JCTC Journal of Chemical Theory and Computation

Efficient Multistate Reactive Molecular Dynamics Approach Based on Short-Range Effective Potentials

Hanning Chen,[†,‡] Pu Liu,[†,§] and Gregory A. Voth*[,∥]

Department of Chemistry, James Franck Institute, and Computation Institute, University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637, Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208-3113, and Johnson & Johnson Pharmaceutical Research & Development, 665 Stockton Drive, Exton, Pennsylvania 19341

Received June 11, 2010

Abstract: Nonbonded interactions between molecules usually include the van der Waals force and computationally expensive long-range electrostatic interactions. This article develops a more efficient approach: the effective-interaction multistate empirical-valence-bond (EI-MS-EVB) model. The EI-MS-EVB method relies on a mapping of all interactions onto a short-range and thus, computationally efficient effective potential. The effective potential is tabulated by matching its force to known trajectories obtained from the full-potential empirical multistate empirical-valence-bond (MS-EVB) model. The effective pairwise interaction depends on and is uniquely determined by the atomic configuration of the system, varying only with respect to the hydrogen-bonding topology. By comparing the EI-MS-EVB and full MS-EVB calculations of several equilibrium and dynamic properties important to hydrated excess proton solvation and transport, we show that the EI-MS-EVB model produces very accurate results for the specific system in which the tabulated potentials were generated. The EI-MS-EVB potential also transfers reasonably well to similar systems with different temperatures and box sizes. The EI-MS-EVB method also reduces the computational cost of the nonbonded interactions by about 1 order of magnitude in comparison with the full algorithm.

1. Introduction

An empirical force field must either explicitly or implicitly account for all of the electrostatic interactions between charged particles. The long-range electrostatic interaction,[1−5] which decays as an inverse function of the interparticle distance r, is a crucial element of many molecular simulations—especially for highly charged biological systems such as DNA.[6−8] Reliable and accurate calculations of the Coulomb force are possible using lattice summation methods, including the original Ewald summation,[9] particle−particle particle−mesh (P3M),[10,11] and particle−mesh Ewald (PME) algorithms.[12] Numerous simulations based on these techniques have produced results in good agreement with experimental data. Even with carefully chosen parameters, however, the computational cost of the original Ewald summation is $O(N^{3/2})$,[13] with N being the number of charged particles in the system. By employing the fast Fourier transform, P3M and PME significantly accelerate long-range electrostatic calculations and reduce the cost to $O(N \log N)$.[10,11] Although all three methods are much faster than a direct evaluation of all possible pairs, they are still very computationally expensive for large systems,[14] and thus may not be well suited for highly scalable computations.

Fast multipole methods (FMMs), an alternative approach to the Ewald algorithms, are based on the multipole expansion of the electrostatic potential.[15−18] Although their asymptotic computational scaling is claimed to be $O(N)$, much better than that of particle−mesh methods, in a trial

* Corresponding author e-mail: gavoth@uchicago.edu.
[†] These authors contributed equally to this work.
[‡] Northwestern University.
[§] Johnson & Johnson Pharmaceutical Research & Development.
[∥] University of Chicago.

10.1021/ct100318f
Published on Web 08/31/2010

conducted by Pollock and Glosli, the latter actually outperformed FMMs for all systems studied.[17] Furthermore, it is not easy to obtain a reasonably good conservation of the total system energy using FMMs. A very high-order expansion must be used, at the cost of significantly more CPU (central processing unit) cycles. However, as pointed out by one of the reviewers, some recent progress has substantially reduced the prefactor implicit in the O(N) notation,[19,20] making the improved FMMs quite competitive.

Among the various ways to improve the efficiency of long-range force calculations, the most straightforward and convenient one is to truncate the interaction beyond a certain (usually fairly short) distance. In other words, in this approach, any nonbonded forces between particles separated by more than a certain cutoff distance are simply neglected.[21,22] At the cutoff distances typically used in simulations, however, the long-range force is not yet sufficiently small to be ignored. The resulting discontinuity is known to produce severe artifacts in various properties of the simulated systems.[23−36] Although the discontinuity can be diminished by smoothing or shifting functions, such methods cannot completely remove the artifacts.[37−40]

Alternatively, interactions between particles at large length scales can be treated as a homogeneous dielectric medium in the mean-field sense. The contribution from all molecules beyond the cutoff distance can then be approximated using a Barker−Watts reaction-field (RF) correction.[24,29,41−45] The problem with this approach is that many systems of interest (e.g., aqueous solutions of large proteins) exhibit intrinsically heterogeneous dielectric properties. In such cases, RF corrections with a short cutoff will induce significant artifacts. These errors can be reduced by setting a large cutoff, but this solution increases the computational expense, thereby limiting the applicability of the RF method.

The above summary is consistent with the concept that expensive, explicit computations are both necessary and inevitable for an accurate description of long-range interactions. However, recent thinking suggests that the effective electrostatic interaction in condensed systems might decay substantially faster than an inverse power of r. Indeed, several short-range potentials have been constructed that appear to account for the long-range electrostatic interaction.[46−48] The damped, force-shifted (DFS) potential[46] is one good example; not only is it computationally efficient, with $O(N)$ computational cost, but it satisfactorily reproduces some thermostatic and dynamic properties for a variety of systems, including argon in water, liquid water, crystalline water, NaCl crystals, and NaCl aqueous solutions. A weakness of the DFS potential is that some of its key variables (such as the damping parameter itself) can be optimized only by trial and error. Moreover, the fact that it relies on a predefined analytical function might limit the DFS potential's ability to represent the "best" effective short-range interaction in a wide range of systems.

Very recently, a different and systematic method for determining the optimal effective short-range potential was developed.[48,49] This method is called coarse-graining in interaction space (CGIS). The goal is to minimize differences between the forces generated by a candidate effective short-range potential and those generated by the full Coulomb potential. This innovative approach represents a "bottom-up" strategy, fundamentally different from the more common "top-down" strategies described previously. This force-matching (FM) algorithm[50,51] takes into account the effects of the full Coulomb interaction, generating a statistically accurate effective[48,49] potential. Furthermore, the new potential need not be a predefined analytical function. This method has successfully reproduced several important structural, energetic, and dynamical properties of bulk water systems.[48] The subsequent work by Shi et al.[49] took the theory a step further by analytically demonstrating that the effective short-range force approaches zero naturally at the FM cutoff and by deriving its corresponding analytic approximation.

Turning now to a more specific focus, the multistate empirical-valence-bond (MS-EVB)[52−57] method represents a significant theoretical and computational advance toward understanding proton solvation and transport (PS&T). This phenomenon is critical to many chemical, biochemical, and biophysical systems. The MS-EVB method has provided statistically reliable and accurate descriptions of excess hydrated proton behavior in a wide variety of environments, including small water clusters[58] and the bulk and interfacial water phase,[54,55,59] the aquaporin channels,[60,61] the influenza A virus M2 channel,[62] cytochrome c oxidase,[63] and liquid-phase imidazole.[64] The MS-EVB approach describes proton transport as a multipathway and multistep reaction, wherein distinct reactant-like (or product-like) intermediates with different chemical and hydrogen-bonding topologies are represented as individual states of an MS-EVB Hamiltonian matrix. The diagonal elements of the matrix represent the diabatic energies of the MS-EVB states, and the off-diagonal elements provide the coupling between any pair of states. The MS-EVB method is a sort of multistate or multiconfiguration molecular dynamics (MD) approach that allows chemically reactive processes such as Grotthuss proton shuttling[56,57] to be modeled.

After diagonalization of the MS-EVB Hamiltonian and identification of the eigenvector with the lowest eigenvalue, the system nuclei are propagated with MD in accordance with the Hellmann−Feynman theorem. Empirical potentials and atomistic forces (hereafter denoted E&F) need to be calculated for each element in the EVB matrix, except for those off-diagonal elements that correspond to noncoupling MS-EVB states.[52,54,55] In aqueous solutions, for example, approximately 30 MS-EVB states and roughly the same number of MS-EVB coupling pairs are typically required to describe the delocalized excess proton charge defect and to satisfactorily conserve total system energy. Thus, one should expect about 60 computationally expensive long-range electrostatic lattice summations to be carried out for each time step. This calculation is the dominant source of CPU cycles in an MS-EVB calculation. A new approach for accurately and efficiently evaluating the long-range electrostatic interactions in the MS-EVB method is therefore highly desirable for many interesting applications, especially those involving an implementation within a highly scalable computing environment.

The CGIS algorithm folds long-range electrostatic interactions into short-range effective potentials.[48,49] In the present study, we demonstrate that this algorithm can be generalized to greatly accelerate MS-EVB calculations by simplifying most of the nonbonded interactions. This demonstration is important because proton transport as described by the MS-EVB model has various exponentially sensitive properties (such as diffusion barriers) that might not be properly modeled by more ad hoc effective interaction schemes. For a given atomistic configuration, distinct MS-EVB states differ only in their hydrogen-bonding topology. Thus, once the E&F have been determined for one EVB state, those of other states can be calculated conveniently and rapidly by identifying the few atom pairs that differ between the states. It is therefore not surprising that such short-range effective potentials are a particularly efficient tool for MS-EVB calculations. As will be demonstrated later, this "effective-interaction" MS-EVB (EI-MS-EVB) approach can substantially reduce the cost of nonbonded calculations in the MS-EVB calculations while producing structural, energetic, and dynamical properties comparable to those of the original full-potential (FP) MS-EVB model. More importantly, the improvement in computational efficiency becomes pronounced in larger systems.

The remaining sections of this article are organized as follows: Section 2 provides an overview of the CGIS FM algorithm and the MS-EVB methodology, followed by a detailed explanation of the construction of effective short-range potentials. Section 3 compares some physical properties of systems simulated using the EI-MS-EVB method and the regular (FP) MS-EVB method and discuss the efficiency of the algorithms. Conclusions and future directions are given in section 4.

2. Methods

2.1. Force-Matching Algorithm. The FM algorithm[50,51] has previously been used to construct short-range effective potentials via the CGIS methodology.[48,49] Given that the system itself is not coarse-grained and that nonbonded atomistic potentials are usually pairwise and additive, this procedure can produce very accurate effective potentials. Indeed, this accuracy has been demonstrated for various systems.[48]

Briefly, the optimal parameters $\{g_m\}$ of an effective short-range interaction can be systematically derived from a known trajectory by variationally minimizing the residual function

$$\chi^2(\{g_m\}) = \frac{1}{N}\sum_{i,l} \| \vec{F}_{i,l}^{\,\mathrm{ref}} - \vec{F}_{i,l}^{\,\mathrm{eff}}(g_m, r_c)\|^2$$

In this formula, N is the number of atoms in the system, $\vec{F}_{i,l}^{\,\mathrm{ref}}$ is the known force acting on atom i in frame l of the reference trajectory, and

$$\vec{F}_{i,l}^{\,\mathrm{eff}}(g_m, r_c) = \sum_{d=1}^{N_p} \phi_{i,l}^{d}(r^n) g_d$$

is the effective force calculated from the configuration r^n given the parameter set $\{g_m\}$. In the effective force formula, N_p is the number of elements in $\{g_m\}$, r^n represents the coordinates of the atoms, and $\phi_{i,l}^{d}(r^n)$ is a vector whose elements depend on the system configuration. Beyond the cutoff distance r_c, the effective force $\vec{F}_{i,l}^{\,\mathrm{eff}}(g_m,r_c)$ is defined to be zero. A more detailed description of the force-matching procedure can be found elsewhere.[48,50] It is worth emphasizing that, because $\vec{F}_{i,l}^{\,\mathrm{ref}}$ is deduced from a real MD trajectory, it includes contributions from the whole system. Even though the effective force $\vec{F}_{i,l}^{\,\mathrm{eff}}(g_m,r_c)$ includes contributions only from particles within a spherical region around atom i, it is constructed to imitate $\vec{F}_{i,l}^{\,\mathrm{ref}}$ as closely as possible and therefore also includes information on the long-range interactions.

Mathematically, the residual function can be rewritten as a multidimensional quadratic function of the parameter set $\{g_m\}$[65−67]

$$\chi^2 = \sum_{d,d'} G_{dd'} g_d g_{d'} - 2\sum_d b_d g_d + \chi_0^{\,2} \tag{1}$$

where

$$\begin{aligned} G_{dd^1} &= \frac{1}{N}\left\langle \sum_i \sum_l \varphi_{i,l}^{d}(r^n)\cdot\varphi_{i,l}^{d^1}(r^n)\right\rangle \\ b_d &= \frac{1}{N}\left\langle \sum_i \sum_l \varphi_{i,l}^{d}(r^n)\cdot\vec{F}_{i,l}^{\,\mathrm{ref}}\right\rangle \end{aligned} \tag{2}$$

$$\chi_0^{\,2} = \frac{1}{N}\left\langle \sum_i \sum_l \vec{F}_{i,l}^{\,\mathrm{ref}}\cdot\vec{F}_{i,l}^{\,\mathrm{ref}}\right\rangle \tag{3}$$

Differentiating the residual function with respect to the parameter set yields a normal system of equations from which the effective short-range force parameters can be determined as

$$\sum_{d^1=1}^{N_p} G_{dd^1} g_{d^1} = b_d \tag{4}$$

for $d = 1, ..., N_p$.

Note that $\vec{F}_{i,l}^{\,\mathrm{ref}}$ and $\vec{F}_{i,l}^{\,\mathrm{eff}}(g_m,r_c)$ need not include all of the force components. As the present work will demonstrate, the formalism is applicable even when only part of the force is considered.

2.2. MS-EVB Multistate MD Method. Because a hydrated excess proton is shared among several solvating water molecules (charge defect delocalization), it is not appropriate to simulate PS&T using conventional empirical force fields with MD. In a manner similar to quantum mechanics, the MS-EVB framework describes delocalized hydrated excess protons as a linear combination of basis states $|i\rangle$ corresponding to distinct hydrogen-bonding topologies.[53−55] Given a configuration of system nuclei $\mathbf{x}$, the state function $|\Psi\rangle$ of a delocalized excess proton is written as

$$|\Psi\rangle = \sum_i^N c_i(\mathbf{x})|i\rangle \tag{5}$$

where N is the total number of basis states and c_i represents the normalized state coefficients. The MS-EVB state am-

plitude squared, c_i^2, represents the probability of finding the system in state $|i\rangle$. The state function $|\Psi\rangle$ at a given instant is determined by a Hamiltonian matrix $\mathbf{H}$ whose elements h_{ij} are typically described in terms of an empirical force field. Given $\mathbf{H}$, the ground state $|\Psi\rangle$ and lowest-energy $\mathbf{E}(\mathbf{x})$ eigenstate can be obtained by solving the eigenvalue problem

$$\mathbf{c}^{\mathrm{T}}\mathbf{H}\mathbf{c} = \mathbf{E}(\mathbf{x}) \tag{6}$$

where $\mathbf{c}$ is the N-dimensional eigenvector with elements c_i, $i = 1, 2, ..., N$, as described in eq 5. According to the Hellmann−Feynman theorem, the force exerted on atom i for a given nuclear configuration can be expressed as

$$\mathbf{F}_i(\mathbf{x}) = -\left\langle \Psi_0 \left| \frac{\partial \mathbf{H}}{\partial \mathbf{x}_i} \right| \Psi_0 \right\rangle = -\sum_{m,n} c_m c_n \frac{\partial h_{mn}(\mathbf{x})}{\partial \mathbf{x}_i} \tag{7}$$

2.3. Effective Short-Range Forces for the EI-MS-EVB Model. Under a full-potential model (denoted here simply as the MS-EVB model), the diagonal elements $h_{ii}(\mathbf{x})$ of the Hamiltonian matrix $\mathbf{H}$ are described by the potential energy function[55]

$$h_{ii} = V_{\mathrm{H_3O^+}}^{\mathrm{intra}} + \sum_{k}^{N_{\mathrm{H_2O}}} V_{\mathrm{H_2O}}^{\mathrm{intra},k} + \sum_{k}^{N_{\mathrm{H_2O}}} V_{\mathrm{H_3O^+,\,H_2O}}^{\mathrm{inter},k} + \sum_{k<k'}^{N_{\mathrm{H_2O}}} V_{\mathrm{H_2O}}^{\mathrm{inter},kk'} \tag{8}$$

The term $V_{\mathrm{H_3O^+}}^{\mathrm{intra}}$ represents the intramolecular potentials for hydronium, and $\sum_k^{\mathrm{N_{H_2O}}} V_{\mathrm{H_2O}}^{\mathrm{intra},k}$ is the intramolecular potential of the flexible, simple point-charge underlying water (SPC/Fw) model developed by Wu et al.[55] The term $\sum_{k<k'}^{N_{\mathrm{H_2O}}} V_{\mathrm{H_2O}}^{\mathrm{inter},kk'}$ represents all nonbonded interactions between water molecules, and the term $V_{\mathrm{H_3O^+,H_2O}}^{\mathrm{inter},k}$ represents the intermolecular potential between hydronium and water molecules and can be written as[55]

$$V_{\mathrm{H_3O^+,H_2O}}^{\mathrm{inter},k} = 4\varepsilon_{\mathrm{OO_w}}\left[\left(\frac{\sigma_{\mathrm{OO_w}}}{R_{\mathrm{OO}_k}}\right)^{12} - \left(\frac{\sigma_{\mathrm{OO_w}}}{R_{\mathrm{OO}_k}}\right)^{6}\right] + 4\varepsilon_{\mathrm{HO_w}}\left[\left(\frac{\sigma_{\mathrm{HO_w}}}{R_{\mathrm{HO}_k}}\right)^{12} - \left(\frac{\sigma_{\mathrm{HO_w}}}{R_{\mathrm{HO}_k}}\right)^{6}\right] + \sum_{m}^{4}\sum_{n_k}^{3} \frac{q_m^{\mathrm{H_3O^+}} q_{n_k}^{\mathrm{H_2O}}}{R_{mn_k}} + V_{\mathrm{OO}_k}^{\mathrm{rep}} + V_{\mathrm{HO}_k}^{\mathrm{rep}} \tag{9}$$

In addition to the standard Lennard-Jones (LJ) and Coulomb potentials, two repulsive terms ($V_{\mathrm{OO}_k}^{\mathrm{rep}}$ and $V_{\mathrm{HO}_k}^{\mathrm{rep}}$) are required to correctly describe interactions between hydronium ion and the water molecules in its first solvation shell. In practical terms, the repulsive terms improve consistency with the high-level ab initio potential energy surface. For the sake of simplicity, only $\sum_{k<k'}^{N_{\mathrm{H_2O}}} V_{\mathrm{H_2O}}^{\mathrm{inter},kk'}$ and the first three terms of $V_{\mathrm{H_3O^+,H_2O}}^{\mathrm{inter}.k}$ (eq 9) are included in the effective interaction FM procedure. The intramolecular terms $V_{\mathrm{H_3O^+}}^{\mathrm{intra}}$ and $V_{\mathrm{H_2O}}^{\mathrm{intra},k}$, as well as the repulsive terms $V_{\mathrm{OO}_k}^{\mathrm{rep}}$ and $V_{\mathrm{HO}_k}^{\mathrm{rep}}$, are many-body, already short-range, and computationally efficient. They can therefore be calculated directly from the empirical functions, by means described in another work.[55]

The off-diagonal (coupling) elements $h_{ij}(\mathbf{x})$ ($i \neq j$) in eq 6 are defined to be nonzero only when the hydronium ions of states $|i\rangle$ and $|j\rangle$ share a transferring proton. These elements can be expressed as[55]

$$h_{ij} = (V_{\mathrm{const}}^{ij} + V_{\mathrm{ex}}^{ij})A(R_{\mathrm{OO}}, \mathbf{q}) \tag{10}$$

where V_{const}^{ij} is a constant and V_{ex}^{ij} is the electrostatic potential between the $H_5O_2^+$ Zundel complex and the remaining water molecules.[55] The latter term is given by

$$V_{\mathrm{ex}}^{ij} = \sum_{m}^{7}\sum_{k}^{N_{\mathrm{H_2O}}-1}\sum_{n_k}^{3} \frac{q_{n_k}^{\mathrm{H_2O}} q_m^{\mathrm{ex}}}{R_{mn_k}} \tag{11}$$

where $q_{\mathrm{m}}^{\mathrm{ex}}$ represents exchange charges of the $H_5O_2^+$ complex and $q_{n_k}^{\mathrm{H_2O}}$ is the atomic charge of water molecule k. The term $A(R_{\mathrm{OO}},\mathbf{q})$ in eq 10 is a geometric scale factor dependent on the positions of the atoms forming the $H_5O_2^+$ complex. Thus, only the exchange charge term V_{ex}^{ij} is involved in the construction of an effective short-range force for off-diagonal terms.

Unless otherwise specified, the construction of effective short-range potentials and all simulations (using both the EI-MS-EVB and MS-EVB models) were carried out in a cubic volume with 216 water molecules and one excess proton at 298.15 K. The box size was 18.621 Å, yielding an average density of 1.0 g/cm^3. For thermostatic calculations, a constant-*NVT* ensemble of simulations with a Nosé−Hoover themostat[68] was used. For dynamical properties, the constant-*NVE* ensemble was employed.

In the MS-EVB simulations, long-range electrostatic forces were treated by Ewald summation with a relative tolerance of 10^{-6}. A spherical cutoff of 9.0 Å was chosen for the LJ interactions. In the EI-MS-EVB simulations, the short-range effective potentials employed a slightly longer spherical cutoff radius of 9.24 Å, which is one-half of the simulation box size of the atomistic system for which the effective potential is derived (i.e., the largest possible cutoff that could be used). A leapfrog algorithm was applied to integrate the equations of motion, with a time step of 1 fs. In total, 48 ns of simulation data were produced, consisting of six independent 4-ns simulations for both MS-EVB and EI-MS-EVB cases. All simulations were performed using the DL_EVB program,[69] derived from the DL_POLY package.[70]

3. Results

3.1. Instantaneous Atomistic Forces. Because the present study aims to improve the efficiency of MS-EVB simulations without sacrificing accuracy, the short-range effective potentials should be assessed in terms of force deviations between the EI-MS-EVB and MS-EVB models for a given configuration of the system nuclei. The force deviation ΔF_{ix} for an atom i along the x axis can be expressed as

$$\Delta F_{ix} = \mathrm{FM}_{ix} - \mathrm{FP}_{ix} \tag{12}$$

where the forces FM_i and FP_i are calculated under the EI-MS-EVB and MS-EVB models, respectively. As all systems in the present study are isotropic, the observed distributions of ΔF_i are indistinguishable along the three Cartesian axes (data not shown). Consequently, all distributions of ΔF_i presented in this article have been averaged over the three axes. The standard deviation $\delta(\Delta F)$ is calculated by the equation

$$\delta(\Delta F) = \sqrt{\frac{1}{M}\frac{1}{N}\sum_{j=1}^{M}\sum_{i=1}^{N}(\Delta F_{i,j} - \overline{\Delta F_{i,j}})^2} \qquad (13)$$

where M is the number of atomistic configurations and N is the number of atoms in the system. To ensure statistical reliability, 6000 configurations were randomly selected from each of the six EI-MS-EVB trajectories, which were obtained at temperatures of 260, 270, 280, 298.15, 310, and 320 K. The resulting ΔF_i distributions and the corresponding $\delta(\Delta F)$ values are shown in Figure 1. Although the wings of the force deviation distributions are slightly wider than the standard Gaussian, the distributions have a zero mean for all six temperatures, demonstrating the absence of systematic error in the effective short-range potentials. Moreover, the standard deviations indicate that about 90% of the individual ΔF_i values are less than 0.5 kcal·mol^{-1}·Å^{-1} in absolute value. (The probability of obtaining a ΔF_i magnitude larger than 1.0 kcal·mol^{-1}·Å^{-1} is only about 1%.) The average value of $\delta(\Delta F)$ is approximately 0.3 kcal·mol^{-1}·Å^{-1}, only 2% of the average force (14 kcal·mol^{-1}·Å^{-1}) exerted on the atoms.

To further validate the effective short-range potentials, several systems were simulated with larger box sizes at 298.15 K (but otherwise similar to those described above). The greater volume substantially increases the number of atoms beyond the cutoff radius. As depicted in Figure 2, the variance of ΔF_{ix} generally increases with box size, except for the largest system with a box size of 37.106 Å, which exhibits a $\delta(\Delta F_{ix})$ value similar to that of the medium-sized systems. The moderate increase is expected because the effective short-range potentials were based solely on information available within the original-size system, which was relatively small. Nevertheless, $\delta(\Delta F)$ increased by only 0.11 kcal·mol^{-1}·Å^{-1} (Figure 2) when the box size nearly doubled from 18.621 to 37.106 Å. Given the intrinsic limitations of the accuracy of the underlying empirical force field, errors in the range of a few percent can be viewed as being essentially negligible. The effective potentials generated by the present force-matching scheme should therefore be reasonably transferable to other temperatures and box sizes.

3.2. Radial Distribution Functions. The solvation structure of a hydrated excess proton essentially governs the extent of its charge defect delocalization, which can be characterized by a radial distribution function (RDF). Figure 3 compares the RDFs obtained in EI-MS-EVB and MS-EVB simulations. With the exception of RDF(O*−H), where the EI-MS-EVB and MS-EVB models deviate in the rarely sampled core region (<3.0 Å), the two approaches produce almost indistinguishable descriptions of excess proton solvation in water. The EI-MS-EVB method generates a somewhat lower probability of finding water hydrogen atoms close to the hydronium oxygen at short distances (Figure 3b), but this feature is not directly related to the process of proton transport.

3.3. Free Energy Profile of Proton Transfer. The free energy profile of proton transfer between two water molecules can be calculated by the equation

$$\Delta E(c_1^2 - c_2^2) = -RT\ln[P(c_1^2 - c_2^2)] - C \qquad (14)$$

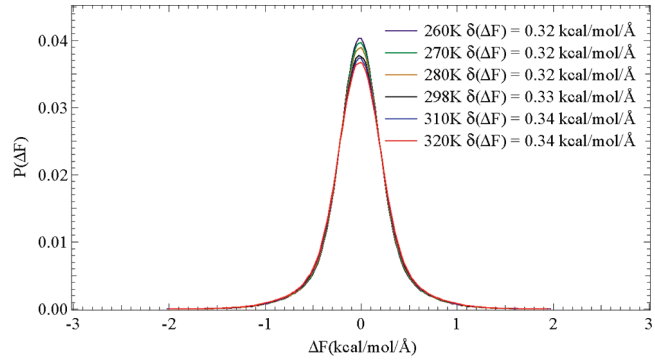

Figure 1. Distribution and standard deviation of the force difference ΔF_{ix} as a function of temperature.

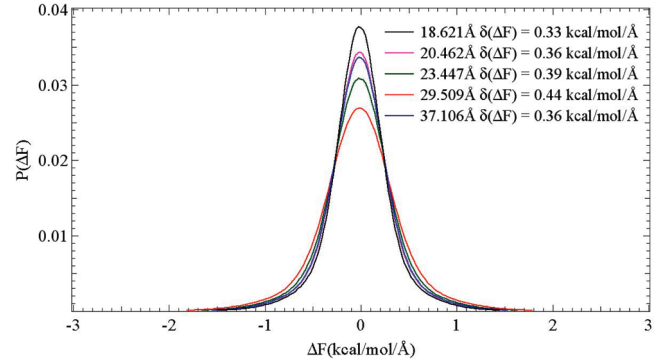

Figure 2. Distribution and standard deviation of the force difference ΔF_{ix} as a function of box size.

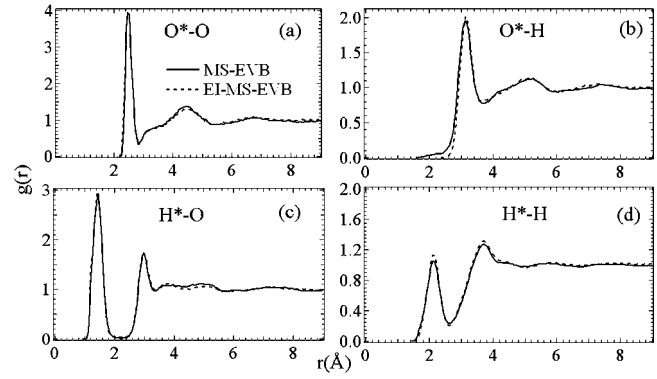

Figure 3. Comparison of the radial distribution functions (RDFs) calculated by the EI-MS-EVB and the full-potential MS-EVB methods. (a) RDFs of water oxygen atoms O around the hydronium oxygen atom O*, (b) RDFs of water hydrogen atoms H around the hydronium oxygen atom O*, (c) RDFs of water oxygen atoms O around the hydronium hydrogen atoms H*, (d) RDFs of water hydrogen atoms H around the hydronium hydrogen atoms H*.

where a "coordinate" relevant to the proton transfer process can be defined as the difference between the largest and second-largest MS-EVB amplitudes, c_1^2 and c_2^2. The function $P(c_1^2 - c_2^2)$ is thus the probability distribution of that coordinate. In this expression, R is the molar gas constant (8.314 J mol^{-1} K^{-1}), T is the system temperature (298.15 K), and C is an arbitrary constant that can be adjusted to define the point of zero free energy. Typical values for reference[52,54−57] are $c_1^2 - c_2^2 \approx 0.45$ for the Eigen cation ($H_9O_4^+$) and $c_1^2 - c_2^2 \approx 0.0$ for the Zundel cation ($H_5O_2^+$);

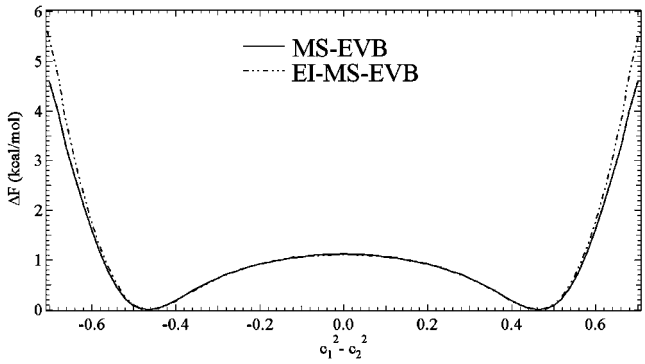

Figure 4. Free energy profile of the proton-transfer coordinate for the EI-MS-EVB model versus the full-potential MS-EVB calculation.

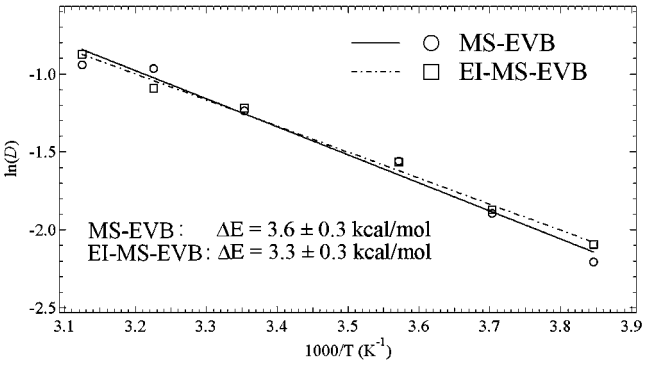

Figure 5. Temperature dependence of the proton diffusion coefficient for the EI-MS-EVB model versus the full-potential MS-EVB calculation.

the charge defect is more localized in the former case. As demonstrated in Figure 4, the EI-MS-EVB simulations nearly perfectly reproduce not only the free energy barrier of proton transfer, but also the probability distribution of the reactive coordinate. The central peak is slightly lower compared to the MS-EVB case, by only about 0.02 kcal/mol.

3.4. Temperature Dependence of the Diffusion Coefficient. The activation energy for proton transport, ΔE_a, can be described according to the classical Arrhenius equation

$$D_{H^+} = A \exp(-\Delta E_a/RT) \quad (15)$$

where A is a temperature-independent constant. By plotting $\ln(D_{H^+})$ against $1/T$, the value of ΔE_a can be determined from a simple linear fit. Simulations were carried out at 260, 270, 280, 310, and 320 K to obtain the temperature dependence of D_{H^+}, as shown in Figure 5. On average, the deviation between D_{H^+} values obtained in EI-MS-EVB and MS-EVB simulations is about 0.03 Å^2/ps for all temperatures. The two approaches are essentially indistinguishable with respect to this property, as the error bar for D_{H^+} is ±0.03 Å^2/ps. The ΔE_a value for the EI-MS-EVB model is 3.6 ± 0.3 kcal/mol, whereas that for the MS-EVB model is 3.3 ± 0.3 kcal/mol. Again, the two values are statistically consistent.

3.5. Effect of Box Size on the Proton Diffusion Coefficient. This section reports the transferability of the effective short-range potentials in the EI-MS-EVB method to other box sizes. In particular, we investigate the variation in D_{H^+} with system size. EI-MS-EVB and MS-EVB simulations

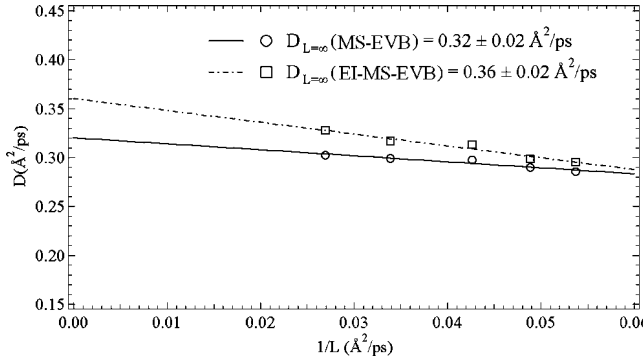

Figure 6. Effect of box size on the proton diffusion coefficient for the EI-MS-EVB model versus the full-potential MS-EVB calculation.

were performed on cubic systems with box sizes L of 20.462, 23.447, 29.509, and 37.106 Å in addition to the original system. Figure 6 demonstrates that the error ranges of all corresponding D_{H^+} pairs overlap, except for the largest box size. The actual D_{H^+} values calculated with the EI-MS-EVB method are consistently somewhat greater than those calculated with the MS-EVB method. Following the procedure of Yeh and Hummer,[71] the D_{H^+} value for an infinite volume can be estimated by linearly extrapolating D_{H^+} to the point $1/L = 0$. The asymptotic value of D_{H^+} with the FP EI-MS-EVB model turns out to be 0.36 ± 0.02 Å^2/ps, only 0.04 Å^2/ps larger than that determined with the MS-EVB approach. The trend toward increasing deviations with box size reflects the slight inconsistency between instantaneous atomistic forces and the effective potential, as discussed in section 3.1. Despite this small discrepancy, the EI-MS-EVB simulations correctly predict the nature of the relationship between box size and D_{H^+}. Moreover, the projected values of D_{H^+} remain in good agreement even for the larger systems.

3.6. Computational Efficiency. This section summarizes the most important result from the present article—that of increased computational efficiency wth the EI-MS-EVB model. Two simulations were performed to compare the efficiency of nonbonded interaction calculations. In the first simulation, electrostatic forces were calculated using the smooth particle−mesh Ewald (SPME)[12] algorithm with a tolerance of 10^{-6}, and LJ interactions were calculated using a simple spherical cutoff of radius 9.0 Å. In the second simulation, effective short-range potentials were applied with a simple spherical cutoff at 9.26 Å. The two approaches are denoted FP (full-potential MS-EVB model) and EI (EI-MS-EVB model), respectively.

Both runs were performed on a single core of a computer equipped with IBM PowerPC970 quad-core processors and 8 Gb of memory. To avoid introducing additional complexities related to parallel computing, both runs were carried out on a single CPU. The results for four different system sizes are summarized in Table 1, which reports the total CPU time spent on calculations of nonbonded interactions over 10000 time steps. At all system sizes, EI is significantly faster than FP, being nearly 4 times faster than FP even for a relatively small system with 5185 atoms. Very encouragingly, but not surprisingly, the efficiency improvement increases with

Table 1. Computational Efficiency of Nonbonded Interaction Calculations as a Function of System Size, Comparing the Full-Potential MS-EVB (FP) and EI-MS-EVB (EI) Approaches

number of atoms	FP time (s)	EI time (s)	ratio (FP/EI)
649	9912	3414	2.9
5185	41229	8267	5.0
12427	104801	13889	7.5
34252	313810	27267	11.5

system size. For the system with 34252 atoms, the EI approach is roughly 10 times faster than the FP approach.

Although the present results are for only a single processor, when considering the difficulty in constructing an efficient global parallelization of the fast Fourier transform (used by the SPME calculation), the computational efficiency of the EI-MS-EVB approach will become even more pronounced when the scalability of the algorithm is considered. Molecular simulations of large-scale biological systems are typically performed on many CPUs, requiring such parallelization. The EI-MS-EVB approach might therefore be better suited to such applications than the full-potential MS-EVB method, where calculations of long-range, nonbonded interactions (particularly electrostatics) are computationally less scalable. It has recently been demonstrated,[72] for example, that highly scalable biomolecular MD simulations are within reach when implementing short-range (cutoff) electrostatic potentials similar in spirit to those in the EI-MS-EVB approach.

4. Conclusions

Depending on the system of interest, the time and length scales of importance to reactive events such as those involved in PS&T can vary widely: from picoseconds to microseconds and from nanometers to micrometers, respectively. Because the system coordinates are usually propagated on a time scale of femtoseconds in an MD allgorithm, MS-EVB applications can be extremely computationally challenging. The EI-MS-EVB approach utilizes an effective short-range potential[48,49] and accurately reproduces reference trajectories from the full MS-EVB method. By comparing the instantaneous forces on single atoms from the same configuration of nuclei, we have shown that EI-MS-EVB model is not only accurate but also transferable over a range of system temperatures and box sizes. The accuracy of the EI-MS-EVB method was further confirmed by evaluating several key properties: RDFs, free energy profiles, activation energies of proton transport, and proton diffusion coefficients. Most importantly, the EI-MS-EVB model was found to outperform the full-potential MS-EVB model in terms of computational efficiency. For larger systems, the EI-MS-EVB approach is approximately 10 times faster than the SPME-based FP MS-EVB calculation on a single processor. In addition, the EI-MS-EVB method of calculating short-range interactions should be much easier to decompose and parallelize than the fast Fourier transform of the SPME method, which is intrinsically a global operation. This is an important consideration for complex applications and large systems, which typically require many thousands of CPUs.

Very recently, the FM algorithm has even been used to construct effective short-range forces from explicit evaluations of the long-range Coulomb interaction.[48] The resulting effective interaction between unit charges successfully reproduces many structural, dynamic, and thermodynamic properties of various systems, including liquid water, solvated ions, and hydrophobic solutes. This methodology[48] and its analytic approximation[49] is closely related to the present approach. Provided the van der Waals interaction is correctly accounted for, the effective nonbonded interaction obtained by force matching should be recoverable from the "universal" effective short-range electrostatic interaction identified previously.[48] Furthermore, integrating all the nonbonded interactions in this manner can reduce the number of operations per particle pair to just one table lookup in both the EI and charge-scaled schemes.

In future work, one goal will be to extend the EI-MS-EVB approach to the self-consistent, iterative MS-EVB (SCI-MS-EVB)[73] method. This approach determines the EVB amplitude separately for each protonated complex, enforcing consistency with the EVB amplitudes of other protonated complexes in an iterative manner until the total potential energy of the multiproton system has converged. The SCI-MS-EVB method demands significantly more CPU cycles for nonbonded interactions than does the single-proton MS-EVB. The SCI-MS-EVB approach also typically requires five or more iterations for the EVB amplitudes to converge. Incorporating the EI-MS-EVB model into the SCI-MS-EVB framework should therefore greatly facilitate the investigation of PS&T behavior in highly acidic systems such as the proton-exchange membranes used in fuel cells.[74]

Acknowledgment. This research was supported by the National Science Foundation (CHE-0719522) and the Department of Energy (DE-FG02-05ER15724). The computational resources utilized in this research were provided by the following NSF programs: Partnerships for Advanced Computational Infrastructure, Distributed Terascale Facility (DFT), and Terascale Extensions: Enhancements to the Extensible Terascale Facility.

Supporting Information Available: Atom types in the EI-MS-EVB model and a table of numerical pairwise nonbonded forces fitted by the force-matching method. This information is available free of charge via the Internet at http://pubs.acs.org/.

References

(1) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1–218.

(2) Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155–179.

(3) Tobias, D. J. *Curr. Opin. Struct. Biol.* **2001**, *11*, 253–261.

(4) Harvey, S. C. *Proteins* **1989**, *5*, 78–92.

(5) Smith, P. E.; van Gunsteren, W. F. In *Computer Simulations of Biomolecular Systems: Theoretical and Experimental Applications*; ESCOM: Leiden, The Netherlands, 1993; pp 182−212.

(6) Norberg, J.; Nilsson, L. *Biophys. J.* **2000**, *79*, 1537–1553.

(7) Cheathan, T. E., III; Brooks, B. R. *Theor. Chem. Acc.* **1999**, *99*, 279–288.

(8) Auffinger, P.; Westhof, E. In *Encyclopedia of Computational Chemistry*; John Wiley & Sons: New York, 1998; pp 1628−1639.

(9) Ewald, P. P. *Ann. Phys.* **1921**, *64*, 253–287.

(10) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; McGraw-Hill: New York, 1981.

(11) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089–92.

(12) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(13) Perram, J. W.; Petersen, H. G.; DeLeeuw, S. W. *Mol. Phys.* **1988**, *65*, 875–893.

(14) Toukmaji, A.; Board, J. A. *Comput. Phys. Commun.* **1996**, *95*, 78–92.

(15) Greengard, L.; Rokhlin, V. *J. Comput. Phys.* **1987**, *73*, 325–348.

(16) Board, J. A.; Causey, J. W.; Leathrum, J. F.; Windemuth, A.; Shulten, K. *Chem. Phys. Lett.* **1992**, *198*, 89–94.

(17) Pollock, E.; Glosli, J. *Comput. Phys. Commun.* **1996**, *95*, 93–110.

(18) Figuerido, F.; Levy, R.; Zhou, R.; Berne, B. J. *J. Chem. Phys.* **1997**, *106*, 9835–9849.

(19) Cheng, H.; Greengard, L.; Rokhlin, V. *J. Comput. Phys.* **1999**, *155*, 468–98.

(20) Greengard, L. F.; Huang, J. *J. Comput. Phys.* **2002**, *180*, 642–58.

(21) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, U.K., 1987.

(22) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.

(23) Brooks, C. L., III; Pettitt, B. M.; Karplus, M. *J. Chem. Phys.* **1985**, *83*, 5897–5908.

(24) Hunenberger, P. H.; van Gunsteren, W. F. *J. Chem. Phys.* **1998**, *108*, 6117–6134.

(25) Madura, J. D.; Pettitt, B. M. *Chem. Phys. Lett.* **1988**, *150*, 105–108.

(26) Schreiber, H.; Steinhauser, O. *Chem. Phys.* **1992**, *168*, 75–89.

(27) Neumann, M.; Steinhauser, O.; Pawley, G. S. *Mol. Phys.* **1984**, *52*, 97–113.

(28) Baker, N. A.; Hünenberger, P. H.; McCammon, J. A. *J. Chem. Phys.* **1999**, *110*, 10679–10692.

(29) Wood, R. H. *J. Chem. Phys.* **1995**, *103*, 6177–6187.

(30) Brooks, C. L., III. *J. Chem. Phys.* **1987**, *86*, 5156–5162.

(31) Straatsma, T. P.; Berendsen, H. J. C. *J. Chem. Phys.* **1988**, *89*, 5876–5886.

(32) Kalko, S. G.; Sese, G.; Padro, G. A. *J. Chem. Phys.* **1996**, *104*, 9578–9585.

(33) Resat, H.; McCammon, J. A. *J. Chem. Phys.* **1996**, *104*, 7645–7651.

(34) Dang, L. X.; Pettitt, B. M.; Rossky, P. J. *J. Chem. Phys.* **1992**, *96*, 4046–4047.

(35) Badert, J. S.; Chandler, D. *J. Chem. Phys.* **1992**, *96*, 6423–4427.

(36) Del Buono, G. S.; Figueirido, F. E.; Levy, R. M. *Chem. Phys. Lett.* **1996**, *263*, 521–529.

(37) Loncharich, R. J.; Brooks, B. R. *Proteins* **1989**, *6*, 32–45.

(38) Lau, K. F.; Alper, H. E.; Thacher, T. S.; Stouch, T. R. *J. Phys. Chem.* **1994**, *98*, 8185–8792.

(39) Prevost, M.; van Belle, D.; Lippens, G.; Wodak, S. *Mol. Phys.* **1990**, *76*, 587–603.

(40) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667–683.

(41) Barker, J. A.; Watts, R. O. *Mol. Phys.* **1973**, *26*, 789–792.

(42) Hummer, G.; Soumpasis, D. M.; Neumann, M. *Mol. Phys.* **1992**, *77*, 769–785.

(43) Chipot, C.; Millot, C.; Maigret, B.; Kollman, P. A. *J. Chem. Phys.* **1994**, *101*, 7953–7962.

(44) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451–5459.

(45) Daura, X.; Hünenberger, P. H.; Mark, A. E.; Querol, E.; Avilés, F. X.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **1996**, *118*, 6285–6294.

(46) Fennell, C. J.; Gezelter, J. D. *J. Chem. Phys.* **2006**, *124*, 234104.

(47) Wolf, D.; Keblinski, P.; Phillpot, S. R.; Eggebrecht, J. *J. Chem. Phys.* **1999**, *110*, 8254–8282.

(48) Izvekov, S.; Swanson, J. M.; Voth, G. A. *J. Phys. Chem. B* **2008**, *112*, 4711–4724.

(49) Shi, Q.; Liu, P.; Voth, G. A. *J. Phys. Chem. B* **2008**, *112*, 16230–7.

(50) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896–10913.

(51) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 6573–86.

(52) Schmitt, U. W.; Voth, G. A. *J. Phys. Chem. B* **1998**, *102*, 5547–5551.

(53) Schmitt, U. W.; Voth, G. A. *J. Chem. Phys.* **1999**, *111*, 9361–9381.

(54) Day, T. J. F.; Soudackov, A. V.; Cuma, M.; Schmitt, U. W.; Voth, G. A. *J. Chem. Phys.* **2002**, *117*, 5839–5849.

(55) Wu, Y.; Chen, H.; Wang, F.; Paesani, F.; Voth, G. A. *J. Phys. Chem. B.* **2007**, *112*, 467–482.

(56) Voth, G. A. *Acc. Chem. Res.* **2006**, *39*, 143–50.

(57) 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*, 4300–14.

(58) Iyengar, S. S.; Petersen, M. K.; Day, T. J. F.; Burnham, C. J.; Teige, V. E.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 084309.

(59) Day, T.; Schmitt, U.; Voth, G. *J. Am. Chem. Soc.* **2000**, *122*, 12027–12028.

(60) Chen, H.; Wu, Y.; Voth, G. A. *Biophys. J.* **2006**, *90*, L73–75.

(61) Chen, H.; Ilan, B.; Wu, Y.; Zhu, F.; Schulten, K.; Voth, G. A. *Biophys. J.* **2007**, *92*, 46–60.

(62) Chen, H.; Wu, Y.; Voth, G. A. *Biophys. J.* **2007**, *93*, 3470–3479.

(63) Xu, J.; Voth, G. A. *Proc. Natl. Acad. Sci.* **2005**, *102*, 6795–6800.

(64) Chen, H.; Yan, T.; Voth, G. A. *J. Phys. Chem. A* **2009**, *113*, 4507–17.

(65) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 4116–4127.

(66) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 244114.

(67) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2008**, *128*, 244115.

(68) Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.

(69) Smondyrev, A. M.; Voth, G. A. *Biophys. J.* **2002**, *83*, 1987–1996.

(70) Smith, W.; Forester, T. R. *J. Mol. Graphics* **1996**, *14*, 136–141.

(71) Yeh, I. C.; Hummer, G. *J. Phys. Chem. B* **2004**, *108*, 15873–15879.

(72) Schulz, R.; Lindner, B.; Petridis, L.; Smith, J. C. *J. Chem. Theory Comput.* **2009**, *5*, 2798–808.

(73) Wang, F.; Voth, G. A. *J. Chem. Phys.* **2005**, *122*, 144105–9.

(74) Petersen, M. K.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 18594–18600.

CT100318F