_{b_j} + \mathbf{L}|$ . The geometric properties, partial charges on the H and O atoms as well as the parameters for the Lennard-Jones interaction is between the oxygen centers are listed in Table S1. (the O-H and H-H Lennard-Jones interaction is zero).

### 1.4 Water-water interaction potential for the MM region of the QM/MM MD simulations

A modification of the rigid simple point charge SPC<sup>19</sup> pair potential for water to include intramolecular interactions by Toukan and Rahman<sup>20</sup> was also employed for the classical MM region in all QM/MM simulations. This potential was employed because it produces a realistic intramolecular vibrational spectrum of water. It must be noted that the intermolecular form for SPC potential is the same as that for the SPC/E potential given in equation (8) above but the O and H partial charges are different for the two models (see Table S1). The Toukan-Rahman water intramolecular interaction potential,  $U_{\text{intra}}$ , is given by

$$U_{\text{inter}} = \sum_{i=1}^M (U_{\text{OH}}^i + U_{\text{HOH}}^i) \quad (9)$$

where the first term in the summation  $U_{\text{OH}}^i$  is the OH bond stretching Morse potential

$$U_{\text{OH}}^i = D_{\text{OH}} \left[ \left( 1 - \exp(-\rho \Delta r_{12}^i) \right)^2 + \left( 1 - \exp(-\rho \Delta r_{13}^i) \right)^2 \right] \quad (10)$$

and the second term in the summation  $U_{\text{HOH}}^i$  is the harmonic HOH bond bending potential

$$U_{\text{HOH}}^i = k_{rr} \Delta r_{12}^i \Delta r_{13}^i + k_{r\theta} \Delta r_{23}^i (\Delta r_{12}^i + \Delta r_{13}^i) + \frac{1}{2} k_{\theta\theta} (\Delta r_{12}^i)^2 \quad (11)$$

The subscript 1 labels O and 2,3 label H,  $\Delta r_{ab}^i = |\mathbf{r}_{a_i} - \mathbf{r}_{b_i} + \mathbf{L}| - r_{ab}^{eq}$  is the displacement of the bond distance  $r_{ab}$  from its equilibrium value  $r_{ab}^{eq}$ . The values of the parameters  $D_{\text{OH}}$ ,  $r_{ab}^{eq}$ ,  $\rho$ ,  $k_{rr}$ ,  $k_{r\theta}$ , and  $k_{\theta\theta}$  are listed in Table S1.

## 2. Computation of mean residence time

We computed MRT using the direct approach<sup>21</sup> and the probabilistic approach<sup>22</sup>. In the direct approach<sup>21</sup>, the MRT  $\tau_D^{t*}$  of a ligand in a specific coordination shell is defined as

$$\tau_D^{t*} = \frac{n_{CN} t_{sim}}{N_{ev}(t^*)} \quad (12)$$

where  $n_{CN}$  is the average coordination number in the shell,  $t_{sim}$  is the total simulation time,  $N_{ev}(t^*)$  is the total number events during the entire simulation, and  $t^*$  is the *persistence time* defined earlier. Thus in the definition for the direct MRT, only events lasting longer than  $t^*$  contribute to  $N_{ev}(t^*)$ . The probabilistic approach<sup>22</sup> for computing the MRT utilizes the concept of probability survival function  $p_i(t_0, t + t_0; t^*)$ .  $p_i(t_0, t + t_0; t^*) = 1$  if water molecule  $i$  is located in the coordination shell at both times  $t_0$  and  $t + t_0$  and within the time interval  $[t_0, t + t_0]$  does not leave the shell for a period longer than  $t^*$ , otherwise  $p_i(t_0, t + t_0; t^*) = 0$ . The survival function is used to compute of a time correlation function  $C(t)$  defined by

$$C(t) = \left\langle \frac{1}{N_C} \sum_{i=1}^{N_w} p_i(t_0, t + t_0; t^*) \right\rangle_{t_0} \quad (13)$$

where  $N_w$  is the total number of water molecules,  $N_C$  is the average coordination number of the shell and  $\langle \cdot \rangle_{t_0}$  denotes average over all time origins  $t_0$ . It is easy to see that  $C(0) = 1$ . Excluding the initial period during  $C(t)$  decays rapidly,  $\tau_i^{t*}$  is estimated by fitting  $C(t)$  in equation (13) to an exponential decay function

$$C(t) = \exp\left(-\frac{t}{\tau_I^{t^*}}\right) \quad (14)$$

Fitting equation (13) to (14) in short time simulations could lead to poor estimates for  $\tau_I^{t^*}$  due to the lack of sufficient statistics. Thus  $\tau_I^{t^*}$  was computed for only the classical molecular dynamics simulation. We must mention that the function of the parameter  $t^*$  in both the direct and probability survival function estimation of MRT is prevent short-lived visitations to and departures from a shell from being counted as real events.



Table S1: Partial O and H atomic charges and values of the parameters for the rigid body SPC/E water potential, and the flexible intramolecular and intermolecular SPC-Toukan-Rahman (SPC-TR) water potential.

Intermolecular parameters							
	$r_{\text{OH}}$ (Å)	$\angle_{\text{H-O-H}}$ (rad)	$q_{\text{O}}$ (e)	$q_{\text{H}}$ (e)	$\varepsilon_{\text{OO}}$ (kcal/mol)	$\sigma_{\text{OO}}$ (Å)	
SPC/E	1.0	arcsin(-1/3)	-0.8476	0.4238	0.1554	3.1655	
SPC-TR	1.0	arcsin(-1/3)	-0.82	0.41	0.1554	3.1655	
Intramolecular parameters							
	$D_{\text{OH}}$ (kcal/mol)	$\rho$ (Å <sup>-1</sup> )	$r_{\text{OH}}^{\text{eq}}$ (Å)	$r_{\text{HH}}^{\text{eq}}$ (Å)	$k_{rr}$ (kcal/mol/Å <sup>2</sup> )	$k_{r\theta}$ (kcal/mol/Å <sup>2</sup> )	$k_{\theta\theta}$ (kcal/mol/Å <sup>2</sup> )
SPC-TR	101.904	2.511	1.0	1.633	111.69	-211.44	328.60

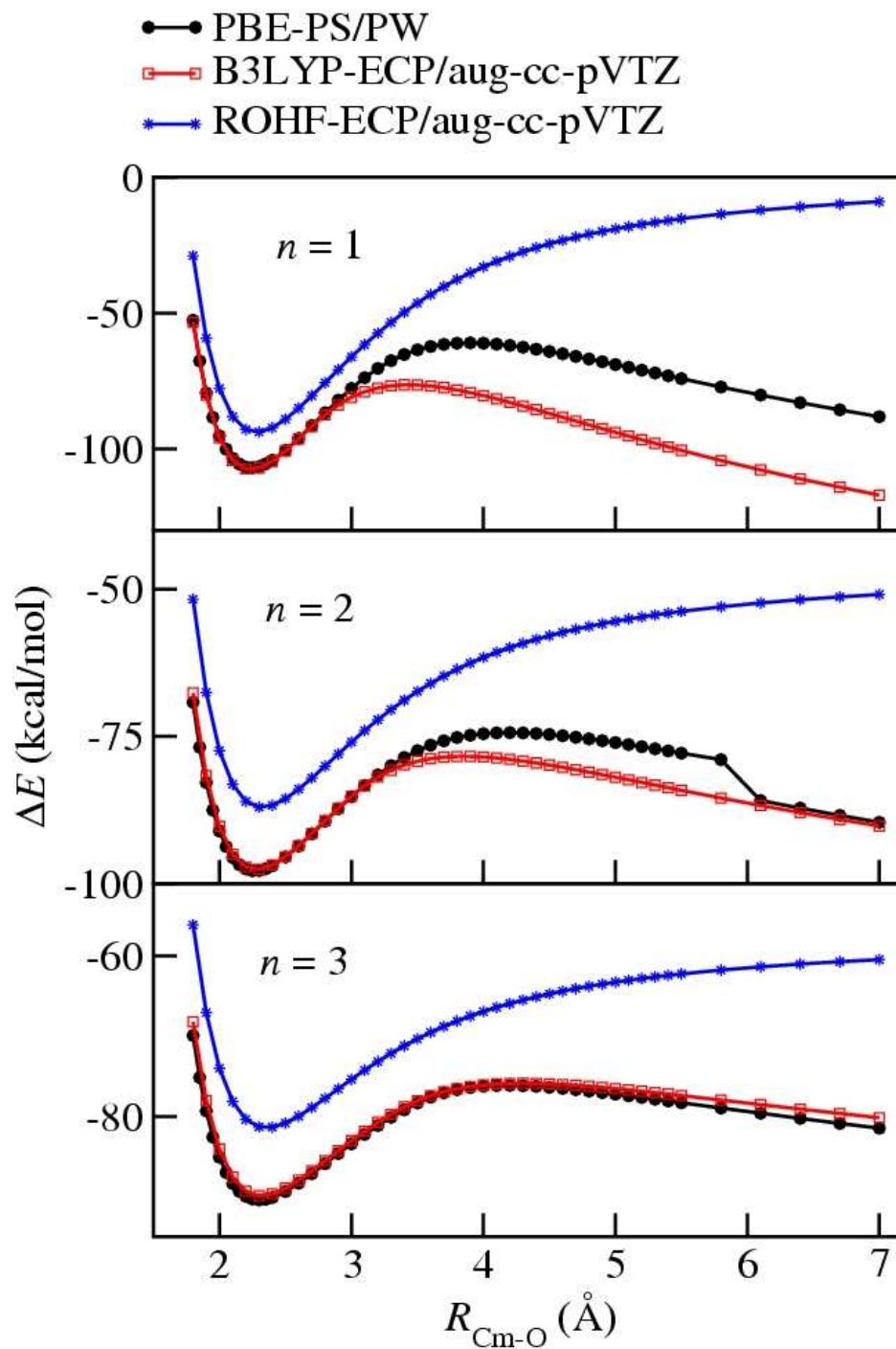


FIGURE S1: Potential energy curves of a single water molecule radially detached from  $[\text{Cm}(\text{OH}_2)_n]^{3+} (g)$  ( $n=1-3$ ) cluster along the Cm-O bond.

### 3. Fortran code for generating the initial configurations

```

C*****
C      THIS CODE GENERATES THE POSITIONS OF WATER MOLECULES (RANDOMLY) *
C      IN A CUBIC PERIODIC CELL WITH A METAL ION AT THE CENTER.      *
C      CONSTRAINTS ARE IMPOSED ON THE MINIMUM METAL-OXYGEN AND      *
C      OXYGEN-OXYGEN SEPARATION (SEE THE DESCRIPTION INSIDE THE CODE). *
C                                                                    *
C      AUTHORS: RAYMOND ATTA-FYNN, ERIC J. BYLASKA, WIBE A. DEJONG    *
C      AFFILIATION: EMSL, PACIFIC NORTHWEST NAT. LAB., RICHLAND WA 99352 *
C                                                                    *
C      KINDLY CITE OUR PAPER IF YOU USE THIS CODE TO GENERATE INITIAL *
C      STRUCTURES FOR YOUR SIMULATION                                *
C                                                                    *
C      DEFINITION OF VARIABLES                                       *
C      _____                                                    *
C                                                                    *
C      ATOM_SYMB      : ATOMIC SYMBOLS                               *
C      H_MASS         : ATOMIC MASS OF HYDROGEN                     *
C      O_MASS         : ATOMIC MASS OF OXYGEN                       *
C      S_MASS         : ATOMIC MASS OF METAL ION                    *
C      H2O_MASS       : MOLAR MASS OF WATER                         *
C      AVOGADRO       : AVOGADRO'S NUMBER                          *
C      RHO            : DENSITY OF THE ENTIRE SYSTEM (IN G/CM^3)    *
C      BOX_L         : BOX LENGTH (IN ANGSTROMS)                    *
C      NAT            : TOTAL NUMBER OF ATOMS                       *
C      NWM            : TOTAL NUMBER OF WATER MOLECULES             *
C      POS(3,NMAX)    : ATOMIC POSITION (IN ANGSTROMS)               *
C      D_OH           : O-H BOND DISTANCE (IN ANGSTROMS)            *
C      ANG_HOH        : H-O-H BOND ANGLE (IN DEGREES)               *
C      ANG_EPS        : H-O-H BOND ANGLE TOLERANCE (IN DEGREES)     *
C      RCUT_MO        : MINIMUM METAL-O SEPARATION (IN ANGSTROM)    *
C      RCUT_OO        : MINIMUM O-O SEPARATION (IN ANGSTROM)        *
C      RAN2           : RANDOM NUMBER GENERATOR FUNCTION (FROM NUMERICAL *
C                      RECIPES)                                     *
C      ISEED          : RANDOM NUMBER SEED                           *
C      DIST(A,B,L)    : FUNCTION WHICH COMPUTES THE DISTANCE BETWEEN *
C                      POINTS A AND B IN BOX OF LENGTH L SUBJECT TO  *
C                      PERIODIC BOUNDARY CONDITIONS                 *
C      ICOUNT         : ATOM COUNTER                                *
C      SOLUTE         : NAME OF METAL ION                           *
C      OUTPUT_FILE    : NAME OF FILE IN WHICH OUTPUT IS WRITTEN    *
C*****

PROGRAM          BUILD_CONFIG
IMPLICIT DOUBLE PRECISION (A-H,O-Z)
PARAMETER        (N1=5000)
CHARACTER        ATOM_SYMBOL(N1)*2, SOLUTE*2,
>               OUTPUT_FILE*30
DIMENSION        POS(3,N1), ATOM_MASS(N1)
DIMENSION        TMP0(3), TMP1(3)
EXTERNAL         RAN2, DISTANCE
DATA             ISEED/-1/, PI/3.14159265358979D0/
DATA             AVOGADRO/6.02214179D23/

```

```

C*****
C
C          EDIT THIS SECTION ACCORDING TO YOUR NEEDS
C
C*****

C      (A) FIXED INPUT PARAMETERS
C      NOTE: TO SIMULATE DEUTERIUM SET H_MASS = 2.01410178D0

      H_MASS      = 1.00794D0
      O_MASS      = 15.9994D0
      H2O_MASS    = O_MASS + 2.D0*H_MASS

C      (B) VARIABLE INPUT PARAMETERS: MAKE USE OF APRIORI INFORMATION
C      ABOUT YOUR SYSTEM. FOR EXAMPLE, YOU CAN USE THE FIRST PEAK
C      IN RADIAL DISTRIBUTION TO SET THE VALUES RCUT_MO AND RCUT_OO

      SOLUTE      = 'Cm'
      S_MASS      = 247.0704D0
      NWM         = 1000
      RHO         = 1.D0
      D_OH        = 1.D0
      ANG_HOH     = 109.47D0
      ANG_EPS     = 1.D-3
      RCUT_MO     = 2.5D0
      RCUT_OO     = 2.7D0
      OUTPUT_FILE = 'Cm_1000Water.xyz'

C      (C) OPTION TO COMPUTE THE BOX LENGTH:

C      L_OPTION DETERMINES HOW TO SET THE SYSTEM DENSITY RHO
C      IF L_OPTION = 0, RHO TO CORRESPOND TO ONLY WATER
C      IF L_OPTION = 1, RHO TO CORRESPOND WATER PLUS THE METAL ION

      L_OPTION    = 0

C*****
C
C          EDITABLE SECTION ENDS HERE
C
C*****

C*****
C          COMPUTING THE LENGTH OF THE BOX
C*****

C      THE TOTAL MASS IS DETERMINED BASED ON THE VALUE OF L_OPTION
      IF(L_OPTION.EQ.0)TOTAL_MASS = DFLOAT(NWM)*H2O_MASS
      IF(L_OPTION.EQ.1)TOTAL_MASS = DFLOAT(NWM)*H2O_MASS + S_MASS
      IF(L_OPTION.GT.1.OR.L_OPTION.LT.0)STOP 'ERROR IN L_OPTION'

C      COMPUTE THE BOX-LENGTH IN ANGSTROM
      BOX_L = 1.D8*((TOTAL_MASS/(AVOGADRO*RHO))**(1.D0/3.D0))

```

```

C*****
C          BOX LENGTH COMPUTATION DONE          *
C*****

C*****
C          SET THE PROPERTIES OF METAL ION IN THE BOX          *
C*****

C      SET THE POSITION OF THE THE METAL ION AT (0,0,0)
      ICOUNT = 1
      DO 10 I = 1, 3
        POS(I,ICOUNT) = 0.D0
10    CONTINUE

C      RECORD THE ATOMIC SYMBOL AND MASS
      ATOM_SYMBOL(ICOUNT) = SOLUTE
      ATOM_MASS(ICOUNT) = S_MASS

C      PRINT SUCCESSFUL GENERATION ON THE SCREEN
      WRITE(*,31)

C*****
C          DONE GENERATING PROPERTIES OF METAL ION IN THE BOX          *
C*****

C*****
C                      MAIN OF SECTION          *
C          SUCCESSIVE GENERATION OF WATER MOLECULES          *
C*****

C      IN THIS SECTION OF THE CODE, WATER MOLECULES WILL BE RANDOMLY
C      GENERATED IN A CUBIC BOX WITH PREDIODIC BOUNDARY CONDITIONS AND
C      FOLLOWING CONSTRAINTS:
C      (a) NO METAL-OXYGEN DISTANCE IS LESS THAN RCUT_CmO
C      (b) NO O-O DISTANCE IS LESS THAN RCUT_OO.

C      IMPORTANT NOTES:
C      (a) ANY FAIRLY GOOD RANDOM NUMBER GENERATOR IS OKAY. THE RANDOM
C          NUMBER USED HERE WAS TAKEN ALMOST VERBATIM FROM NUMERICAL
C          RECIPES(F77 VERSION).
C      (b) IN THIS CODE NO CONSTRAINT WAS IMPOSED ON THE ORIENTATION OF
C          THE WATER DIPOLE VECTORS; THEY ARE RANDOM. IN SOME CASES
C          HOWEVER, IT IS USEFUL TO IMPOSE CONSTRAINTS ON THE DIPOLE
C          MOMENT ORIENTATION BASED ON THE CHARGE OF THE METAL ION.
C      (c) USERS CAN SEND THE AUTHORS AN E-MAIL FOR A VERSION WHICH
C          ENFORCES THE DIPOLE ORIENTATION CONSTRAINT IN THE NEIGHBORHOOD
C          OF THE METAL ION.

C      NAT IS THE TOTAL NUMBER OF ATOMS
      NAT = 3*NWM + 1

      DO WHILE (ICOUNT.LT.NAT)

C          RANDOMLY GENERATE AN O ATOMIC POSITION IN THE

```

```

C      INTERVAL  $[-L/2, L/2]^3$ 

      DO 11 I = 1, 3
        TMP0(I) = BOX_L*(RAN2(ISEED)-0.5D0)
11     CONTINUE

C      TEMPORARILY STORE THE SOLUTE POSITION IN TMP1
      DO 12 I = 1, 3
        TMP1(I) = POS(I,1)
12     CONTINUE

C      CHECK IF ION-O DISTANCE CONSTRAINT IS SATISFIED, IF NOT REJECT
      DMO = DISTANCE(TMP0,TMP1,BOX_L)
      IF(DMO.LT.RCUT_MO)GOTO 27

      IF(ICOUNT.LE.2)GOTO 16

C      LOOP OVER ALL PREVIOUSLY GENERATED O-ATOMS UP TILL NOW
      DO 14 I = 2, ICOUNT

C          MAKE SURE H ATOMS ARE NOT USED IN THE CONSTRAINT CHECK
          IF(MOD(I+1,3).NE.0)GOTO 14

C          TEMPORARILY STORE THE O POSITION IN TMP1
          DO 15 J = 1, 3
            TMP1(J) = POS(J,I)
15         CONTINUE

C          CHECK IF O-O DISTANCE CONSTRAINT IS SATISFIED, IF NOT REJECT
          DOO = DISTANCE(TMP0,TMP1,BOX_L)
          IF(DOO.LT.RCUT_OO)GOTO 27

14         CONTINUE

16         CONTINUE

C      ACCEPT THE GENERATED OXYGEN ATOM
      ICOUNT = ICOUNT + 1
      DO 17 I = 1, 3
        POS(I,ICOUNT) = TMP0(I)
17     CONTINUE

C      RECORD THE ATOMIC SYMBOL AND MASS
      ATOM_SYMBOL(ICOUNT) = 'O'
      ATOM_MASS(ICOUNT) = O_MASS

C*****FIRST H POSITION WILL BE GENERATED HERE*****

C      THE PROCEDURE IS STRAIGHT-FOWARD: SIMPLY GENERATE A RANDOM
C      VECTOR OF LENGHT D_OH. WITH O AS THE ORIGIN OF THE VECTOR,
C      SIMPLY ADD H TO THE OTHER END

18     CONTINUE

C      RANDOMLY GENERATE A VECTOR ALONG WHICH THE OH BOND WILL BE

```

```

C      FORMED.

      V_MODULUS = 0.D0
      DO 19 I = 1, 3
          TMP0(I) = 2.D0*RAN2(ISEED)-1.D0
          V_MODULUS = V_MODULUS + TMP0(I)*TMP0(I)
19      CONTINUE
      V_MODULUS = DSQRT(V_MODULUS)

C      ENSURE THAT THE MODULUS OF THE VECTOR IS NOT CLOSE TO ZERO
      IF(V_MODULUS.LT.1.D-5)GOTO 18

C      MAKE THE LENGTH OF THE VECTOR 1
      DO 20 I = 1, 3
          TMP0(I) = TMP0(I)/V_MODULUS
20      CONTINUE

C      GENERATE THE FIRST H ATOMIC POSITION USING TMP0
      ICOUNT = ICOUNT + 1
      DO 21 I = 1, 3
          POS(I,ICOUNT) = POS(I,ICOUNT-1)+D_OH*TMP0(I)
21      CONTINUE

C      RECORD THE ATOMIC SYMBOL AND MASS
      ATOM_SYMBOL(ICOUNT) = 'H'
      ATOM_MASS(ICOUNT) = H_MASS

C*****GENERATION OF FIRST H POSITION DONE*****

C*****SECOND H POSITION WILL BE GENERATED HERE*****

C      THE PROCEDURE TO GENERATE THE SECOND-HYDROGEN IS SIMILAR TO
C      THE FIRST. HOWEVER, THE WATER BOND ANGLE CONSTRAINT MUST BE
C      SATISFIED.

C      RANDOMLY GENERATE A VECTOR ALONG WHICH THE OH BOND WILL BE
C      FORMED.

22      CONTINUE
      V_MODULUS = 0.D0
      DO 23 I = 1, 3
          TMP0(I) = 2.D0*RAN2(ISEED)-1.D0
          V_MODULUS = V_MODULUS + TMP0(I)*TMP0(I)
23      CONTINUE
      V_MODULUS = DSQRT(V_MODULUS)

C      ENSURE THAT THE MODULUS OF THE VECTOR IS NOT CLOSE TO ZERO
      IF (V_MODULUS.LT.1.D-5)GOTO 22

C      MAKE THE LENGTH OF THE VECTOR 1
      DO 24 I = 1, 3
          TMP0(I) = TMP0(I)/V_MODULUS

```

```

24      CONTINUE

C      TEMPORARILY GENERATE THE SECOND H POSTION (STORED IN TMP0)
C      ALSO STORE THE FIRST H POSITION (GENERATED IN THE PREVIOUS
C      BLOCK) IN TMP1

      DO 25 I = 1, 3
          TMP0(I) = POS(I, ICOUNT-1)+D_OH*TMP0(I)
          TMP1(I) = POS(I, ICOUNT)
25      CONTINUE

C      COMPUTE THE H-O-H BOND ANGLE USING THE LAW OF COSINES
      DHH = DISTANCE(TMP0, TMP1, BOX_L)
      ANG_TMP = DACOS((D_OH**2 + D_OH**2-DHH**2)/(2.D0*D_OH**2))
      ANG_TMP = 180.D0*ANG_TMP/PI

C      CHECK IF BOND ANGLE IS WITHIN TOLERANCE, IF NO REJECT
      ANG_TMP = ANG_TMP-ANG_HOH
      IF(DABS(ANG_TMP).GT.ANG_EPS)GOTO 22

C      GENERATE THE SECOND H ATOMIC POSITION
      ICOUNT = ICOUNT + 1
      DO 26 I = 1, 3
          POS(I, ICOUNT) = TMP0(I)
26      CONTINUE

C      RECORD THE ATOMIC SYMBOL AND MASS
      ATOM_SYMBOL(ICOUNT) = 'H'
      ATOM_MASS(ICOUNT) = H_MASS

C      PRINT SUCCESSFUL GENERATION WATER MOLECULE ON THE SCREEN
      WRITE(*,32) ICOUNT/3

27      CONTINUE

      END DO

C*****
C      GENERATION OF WATER MOLECULES DONE      *
C*****

C*****
C      PRINTING OUTPUT      *
C*****

      OPEN(28, FILE=OUTPUT_FILE, STATUS='UNKNOWN')
      WRITE(28, '(I7)') NAT
      WRITE(28, '(F12.4)') BOX_L

C      SHIFT ALL ATOMS TO [-L/2, L/2]^3

      BOX_L_INV = 1.D0/BOX_L

      DO 29 I = 1, NAT

```



```

        DO 30 J = 1, 3
            TMP = POS(J,I)
            TMP = TMP - BOX_L*DNINT(TMP*BOX_L_INV)
            POS(J,I)=TMP
30      CONTINUE
        WRITE(28,33)ATOM_SYMBOL(I), (POS(J,I), J=1,3), ATOM_MASS(I)
29      CONTINUE
        CLOSE(28)

C*****
C                                OUTPUT PRINTING DONE                                *
C*****

        WRITE(*,34)OUTPUT_FILE

31      FORMAT('GENERATED METAL ION')
32      FORMAT('GENERATED WATER MOLECULE',2X,I7)
33      FORMAT(A,4F12.6)
34      FORMAT(//,'TASK COMPLETED',//
>          'OUTPUT XYZ FILE IS:',2X,A,//)

        STOP

        END

C      END OF PROAGRAM

C*****
C                                DISTANCE FUNCTION                                *
C*****

        FUNCTION                DISTANCE(A, B, BOX_L)
        IMPLICIT DOUBLE PRECISION (A-H,O-Z)
        DIMENSION                A(3), B(3)

        BOX_L_INV = 1.D0/BOX_L
        DISTANCE = 0.D0

        DO 10 I = 1, 3
            TMP = A(I)-B(I)
            TMP = TMP - BOX_L*DNINT(TMP*BOX_L_INV)
            DISTANCE = DISTANCE + TMP**2
10      CONTINUE

        DISTANCE = DSQRT(DISTANCE)

        RETURN

        END

```

```

C*****
C                                DISTANCE FUNCTION ENDS HERE                                *
C*****

C*****
C    FORTRAN 77 RANDOM NUMBER GENERATOR                                                    *
C    REFERENCE: NUMERICAL RECIPES IN FORTRAN 77: THE ART OF                               *
C    SCIENTIFIC COMPUTING, SECOND EDITION                                                  *
C    (CAMBRIDGE UNIVERSITY PRESS, 1992)                                                    *
C*****

C    Long period (> 2 x 10^18) random number generator of L'Ecuyer
C    with Bays-Durham shuffle and added safeguards. Returns a uniform
C    random deviate between 0.0 and 1.0 (exclusive of the endpoint
C    values). Call with IDUM a negative integer to initialize;
C    thereafter, do not alter IDUM between successive deviates in a
C    sequence. RNMIX should approximate the largest floating value that
C    is less than 1.

FUNCTION RAN2(IDUM)
  INTEGER IDUM, IM1, IM2, IMM1, IA1, IA2, IQ1, IQ2, IR1, IR2, NTAB, NDIV
  DOUBLE PRECISION ran2, AM, EPS, RNMIX
  PARAMETER (IM1=2147483563, IM2=2147483399, AM=1.D0/IM1,
>IMM1=IM1-1, IA1=40014, IA2=40692, IQ1=53668, IQ2=52774, IR1=12211,
>IR2=3791, NTAB=32, NDIV=1+IMM1/NTAB, EPS=1.2D-7, RNMIX=1.D0-EPS)
  INTEGER IDUM2, J, K, IV(NTAB), IY
  SAVE IV, IY, IDUM2
  DATA IDUM2/123456789/, IV/NTAB*0/, IY/0/

  IF (IDUM.LT.0) THEN                                !Initialize.

    IDUM=MAX(-IDUM,1)                                !Be sure to prevent IDUM = 0.
    IDUM2=IDUM
    DO 11 J=NTAB+8,1,-1                               !Load the shuffle table
      IDUM=IA1*(IDUM-K*IQ1)-K*IR1 ! (after 8 warm-ups).

      IF (IDUM.LT.0) IDUM=IDUM+IM1
      IF (J.LE.NTAB) IV(J)=IDUM

11    CONTINUE
    IY=IV(1)

  END IF

  K=IDUM/IQ1                                           !Start here when not initializing

  IDUM=IA1*(IDUM-K*IQ1)-K*IR1                          !Compute IDUM=mod(IA1*IDUM, IM1)
  IF (IDUM.LT.0) IDUM=IDUM+IM1                          !without overflows by
  K=IDUM2/IQ2                                           !Schrage's method.

  IDUM2=IA2*(IDUM2-K*IQ2)-K*IR2                        !Compute IDUM2=mod(IA2*IDUM2, IM2)
                                                         !likewise.

  IF (IDUM2.LT.0) IDUM2=IDUM2+IM2

```

```

J=1+IY/NDIV
IY=IV(J)-IDUM2

IV(J)=IDUM

IF(IY.LT.1) IY=IY+IMM1
RAN2=MIN(AM*DFLOAT(IY),RNMX)

RETURN
END

```

```

!Will be in the range 1:NTAB.
!Here IDUM is shuffled, IDUM and
!IDUM2 are combined to generate output

!Because users don't expect
!endpoint values.

```

```

C*****
C      RANDOM NUMBER GENERATOR ROUTINE ENDS HERE      *
C*****

```

- 
- <sup>1</sup> M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, and W. A. De Jong, *Comput. Phys. Commun.* **181**, 1477 (2010); <http://www.nwchem-sw.org>.
- <sup>2</sup> J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- <sup>3</sup> T. H. Dunning Jr., *J. Chem. Phys.* **90**, 1007 (1988).
- <sup>4</sup> W. Kuchle, M. Dolg, H. Stoll, and H. Preuss, *J. Chem. Phys.* **100**, 7535 (1994).
- <sup>5</sup> A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993); C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988), S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* **58**, 1200 (1980); P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.* **98**, 11623 (1994).
- <sup>6</sup> S. Boys and F. Bernardi, *Mol. Phys.* **19**, 553 (1970).
- <sup>7</sup> D. Hagberg, E. Bednarz, N. M. Edelstein, and L. Gagliardi, *J. Am. Chem. Soc.* 14136 (2007)
- <sup>8</sup> E. Wasserman, J. R. Rustad, and S. S. Xantheas, *J. Chem. Phys.* **106**, 9769 (1997)
- <sup>9</sup> A. Bakker, K. Hermansson, J. Lindgren, M. M. Probst, and P. A. Bopp, *Int. J. of Quant. Chem.* **75**, 659 (1999)
- <sup>10</sup> Y. Inada, A. M. Mohammed, H. H. Loeffler, and B. M. Rode, *J. Phys. Chem. A* **106**, 6783 (2002)
- <sup>11</sup> E. Clementi, H. Kistenmacher, and W. Kolos, *Theoret. Chim. Acta. (Berl.)* **55**, 257 (1980).
- <sup>12</sup> L. X. Dang, *J. Chem. Phys.* **96**, 6790 (1992)
- <sup>13</sup> T. P. Lybrand and P. A. Kollman, *J. Chem. Phys.* **83**, 2329 (1985)
- <sup>14</sup> T. M. C. Faro, G. P. Thim, and M. S. Skaf, *J. Chem. Phys.* **132**, 114509 (2010)
- <sup>15</sup> A. S. de Araujo, M. T. Sonoda, O. E. Piro, and E. E. Castellano, *J. Phys. Chem. B* **111**, 2219 (2007)
- <sup>16</sup> A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff, *J. Amer. Chem. Soc.* **114**, 10024 (1992).
- <sup>17</sup> H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *J. Phys. Chem.* **91**, 6269 (1987).

- 
- <sup>18</sup> J. E. Lennard-Jones, *Proc. R. Soc. Lond. A* **106**, 463 (1924)
- <sup>19</sup> H. J. C. Berendsen, J. P. M. Postma, W. F. van Gustren, and J. Hermans In *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, **1981**), page 331.
- <sup>20</sup> K. Toukan and A. Rahman, *Phys. Rev. B* **31**, 2643 (1985).
- <sup>21</sup> T. S. Hofer, H. T. Tran, C. F. Schwenk, and B. M. Rode, *J. Comput. Chem.* **25**, 211 (2004).
- <sup>22</sup> R. W. Impey, P. A. Madden, and I. R. McDonald, *J. Phys. Chem.* **87**, 5071 (1983).