JCTC Journal of Chemical Theory and Computation

Defining Condensed Phase Reactive Force Fields from ab Initio Molecular Dynamics Simulations: The Case of the Hydrated Excess Proton

Chris Knight,[†] C. Mark Maupin,[‡] Sergei Izvekov,[‡] and Gregory A. Voth*[,†,‡]

Department of Chemistry, James Franck Institute, and Computation Institute, University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637, United States, and Center for Biophysical Modeling and Simulation and Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States

Received August 9, 2010

Abstract: In this report, a general methodology is presented for the parametrization of a reactive force field using data from a condensed phase ab initio molecular dynamics (AIMD) simulation. This algorithm allows for the creation of an empirical reactive force field that accurately reproduces the underlying ab initio reactive surface while providing the ability to achieve long-time statistical sampling for large systems not possible with AIMD alone. In this work, a model for the hydrated excess proton is constructed where the hydronium cation and proton hopping portions of the model are statistically force-matched to the results of Car−Parrinello Molecular Dynamics (CPMD) simulations. The flexible nature of the algorithm also allows for the use of the more accurate classical simple point-charge flexible water (SPC/Fw) model to describe the water−water interactions while utilizing the ab initio data to create an overall multistate molecular dynamics (MS-MD) reactive model of the hydrated excess proton in water. The resulting empirical model for the system qualitatively reproduces thermodynamic and dynamic properties calculated from the ab initio simulation while being in good agreement with experimental results and previously developed multistate empirical valence bond (MS-EVB) models. The present methodology, therefore, bridges the AIMD technique with the MS-MD modeling of reactive events, while incorporating key strengths of both.

1. Introduction

The simulation of certain classes of chemical reactions and charge transport processes in the condensed phase is a challenging problem. This is because of the large system sizes necessary to eliminate system size dependent artifacts from long-ranged electrostatics as well as the long time scales necessary to accurately simulate competing or intertwined dynamical processes such as diffusion. While simple ion transport is difficult in its own right, the hydrated excess proton, which is a delocalized net positive charge defect spanning over multiple water molecules, is considered one of the most difficult ions to simulate due to the variable bonding topology that accompanies transport via the Grotthuss mechanism.[1,2] For modeling of complex chemical processes in the condensed phase such as the hydrated excess proton, one can utilize ab initio molecular dynamics (AIMD) simulations (refs 3−5 and those cited therein). However, the computational cost of AIMD simulations typically limits the system size and simulation time length when applied to condensed phase systems. In addition, one is also currently limited by the accuracy of density functionals, at least until high-level Schrödinger wave function methods, such as MP2,[6,7] become tractable for condensed phase simulations. Therefore, it is advantageous to develop a general procedure for accurately parametrizing empirical reactive force fields that allow for the study of complex chemical processes, i.e., the making and breaking of chemical bonds, and that can reproduce thermodynamic and dynamic properties of an

* Corresponding author e-mail: gavoth@uchicago.edu.
[†] University of Chicago.
[‡] University of Utah.

10.1021/ct1004438
Published on Web 09/23/2010

AIMD simulation while also extending both the length and time scale of the simulation to probe more complex phenomena.

In typical molecular mechanics force fields, a single bonding topology is assumed throughout the entire simulation, which precludes the possibility of modeling reactive events that rely upon the breaking and formation of chemical bonds. To address this fixed bonding topology deficiency, reactive molecular dynamics methodologies that explicitly handle variable coordination and bonding topologies have been utilized to study various chemically reactive systems of interest.[8,9] The multistate molecular dynamics (MS-MD) approach, of which the multistate empirical valence bond (MS-EVB) methodology is an example,[10−14] is a reactive simulation method that dynamically changes the underlying bonding topology of a reactive species in response to the environment. This dynamic evolution of the bonding topology is achieved by constructing a linear combination of several bonding topologies describing various "states" of the acceptor/donor species; the relative weights of which are determined "on the fly" during the course of the simulation. It is the dynamic nature of this reactive force field that allows for an accurate description of bond making/breaking processes, such as those involved in the transport of the hydrated excess proton and hydroxide ions.[2,5,13−15]

The original EVB approach[16−19] was used as an interpolation scheme to describe the reaction path for chemical reactions between well-defined reactant−product pairs. In this original method, a Hamiltonian matrix, commonly 2×2, is constructed where all elements are expressed as functions of geometric coordinates of the system. The diagonal elements of the Hamiltonian matrix correspond to the diabatic states (bonding topologies) representing the donor and acceptor in the chemical reaction of interest. The off-diagonal elements describe the coupling between two diabatic states (donor and acceptor) and provide a mechanism for transitions, allowing for the donor and acceptor to switch identities. This scheme is general enough to be able to describe various types of chemical reactions, such as intramolecular isomerization/rearrangement[20] and nucleophilic substitution.[21] Once the reactant and product states are identified, the Hamiltonian matrix is populated (diagonal and off-diagonal components) and subsequently diagonalized. Upon diagonalization, the coefficients of the lowest energy eigenvector are used to calculate the forces using the Hellman−Feynman theorem and the system is propagated in time using Newtonian equations of motion. The same basic idea with certain extensions has been implemented in the multiconfigurational molecular mechanics approach.[22,23]

The multistate MD generalization of the EVB idea was to dynamically include bonding topologies spanning multiple molecules and larger effective distances (MS-EVB).[13,14] For each simulation step, the multiple diabatic states, used to construct the Hamiltonian matrix, are determined by identifying all possible donor−acceptor pairs proximal to an initial donor−acceptor. The dimension of this Hamiltonian can be kept relatively modest using geometric criteria, such as bond distances and angles, to "weed out" those bonding topologies that are likely to be energetically unfavorable. For the case of the hydrated excess proton in bulk water, it is found to be sufficient to include ∼30 states (up to and including the third solvation shell of the excess proton),[24] while only a few states have significant weight for channel-type systems.[25] By allowing the MS-MD algorithm the freedom to consider all of these possible states, one can properly model the excess proton charge defect delocalization over several molecules adjacent to the donor−acceptor species while accounting for the dynamical rearrangement of the chemical bond topology.

The MS-MD (or specifically MS-EVB) models developed previously have been parametrized against high-level ab initio gas phase calculations[12,26,27] and experimental data. These fitting procedures only provide a partial route for defining an accurate condensed phase potential energy surface for the reactive system. It is, thus, desirable to develop new, robust, and efficient algorithms for the parametrization of accurate reactive MS-MD force fields based instead on explicit condensed phase AIMD simulations. As detailed earlier and below, the MS-MD methodology then enables simulations of large systems over the long time scales necessary to adequately sample the evolving environment, whether it is the fluctuating hydrogen bond (H-bond) rearrangements in aqueous solutions or the slow protein conformational changes in biomolecular environments.

Force matching (FM) algorithms provide a means by which the forces on atoms calculated from some method with a high level treatment of interactions, typically a computationally expensive method such as accurate ab initio calculations, can be used to construct and parametrize an empirical model that best reproduces the original forces.[28−31] The result of a force matching calculation is an empirical force field that is capable of accurately reproducing the physical properties of the original method used to build the training set of configurations. When the interactions are expressed as a sum of pairwise functions that are linear with respect to the model parameters,[30,31] one is able to take advantage of linear least-squares fitting procedures to determine the model parameters that best reproduce the original forces. This latter approach dramatically simplifies the FM algorithm for complex condensed phase systems.[30,31] By choosing these effective interactions to be interpolated as cubic spline functions (third order polynomials), one can also take advantage of the favorable property of continuity of the force and its first derivative.

In this work, a force matching algorithm is developed and applied to construct a reactive force field (in this case for aqueous proton transport) using a training set of ab initio data from a condensed phase simulation. Fitting a model to data from a condensed phase simulation is advantageous due to issues related to transferability of potentials parametrized with gas phase calculations. The method presented here also allows the flexibility to force match only those portions of the model that one wants. Instead of going through the procedure of parametrizing all interactions from a training set of AIMD simulation data, one can simply use a previously parametrized model. For example, as discussed below, the previously parametrized simple point-charge flexible water (SPC/Fw) model is used to describe water−water interactions

instead of force matching a water model based on the AIMD simulation.[30]

In the next section, the AIMD simulations that were utilized for the force matching procedure are discussed along with the procedure for the MS-EVB simulations. The use of ab initio simulations in the construction of a nonreactive hydronium−water model and reactive force-matched MS-EVB model (FM-MS-EVB), which describes the dynamic proton hopping events, is then discussed. In the Results and Discussion, the results of the newly created FM-MS-EVB model, the underlying AIMD simulation, and the original MS-EVB3 model of Wu et al.[24] are compared. The Article is concluded with a discussion regarding the future outlook for the present force matching algorithm and its application to complex condensed phase reactions.

2. Methods and Development

2.1. Condensed Phase ab Initio MD Simulations. To generate data to parametrize the reactive force field, AIMD simulations were conducted on a system containing 128 water molecules plus an excess proton. The simulation cell was cubic with side lengths measuring 15.66 Å. The initial configuration for the simulation was taken from an equilibrated MD simulation of bulk water using a force-matched water model.[30] An excess proton was added, and the system was equilibrated for 8 ps using the Car−Parrinello Molecular Dynamics (CPMD) simulation method.[32] The CPMD software package (v3.5) was used for all AIMD simulations.[32,33] The HCTH/120 density functional[34] was used because it has been shown to give an improved description of experimental liquid water as compared to other density functionals.[34−36] The Kohn−Sham orbitals were expanded in a plane wave basis set with a cutoff of 80 Ry, and Troullier−Martins pseudopotentials[37] were used. The equations of motion were integrated with a time step of 3 au (0.073 fs), and the fictitious electron mass was set at 340 au. For hydrogen atoms, it was shown that fictitious electron masses around this value lead to stable CPMD simulations and a proper separation of the electronic and nuclear degrees of freedom.[35,38] During this equilibration phase, the temperature was maintained at 300 K by rescaling the velocities. From the latter portion of this trajectory, two configurations were chosen at random as the initial configurations for simulations in the microcanonical (constant NVE) ensemble. Forces on all atoms in the two NVE ensemble simulations, 27 and 35 ps in length, respectively, were then used as input to force matching algorithms to parametrize both a nonreactive and reactive hydronium−water model. Results from these simulations were discussed in an earlier publication.[39]

2.2. MS-EVB Simulations. The MS-EVB simulations were run with systems containing 216 water molecules plus an excess proton. The simulation cells were cubic with side lengths measuring 18.62 Å. The equations of motion for the FM-MS-EVB simulations were integrated with a time step of 0.5 fs using the leapfrog algorithm in a modified version of the DL_POLY (v2) software package.[40] Simulations with the nonreactive hydronium−water model and the MS-EVB3 model were run in a modified version of the LAMMPS software package using the velocity verlet algorithm and a time step of 1.0 fs.[41,42] The Ewald method was used to calculate the long-range electrostatics of the system. For each parameter set, as discussed below, the system was equilibrated using a Nose−Hoover thermostat[43] with a 0.5 ps relaxation time to maintain the temperature at 298.15 K. Once equilibrated, the final configuration from a simulation was then used to initialize a subsequent simulation in the constant NVE ensemble from which physical properties were calculated.

2.3. Force Matching the Nonreactive Hydronium Model. A reactive model describing transport of the excess proton in bulk water was constructed in several stages. First, a classical nonreactive hydronium cation model solvated by water was developed. This model was used to parametrize the diabatic states of the resulting reactive FM-MS-EVB model. The intramolecular model for the hydronium cation, H_3O^+, was based on the MS-EVB2 hydronium model[12] with the Morse bonding potential refitted to gas phase ab initio calculations. The hydronium OH stretch was modeled with a Morse oscillator potential, eq 1, with parameters D_e = 143.81 kcal/mol, $\alpha = 1.7$ Å^{-1}, and $r_0 = 1.0$ Å.

$$V(r) = D_e[(1 - e^{-\alpha(r-r_0)})^2 - 1] \quad (1)$$

A harmonic bending potential, eq 2 below, was used for the hydronium HOH angle with parameters $k_b = 73.269$ kcal/mol rad^{-1} and $\theta_0 = 116°$.

$$V(\theta) = \frac{k_b}{2}(\theta - \theta_0)^2 \quad (2)$$

Charges were fixed at −0.5e and +0.5e for the oxygen and hydrogen atoms of the hydronium cation, respectively. In ab initio simulations of water with gradient-corrected functionals, nonergodic sampling has been observed at temperatures below 330 K.[36] The melting temperature for water using the PBE and BLYP density functionals was also determined to be ∼100 K higher than the experimental value, indicating that one is actually sampling a supercooled phase at a density of 1.0 g/cm^3 and temperatures near 300 K.[44] Instead of force matching a water model based on the CPMD trajectories, the more accurate SPC/Fw[45] water model was, therefore, used to describe the water−water interactions. This is a flexible point-charge model that has been shown to reproduce a number of thermodynamic and dynamic properties as compared to experiment.[45] The remaining nonbonded interactions between the hydronium cation and water molecules were obtained from a force matching calculation using data from the ab initio trajectories.

The ab initio simulation data used for the force matching calculation was selected from configurations of the CPMD trajectory where the hydronium and its first solvation shell closely resembled the Eigen cation, $H_9O_4^+$, which donates three strong H-bonds, as opposed to the Zundel cation, $H_5O_2^+$. The Eigen cation was selected by identifying the configurations that correspond to a "special pair" coordinate $\delta_{O1x} > 0.2$ Å.[46] The δ_{O1x} coordinate is a measure of the difference between the H-bond and covalent bond distances for a specific hydronium−water pair, as illustrated in Figure

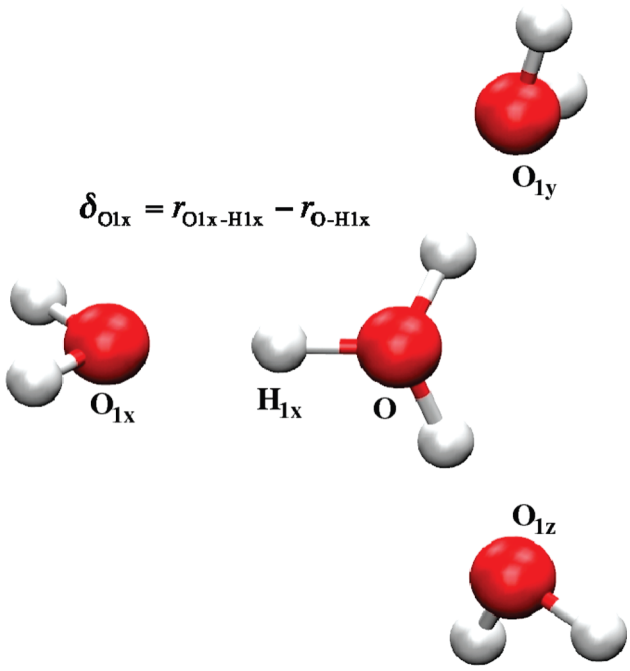

Figure 1. Hydronium cation and three H-bond accepting waters in the first solvation shell. The three water molecules are ordered according to the value of the delta parameter ($\delta_{O1x} < \delta_{O1y} < \delta_{O1z}$) for each H-bond defined as the H-bond distance minus the covalent bond distance.

1. The "O1x" labeled water with the smallest δ value identifies the water molecule that is the most likely candidate to accept a proton from the hydronium, thereby facilitating proton transfer through the Grotthus mechanism.[46] This subset of configurations with $\delta_{O1x} > 0.2$ Å was then used as the training set for the force matching calculation[31] that derived a set of effective pairwise potentials for the short-ranged nonbonded interactions between the atoms of a hydronium cation and those of a water molecule.

With this method, tabulated potentials expressed as a pairwise sum of piecewise spline functions were obtained for interactions between the atoms of the hydronium cation and water molecules. The forces determined from the hydronium intramolecular model, SPC/Fw water model, and the electrostatic interactions between the hydronium cation and water molecules, calculated using Ewald summation and metallic boundary conditions,[47] were first subtracted from the ab initio forces. The remaining contribution to the ab initio force on each atom was then used to fit the hydronium−water pairwise potentials for all oxygen and hydrogen interatomic pairs. Each pairwise potential was represented by a spline curve on a grid with a spacing of 0.1 au (~0.05 Å). The spline force curve and its first derivative were constrained to be zero at the cutoff distance for the pairwise interactions, $r_{cut} = 9.3$ Å. At short interatomic separations that were not sampled in the simulations, the force was set to a constant (the potential was linearly extrapolated) starting at values of the "core" cutoff r_{core}: 2.275 Å for OH−OW, 2.169 Å for OH−HW, 0.9525 Å for HH−OW, and 1.5345 Å for HH−HW interatomic potentials. The resulting force curves are shown as the solid black lines in Figure 2 for each interatomic pair of atoms. The first letter of the atom label defines the element, and the second letter indicates the molecule type. For example, "OH" denotes the oxygen of the hydronium cation while "HW" defines the hydrogen of a water molecule. The original spline tabulated force curves were then fit to a polynomial series to smooth out roughness due to limited statistical sampling of interatomic separations from the CPMD simulations. A least-squares minimization was used, and the fitted

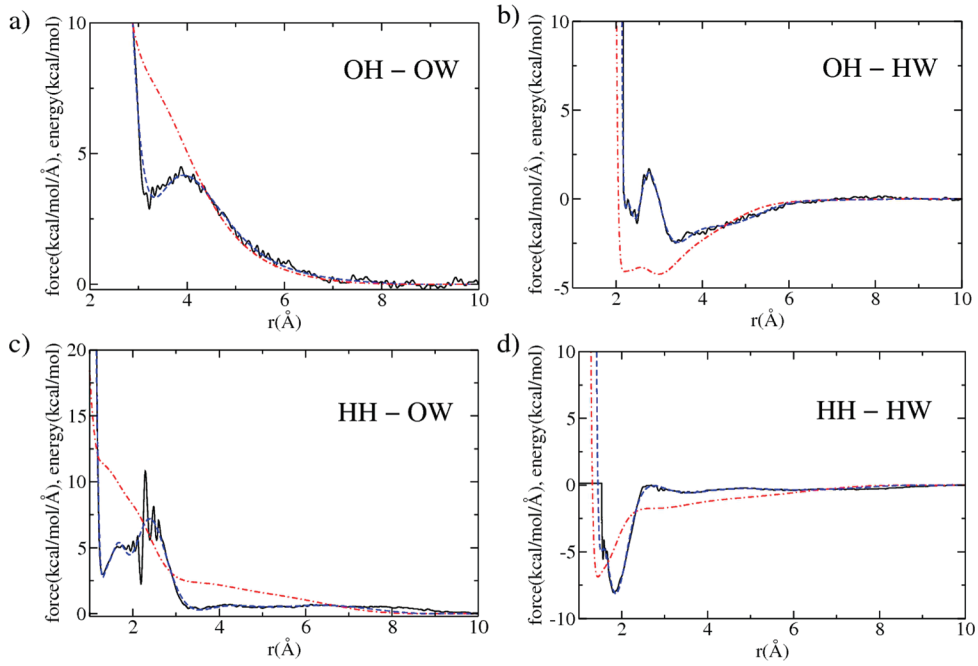

Figure 2. Pairwise forces and potentials for the hydronium−water interactions as a function of distance obtained using forces from the CPMD trajectories as input to a force matching calculation. In each plot, the solid black line is the original spline tabulated force curve as a function of interatomic distance, the dashed blue line is the smoothed force curve, and the dashed-dotted red line is the potential obtained by integrating the smoothed force.

Table 1. Coefficients from a Least Squares Fit of the Original Spline Tabulated Force Curves Generated from the Force Matching Calculation for Each Hydronium−Water Interatomic Pair[a]

order	OH−OW	OH−HW	HH−OW	HH−HW
0	−17.75179440	−245.77772117	765.85959888	−20.12874479
1	39.99415051	512.74997783	−1471.08432228	30.95649268
2	−32.19721078	−470.26747909	1298.96366415	−32.77387486
3	25.71041189	447.91419530	−1051.04732640	23.76357441
4	−18.85323611	−391.49346573	779.48446553	−18.42840895
5	12.38193946	345.09130440	−523.30121181	12.73151116
6	−6.09658945	−288.91345247	315.98084927	−6.71330105
7	3.13297827	231.82017615	−168.82081489	3.74134680
8	−1.07030405	−181.81971455	77.78832254	−1.48022490
9	0.14352517	134.02526896	−29.96910784	0.22780730
10		−96.50667993	8.76791113	
11		65.83383950	−1.61090267	
12		−41.22269725	0.13169694	
13		24.56642398		
14		−13.17741630		
15		6.82225054		
16		−2.86081399		
17		0.54515725		

[a] The resulting curves are shown in Figure 2.

curves were subject to constraints that the force and its first derivative are zero at the cutoff, $r_{cut} = 9.3$ Å. Also, the values for the core cutoff were extended to shorter distances to enable smooth extrapolation in this region: 1.5 Å for OH−OW, 1.5 Å for OH−HW, 0.7 Å for HH−OW, and 1.0 Å for HH−HW interatomic potentials.

In previous work, rapid convergence of the least-squares fitting of force curves was found when a truncated expansion of Chebyshev polynomials of the first kind as a function of the inverse interatomic distance was used to approximate the original spline tabulated data for the force:[31]

$$f(r) = \sum_{n=0}^{N} A_n T_n(\bar{r}) \tag{3}$$

where $T_n(\bar{r})$ is an nth order Chebyshev polynomial, A_n is a least-squares coefficient, and

$$\bar{r} = \frac{2/r - 1/r_{core} - 1/r_{cut}}{1/r_{core} - 1/r_{cut}} \tag{4}$$

For each curve, the highest order term included in the fit was the lowest one that yielded an overall good fit and for which the inclusion of higher order terms did not significantly reduce the fitting residual. The least-squares coefficients for the effective short-ranged nonbonded force for each interatomic pair are shown in Table 1. The fitted force curves and the resulting potential curves are shown in Figure 2. Results from simulations with the nonreactive hydronium model are discussed below and compared with results from the CPMD simulations and the MS-EVB3 model. The nonreactive hydronium model was also used to describe the diabatic states in the full FM-MS-EVB model, the parametrization of which is discussed in the next subsection.

2.4. Force Matching the Reactive MS-EVB Model. The FM-MS-EVB model developed here uses the state search algorithm and underlying functional forms of the recently developed MS-EVB3 model for proton transport in water.[24] The functions taken from the MS-EVB3 model describe the

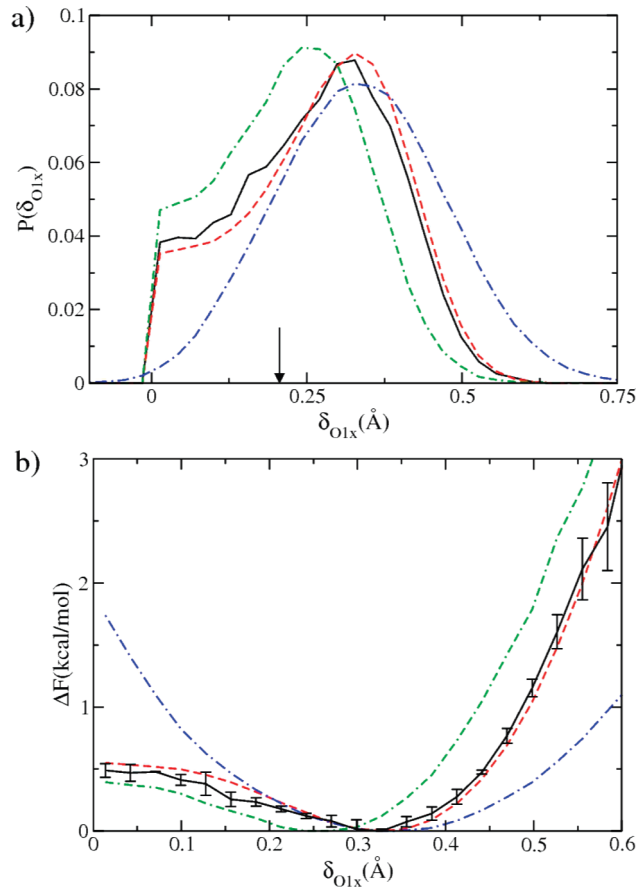

Figure 3. Probability (a) and potential of mean force (b) along the δ_{O1x} coordinate. The solid black lines were calculated from the CPMD trajectories. The dashed red lines are from the FM-MS-EVB model; the dashed-dotted blue lines are from the nonreactive hydronium model, and the double-dashed-dotted green lines are from the MS-EVB3 model. Configurations with $\delta_{O1x} > 0.2$, as indicated by the arrow, were included in the force matching calculation to determine the hydronium−water pairwise interactions for the nonreactive model. For the CPMD potential of mean force, the error bars indicate one standard deviation. Typical standard deviations for the FM-MS-EVB curve are 0.005−0.01 kcal/mol.

off-diagonal coupling of states which are overall attractive and promote successful chemical reactions. To compensate an overestimated attraction at short interatomic separations and prevent unphysical configurations from being sampled, the underlying nonreactive model is supplemented with a set of short-ranged repulsive interactions for the description of the diabatic states in the MS-EVB simulations. Using forces from the ab initio trajectories, a new set of MS-EVB3 model parameters are determined via the force matching algorithm. To ensure a sufficient sampling of all possible excess proton solvation structures, 4000 configurations were randomly sampled from the ab initio trajectories with weights determined from the δ_{O1x} probability distribution (Figure 3a). A residual, χ^2, was then constructed as a function of the forces on all atoms in the system,

$$\chi^2 = \frac{1}{3N_C N_A} \sum_{j=1}^{N_C} \sum_{i=1}^{N_A} w(r_{ij}) |\mathbf{F}_{ij} - \mathbf{F}_{ij}^{CPMD}|^2 \tag{5}$$

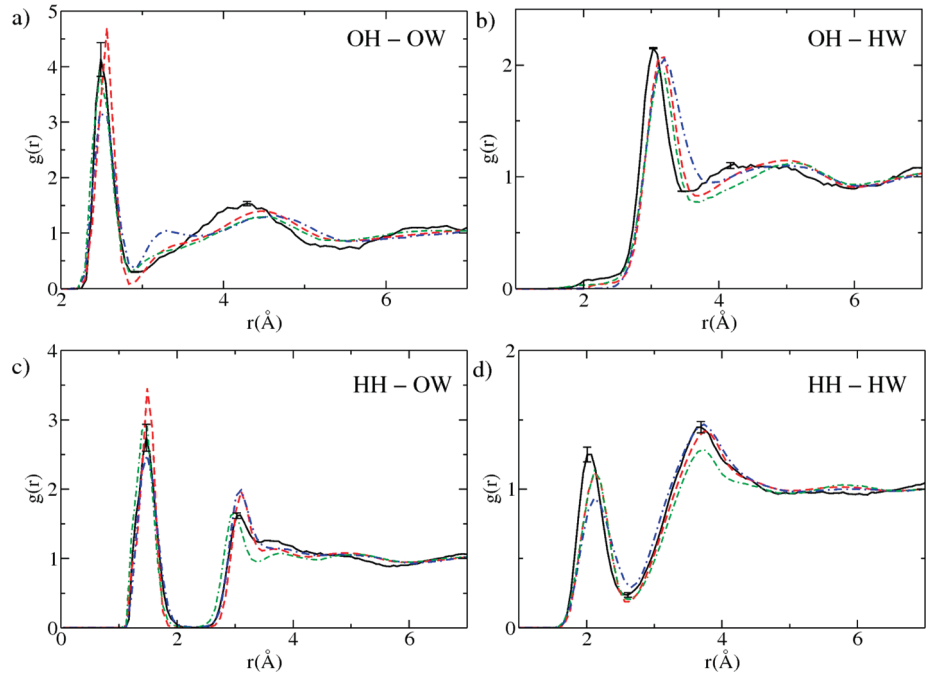

Figure 4. Radial distribution functions between atoms of the hydronium cation and water molecules. In each plot, the solid black line is calculated from the CPMD trajectories, the dashed red line is obtained from the FM-MS-EVB model, the dashed-dotted blue line is for the nonreactive hydronium model, and the double-dashed-dotted green line is for the MS-EVB3 model. Error bars for the first two peaks of the CPMD curves indicate one standard deviation.

where N_C is the number of configurations, N_A is the number of atoms, $\mathbf{F}_{ij}^{CPMD}$ are the atomic forces from the CPMD simulation, and $\mathbf{F}_{ij}$ are the FM-MS-EVB model forces for a given set of model parameters. In eq 5, a simple weighting function, $w(r)$, was used to ensure that those atoms proximal to the excess proton carried the largest weight in the calculated residual:

$$w(r) = \begin{cases} 1 & r \leq r_w \\ e^{-\alpha(r-r_w)} & r > r_w \end{cases} \quad (6)$$

where $r_w = 3.0$ Å and $\alpha = 1.54$ Å^{-1}. The parameters for the weighting function were chosen on the basis of the peak positions in the radial distribution functions (RDFs) calculated from the ab initio trajectories, shown in Figure 4. Those atoms within the first solvation shell of the excess proton have full weight, and those atoms beyond the second solvation shell have reduced weights. The use of the weighting function prevented the atoms far from the excess proton, where the main contribution to the atomic forces is due to the underlying water model, from contributing significantly to the residual. It should be noted that the MS-EVB parameter determination from the minimization of the residual in eq 5 is a nonlinear optimization, which is different from the linear algorithm described previously.[30,31]

In total, the reactive portion of the FM-MS-EVB model to be parametrized requires the determination of 24 parameters. The Nelder−Mead downhill simplex algorithm[48] was used for the simultaneous optimization of all model parameters to minimize the nonlinear residual in eq 5. Twenty sets of optimizations were carried out, and the initial solution vectors for each one were chosen by randomly sampling parameter values within specified ranges. These ranges, along with quadratic restraints, were employed to prevent the sampling of wildly unphysical solutions. For example, these restraints prevented significant sampling of a positive partial charge on oxygen atoms and exceptionally large values (>10 Å) for parameters that corresponded to nearest-neighbor oxygen−oxygen distances. Each of these separate optimizations were continued as necessary until the termination criterion (relative difference between residual of best and worst solution vectors) was less than 10^{-4}. The final results from these calculations were then used as the starting point for a new series of calculations in which the initial solution vectors were chosen as random perturbations from the previous solution. This process was repeated until further optimization lead to no significant changes in the model parameters. For all models found in this manner, molecular dynamics simulations were used to calculate physical properties, such as RDFs and potentials of mean force along the δ_{O1x} coordinate, which were compared against results from the ab initio trajectories. From this final set of solutions, the best model was identified as the one that had the best agreement with the physical properties calculated from the CPMD trajectories. The parameters for this model are listed in Table 2.

A comparison of the forces from the ab initio trajectories and those calculated with the FM-MS-EVB model is shown in Figure 5. For most data points, it can be seen that the deviations are a small fraction of the range of sampled forces. With this model, the average root-mean-squared-deviation (rmsd) for all components of the total forces on all atoms was 0.02 au. The maximum deviation in any component of a total force was 0.36 au with 0.3% of forces having deviations larger than 0.05 au. Of all the outliers in Figure 5, with deviations larger than 0.05 au, only 10% of those

Table 2. Optimized Force Field Parameters for the FM-MS-EVB Model Obtained by Minimizing the Residual in Eq 5[a]

B	3.3732 kcal/mol	γ	0.3628 Å^{-2}
b	8.3494 Å^{-1}	P	1.1066
d^0_{OO}	2.1165 Å	k	7.5707 Å^{-2}
b'	8.1908 Å^{-2}	D_{OO}	2.7128 Å
C	11.8746 kcal/mol	β	12.4834 Å^{-1}
c	9.6546 Å^{-1}	R^0_{OO}	3.1366 Å
d^0_{OH}	0.4073 Å	P'	4.2352
V^{ij}_{const}	−11.1268 kcal/mol	α	13.7322 Å^{-1}
q^{ex}_O	−0.0500 e	r^0_{OO}	2.0545 Å
q^{ex}_H	0.0167 e		
$q^{ex}_{H^*}$	0.0332 e		

[a] The switching ranges for the smooth-cutoff function for the V^{ep}_{OO} and V^{ep}_{HO} short-ranged repulsions are 2.414−2.610 and 1.335−1.536 Å, respectively. The parameters and corresponding equations are defined in the original MS-EVB3 publication.[24]

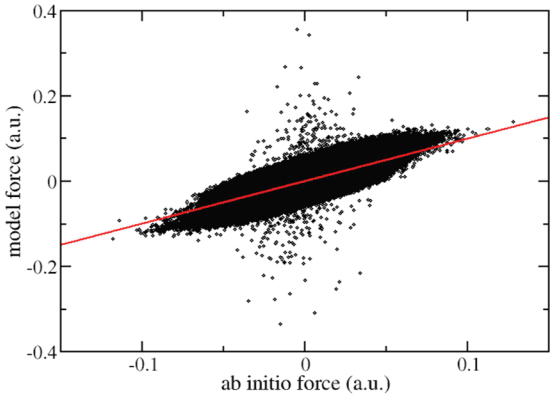

Figure 5. Comparison of total forces from the CPMD simulation calculated with the MS-EVB model for all atoms. A line of slope unity is shown to indicate where points would lie for perfect agreement. As discussed in the text, the majority of outliers are associated with water molecules beyond the first solvation shell of the excess proton.

involve components of an atomic force on the excess proton or atoms within the first solvation shell.

3. Results and Discussion

The FM-MS-EVB model was used to generate ten independent trajectories for 3 ns each in the constant NVE ensemble. The equations of motion for these FM-MS-EVB simulations were integrated with a time step of 0.5 fs in a modified version of the DL_POLY (v2) software package.[40] The MS-EVB3 state search algorithm was used. The average temperature over all simulations was 300.6 ± 1.4 K. The energy drift observed in the FM-MS-EVB simulations was 11.4 ± 2.0 kcal/mol per ns. This is smaller than the drift in the MS-EVB2 model but three times larger than the drift observed in the MS-EVB3 model, 13.1 and 3.4 kcal/mol per ns, respectively.[24] The increase in the drift of total energy as compared to the MS-EVB3 model is due in part to the hydronium−water pairwise interaction potentials which extend beyond the second solvation shell (unlike in the MS-EVB3 model) and the MS-EVB3 state search algorithm. An improved state search algorithm (work currently in progress) would need to be insensitive to the range of interaction

Table 3. Positions (Å) and Heights of the First Maximum and Minimum for the OH−OW and OH−HW Hydronium−Water RDFs[a]

OH−OW	r_{max}	g_{max}	r_{min}	g_{min}	n
CPMD	2.49	4.16	2.91	0.30	3.1
FM nonreactive	2.48	3.15	2.93	0.36	3.0
FM-MS-EVB	2.57	4.68	2.84	0.09	3.0
MS-EVB3	2.48	3.87	2.84	0.31	3.0

OH−HW	r_{max}	g_{max}	r_{min}	g_{min}	n
CPMD	3.02	2.12	3.47	0.87	9.4
FM nonreactive	3.20	2.05	3.92	0.95	15.6
FM-MS-EVB	3.20	2.06	3.65	0.83	11.2
MS-EVB3	3.11	1.95	3.74	0.77	11.4

[a] Integrated coordination numbers (n) for the first peak are also given.

potentials used in the model and the addition/removal of MS-EVB states over the course of a simulation to eliminate any drift in the total energy and ensure a proper sampling of ensemble distributions. As discussed below, good agreement was found when comparing results with the CPMD simulations for a number of equilibrium and dynamic properties. A comparison of the calculated properties for the CPMD, FM-MS-EVB, and MS-EVB3 models is discussed next.

The hydronium−water RDFs calculated with both the force-matched nonreactive and FM-MS-EVB models are shown in Figure 4. For all curves, there is general agreement with respect to the peak positions and heights between the CPMD simulations and the force-matched models. Deviations can be seen in the OH−OW RDF, Figure 4a, where there is a small, enhanced depletion, near 2.9 Å, between the first and second peaks in the FM-MS-EVB model compared to the CPMD results. This depletion is partially a result of waters accepting H-bonds from the hydronium that are somewhat too strong (short). The contribution to the RDF from these waters, particularly the waters labeled "O1z", is shifted closer to the first peak. This shift in density has the effect of increasing the height of the first peak and shifting the peak maximum toward larger separations, ∼0.1 Å, compared to the RDF calculated from the CPMD simulation (Table 3). This effect can also be seen in the first peak of the HH−OW RDF (Figure 4c) where again there is a slight shift of density toward shorter distances, increasing the height of the first peak. The integrated coordination number for the first peak of the OH−OW RDF for CPMD and the force-matched models are listed in Table 3. In line with the small shifts in density, the coordination numbers calculated from the CPMD and FM-MS-EVB simulations agree quite well with one another, 3.1 and 3.0, respectively. The coordination number calculated from the MS-EVB3 model also similarly agrees with the CPMD simulation.

The depletion between the first two peaks in the OH−OW RDF also has a contribution from waters that donate weak H-bonds to the hydronium being repelled too strongly. A signature for this effect is the presence of a peak/shoulder at separations just longer than 3 Å. This effect is much more pronounced, however, in the nonreactive hydronium model than the FM-MS-EVB model. In the OH−HW RDF, more evidence for water molecules donating H-bonds somewhat

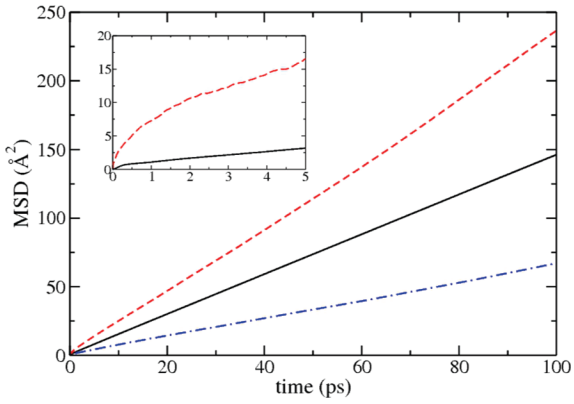

Figure 6. Mean squared displacements as a function of time for water (solid black line) and hydronium in the FM-MS-EVB (dashed red line) and nonreactive models (dashed-dotted blue line). The inset shows data from the longer CPMD trajectory for water (solid black line) and hydronium (dashed red line).

too weakly to the hydronium can be seen in the shoulder region near 2.0 Å just before the first major peak, as well as in the shift of the first peak toward larger separations. The FM-MS-EVB model recovers some of the density in this shoulder region which integrates to ∼0.1 in the CPMD simulations. This shoulder is absent in simulations with the nonreactive hydronium model. The main contribution to the OH−HW RDF from waters that donate a H-bond to the hydronium accumulates at larger separations, ∼3 Å, which gives rise to a slight shift of the first peak and first minimum toward larger separations. Again, the effect is more pronounced in the nonreactive hydronium model. The shifting of density in the OH−HW RDF compared to the CPMD simulations has a slightly larger effect on the integrated coordination numbers for the first peak. The coordination number calculated from the CPMD simulation was 9.4. The value calculated from the FM-MS-EVB and MS-EVB3 models, 11.2 and 11.4, respectively, are larger than the CPMD result, consistent with a shifting of the first peak maximum and minimum toward larger separations. The increased coordination number of the MS-EVB models as compared to CPMD is partly due to those waters donating H-bonds to hydronium and those waters that do not directly accept a H-bond from hydronium.

The peak positions and intensities for the other hydronium−water RDFs calculated from the FM-MS-EVB model similarly match well when compared to those calculated from the CPMD simulations. For comparisons between CPMD and the FM-MS-EVB model in the second solvation shell and beyond, it should be noted that the underlying water model from the CPMD simulation was replaced with the SPC/Fw model. This can make an exact comparison of the RDFs far from the hydronium cation difficult. Both the FM-MS-EVB and MS-EVB3 models use the SPC/Fw water model, and it can be seen in Figure 4 that all RDFs track one another very well in these two models beyond the second solvation shell, >5 Å.

The mean-squared displacements as a function of time for both water and hydronium are shown in Figure 6 for the force-matched nonreactive model, FM-MS-EVB model, and the CPMD simulations (inset). For water, the calculated self-diffusion coefficient of 0.22 ± 0.01 Å^2/ps from the FM-MS-EVB simulations agrees with the previously reported value for the SPC/Fw[45] model (0.23 ± 0.05 Å^2/ps) and the experimental value, 0.23 Å^2/ps.[49] The water self-diffusion coefficient calculated from the CPMD trajectories is 0.1 Å^2/ps, in agreement with previously published results.[36,50] The hydronium self-diffusion coefficient for the nonreactive model, which is solely vehicular in character (no Grotthuss shuttling), is 0.11 ± 0.01 Å^2/ps, which is a factor of 2 smaller than the diffusion of the underlying SPC/Fw water model. The net excess proton diffusion constant (including the effect of Grotthuss shuttling) calculated from the FM-MS-EVB and the longer CPMD trajectory is 0.35 ± 0.06 and 0.31 Å^2/ps, respectively. The same diffusion coefficient calculated from the MS-EVB3 model, 0.29 ± 0.03 Å^2/ps,[24] is smaller than the value from both the FM-MS-EVB and CPMD simulations. The CPMD value reported here is larger than the previously reported value of 0.21 Å^2/ps for the HCTH/120 density functional[50] that was calculated from a simulation containing 64 water molecules. All of the calculated excess proton diffusion coefficients are smaller than the experimental value of 0.94 Å^2/ps,[51] but it should be noted that the present simulations are classical and, thus, have no effects included from the quantization of the nuclear motion.[24]

The proton hopping rate was calculated as the slope of the forward proton hop accumulation function.[24] The rate calculated from the FM-MS-EVB model was 0.12 ± 0.01 ps^{-1}. This value is somewhat smaller than that estimated from the CPMD simulations, 0.4 ± 0.2 ps^{-1}; however, the limited statistics from the CPMD trajectories makes it difficult to determine this value with sufficient accuracy. The proton hopping rate obtained from the MS-EVB3 model, 0.108 ± 0.009 ps^{-1}, is a little smaller than that calculated with the FM-MS-EVB model. The smaller proton hopping rate in the MS-EVB3 model is consistent with a smaller excess proton diffusion as compared with the FM-MS-EVB model.

The probability distribution and potential of mean force along the δ_{O1x} coordinate for the CPMD, MS-EVB3, and FM-MS-EVB simulations are shown in Figure 3. It can be seen that the curves generated from the FM-MS-EVB simulations agree quite well with the CPMD results over the whole range of sampled δ_{O1x} values. The barrier for proton transfer along this coordinate for the CPMD simulation is about 0.5 kcal/mol. The barrier for the FM-MS-EVB model is only slightly higher by 0.05 kcal/mol. The potential of mean force curve for the nonreactive model is also included for comparison. The equilibrium value for δ_{O1x} for the nonreactive model, the FM-MS-EVB model, and CPMD are all near 0.33 Å. The potential of mean force calculated from the MS-EVB3 model does not agree as well with that calculated from the CPMD simulations. The barrier for proton transfer is lower by 0.1 kcal/mol, and the minimum for the well is shifted toward shorter distances by 0.08 Å compared to the CPMD result. This result especially reflects the ability of the FM-MS-EVB methodology to improve the agreement with AIMD data over that obtained with the standard MS-EVB approach alone.

4. Conclusions

In this work, a reactive MS-MD force field was successfully parametrized using data from AIMD simulations as input to a force matching algorithm. The force matching algorithm also provided the freedom to only parametrize select portions of the model to the ab initio data, mainly the hydronium−water interactions and proton hopping portions of the model. The SPC/Fw water model, which agrees well with the experiment for a number of properties, was used in place of also parametrizing a water model based on the CPMD simulations that have been shown to be less accurate.[44] The resulting reactive FM-MS-EVB model then enabled the simulation of the excess hydrated proton in water for several nanoseconds, something not currently possible with AIMD simulation techniques, and it was found to reproduce a number of thermodynamic and dynamic properties calculated from the original CPMD trajectories. A comparison of the FM-MS-EVB model with the previously empirically parametrized MS-EVB3 model was also made where calculated properties were found to qualitatively agree with each other. The excess proton self-diffusion coefficient and proton hopping rate were found to be slightly larger in the FM-MS-EVB model compared to the values calculated from the MS-EVB3 model. The improved agreement of the FM-MS-EVB model, based on parametrization from condensed phase AIMD simulations, over the MS-EVB3 model suggests that the optimal parameters for the particular set of functions used in the MS-EVB3 model can be identified via the force matching approach. As more accurate AIMD methods become tractable for the simulation of condensed phase systems, the force matching algorithm presented in this work will facilitate the generation of empirical MS-MD models that can accurately reproduce the computationally expensive ab initio simulations, thus allowing for reactive MD simulations of much larger systems and for significantly longer time scales.

A key goal of future work will be to increase the flexibility of the force matching procedure discussed here and its application to complex reactions in the condensed phase. Effort toward this goal will be directed at the incorporation of more flexible functional forms for the interactions in the MS-MD models such as interpolated functions to describe the reactive portion of the model. The use of these interpolated functions would replace the use of functional forms that may be nonlinear with respect to the model parameters and, therefore, difficult to generalize to describe other chemical reactions. For example, in the work presented here, these interpolated functions would replace the empirical functions taken from the MS-EVB3 functional form, which constrain the range of accuracy of the resulting FM-MS-EVB model. The optimization of a nonlinear function of several variables can also be a difficult task due to the possible presence of many unphysical local minima and poor convergence properties. By constructing a model where the atomic forces are linear (or close too) with respect to all model parameters, one will be able to take advantage of linear least-squares algorithms[30,31] to systematically and variationally determine those parameters that best represent the training set of data used to develop a reactive model.

Acknowledgment. This research was supported by the National Science Foundation (CHE-1036464). This work was supported in part by a grant of computer time from the DOD High Performance Computing Modernization Program at the Navy and Army Research Laboratory DOD Supercomputing Resource Centers. This research was also supported in part by the National Science Foundation Teragrid computing resources provided by the Texas Advanced Computing Center under Grant Number TG-MCA94P017. The authors thank Dr. Jessica Swanson for helpful discussions.

References

(1) Grotthuss, C. J. T. d. *Ann. Chim. (Paris)* **1806**, *58*, 54.

(2) Agmon, N. *Chem. Phys. Lett.* **1995**, *244*, 456.

(3) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. *J. Chem. Phys.* **1995**, *103* (1), 150.

(4) Marx, D.; Tuckerman, M. E.; Hütter, J.; Parrinello, M. *Nature* **1999**, *397*, 601.

(5) Marx, D. *ChemPhysChem* **2006**, *7* (9), 1848.

(6) Pisani, C.; Maschio, L.; Casassa, S.; Halo, M.; Schütz, M.; Usvyat, D. *J. Comput. Chem.* **2008**, *29* (13), 2113.

(7) Marsman, M.; Grüneis, A.; Paier, J.; Kresse, G. *J. Chem. Phys.* **2009**, *130* (18), 184103.

(8) Duin, A. C. T. v.; Dasgupta, S.; Lorant, F.; Goddard, W. A., III *J. Phys. Chem. A* **2001**, *105* (41), 9396.

(9) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48* (7), 1198.

(10) Schmitt, U. W.; Voth, G. A. *J. Phys. Chem. B* **1998**, *102* (29), 5547.

(11) Schmitt, U. W.; Voth, G. A. *J. Chem. Phys.* **1999**, *111* (20), 9361.

(12) Day, T. J. F.; Soudackov, A. V.; Čuma, M.; Schmitt, U. W.; Voth, G. A. *J. Chem. Phys.* **2002**, *117* (12), 5839.

(13) Voth, G. A. *Acc. Chem. Res.* **2006**, *39* (2), 143.

(14) Swanson, J. M. J.; Maupin, C. M.; Chen, H.; Petersen, M. K.; Xu, J.; Wu, Y.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111* (17), 4300.

(15) Marx, D.; Chandra, A.; Tuckerman, M. E. *Chem. Rev.* **2010**, *110*, 2174.

(16) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218.

(17) Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523.

(18) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425.

(19) Warshel, A. In *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley and Sons: New York, NY, 1991; pp 1−236.

(20) Guo, Y.; Thompson, D. L. *J. Chem. Phys.* **2003**, *118* (4), 1673.

(21) Nelson, K. V.; Benjamin, I. *J. Phys. Chem. C* **2010**, *114* (2), 1154.

(22) Kim, Y.; Corchado, J. C.; Villà, J.; Xing, J.; Truhlar, D. G. *J. Chem. Phys.* **2000**, *112* (6), 2718.

(23) Higashi, M.; Truhlar, D. G. *JCTC* **2009**, *5* (11), 2925.

(24) Wu, Y.; Chen, H.; Wang, F.; Paesani, F.; Voth, G. A. *J. Phys. Chem. B* **2008**, *112*, 7146.

(25) Brewer, M. L.; Schmitt, U. W.; Voth, G. A. *Biophys. J.* **2001**, *80* (4), 1691.

(26) Maupin, C. M.; Wong, K. F.; Soudackov, A. V.; Kim, S.; Voth, G. A. *J. Phys. Chem. A* **2006**, *110*, 631.

(27) Maupin, C. M.; McKenna, R.; Silverman, D. N.; Voth, G. A. *J. Am. Chem. Soc.* **2009**, *131* (22), 7598.

(28) Ercolessi, F.; Adams, J. B. *Eruophys. Lett.* **1994**, *26* (8), 583.

(29) Tangney, P.; Scandolo, S. *J. Chem. Phys.* **2002**, *117* (19), 8898.

(30) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120* (23), 10896.

(31) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109* (14), 6573.

(32) Marx, D.; Hutter, J. Ab initio molecular dynamics: Theory and Implementation. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing (NIC); Forschungszentrum Jülich: Jülich,Germany, 2000; Vol. 1, pp 301−449.

(33) CPMD, http://www.cpmd.org, Copyright IBM Corp 1990-2008, Copyright MPI für Festkörperforschung Stuttgart 1997-2001 (accessed March 1, 2002).

(34) Boese, A. D.; Doltsinis, N. L.; Handy, N. C.; Sprik, M. *J. Chem. Phys.* **2000**, *112* (4), 1670.

(35) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 044505.

(36) VandeVondele, J.; Mohamed, F.; Krack, M.; Hutter, J.; Sprik, M.; Parrinello, M. *J. Chem. Phys.* **2005**, *112*, 014515.

(37) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43* (3), 1993.

(38) Grossman, J. C.; Schwegler, E.; Draeger, E. W.; Gygi, F.; Galli, G. *J. Chem. Phys.* **2004**, *1120*, 300.

(39) Maupin, C. M.; Aradi, B.; Voth, G. A. *J. Phys. Chem. B* **2010**, *114* (20), 6922.

(40) Smith, W.; Forester, T. R. *J. Mol. Graphics* **1996**, *14* (3), 136.

(41) LAMMPS, http://lammps.sandia.gov (accessed February 1, 2010).

(42) Plimpton, S. J. *J. Comput. Phys.* **1995**, *117*, 1.

(43) Nosé, S. *J. Chem. Phys.* **1984**, *81* (1), 511.

(44) Yoo, S.; Zeng, X. C.; Xantheas, S. S. *J. Chem. Phys.* **2009**, *130* (22), 221102.

(45) Wu, Y.; Tepper, H. L.; Voth, G. A. *J. Chem. Phys.* **2006**, *124* (2), 024503.

(46) Markovitch, O.; Chen, H.; Izvekov, S.; Paesani, F.; Voth, G. A.; Agmon, N. *J. Phys. Chem. B* **2008**, *112* (31), 9456.

(47) Frenkel, D.; Smit, B. In *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: San Diego, CA, 2002; pp 291−316.

(48) Nelder, J. A.; Mead, R. *Comput. J.* **1965**, *7* (4), 308.

(49) Krynicki, K.; Green, C. D.; Sawyer, D. W. *Faraday Discuss. Chem. Soc.* **1978**, *66*, 199.

(50) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123* (4), 044505.

(51) Roberts, N. K.; Northey, H. L. *J. Chem. Soc., Faraday Trans.* **1974**, *1* (70), 253.

CT1004438