Hydration Free Energy from Orthogonal Space Random Walk and Polarizable Force Field

Jayvee R. Abella,[†,§] Sara Y. Cheng,[‡,§] Qiantao Wang,[†] Wei Yang,[⊥,‖] and Pengyu Ren[*,†]

[†]Department of Biomedical Engineering and [‡]Department of Physics, The University of Texas at Austin, Austin, Texas 78712, United States

[⊥]Institute of Molecular Biophysics and [‖]Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32306, United States

Ⓢ Supporting Information

ABSTRACT: The orthogonal space random walk (OSRW) method has shown enhanced sampling efficiency in free energy calculations from previous studies. In this study, the implementation of OSRW in accordance with the polarizable AMOEBA force field in TINKER molecular modeling software package is discussed and subsequently applied to the hydration free energy calculation of 20 small organic molecules, among which 15 are positively charged and five are neutral. The calculated hydration free energies of these molecules are compared with the results obtained from the Bennett acceptance ratio method using the same force field, and overall an excellent agreement is obtained. The convergence and the efficiency of the OSRW are also discussed and compared with BAR. Combining enhanced sampling techniques such as OSRW with polarizable force fields is very promising for achieving both accuracy and efficiency in general free energy calculations.
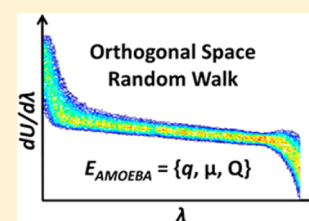
INTRODUCTION

Water is a substantial component of living organisms, forming an environment where biological processes such as the transportation of ions, the folding of proteins, and the activation/deactivation of signaling pathways can take place. The interactions between water and physiologically relevant molecules, such as monatomic ions, small molecules, and macromolecules, are crucial to the efforts of understanding such biological processes and applications such as protein engineering and drug discovery. Therefore, accurately modeling the hydration process is arguably the first step in modeling these biological processes and developing accurate physical models and robust computational approaches. For instance, not only is the hydration free energy (HFE) a key property in predicting the solubility of organic molecules and their binding to proteins,[1−4] but hydration free energy is also an important measure in the development and evaluation of the accuracy of force fields[5−10] and sampling methods.[11−16] The hydration free energy of a molecule can be calculated by using explicit solvent models, e.g., TIP3P water[17] and AMOEBA water,[18] in combination with alchemical approaches, such as thermodynamic integration (TI; see review by Kollman[1]), Bennett acceptance ratio (BAR),[19] or the orthogonal space random walk (OSRW).[4,20−23] Once the force field is well-defined, the accuracy and precision of the alchemical results are somewhat more predictable.

Although the importance of including explicit polarization in molecular modeling have been demonstrated in previous studies,[24−26] the routine application of polarizable force fields, such as AMOEBA,[18,27−30] to obtain accurate thermodynamic properties is still hindered by the high computational cost of traditional alchemical approaches. Thus, enhanced sampling methods such as the OSRW method are more appealing in such simulations. Unlike BAR or TI, which requires a number of arbitrary, fixed order parameter λ to connect the two end states, the OSRW method[4,20−23] described in later sections utilizes dynamic order parameters, λ, and $dU/d\lambda$, coupled with the metadynamics approach[31] to sample the two dimensions. In this way, the alchemical perturbation between the two end states can be performed in a single molecular dynamics simulation and with improved efficiency. Previously, we have demonstrated that OSRW allows efficient sampling of configurational spaces of molecular crystals.[4]

In this paper, the OSRW method is implemented with the polarizable multipole based AMOEBA force field in TINKER and applied to compute the hydration free energy of several small organic molecules. The hydration free energy results from OSRW are compared with those computed from the conventional BAR method, which has been utilized to compute free energy of hydration and binding in combination with AMOEBA in previous studies.[5,30,32−39] The results from the two approaches are in excellent agreement (RMSD = 0.49 kcal/mol), with the OSRW method showing a significant advantage in computational efficiency.

METHODS

Theoretical Background of OSRW. Because free energy is a path-independent property, a common approach to calculate

the free energy difference is to define a mixed potential, so that the potential functions of the two end states of interest can be connected analytically. Such a mixed potential is usually defined as

$$U(\lambda, \mathbf{r}) = (1 - \lambda)U_0 + \lambda U_1 \tag{1}$$

where U is the Hamiltonian of the mixed potential, $\mathbf{r}$ is the coordinate, and the scaling parameter $\lambda = 0$ and 1 corresponds to the two end states, U_0 and U_1, respectively. The free energy change from one state to the other can thus be given as

$$\Delta G = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{2}$$

where G is the free energy of each state, ΔG is the change in free energy, and $\langle \ \rangle_\lambda$ is the ensemble average of each λ state.

Such construction of the approach, however, relies on the assumption that sufficient conformational sampling can be done as the system adjusts to the new intermediate states. For complex systems, such transition usually requires a significant amount of simulation time, especially when there are larger changes in structure. This is often known as the "Hamiltonian lagging" problem, which exists for methods where λ is a continuous and dynamic variable.[40] For methods that perform simulation at discrete λ "windows," a large number of intermediate steps are required, and long simulations at each step are needed to ensure sufficient equilibration and sampling at each step.

Alternatively, by combining the ideas of the dynamic λ method[40,41] and the metadynamics method,[31] Yang and co-workers proposed an efficient free energy sampling approach, which they referred to as the orthogonal space random walk (OSRW).[4,20−23] In this approach, a random walk is performed in two dimensions, λ and its orthogonal generalized force $F_\lambda = \partial U/\partial \lambda$. The use of dynamic λ itself unnecessarily improves the computational efficiency; however, directly biasing along the $\partial U/\partial \lambda$ dimension can potentially accelerate the free energy calculation since the integral of $\partial U/\partial \lambda$ is exactly the free energy. By repetitively adding a Gaussian-like repulsive potential to λ and F_λ spaces, the low energy wells can be "flooded" to overcome the energy barriers. The potential of the system in the OSRW can be written as

$$U_{OSRW} = U(\lambda, \mathbf{r}) + g(\lambda, F_\lambda) - G(\lambda) \tag{3}$$

where g is the biasing potential that can be defined recursively as

$$g(\lambda, F_\lambda) = \sum_{t_i} h \exp\left(-\frac{|\lambda - \lambda(t_i)|^2}{2w_1^2}\right) \exp\left(-\frac{|F_\lambda - F_\lambda(t_i)|^2}{2w_2^2}\right) \tag{4}$$

where h and w are the height and width of the Gaussian, which can be adjusted to balance the accuracy and efficiency of the method, and t_i is the index of states. The free energy along the reaction coordinates can thus be estimated as $-g(\lambda, F_\lambda)$. To move from an initial state to a target state, λ, the free energy change can thus be estimated as

$$G(\lambda) = \int_{\lambda_o}^\lambda \frac{dG}{d\lambda} d\lambda$$
$$= \int_{\lambda_o}^\lambda \frac{\sum_{F_\lambda} F_\lambda \exp[\beta g(\lambda, F_\lambda)]\delta(\lambda - \lambda')}{\sum_{F_\lambda} \exp[\beta g(\lambda, F_\lambda)]\delta(\lambda - \lambda')} d\lambda' \tag{5}$$

By adaptively adding a negative $G(\lambda)$ to the system Hamiltonian as shown in eq 3, the "flooding" of the free energy surface can be accelerated along the λ space.

AMOEBA Based Alchemical Scheme in TINKER. A hybrid Hamiltonian based on eq 1 is implemented to calculate the alchemical free energy using the AMOEBA force field in TINKER. A dual topology approach is used to keep the intramolecular energies of the mutating systems (e.g., solute molecules in solution) throughout the simulation, and the mutating systems A (initial) and B (final) never interact with each other. In case of absolute hydration free energy calculations performed in this study, the two topologies are solute-in-water and water without solute, respectively. In this approach, the mixed potential can be written as

$$U_{dt}(\lambda) = U_{bonded} + U_{vdW}(\lambda) + U_{els}^{real}(\lambda) + U_{els}^{recip}(\lambda) + U_{pol}(\lambda) \tag{6}$$

where the subscript dt indicates the Hamiltonian is for dual topology, and the bonded term is independent of λ. A softcore van der Waals (vdW) potential[4,32] has been adapted to replace the original buffered 14−7 potential[42] in AMOEBA in this implementation:

$$U_{vdW,ij}(\lambda) = \lambda \varepsilon_{ij} \frac{1.07^7}{\alpha(1 - \lambda^2) + (\rho + 0.07)^2} \left(\frac{1.12}{\alpha(1 - \lambda^2) + \rho^7 + 0.12} - 2 \right) \tag{7}$$

where i and j are the indices of the atoms, ε is the well-depth of the potential, α is an adjustable constant, $\rho = r/r^*$, and r^* is the equilibrium distance between two atoms. This equation prevents the numerical instability of the system when λ is small, and it reduces to the original buffered 14−7 potential when $\lambda = 1$. The real space electrostatic potential[18] is written as

$$U_{els,ij}^{real}(\lambda) = \lambda \sum_{l=0}^4 G_{ij}^l(\mathbf{r}) B_l(f) \tag{8}$$

where G is the permanent multipole moment, l is the order of the multipole moment, B is the screening function, and f is the modified distance defined as

$$f = (r_{ij}^2 + \alpha(1 - \lambda)^2)^{1/2} \tag{9}$$

This definition, similar to the softcore potential for the vdW term, can prevent the numerical instability of the system when λ is small and atoms are very close to each other. The final mixed potential for the real space electrostatic potential is thus

$$U_{els}^{real}(\lambda) = U_{els,A}^{real,tot}(1 - \lambda) + \lambda U_{els,B}^{real,tot}(\lambda) + \lambda U_{els,A}^{gas,mut} + (1 - \lambda)U_{els,B}^{gas,mut} \tag{10}$$

where the superscripts tot and mut respectively indicate the energy of the whole system and the energy of the part undergoing alchemical transformation. The reciprocal space part of the permanent multipole interaction energy is mixed linearly:

$$U_{els}^{recip}(\lambda) = (1 - \lambda)U_{els,A}^{recip,tot} + \lambda U_{els,B}^{recip,tot} \tag{11}$$

The polarization energy has a slightly different combination. To eliminate the potential problem[4] of the self-consistent field calculation at unphysically small atomic distances, the polar-

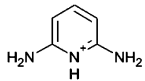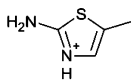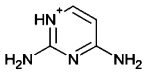**Table 1. Calculated Hydration Free Energy (kcal/mol) of 20 Compounds Using BAR and OSRW**
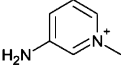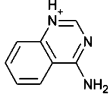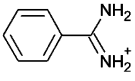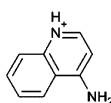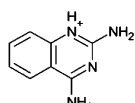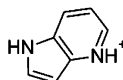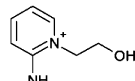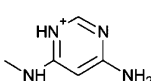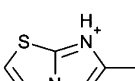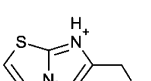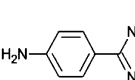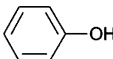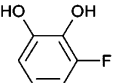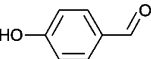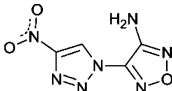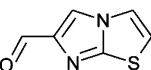
Compounds		BAR	OSRW	Absolute deviation
1		-46.61 ± 0.82	-46.93 ± 0.42	0.32
2		-44.59 ± 0.52	-43.95 ± 0.28	0.64
3		-45.76 ± 0.77[c]	-45.57 ± 0.09	0.19
4		-49.70 ± 0.83	-50.18 ± 0.40	0.48
5		-41.68 ± 0.79[c]	-41.49 ± 0.31	0.19
6		-46.36 ± 0.57	-46.55 ± 0.05	0.19
7		-48.70 ± 0.57	-48.55 ± 0.10	0.15
8		-41.66 ± 0.56	-41.80 ± 0.20	0.14
9		-46.22 ± 0.62	-45.73 ± 0.40	0.50
10		-44.33 ± 0.51	-44.80 ± 0.60	0.47
11		-49.25 ± 0.15	-49.45 ± 0.79	0.20
12		-45.73 ± 0.56	-45.37 ± 0.50	0.36
13		-41.98 ± 0.85	-41.45 ± 0.06	0.53
14		-46.55 ± 0.96	-46.49 ± 0.33	0.06
15		-45.83 ± 0.58	-45.98 ± 0.25	0.15

Table 1. continued

Compounds		BAR	OSRW	Absolute deviation
16	(phenol)	-5.38 ± 0.39	-5.28 ± 0.04	0.10
17	(difluoro catechol structure)	-14.25 ± 0.32[a]	-12.87 ± 0.33[b]	1.38
18	(4-hydroxybenzaldehyde)	-9.08 ± 0.47	-8.43 ± 0.59	0.65
19	(nitro-triazole amino-oxadiazole structure)	-9.99 ± 0.27	-9.70 ± 0.25	0.30
20	(imidazothiazole carbaldehyde structure)	-8.71 ± 0.48	-8.61 ± 0.18	0.10
Mean Unsigned Deviation (without **17**)				0.34
Mean Unsigned Deviation				0.40
RMSD (without **17**)				0.39
RMSD				0.49

[a] -14.06 ± 0.22 kcal/mol after simulations extended from 1 to 5 ns at each of the 25 steps. [b] -12.97 ± 0.39 kcal/mol after simulations extended from 4 to 8 ns, averaged over seven independent simulations. [c] Data collected from 0.5 ns simulation in each step; otherwise 1 ns simulations were performed.

ization is switched off until λ is equal to or greater than 0.75, which is to referred as λ_{pol}^{start}. An artificial scaling factor, given as $\lambda_{pol,A} = (1 - \lambda_{pol}^{start} - \lambda)/(1 - \lambda_{pol}^{start})$ and $\lambda_{pol,B} = (\lambda - \lambda_{pol}^{start})/(1 - \lambda_{pol}^{start})$, is used to smoothly switch the potential across $1 - \lambda_{pol}^{start}$ and λ_{pol}^{start}, respectively. The final mixed polarization potential can be written as

$$
U_{pol}(\lambda) = \begin{cases} \lambda_{pol,A}^3 (U_{pol,A}^{tot} - U_{pol}^{env} - U_{pol,A}^{gas,mut}) + U_{pol}^{env} \\ \quad + U_{pol,A}^{gas,mut} + U_{pol,B}^{gas,mut},\ 0 \leq \lambda \leq 1 - \lambda_{pol}^{start} \\ U_{pol}^{env} + U_{pol,A}^{gas,mut} + U_{pol,B}^{gas,mut},\ 1 - \lambda_{pol}^{start} < \lambda \\ \quad < \lambda_{pol}^{start} \\ \lambda_{pol,B}^3 (U_{pol,B}^{tot} - U_{pol}^{env} - U_{pol,B}^{gas,mut}) + U_{pol}^{env} \\ \quad + U_{pol,A}^{gas,mut} + U_{pol,B}^{gas,mut},\ \lambda_{pol}^{start} \leq \lambda \leq 1 \end{cases}
$$

(12)

Various derivatives, $\partial U/\partial \lambda$, $\partial^2 U/\partial \lambda^2$, and $\partial^2 U/\partial \mathbf{r} \partial \lambda$ have been derived in the same way as reported previously.[4]

■ COMPUTATIONAL DETAILS

In this study, the hydration free energy of 20 small molecule solutes,[43] among which 15 are positively charged compounds and five are neutral compounds, were calculated in an explicit solvent model[18] using the polarizable AMOEBA force field.[27] Parameters of the compounds were obtained using POL-TYPE.[38] Both orthogonal space random walk and Bennett acceptance ratio methods were used to estimate the free energy using the same simulation conditions, such as box size, simulation ensemble, boundary conditions, vdW, and electrostatic cutoffs. Thus, the comparison between OSRW and BAR methods would not be affected by potential artifacts in the calculated hydration free energy due to different boundary conditions or treatment of electrostatic interactions suggested in previous studies.[44,45] All molecular dynamics simulations were conducted with a RESPA integrator,[46] Bussi thermostat,[47] and 2 fs time step using the TINKER software package if not otherwise stated. A cutoff of 12 Å was applied to vdW's interaction with $\alpha = 0.07$ (eq 7), while a cutoff of 7.0 Å was applied in the real space softcore electrostatic calculations with $\alpha = 2.0$ (eqs 8 and 9). Self-consistent induced dipole moments were converged to below 0.00001 D per atom.

Before the alchemical simulations, all the solutes were first soaked in a 30 Å cubic water box followed by a 600 ps relaxation using NPT molecular dynamics simulation at 298 K with a 2 fs time step. The resulting boxes were used in the subsequent NVT simulations with the density fixed at the average from the NPT simulations.

In OSRW, a 2D grid along the λ and F_λ axes was constructed to store the history of the (λ, F_λ) states visited. Each point on the grid represents a bin with finite dimensions. Our implementation has a λ width of 0.005 and an F_λ width of 2.0. For the λ axis, λ ranges from 0 to 1, and for mathematical convenience the first and last λ bins are half size (the total number of the λ bins is 201) and centered at 0 and 1, respectively. For the F_λ axis, F_λ has no clear range, so the range must be dynamically updated if the calculated F_λ falls outside the initial specified range. Also for mathematical convenience, there is always an F_λ bin centered at zero. Throughout the MD simulation, each bin centered at (λ, F_{λ_i}) in the grid represents the number of times a particular state with $\lambda - \Delta\lambda/2 < \lambda < \lambda + \Delta\lambda/2$ and $F_\lambda - \Delta F_\lambda/2 < F_\lambda < F_\lambda + \Delta F_\lambda/2$ was visited, where $\Delta\lambda$ and ΔF_λ represent the λ width and F_λ width, respectively. For each count, a 2D biasing Gaussian centered at (λ, F_λ) is added to the potential. Our implementation uses a Gaussian height of 0.005 with variances of $w_1^2 = (2\Delta\lambda)^2$ and $w_2^2 = (2\Delta F_\lambda)^2$. The Gaussians are cut off after five bins from the central bin. However, simply with the current implementation, the random walk may end up being stuck at the λ end points since λ is between 0 and 1. Thus, mirror conditions are enforced where if a Gaussian has a contributing value outside the λ range, the contribution is mirrored onto a bin with a valid range. For the first and last λ bins, the contributions are automatically doubled since they are centered at 0 and 1, respectively. The statistical error was estimated from three repeat simulations of 4 ns for each compound.

In BAR calculations, a three-step perturbation approach was applied using the AMOEBA force field.[5,30,32] To make the solute disappear in the solvent, the electrostatic and polarization interactions were perturbed in 11 windows, scaled by λ = 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.0. After the electrostatic interactions were scaled to zero, then the vdW's contribution was perturbed using 14 windows with λ = 1.0, 0.9, 0.8, 0.75, 0.7, 0.65, 0.6, 0.55, 0.5, 0.4, 0.3, 0.2, 0.1, 0.0. Then, 1000 ps NVT simulations at 298 K were performed at each window. The recharging of each solute in the gas phase was modeled using 11 windows (with an interval of 0.1 for λ) of 1000 ps molecular dynamics simulation at 298 K, a stochastic integrator, and a 0.1 fs time step. Data collected from 50 to 1000 ps range was then processed using the BAR equations. The errors for BAR results were computed as a sum of errors from the individual alchemical perturbation steps.

■ RESULTS AND DISCUSSION

Hydration Free Energy. The hydration free energy was calculated for the 20 compounds using both OSRW and BAR methods. In general, a good agreement between the two methods was obtained (Table 1). The energy values from the two methods were plotted against each other (Figure 1), and the r^2 correlation coefficient was calculated to be 0.98 for the charged set and 0.99 for the neutral set. The unsigned average difference in the calculated hydration free energy between the two methods is 0.39 kcal/mol, and the root-mean-square difference is 0.49 kcal/mol (Table 1). For illustration, the hydration free energy over time is plotted for the OSRW
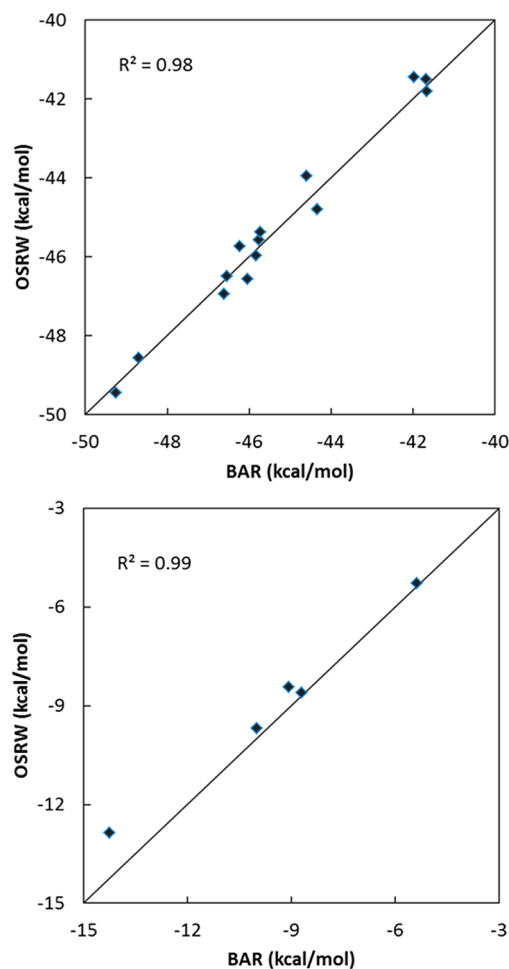

Figure 1. Comparison of the calculated hydration free energy from BAR and OSRW methods. An excellent agreement between the values from the two methods is obtained with an R^2 coefficient of 0.98 ± 0.02 and 0.99 ± 0.01 for the 15 charged (upper panel) and the five neutral molecules (lower panel), respectively.

(Figure 2) and BAR (Figure 3) methods for compounds **1**, **11**, and **17**.

Compound **17** shows the largest difference of 1.38 kcal/mol between the two methods, which reduced to 1.09 kcal/mol after we significantly extended simulations for both methods (Table 1). Compound **17** has a "complex" structure with two hydroxyl groups and one fluorine atom all connecting to adjacent carbons in the benzene ring. The interactions among these groups and with water can be rather complex. For example, from both BAR and OSRW simulations, the two hydroxyl groups were seen as very flexible with the hydrogen atoms either facing away or forming hydrogen bonds with each other, albeit with different frequencies. In addition, compound **17** is the only solute with a fluorine atom in this set. To verify the fluorine parameters, hydration free energy of fluorobenzene and 2-fluorophenol have also been calculated using OSRW and compared with experiment values. For each of the two compounds, three independent OSRW calculations were performed. The average hydration free energy from the simulations is -0.76 ± 0.33 and -5.71 ± 0.16 kcal/mol for fluorobenzene and 2-fluorophenol, respectively. This is in a good agreement with the experimental hydration free energy of -0.80 and -5.29 kcal/mol for the two compounds.[9]
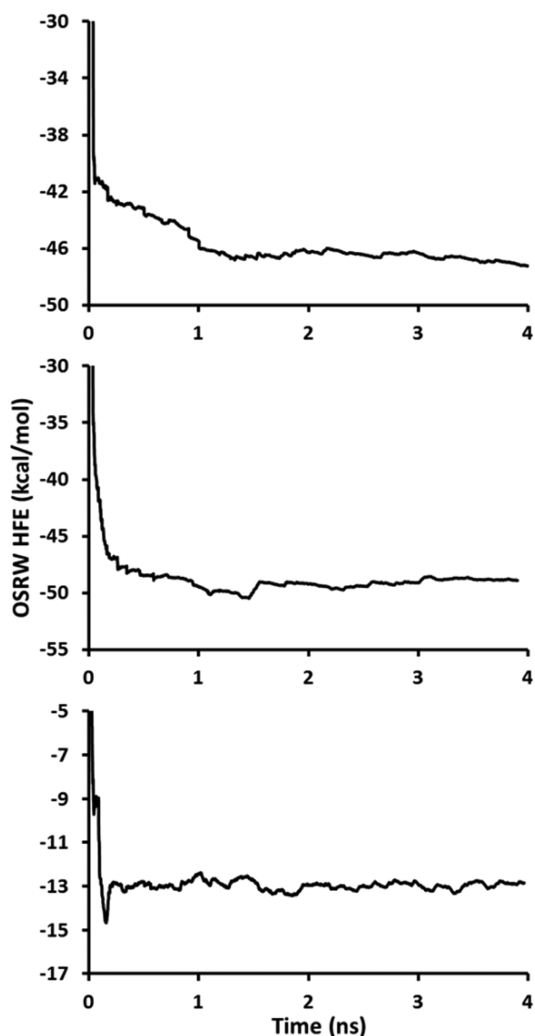
2796

dx.doi.org/10.1021/ct500202q | J. Chem. Theory Comput. 2014, 10, 2792−2801

Figure 2. Plots of hydration free energy (kcal/mol) for compounds **1** (top), **11** (middle), and **17** (bottom) calculated by OSRW as a function of time.

Figure 3. Plots of hydration free energy of compounds **1** (top), **11** (middle), and **17** (bottom) calculated by BAR as a function of time.

We first examined the effect of van der Waals perturbation steps on the hydration free energy calculated from BAR for compound **17**. In our implementation of the BAR approach, the electrostatic interaction between solute and environment was first turned off, and then the vdW interactions were scaled. During this latter stage, water and solute molecules can have significant overlap, resulting in large uncertainty in the free energy as evident by the difference between the forward and backward free energy perturbation results. We added an additional six windows in the middle of the vdW perturbation (λ = 0.775, 0.725, 0.675, 0.625, 0.575, 0.525). With these additional steps and longer simulations, the hydration free energy of compound **17** is merely increased by 0.2 kcal/mol (Supporting Information).

We have further investigated water structure near the solute sampled during the BAR simulations by plotting radial distribution functions (RDFs) between the compound **17** (O atoms in the two hydroxyl groups and F atom) and water (O). It would be best to compare the RDF at the same λ values. However, this is impossible in this study due to the limitation of our implementation. As explained in the Methods section, the BAR method decouples the vdW and electrostatic interactions separately while in the OSRW approach the two are scaled

Figure 4. Plots of the radial distribution function (bin size 0.2 Å) from BAR simulations: (top) the oxygen of water and the fluorine of compound **17**, (middle) oxygen of water and oxygen of compound **17** in the hydroxide group closest to the fluorine, and (bottom) oxygen of water and oxygen of compound **17** in the hydroxide group furthest from fluorine. The RDF was evaluated using the simulated structures from all perturbation windows.

Figure 5. Plots of the radial distribution function (bin size 0.2 Å) from OSRW simulations: (top) oxygen of water and fluorine of compound **17**, (middle) oxygen of water and oxygen of compound **17** in the hydroxide group closest to fluorine in compound **17**, and (bottom) oxygen of water and oxygen of compound **17** in the hydroxide group farthest away from fluorine in compound **17**.

simultaneously. As a result, the same λ in the two methods actually represents different states. Therefore, we plotted the RDF for the three atom pairs using the trajectories of all the λ
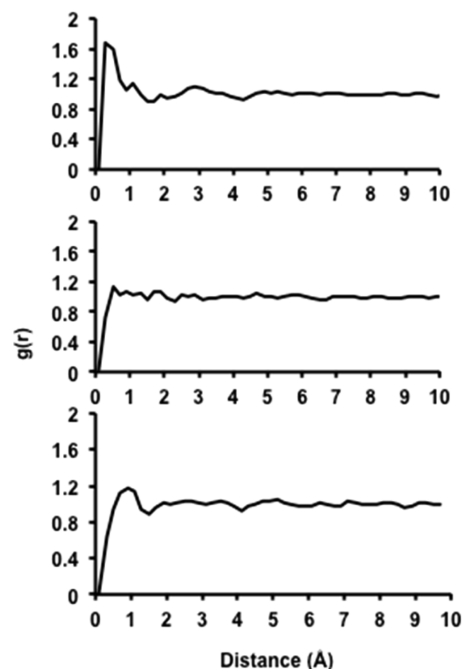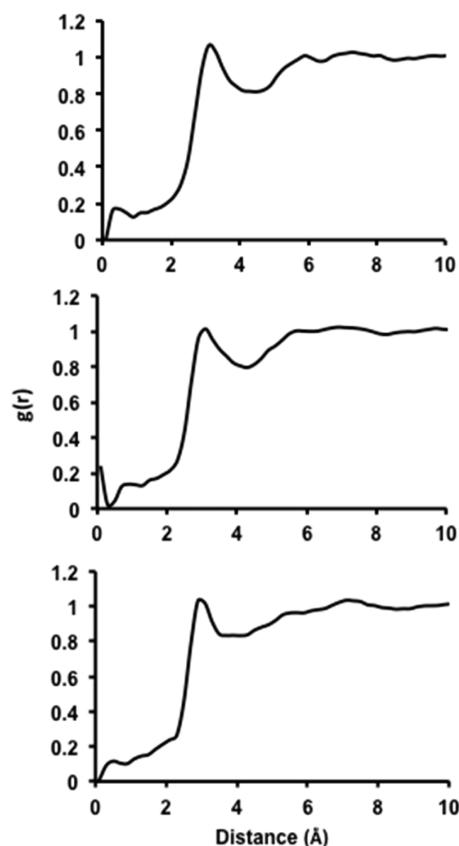
windows (Figure 4). Similar plots for OSRW simulations are shown in Figure 5, which naturally included all λ states visited during the simulations. The first apparent difference between the two methods is that the BAR RDFs have water peaks closer to the solute at around 1 Å. Note that, in addition to the difference in how the vdW and electrostatic solute–water interactions are scaled (sequential in BAR and simultaneous in OSRW), an equal number of coordinate structures were saved for each λ window in BAR simulations while the OSRW had an uneven distribution of λ values given its nature of importance sampling. Both factors would affect the overall RDF. According to these RDFs, there is much less structure in the water around the polar groups of the solute during OSRW simulations, which we attribute to the "flattened" energy surface by the biasing potential introduced in OSRW. Another interesting feature is that there is a consistent but faint "peak" at around 3 Å for the OSRW RDFs.

As seen from the time evolution of hydration free energy for compound **17** (Figure 2 bottom), the OSRW simulations actually first approached −15 kcal/mol in the beginning of the simulation but quickly increased to around −13 kcal/mol. This behavior was also observed in some other repeated OSRW simulations. The results suggest that BAR and OSRW may have sampled different phase space within the limited simulation time in this study. Thus, we have further extended the simulations for both methods. For BAR, we increased the simulation length at each window from 1 to 5 ns, while for OSRW we extended the simulation from 4 ns and 8 ns and added an additional four independent simulation runs for compound **17**. The difference between the hydration free energy from the two methods dropped to 1.09 kcal/mol but was still greater than the statistical uncertainty. This indicates that, for molecules such as compound **17**, substantially long simulations may be necessary to fully converge the answer.

Convergence and Efficiency of the Free Energy in OSRW and BAR. To examine the computational efficiency and convergence of OSRW and BAR methods, we analyze the simulations data for compounds **1**, **11**, and **17** as examples. The cumulative hydration free energies of these compounds
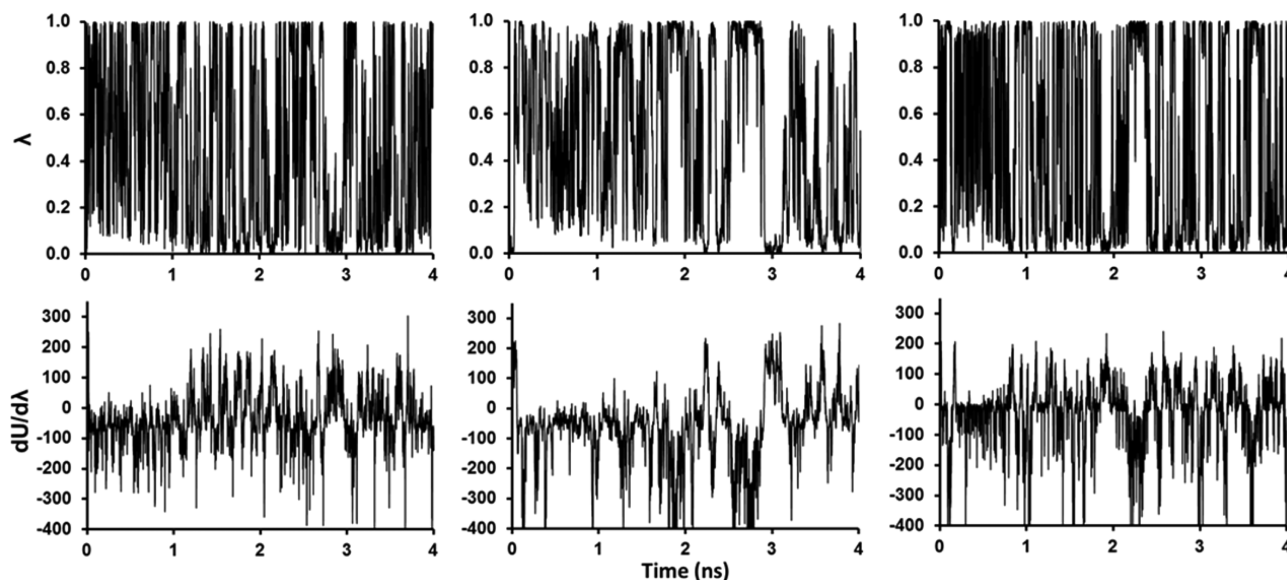

Figure 6. λ and F_λ values over simulation time from the OSRW simulations of compounds **1**, **11**, and **17** (data collected at every 5 ps). The spikes in F_λ correspond to barriers crossed by the system.
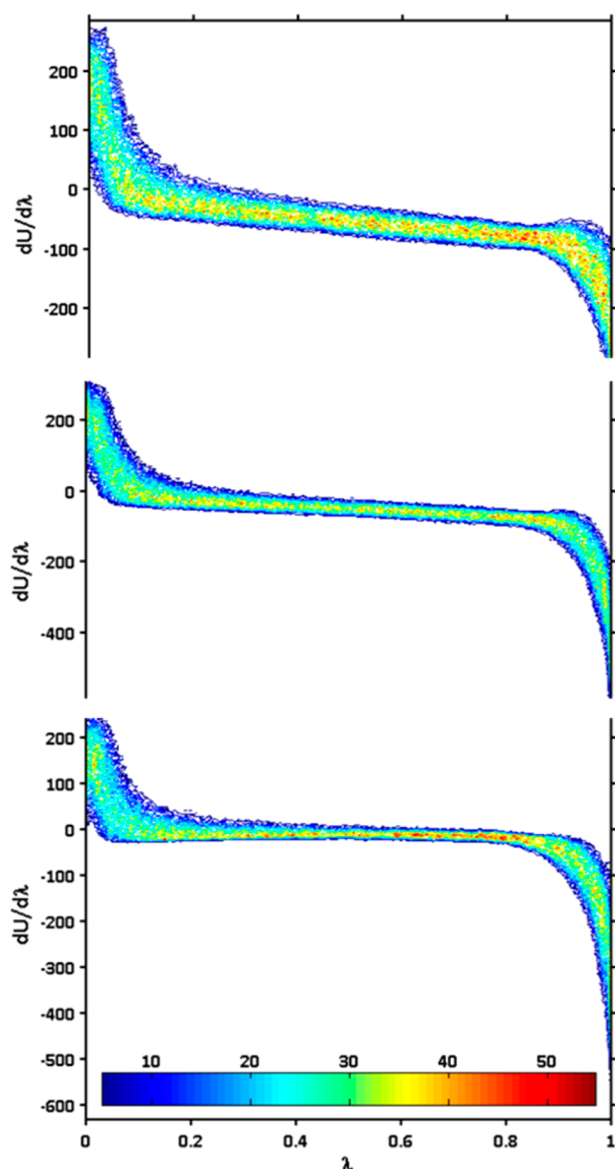
Figure 7. Plots of free energy surfaces for compounds **1** (top), **11** (middle), and **17** (bottom). The color represents the number of visits to a particular state.

calculated by OSRW as a function of time (Figure 2) clearly demonstrate its efficiency and convergence. For the neutral compound **17**, the HFE has reasonably converged in about 300 ps in a single OSRW simulation. For the two charged compounds **1** and **11**, it takes about 1.5 ns to reach a similar level of convergence while the absolute value of the HFE is much large for charged compounds.

Since the free energy is dependent on the λ and $\partial U/\partial \lambda$, a sufficient sampling of the system in both spaces is crucial to the accuracy of the estimated free energy. Distributions of λ and its generalized force $F_\lambda = \partial U/\partial \lambda$ against time are shown in Figure 6. It can be seen that the OSRW method samples the order parameter space multiple times during the simulation. There are frequent peaks and valleys in the $\partial U/\partial \lambda$ plot. By examining the free energy surface with respect to $\partial U/\partial \lambda$ and λ in Figure 7, it can be seen that the large positive values of $\partial U/\partial \lambda$ correspond to the initial state when the solute is mostly decoupled from the environment while the negative attractive $\partial U/\partial \lambda$ occurs toward the end state when $\lambda = 1$. There is a

broad and flat region in between, which corresponds to $\partial U/\partial \lambda$ around zero. A significant feature of the OSRW is that it flattens the energy surface in both λ and $\partial U/\partial \lambda$ dimensions, while the latter is the exact quantity needed to compute free energy ($G = \int \langle \partial U/\partial \lambda \rangle /\mathrm{d}\lambda$). Most other methods that sample only the λ space do not actively handle the energy barriers along $\partial U/\partial \lambda$ at a given λ. Figure 7 shows that the surface along $\partial U/\partial \lambda$ at a given λ can be complicated. In a typical OSRW simulation, the system quickly reaches the minimum (negative) $\partial U/\partial \lambda$ region and biasing Gaussian potentials are added in both λ and $\partial U/\partial \lambda$ dimensions until it reaches the middle flat region ($\partial U/\partial \lambda \sim$ zero) and then glides back and forth along the λ dimension many times. Sufficient sampling near both initial and end states is important as both contribute significantly to the free energy.

It should be noted that our implementation involves the use of metadynamics for the calculation of the biasing potential. The choice of biasing Gaussian heights and λ particle movement has not been fully optimized in this study. In addition, other implementations alternative to metadynamics can be used and have been explored.[23]

The convergence of the free energy calculated using BAR, as seen for compounds **1**, **11**, and **17** in Figure 3, generally takes ~300 ps simulations for each window to converge to a reasonable level. However, if we take the number of windows needed in BAR simulation into account, it is roughly equivalent to 8 ns of a single molecular dynamics simulation. Thus, the superior efficiency of OSRW is apparent.

The OSRW simulation is in general easier to set up and process as compared with the BAR method. In the TINKER implementation, only a single molecular dynamics simulation is required, while the BAR approach requires the maintenance of ~25 simulations (in this study). A way to improve the sampling ability of the BAR method is to allow configuration exchanges among different simulation windows as in the multistate BAR (mBAR) method.[48]

■ CONCLUSION

We have implemented the OSRW method for the polarizable multipole based AMOEBA force field in TINKER and applied it to the hydration free energy calculations. All pairwise interactions including vdW and electrostatics in real space were treated with soft-core. The dual topology approach is applied to all nonpairwise interactions such as the reciprocal component of Ewald, polarization energy, and forces. Our implementation involves the use of metadynamics to introduce the total biasing potential while other methods used to improve the overall robustness of the OSRW method are under consideration.[23] Current implementation in TINKER can be used to evaluate hydration as well as binding free energy of host–guest systems.

The OSRW method has been shown to exhibit superior efficiency for hydration free energy calculations which can be reliably calculated by using traditional methods such as BAR. The hydration free energies for a set of 20 compounds were computed by using both BAR and OSRW. We found excellent agreement between the two, with a RMSD of 0.49 kcal/mol. A noticeable difference between the OSRW and BAR results was observed for compound **17**, which has two hydroxyl groups and a fluorine atom attached to a benzene ring, even after the simulations were extended significantly. This discrepancy shows that even for a relatively small solute, the sampling of solute–solvent configurations can be challenging. In the BAR method, 1 ns at each of 25 steps or more have been performed. For OSRW, one simulation of 4 ns led to similar results within the

statistical uncertainty. While further studies are required to thoroughly evaluate the different free energy methods, our results suggest that OSRW is an efficient alternative sampling method that polarizable force fields can benefit from.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Hydration free energy of compound **17** from BAR simulations with different numbers of perturbation steps for vdW interaction. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: pren@mail.utexas.edu.

Author Contributions
§Contributed equally to this work.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93* (7), 2395–2417.

(2) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303* (5665), 1813–1818.

(3) Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **2006**, *125* (8), 084902–084917.

(4) Schnieders, M. J.; Baltrusaitis, J.; Shi, Y.; Chattree, G.; Zheng, L.; Yang, W.; Ren, P. The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field. *J. Chem. Theory Comput.* **2012**, *8* (5), 1721–1736.

(5) Shi, Y.; Wu, C.; Ponder, J. W.; Ren, P. Multipole electrostatics in hydration free energy calculations. *J. Comput. Chem.* **2011**, *32* (5), 967–977.

(6) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2009**, *5* (2), 350–358.

(7) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003**, *119* (11), 5740–5761.

(8) Shirts, M. R.; Pande, V. S. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.* **2005**, *122* (13), 134508–134520.

(9) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. *J. Chem. Theory Comput.* **2005**, *2* (1), 128–139.

(10) Baker, C. M.; Lopes, P. E. M.; Zhu, X.; Roux, B.; MacKerell, A. D. Accurate Calculation of Hydration Free Energies using Pair-Specific Lennard-Jones Parameters in the CHARMM Drude Polarizable Force Field. *J. Chem. Theory Comput.* **2010**, *6* (4), 1181–1198.

(11) Paliwal, H.; Shirts, M. R. A Benchmark Test Set for Alchemical Free Energy Transformations and Its Use to Quantify Error in Common Free Energy Methods. *J. Chem. Theory Comput.* **2011**, *7* (12), 4115–4134.

(12) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **2010**, *6* (3), 607–624.

(13) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98* (7), 1978–1988.

(14) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127–6129.

(15) Steinbrecher, T.; Joung, I.; Case, D. A. Soft-core potentials in thermodynamic integration: Comparing one- and two-step transformations. *J. Comput. Chem.* **2011**, *32* (15), 3253–3263.

(16) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **2004**, *25* (2), 265–284.

(17) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(18) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107* (24), 5933–5947.

(19) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22* (2), 245–268.

(20) Zheng, L.; Chen, M.; Yang, W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (51), 20227–20232.

(21) Zheng, L.; Chen, M.; Yang, W. Simultaneous escaping of explicit and hidden free energy barriers: Application of the orthogonal space random walk strategy in generalized ensemble based conformational sampling. *J. Chem. Phys.* **2009**, *130* (23), 234105–234114.

(22) Min, D.; Zheng, L.; Harris, W.; Chen, M.; Lv, C.; Yang, W. Practically Efficient QM/MM Alchemical Free Energy Simulations: The Orthogonal Space Random Walk Strategy. *J. Chem. Theory Comput.* **2010**, *6* (8), 2253–2266.

(23) Zheng, L.; Yang, W. Practically Efficient and Robust Free Energy Calculations: Double-Integration Orthogonal Space Tempering. *J. Chem. Theory Comput.* **2012**, *8* (3), 810–823.

(24) Wang, Q.; Bryce, R. A. Accounting for non-optimal interactions in molecular recognition: a study of ion-[small pi] complexes using a QM/MM model with a dipole-polarisable MM region. *Phys. Chem. Chem. Phys.* **2011**, *13* (43), 19401–19408.

(25) Hensen, C.; Hermann, J. C.; Nam, K.; Ma, S.; Gao, J.; Höltje, H.-D. A Combined QM/MM Approach to Protein−Ligand Interactions: Polarization Effects of the HIV-1 Protease on Selected High Affinity Inhibitors. *J. Med. Chem.* **2004**, *47* (27), 6673–6680.

(26) Orozco, M.; Luque, F. J.; Habibollahzadeh, D.; Gao, J. The polarization contribution to the free energy of hydration. *J. Chem. Phys.* **1995**, *102* (15), 6145–6152.

(27) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114* (8), 2549–2564.

(28) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7* (10), 3143–3161.

(29) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046–4063.

(30) Jiao, D.; Zhang, J.; Duke, R. E.; Li, G.; Schnieders, M. J.; Ren, P. Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential. *J. Comput. Chem.* **2009**, *30* (11), 1701–1711.

(31) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562–12566.

(32) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of protein−ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (17), 6290−6295.

(33) Shi, Y.; Zhu, C. Z.; Martin, S. F.; Ren, P. Probing the Effect of Conformational Constraint on Phosphorylated Ligand Binding to an SH2 Domain Using Polarizable Force Field Simulations. *J. Phys. Chem. B* **2012**, *116* (5), 1716−1727.

(34) Zhang, J.; Yang, W.; Piquemal, J. P.; Ren, P. Modeling Structural Coordination and Ligand Binding in Zinc Proteins with a Polarizable Potential. *J. Chem. Theory Comput.* **2012**, *8*, 1314−1324.

(35) Grossfield, A.; Ren, P.; Ponder, J. W. Ion solvation thermodynamics from simulation with a polarizable force field. *J. Am. Chem. Soc.* **2003**, *125* (50), 15671−15682.

(36) Jiao, D.; King, C.; Grossfield, A.; Darden, T. A.; Ren, P. Simulation of Ca2+ and Mg2+ solvation using polarizable atomic multipole potential. *J. Phys. Chem. B* **2006**, *110* (37), 18553−18559.

(37) Rogers, D. M.; Beck, T. L. Quasichemical and structural analysis of polarizable anion hydration. *J. Chem. Phys.* **2010**, *132* (1), 014505−014516.

(38) Wu, J.; Chattree, G.; Ren, P. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012**, *131* (3), 1−11.

(39) Mu, X.; Wang, Q.; Wang, L.-P.; Fried, S. D.; Piquemal, J.-P.; Dalby, K. N.; Ren, P. Modeling Organochlorine Compounds and the σ-Hole Effect Using a Polarizable Multipole Force Field. *J. Phys. Chem. B* **2014**, DOI: 10.1021/jp411671a.

(40) Pearlman, D. A.; Kollman, P. A. The lag between the Hamiltonian and the system configuration in free energy perturbation calculations. *J. Chem. Phys.* **1989**, *91* (12), 7831−7839.

(41) Kong, X.; Brooks, C. L. λ-dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105* (6), 2414−2423.

(42) Halgren, T. A. The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.* **1992**, *114* (20), 7827−7843.

(43) Rocklin, G. J.; Boyce, S. E.; Fischer, M.; Fish, I.; Mobley, D. L.; Shoichet, B. K.; Dill, K. A. Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site. *J. Mol. Biol.* **2013**, *425* (22), 4569−4583.

(44) Kastenholz, M. A.; Hünenberger, P. H. Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation. *J. Chem. Phys.* **2006**, *124* (22), 224501−224520.

(45) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hunenberger, P. H. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. *J. Chem. Phys.* **2013**, *139* (18), 184103−184134.

(46) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **1992**, *97* (3), 1990−2001.

(47) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101−014107.

(48) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129* (12), 124105−124114.