

## Hybrid approach for free energy calculations with high-level methods: Application to the S N 2 reaction of C H Cl 3 and O H – in water

Marat Valiev, Bruce C. Garrett, Ming-Kang Tsai, Karol Kowalski, Shawn M. Kathmann, Gregory K. Schenter, and Michel Dupuis

Citation: *The Journal of Chemical Physics* **127**, 051102 (2007); doi: 10.1063/1.2768343

View online: <http://dx.doi.org/10.1063/1.2768343>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/jcp/127/5?ver=pdfcov>

Published by the AIP Publishing

### Articles you may be interested in

Investigation of the CH<sub>3</sub>Cl + CN<sup>–</sup> reaction in water: Multilevel quantum mechanics/molecular mechanics study  
J. Chem. Phys. **142**, 244505 (2015); 10.1063/1.4922938

Combining ab initio quantum mechanics with a dipole-field model to describe acid dissociation reactions in water: First-principles free energy and entropy calculations  
J. Chem. Phys. **132**, 074112 (2010); 10.1063/1.3317398

Extensive regularization of the coupled cluster methods based on the generating functional formalism: Application to gas-phase benchmarks and to the S N 2 reaction of CHCl<sub>3</sub> and OH<sup>–</sup> in water  
J. Chem. Phys. **131**, 234107 (2009); 10.1063/1.3270957

Gas-phase reaction between calcium monocation and fluoromethane: Analysis of the potential energy hypersurface and kinetics calculations  
J. Chem. Phys. **131**, 144309 (2009); 10.1063/1.3247287

Quantum mechanics/molecular mechanics minimum free-energy path for accurate reaction energetics in solution and enzymes: Sequential sampling and optimization on the potential of mean force surface  
J. Chem. Phys. **128**, 034105 (2008); 10.1063/1.2816557



# NEW Special Topic Sections

**NOW ONLINE**  
Lithium Niobate Properties and Applications:  
Reviews of Emerging Trends

**AIP** Applied Physics Reviews

# Hybrid approach for free energy calculations with high-level methods: Application to the $S_N2$ reaction of $\text{CHCl}_3$ and $\text{OH}^-$ in water

Marat Valiev<sup>a)</sup>

Molecular Sciences Software Group, Environmental Molecular Sciences Laboratory,  
Pacific Northwest National Laboratory, Richland, Washington 99352, USA

Bruce C. Garrett and Ming-Kang Tsai

Chemical and Materials Sciences Division, Pacific Northwest National Laboratory,  
Richland, Washington 99352, USA

Karol Kowalski

Molecular Sciences Software Group, Environmental Molecular Sciences Laboratory,  
Pacific Northwest National Laboratory, Richland, Washington 99352, USA

Shawn M. Kathmann, Gregory K. Schenter, and Michel Dupuis

Chemical and Materials Sciences Division, Pacific Northwest National Laboratory,  
Richland, Washington 99352, USA

(Received 15 May 2007; accepted 10 July 2007; published online 7 August 2007)

We present an approach to calculate the free energy profile along a condensed-phase reaction path based on high-level electronic structure methods for the reactive region. The bulk of statistical averaging is shifted toward less expensive descriptions by using a hierarchy of representations that includes molecular mechanics, density functional theory, and coupled cluster theories. As an application of this approach we study the reaction of  $\text{CHCl}_3$  with  $\text{OH}^-$  in aqueous solution. © 2007 American Institute of Physics. [DOI: 10.1063/1.2768343]

There are many application areas that can greatly benefit from accurate electronic structure description of reaction processes provided by high-level methods, such as coupled cluster (CC) theory.<sup>1</sup> However, in many practical cases, explicit consideration of the complex heterogeneous environment renders such applications nearly intractable because of the need to sample over a large number of degrees of freedom. Previously we have shown<sup>2</sup> how to incorporate high-level *ab initio* methods for calculations of finite temperature averages of excited states. In this work we describe a systematic approach for CC-based calculations of a free energy profile along a reaction path—a quantity central to the analysis of reactions in large complex chemical systems. Similar in spirit to the ONIOM method,<sup>3</sup> our approach combines CC, density functional theory (DFT), and molecular mechanics (MM) methods to construct a sequence of multiple representations of the reactive region allowing the bulk of statistical averaging to be shifted toward less expensive descriptions.

Let us consider the combined solute and solvent system. By construction, the solute contains all the reactive species requiring a quantum-mechanical (QM) description. Adopting a classical description for the solvent, the total energy of the system in the combined quantum mechanical and molecular mechanics (QM/MM) approach<sup>4</sup> can be expressed as a sum of quantum (QM) and classical molecular mechanics (MM) contributions,

$$E = E_{\text{QM}}(\mathbf{r}, \mathbf{R}; \psi) + E_{\text{MM}}(\mathbf{r}, \mathbf{R}), \quad (1)$$

where  $\psi$  represents the electronic degrees of freedom corresponding to the ground state, and  $\mathbf{r}$  and  $\mathbf{R}$  are solute and solvent coordinates, respectively. The QM energy can be further decomposed into internal and external parts,

$$E_{\text{QM}} = E_{\text{QM}}^{\text{int}}(\mathbf{r}; \psi) + E_{\text{QM}}^{\text{ext}}(\mathbf{r}, \mathbf{R}; \rho). \quad (2)$$

Formally the internal part  $E_{\text{QM}}^{\text{int}}$  coincides with the solute QM energy in the gas phase. In a CC-based description, it can be represented as

$$E_{\text{CC}}^{\text{int}} = \langle \Phi | H e^T | \Phi \rangle, \quad (3)$$

where  $H$  is a standard many-electron Hamiltonian and  $\Phi$  is a reference many-electron wave function. The external part of the QM energy contains electrostatic interactions between solute electron density ( $\rho$ ) and solvent classical charges ( $Z_I$ ),

$$E_{\text{QM}}^{\text{ext}}(\mathbf{r}, \mathbf{R}; \rho) = \sum_I \int \frac{Z_I \rho(\mathbf{r}')}{|\mathbf{R}_I - \mathbf{r}'|} d\mathbf{r}'. \quad (4)$$

The van der Waals and Coulomb nuclear solute-solvent interactions are absorbed into the second term in Eq. (1), which also contains the classical energy of the solvent. We assume that the reaction can be fully characterized by a reduced set of coordinates, involving only the solute part of the system. Using an equilibrium solvation model<sup>5</sup> the solvent degrees of freedom can be integrated out leading to the definition of the effective thermodynamic potential,

<sup>a)</sup>Electronic mail: marat.valiev@pnl.gov

$$W(\mathbf{r}, \beta) = -\frac{1}{\beta} \ln \int e^{-\beta E(\mathbf{r}, \mathbf{R}; \psi)} d\mathbf{R}, \quad (5)$$

where  $\beta = 1/kT$ . In many ways the effective thermodynamic potential is reminiscent of the total energy quantity, which facilitates the reaction analysis in terms of concepts developed for gas phase reaction calculations.

Prior to discussing the calculation of  $W(\mathbf{r}, \beta)$  along the reaction pathway in solute coordinate space, let us briefly address an issue of how such a pathway could be generated in the first place. Using gas phase calculations for this purpose may not always be appropriate as the solvent may significantly influence the reaction process. The main challenge is that the direct calculation of the derivatives of  $W(\mathbf{r}, \beta)$  required for path calculations is impractical with CC or even DFT methods because of numerous electronic structure calculations during solvent averaging. The computations, however, can be significantly simplified in the zero temperature limit. In this case the presence of the solvent is taken into account, but the influence of finite temperature fluctuations on the structure of reaction path is neglected. In the zero temperature limit the major contribution to the derivatives of  $W(\mathbf{r}, \beta)$  comes from the region where  $E(\mathbf{r}, \mathbf{R}; \psi)$  is near its minimum with respect to  $\mathbf{R}$ , therefore

$$\lim_{\beta \rightarrow \infty} \partial_{\mathbf{r}} W(\mathbf{r}, \beta) = \partial_{\mathbf{r}} E(\mathbf{r}, \mathbf{R}^*; \psi), \quad (6)$$

where  $\mathbf{R}^*$  is a solvent configuration minimizing  $E(\mathbf{r}, \mathbf{R}; \psi)$ . For a large number of solvent degrees of freedom, many optimization steps may be required to determine  $\mathbf{R}^*$ . As the solute region is kept fixed during this procedure, the computations can be simplified by following the previously suggested method<sup>6</sup> of introducing an effective classical representation for the solute-solvent interactions,

$$E_{\text{QM}}^{\text{ext}} = \sum_{i,l} \frac{Z_l Q_i}{|\mathbf{R}_l - \mathbf{r}_i|} \equiv E_{\text{ESP}}(\mathbf{r}, \mathbf{R}; Q). \quad (7)$$

Here the QM atoms are represented by the effective electrostatic potential (ESP) charges ( $Q_i$ ) such that the electrostatic potential outside the solute region is the same as that produced from the full electron density  $\rho(\mathbf{r})$ . The charges  $Q_i$  do depend on the solvent coordinates, but as long as the optimization cycle is convergent the frequency of their updates is largely irrelevant and can be exploited to achieve the best computational efficiency. The approach presented here allows straightforward incorporation of various reaction pathway methodologies. For example, the nudged elastic band (NEB) method<sup>7</sup> can be implemented in the following way. After generation of an initial trial pathway and determination of effective charges for each of the resulting replicas of the QM region, the self-consistent cycle is carried out consisting of (i) optimization of the solvent MM region in the field of the effective charge representation of the QM region, (ii) calculation of new forces and effective charges for the QM region through the solution of the Schrodinger equation, and (iii) evolution of the QM region according to the NEB methodology.<sup>7</sup> The cycle is repeated until convergence is achieved. This methodology can be applied in conjunction

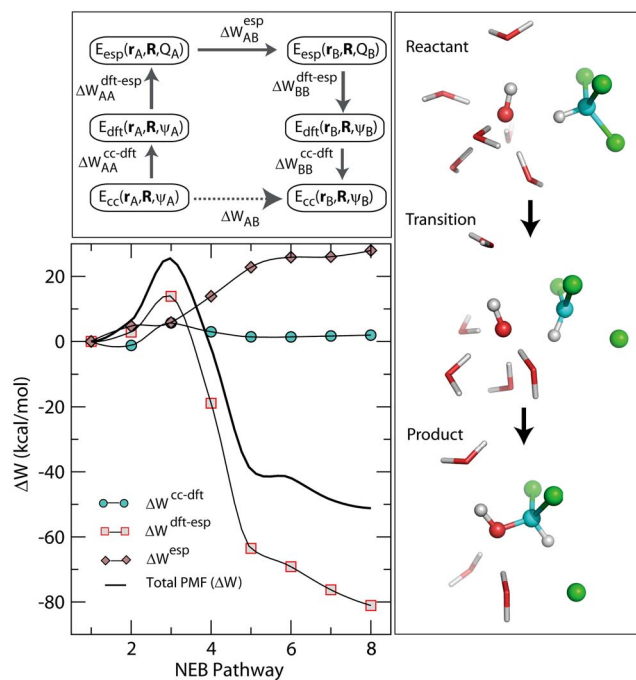


FIG. 1. (Color online) Thermodynamic cycle for free energy calculations (top left); free energy profile (bottom left) and reactant, transition, and product states (right) for the reaction  $\text{CHCl}_3 + \text{OH}^- \rightarrow \text{CHCl}_2\text{OH} + \text{Cl}^-$ .

with both CC or less expensive DFT versions of QM/MM methods.

The free energy difference between points  $\mathbf{r}_A$  and  $\mathbf{r}_B$  on the reaction path can be written as

$$\Delta W_{AB} = -\frac{1}{\beta} \ln \langle e^{-\beta(E(\mathbf{r}_B, \mathbf{R}; \psi_B) - E(\mathbf{r}_A, \mathbf{R}; \psi_A))} \rangle_{\mathbf{r}_A}, \quad (8)$$

where angular brackets denote a statistical average over solvent configurations with fixed solute geometry,

$$\langle \dots \rangle_{\mathbf{r}} = \frac{\int \dots e^{-\beta E(\mathbf{r}, \mathbf{R}; \psi)} d\mathbf{R}}{\int e^{-\beta E(\mathbf{r}, \mathbf{R}; \psi)} d\mathbf{R}}. \quad (9)$$

Direct QM/MM evaluation of Eq. (8) is impractical because of numerous electronic structure calculations during statistical averaging. Taking advantage of a state function property of  $W(\mathbf{r}, \beta)$ , an alternative thermodynamic cycle can be devised where the burden of statistical sampling is shifted to a less expensive representation of the solute-solvent interactions.<sup>6,8</sup> For a CC/MM based description, we suggest a multitiered mapping involving three different representations of varied complexity: CC/MM, DFT/MM, and ESP/MM (see Fig. 1). The ESP/MM representation is obtained by replacing  $E_{\text{QM}}$  in Eq. (1) by the effective charge energy  $E_{\text{ESP}}$  [see Eq. (7)]. The total free energy difference  $\Delta W_{AB}$  can then be represented as a sum of three contributions (see Fig. 1)

$$\Delta W_{AB} = (\Delta W_{AA}^{\text{CC} \rightarrow \text{DFT}} - \Delta W_{BB}^{\text{CC} \rightarrow \text{DFT}}) + (\Delta W_{AA}^{\text{DFT} \rightarrow \text{ESP}} - \Delta W_{BB}^{\text{DFT} \rightarrow \text{ESP}}) + \Delta W_{AB}^{\text{ESP}}. \quad (10)$$

The first and second terms represent free energy difference for changing the description of the *fixed* solute region from the CC to DFT representations and from DFT to a classical

ESP representations [see Eq. (7)], respectively. The third term represents the free energy difference for changing solute configuration from  $(\mathbf{r}_A, Q_A)$  to  $(\mathbf{r}_B, Q_B)$  within classical ESP/MM description. It is this last leg of the cycle that bears the largest burden of statistical sampling, which can now be accommodated easily.

Since the position of the solute region is fixed during the change of representations (CC to DFT, DFT to ESP) it is reasonable to assume that there is a sufficient overlap between accessible solvent phase regions. In this case, free energy perturbation theory can be used to calculate corresponding free energy differences,

$$\Delta W_{AA}^{\text{CC} \rightarrow \text{DFT}} = -\frac{1}{\beta} \ln \langle e^{-\beta \Delta E_{AA}^{\text{CC} \rightarrow \text{DFT}}} \rangle_{\text{DFT/MM}}, \quad (11)$$

$$\Delta W_{AA}^{\text{DFT} \rightarrow \text{ESP}} = -\frac{1}{\beta} \ln \langle e^{-\beta \Delta E_{AA}^{\text{DFT} \rightarrow \text{ESP}}} \rangle_{\text{ESP/MM}}, \quad (12)$$

where

$$\begin{aligned} \Delta E_{AA}^{\text{CC} \rightarrow \text{DFT}} &= E_{\text{CC}}(\mathbf{r}_A, \mathbf{R}; \psi_A) - E_{\text{DFT}}(\mathbf{r}_A, \mathbf{R}; \psi_A), \\ \Delta E_{AA}^{\text{DFT} \rightarrow \text{ESP}} &= E_{\text{DFT}}(\mathbf{r}_A, \mathbf{R}; \psi_A) - E_{\text{ESP}}(\mathbf{r}_A, \mathbf{R}; Q_A). \end{aligned} \quad (13)$$

The benefit of using an intermediate DFT description is that in the ground state the DFT electron density should be fairly close<sup>9</sup> to the nearly exact density generated by CC. Since the electron density is the sole coupling parameter between the solute electronic degrees of freedom and the solvent [see Eq. (4)], the derivatives  $\partial E_{\text{CC}}/\partial \mathbf{R}$  and  $\partial E_{\text{DFT}}/\partial \mathbf{R}$  produce only density- (not wave function) dependent terms and therefore

$$\frac{\partial \Delta E_{AA}^{\text{CC} \rightarrow \text{DFT}}}{\partial \mathbf{R}} \approx 0 \Rightarrow \Delta W_{AA}^{\text{CC} \rightarrow \text{DFT}} \approx \Delta E_{AA}^{\text{CC} \rightarrow \text{DFT}}. \quad (14)$$

In the evaluation of  $\Delta W_{AA}^{\text{DFT} \rightarrow \text{ESP}}$  we can follow a resampling strategy<sup>10</sup> and evaluate  $\Delta E_{AA}^{\text{DFT} \rightarrow \text{ESP}}$  at a certain interval over the solvent trajectories expediently generated at the ESP/MM level of theory.

There are several ways to evaluate the classical  $\Delta W_{AB}^{\text{ESP}}$  term. Based on the smooth transformation between A and B configurations,

$$\begin{aligned} \mathbf{r}_\lambda &= (1 - \lambda)\mathbf{r}_A + \lambda\mathbf{r}_B, \\ Q_\lambda &= (1 - \lambda)Q_A + \lambda Q_B, \end{aligned} \quad (15)$$

$\Delta W_{AB}^{\text{ESP}}$  can be calculated as a sum of differences,

$$\Delta W_{AB}^{\text{ESP}} = -\sum_i \frac{1}{\beta} \ln \langle e^{-\beta \Delta E_{\lambda_i \rightarrow \lambda_{i+1}}^{\text{ESP}}} \rangle_{\lambda_i}, \quad (16)$$

where  $\Delta E_{\lambda_i \rightarrow \lambda_{i+1}}^{\text{ESP}}$  is the ESP energy difference between  $\lambda_{i+1}$  and  $\lambda_i$  windows that partition  $\lambda \in [0, 1]$  interval.

We implemented the above approach into NWChem (Ref. 11) and applied it to the  $\text{S}_\text{N}2$  reaction of  $\text{CHCl}_3$  and  $\text{OH}^-$  in aqueous solution. Only gas phase studies of this reaction have been performed<sup>3,12</sup> so far, and there is some uncertainty whether the  $\text{S}_\text{N}2$  mechanism is a viable option for base-catalyzed hydrolysis of  $\text{CHCl}_3$  in aqueous solution. Our solute QM region contained  $\text{CHCl}_3$  and  $\text{OH}^-$  molecules. At the DFT level it was described using the B3LYP exchange-

correlation functional with the 6-31+G\* basis set, and the CC description was based on the locally renormalized variant of CCSD(T)(LR-CCSD(T)) (Ref. 13) with the aug-cc-pVDZ (Ref. 14) basis set. The van der Waals parameters for the QM region were taken from standard Amber force field definitions.<sup>15</sup> The solute was embedded into a 30 Å cubic box of 888 SPC/E (Ref. 16) solvent water molecules representing the MM region.

Reaction pathway calculations were performed using the DFT/MM representation of the system using the zero temperature methodology described earlier. First we determined the transition state structure, characterized by a single imaginary mode of 238  $\text{cm}^{-1}$ . Then we optimized reactant and product states with the initial structures generated through the displacement of the transition state in the direction of the negative mode. To ensure full relaxation of the solvent, optimization of reactant, transition, and product states were performed in combination with molecular dynamics equilibration (40 ps at  $T=298.15$  K) of the solvent region. The initial pathway for NEB calculations was calculated by linear interpolation between the reactant, transition, and product states of the entire solute-solvent system with six beads/replicas for each segment. In the reactant state the  $\text{OH}^-$  group is positioned 3.1 Å away from the carbon on the  $\text{CHCl}_3$  group serving as a hydrogen bond acceptor to chloroform. During the reaction it makes a symmetrical approach to chloroform such that the OH bond bisects one of the Cl-C-Cl angles. The water solvation shell of  $\text{OH}^-$  changes from six-fold coordination in the reactant state to five-fold and three-fold coordination in the transition and product states, respectively. In the product state the  $\text{Cl}^-$  ion, located 3.23 Å from the  $\text{CHCl}_3\text{OH}$  complex, is coordinated by four waters.<sup>17</sup>

The free energy profile, including individual contributions, is shown in Fig. 1. For  $\Delta W_{AB}^{\text{ESP}}$  calculations we used effective charges generated during the NEB procedure. The  $\lambda$ -summation interval was partitioned in 0.1 Å increments ( $\lambda_i=0, 0.1, \dots, 1.0$ ). Using a double wide sampling strategy,<sup>18</sup> the free energy differences for  $\lambda_i \rightarrow \lambda_{i-1}$  and  $\lambda_i \rightarrow \lambda_{i+1}$  were obtained simultaneously by sampling at  $\lambda_i$  over 20 ps of molecular dynamics simulation at constant temperature ( $T=298.15$  K). The calculation of  $\Delta W_{AB}^{\text{DFT} \rightarrow \text{ESP}}$  was done by resampling the 100 ps solvent trajectory generated at the ESP/MM level with the interval of 0.5 ps. The free energy difference  $\Delta W_{AA}^{\text{CC} \rightarrow \text{DFT}}$  was calculated according to Eq. (14) over the NEB reaction path. To estimate the accuracy of this approximation we also performed a series of CC and DFT calculations over the 10 ps solvent trajectory with 0.5 ps interval for the reactant and transition states. The calculated standard deviations for  $\Delta E_{AA}^{\text{CC} \rightarrow \text{DFT}}$  (0.45 kcal/mol for reactant and 0.13 kcal/mol for transition states) and the estimated barrier error of 0.5 kcal/mol support the validity of our approximation. Overall we find that the CC/MM approach gives a free energy reaction barrier of 29.3 kcal/mol and an overall reaction free energy of -46.7 kcal/mol. We note that DFT/MM treatment of the same process underestimates the reaction barrier (24.6 kcal/mol) but agrees well on the reaction free energy (-47 kcal/mol).

In conclusion we have developed an efficient and systematic approach for free energy calculations along a



condensed-phase reaction path based on high-level electronic structure methods for the reactive region. The approach can be used in conjunction with not only QM/MM but also ONIOM (Ref. 3) and other methodologies.<sup>19</sup> We have illustrated our approach by studying the  $S_N2$  reaction of  $\text{CHCl}_3$  with  $\text{OH}^-$  in aqueous phase. Significant differences between free energy reaction barriers between CC and DFT descriptions illustrate the importance of an accurate description of electron correlation effects in free energy calculations. In agreement with experiment,<sup>20</sup> the high reaction barrier obtained in our calculation indicates that the  $S_N2$  mechanism is unlikely to play a major role in the hydrolysis of  $\text{CHCl}_3$  in aqueous solution.

Three of the authors (M.V., B.C.G., and S.M.K.) would like to acknowledge support from Office of Naval Research. Three of the authors (M.K.T., G.K.S., and M.D.) would like to acknowledge support from the Division of Chemical Sciences, Office of Basic Energy, DOE. This research was performed in part using the MSCF in EMSL, a national scientific user facility sponsored by the U.S. DOE, OBER and located at PNNL. Battelle operates PNNL for DOE.

<sup>1</sup>R. J. Bartlett and M. Musial, *Rev. Mod. Phys.* **79**, 291 (2007).

<sup>2</sup>M. Valiev and K. Kowalski, *J. Chem. Phys.* **125**, 211101 (2006).

- <sup>3</sup>S. Y. Re and K. Morokuma, *Theor. Chem. Acc.* **112**, 59 (2004).
- <sup>4</sup>J. Gao and D. G. Truhlar, *Annu. Rev. Phys. Chem.* **53**, 467 (2002); R. A. Friesner and V. Guallar, *ibid.* **56**, 389 (2005).
- <sup>5</sup>B. C. Garrett and G. K. Schenter, *Int. Rev. Phys. Chem.* **13**, 263 (1994).
- <sup>6</sup>Y. K. Zhang, H. Y. Liu, and W. T. Yang, *J. Chem. Phys.* **112**, 3483 (2000).
- <sup>7</sup>G. Henkelman, B. P. Uberuaga, and H. Jonsson, *J. Chem. Phys.* **113**, 9901 (2000).
- <sup>8</sup>M. Dupuis, G. K. Schenter, B. G. Garrett, and E. E. Arcia, *J. Mol. Struct.: THEOCHEM* **632**, 173 (2003).
- <sup>9</sup>W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- <sup>10</sup>R. H. Wood, E. M. Yezdimer, S. Sakane, J. A. Barriocanal, and D. J. Doren, *J. Chem. Phys.* **110**, 1329 (1999).
- <sup>11</sup><http://www.emsl.pnl.gov/docs/nwchem/>
- <sup>12</sup>Y. A. Borisov, E. E. Arcia, S. L. Mielke, B. C. Garrett, and T. H. Dunning, *J. Phys. Chem. A* **105**, 7724 (2001).
- <sup>13</sup>K. Kowalski and P. Piecuch, *J. Chem. Phys.* **122**, 074107 (2005); K. Kowalski, *ibid.* **123**, 014102 (2005).
- <sup>14</sup>T. H. Dunning, Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- <sup>15</sup>T. Fox and P. A. Kollman, *J. Phys. Chem. B* **102**, 8070 (1998).
- <sup>16</sup>H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *J. Phys. Chem.* **91**, 6269 (1987).
- <sup>17</sup>See EPAPS Document No. E-JCPSA6-127-310730 for detailed structural information. This document can be reached through a direct link in the online article's HTML reference section or via the EPAPS homepage (<http://www.aip.org/pubservs/epaps.html>).
- <sup>18</sup>W. L. Jorgensen and C. Ravimohan, *J. Chem. Phys.* **83**, 3050 (1985).
- <sup>19</sup>S. Hirata, M. Valiev, M. Dupuis, S. S. Xantheas, S. Sugiki, and H. Sekino, *Mol. Phys.* **103**, 2255 (2005).
- <sup>20</sup>J. Hine, *J. Am. Chem. Soc.* **72**, 2438 (1950).