JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

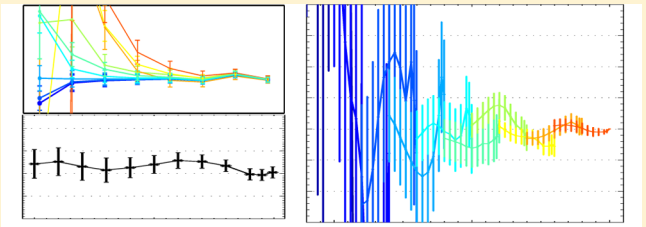
Using Multistate Reweighting to Rapidly and Efficiently Explore Molecular Simulation Parameters Space for Nonbonded Interactions

Himanshu Paliwal and Michael R. Shirts*

Department of Chemical Engineering, University of Virginia, Charlottesville, Virginia 22094, United States

Supporting Information

ABSTRACT: Multistate reweighting methods such as the multistate Bennett acceptance ratio (MBAR) can predict free energies and expectation values of thermodynamic observables at poorly sampled or unsampled thermodynamic states using simulations performed at only a few sampled states combined with single point energy reevaluations of these samples at the unsampled states. In this study, we demonstrate the power of this general reweighting formalism by exploring the effect of simulation parameters controlling Coulomb and Lennard-Jones cutoffs on free energy calculations and other observables.



ΔΔG as a multidimensional function of seven nonbonded interaction parameters

Using multistate reweighting, we can quickly identify, with very high sensitivity, the computationally least expensive nonbonded parameters required to obtain a specified accuracy in observables compared to the answer obtained using an expensive "gold standard" set of parameters. We specifically examine free energy estimates of three molecular transformations in a benchmark molecular set as well as the enthalpy of vaporization of TIP3P. The results demonstrates the power of this multistate reweighting approach for measuring changes in free energy differences or other estimators with respect to simulation or model parameters with very high precision and/or very low computational effort. The results also help to identify which simulation parameters affect free energy calculations and provide guidance to determine which simulation parameters are both appropriate and computationally efficient in general.

1. INTRODUCTION

In order to understand and predict molecular processes, we frequently need to estimate thermodynamic properties such as free energies, enthalpies, entropies, and ensemble averages of structural quantities. In some cases, we can use statistical mechanical techniques to construct analytic expressions for thermodynamic variables with various levels of approximation. However, for properties that require more molecular detail, we require molecular simulation techniques. Molecular dynamics (MD) simulations and Monte Carlo (MC) simulations are now used for *in silico* molecular design tasks such as engineering molecular recognition[1] and drug design.[2−5] For such problems, we need methods to directly estimate the true thermodynamic properties of a given model molecular system, as efficiently as possible, with no assumptions beyond those of the model itself.

Tasks such as computing the properties of many similar systems or refining force fields using very large and diverse molecular sets require high throughput calculations that are both accurate and precise.[6−9] Free energy estimators, for example, typically require collecting samples from every state for which the thermodynamic observables are to be estimated, and potentially nonphysical intermediate states as well. As the number of states run into thousands, generating sufficient number of samples for all states becomes impractical.

However, the sampled states contain information about nearby states in parameter space, as they sample the same set of configurations with similar probability as would occur during direct simulations of these other similar choices of parameters. Such states are typically referred to as having high phase space overlap. The differences between the probability of any given configuration can be computed exactly using the ratio of the Boltzmann distributions resulting from the choices of parameter set. Standard single reference reweighting techniques can be used to compute free energies and expectation values between thermodynamic states with high phase space overlap.[10] A number of useful methods have been developed around reweighting from a single reference state,[11−13] though they have clear limitations when used to calculate the properties of states with little phase space overlap from the reference state.[7,14] However, it is possible to generate significantly improved estimates of thermodynamic observables if reweighting is performed from multiple sampled reference states.[15−17] For example, the multistate Bennett acceptance ratio (MBAR)[17] is a statistically optimal estimator that reweights samples from all sampled states and predict thermodynamic properties at any of these or any other unsampled states.

In this study, we demonstrate the power of this multistate reweighting approach by scanning over the space of simulation parameters controlling nonbonded interactions without actually simulating these many alternate states. We scan this space to identify computationally efficient parameter choices that give

Received: June 14, 2013
Published: September 27, 2013

   dx.doi.org/10.1021/ct4005068 | J. Chem. Theory Comput. 2013, 9, 4700−4717

statistically indistinguishable free energy differences and thermodynamic expectations from those obtained with the most rigorous, computationally expensive parameters. This reweighting is performed using a free energy calculation carried out at a single reference set of parameters. However, it is still multistate reweighting, rather than single state reweighting, since the free energy calculation itself involves simulations at a series of alchemical states interpolating between the two physical end states.

This approach is appropriate for use in other such sensitivity analysis of many parameters simultaneously. For example, this approach is appropriate for computing changes in free energies with respect to model parameters, as noted in a study of changes in geometry parameters for water,[18] although in that case only small numbers of parameters were examined rather than the large-scale scan presented here.

The choice of statistical estimator used to calculate equilibrium expectation values of thermodynamic observables and free energy differences affects both the accuracy and precision of these estimates.[7,14,19−27] The differences in bias and uncertainty in the observables are primarily due to the way each estimator operates on the sampled energies as well as the assumptions made in the algorithm. MBAR estimates of thermodynamic observables generally have lowest bias and provably have the lowest statistical error among all general analysis methods for typical molecular transformations.[14,17] MBAR is particularly useful for the present study because of its correct treatment of statistical uncertainty, especially when examining differences between quantities that are computed with the same set of sampled data and thus are highly correlated. In this study, we will focus only on issues of sample analysis, rather than sample generation, and so, standard molecular dynamics is used to generate the samples.

The accuracy of any thermodynamic property estimator is limited by the quality of potential energy calculations of the samples, as these energies are inputs to the estimator. If the potential energies are not accurately calculated for the model, statistical errors may be swamped by biases due to approximations in the potential. There are of course always approximations in a given model, but there are also approximations in the simulation methods used to generate samples of the model, usually to improve simulation speed, and we must make sure that we avoid errors due to such approximations. Otherwise, simulations of putatively the same model performed with different simulation codes will give different results.

One important caveat is that for our multistate reweighting method (or *any* reweighting method for that matter) to work, we are limited to varying nonbonded parameters that still result in well-defined, usually Boltzmann distributions. For molecular dynamics, this means that the approximated forces must still be the gradient of the approximated potential. Approximations such as the convergence tolerance of a self-consistent integrator or a large number of time steps between neighbor-list updates result in a distribution that is not Boltzmann, and thus cannot be treated via reweighting, at least not with Boltzmann distribution statistics. This also means methods such as Berendsen thermostats and barostats, which give incorrect potential energy and volume distributions,[28] are also inappropriate to use when generating samples for single or multistate reweighting applications.

By far, the most common type of force field in biomolecular or materials simulations, due to both the computational efficiency and ease of implementation, is a pairwise additive force field. Most commonly, nonbonded energy terms in these force fields consist of fixed partial charges and Lennard-Jones nonbonded interactions. Short-range, bonded terms of a large variety are used to approximate local quantum mechanical behavior. The bonded terms are usually calculated without additional approximations as the cost is negligible. However, to accelerate the calculations, nonbonded interaction potentials are usually evaluated as sums of short-range and long-range contributions, frequently requiring a number of assumptions, approximations, and tunable parameters to accelerate the long-range contribution with acceptable losses of accuracy. What 'acceptable' means in this context must generally be determined by the user and the application of interest.

Considerable research has been done to increase the efficiency and accuracy of Coulomb and Lennard-Jones long-range calculations. Efficient algorithms such as particle mesh Ewald (PME),[29,30] reaction-field,[31] and particle−particle-particle-mesh (P3M)[32] accelerate the calculation of Coulomb contributions using different assumptions and approximations. All of these methods have been a matter of significant study. Smith and Pettitt[33] suggest that for solvents with high relative permittivity, such as water, Ewald artifacts are negligible but are more significant for solvents with low permittivity. Mark and Nilson[34] found that PME was more reliable than group based cutoff methods for quantities such as the diffusion coefficient, radial distribution function, distance-dependent Kirkwood G factor and average potential energy. Henrik[35] found that PME is more computationally efficient at the same level of accuracy than the original Ewald method and the fast multipole methods. However, Hünenberger and McCammon found that Ewald methods can introduce finite-size artifacts by imposing artificial periodicity on systems which are inherently nonperiodic.[36] Overall, despite some potential drawbacks, Ewald methods, especially implementations of PME, seem to be the most robust method to handle long-range electrostatics in molecular simulations and have become the most common electrostatics method in molecular simulation.

The long-range component of Lennard-Jones dispersion, usually treated using an r^{-6} functional form, is most typically calculated using an analytical correction that assumes the radial distribution function $g(r) = 1$ at long-range. A proper treatment of long-range dispersion energy is required to get proper enthalpies, densities, and free energies from a simulation.[37] A homogeneous analytical correction is sufficient to properly capture thermodynamic properties for pure fluids and small molecule solvation.[27,37] However, nonistropic systems require additional nonistropic corrections, either as inclusion of very long-range interactions at infrequent intervals[27,38] or Ewald summation in the dispersion term.[27,39]

Less work has gone into the specific choice of the tunable parameters associated with techniques for long-range treatment of electrostatic and Lennard-Jones energies. The choice of these parameters can significantly effect the speed and accuracy of the nonbonded potential calculations. Abraham[40] and Wang[41] have proposed strategies for selecting PME parameters, tuning them by minimizing the computational cost for a given tolerance in force calculations. In both studies, a trajectory is generated with a certain set of nonbonded parameters. The same trajectory is then rerun with different parameters, scanning the entire parameter space until a desired tolerance in the average force calculations is reached. Wang et al.[41] calculate the error estimates using both the analytical and *ik*-differentiation

(differentiation in Fourier space) schemes to obtain an optimized set of PME parameters. They observed that the optimization scheme based on force calculation can be unreliable when used for nonhomogeneous systems such as proteins in water or lipid bilayers.

Even fewer papers have looked at the tuning of Lennard-Jones cutoffs. Shirts et al.[27] found that analytical dispersion corrections were very accurate for small molecule solvation or small molecule liquids for cutoffs as low as 0.8 nm but failed badly for protein−ligand binding calculations for cutoffs as long as 1.5 nm. In't Veld et al.[39] found serious issues in computation of surface tension when using analytical corrections for Lennard-Jones cutoffs at most cutoff lengths.

The effect of these parameters on free energy calculations, which involve simulations at multiple states, has received very little attention in the literature. Errors in potential energy that are dependent on the coupling parameter λ in free energy calculations could have significant effects that may not show up when computing properties at a single end state. Measuring the effects of changing simulation parameters for quantities such as ensemble averages and especially free energies can be challenging because the statistical error must be reduced to a low enough level to observe any change between two independent simulations. This can require extremely long simulations to accurately detect small effects. For free energies, these calculations require converging a series of intermediate simulations, rather than just a single simulation. It is therefore unclear exactly how the parameters controlling the long-range terms must be treated in order to accurately predict thermodynamic observables such as enthalpy and free energy differences. Given the statistical error in any such computation of observables, are there levels of approximation that give negligible effect on the statistical accuracy but still significantly increase simulation throughput?

To show the power of this multistate reweighting method, we identify the most computationally efficient nonbonded parameters that are statistically indistinguishable from results converged to the effectively infinite cutoff limit, examining both free energies of solvation and enthalpies of vaporization. We calculate these quantities with very high precision by rapidly estimating the difference between the observable estimated using a particular combination of nonbonded simulation parameters and the converged, expensive estimate of the observable over a large multidimensional grid of parameters. Importantly, we find that because of the high overlap in phase space between changes in nonbonded interactions, this calculation can be carried out with samples collected from a single relatively inexpensive set of nonbonded parameters. The information obtained in this process provides quantitative guidance for the choice of Lennard-Jones and PME cutoff parameters.

Varying all the nonbonded simulation parameters independently generates a combinatorially large number of possible settings. We can simplify the process by noting that the optimized Lennard-Jones and Coulomb simulation parameters must both independently be converged. It may be possible to make choices of the two sets of the parameters with offsetting errors. However, any such offset is likely to be system dependent, involving a balance of the dispersion, repulsion and dipolar density; we should therefore minimize the error for each of the two interaction types independently.

We assume that we can perform the two optimizations serially, first optimizing the electrostatic parameters while keeping Lennard-Jones parameters set at the initial reference choice, and then optimizing the Lennard-Jones parameters. In theory, this process would need to be self-consistently iterated to guarantee statistical convergence. In the end, we find that a single pass of this two step process converges to an optimized parameter set that is statistically indistinguishable from essentially infinite cutoff parameters, with a final refinement that marginally improves performance.

We specifically evaluate free energies for the three molecular transformations proposed in our benchmark set[14] as test systems, as they represent difficult versions of standard problems occurring in small molecule free energy calculations. The three benchmark systems are (a) united atom methane solvation in TIP3P water, (b) the dipole inversion null transformation (switching from +1/−1 to −1/+1 charges in a diatomic molecule with UA ethane parameters) in TIP3P water, and (c) united atom anthracene solvation in TIP3P water. The observables we choose to evaluate for this study are (a) the free energy estimate of the molecular transformations in the benchmark molecular set, (b) the enthalpy change during molecular transformations, which for methane and anthracene solvation is simply the enthalpy of solvation, and (c) the molar heat of vaporization of TIP3P water. Force field parameters are given in the previous paper.[14] Because of previous mentioned ambiguities and artifacts in treatment of PME when the total charge changes during the transformation, we do not test parameters for ion solvation because of the lack of a correct reference. We validate our predictions by comparing the predictions using only the samples from benchmark states with results generated using samples from simulations run at the expensive and the proposed optimized parameter sets directly.

2. METHODS

2.1. Parameters in Short and Long-Range Nonbonded Interaction Calculation. We first describe the details of the different nonbonded interaction parameters involved in the calculation of long-range contributions in GROMACS.[42,43] These details are generally but not always completely representative of the methods used in other simulation codes; dealing with all possible methods of computing long-range interactions is beyond the scope of this study.

Lennard-Jones Interactions. Lennard-Jones (LJ) interactions are calculated pairwise for the short-range out to $r_{c,LJ}$, the LJ cutoff (in GROMACS, rvdw). The long-range contribution from $r_{c,LJ}$ to $r = \infty$ is estimated using an analytical correction integrating the pair potential out to infinity by assuming the radial distribution function $g(r) = 1$ beyond the LJ cutoff.

$$\begin{aligned} U_{LJ} &= U_{r<r_{c,LJ}} + U_{r\geq r_{c,LJ}} \\ &= U_{LJ,short} + \frac{N}{2}\rho \int_{r_{c,LJ}}^{\infty} U(r)g(r)4\pi r^2\, dr \\ &= U_{LJ,short} + 2\pi N\rho \int_{r_{c,LJ}}^{\infty} 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] dr \\ &= U_{LJ,short} + 8\pi N\rho\varepsilon\sigma^3\left[\frac{1}{9}\left(\frac{\sigma}{r_{c,LJ}}\right)^{9} - \frac{1}{3}\left(\frac{\sigma}{r_{c,LJ}}\right)^{3}\right] \end{aligned} \quad (1)$$

This correction is typically known as the *dispersion correction* as the r^{-12} repulsion contribution to the integral almost completely disappears at typical cutoff distances of 3σ, where

dx.doi.org/10.1021/ct4005068 | J. Chem. Theory Comput. 2013, 9, 4700−4717

σ is the Lennard-Jones radius. The free energy correction is equal to the energy correction for fixed density, as all configurations have the exact same energy correction term. This approximation is easily generalized to systems with multiple LJ sites on the solute or different compositions of solvent.[27]

The dispersion correction must be averaged over all configurations in the case of constant pressure simulations, as each configuration has a different density, and a corresponding correction to the virial must also be added for constant pressure MD algorithms.[27,37] For solvation in homogeneous liquids, this analytical correction will generally be sufficient using cutoffs around $3-5\sigma$ of the largest particles.[27] For heterogeneous systems such as membrane simulations or very large solute molecules, extremely long cutoffs or a more sophisticated long-range treatment is required,[27,39] and the optimized parameters derived in this study will not necessarily be applicable in those situations.

Coulomb Interactions. Individual Coulombic interactions are proportional to $1/r$ and thus converge slowly with distance. They must be summed over multiple periodic copies of the system because of this slow convergence in order to approximate bulk behavior. This calculation can be accelerated by taking the infinite sum over all periodic copies, and splitting it into a pairwise real space short-range contribution and a Fourier space long-range contribution. In the particle mesh Ewald (PME)[29] technique and its smooth variant SPME,[30] the "smeared" long-range charge distribution is interpolated onto a grid where a fast Fourier transform (FFT) can be performed, and the self-interaction portion of the "smearing" is analytically removed.

$$
\begin{aligned}
U_{\text{Coulomb}} &= U_{\text{real}} + U_{\text{Fourier}} + U_{\text{self}} \\
U_{\text{Coulomb}} &= \frac{f}{2}\sum_{n_x}\sum_{n_y}\sum_{n_z}\sum_{i}^{N}\sum_{j}^{N}\frac{q_i q_j}{r_{ij,n}} \\
U_{\text{real}} &= \frac{f}{2}\sum_{i,j}^{N}\sum_{n_x}\sum_{n_y}\sum_{n_z}' q_i q_j \frac{\operatorname{erfc}(\beta r_{ij,n})}{r_{ij,n}} \\
U_{\text{Fourier}} &= \frac{f}{2\pi V}\sum_{i,j}^{N} q_i q_j \times \\
&\qquad \sum_{m_x}\sum_{m_y}\sum_{m_z}' \frac{\exp(-(\pi \mathrm{m}/\beta)^2 + 2\pi i \mathrm{m}\cdot(r_i - r_j))}{\mathrm{m}^2} \\
U_{\text{self}} &= -\frac{f\beta}{\sqrt{\pi}}\sum_{i}^{N} q_i^2
\end{aligned}
\tag{2}
$$

The primary contribution to the Fourier space sum are the first M_c wave vectors in each of the m_x, m_y, and m_z directions,[44] corresponding to the longest wavelengths. For PME calculations, we must specify an order of interpolation of the charge distribution onto a grid (in GROMACS, pme_order), typically using a B-spline interpolation function, which is smoother than polynomial interpolation and allows higher accuracy with increasing order of interpolation.[30] B-spline interpolation also allows the force expressions to be evaluated analytically by differentiating the real and reciprocal energy equations, rather than using finite difference techniques. While calculating Coulomb potential using the Ewald sums, we assume that the real space sum becomes negligible before the Coulomb cutoff, $r_{c,\text{coul}}$ (in GROMACS, rcoul). The Ewald tolerance, etol (ewald_rtol in GROMACS), is the relative value of the real space potential at the Coulomb cutoff compared to the total electrostatic energy at that distance. The Ewald tolerance parameter used in GROMACS is related to the Gaussian β parameter in eq 2 used in some other simulation programs by

$$\text{etol} = \operatorname{erfc}(\beta r_{c,\text{coul}}) \tag{3}$$

A lower tolerance will give a more accurate direct sum but the number of included wave vectors for the reciprocal sum must be increased to compensate. The Fourier spacing is the maximum grid spacing for the fast Fourier transform grid used in PME. In GROMACS, this grid spacing is used to specify the number of grid points in each of the x, y, and z directions.

An additional parameter that must be set in GROMACS is rlist, the neighbor list distance. The neighbor list must be larger than the distance between atoms in the same charge group, or errors in energy can appear between neighbor list updates but has negligible effect on the energy as long as it is longer than this distance. For the systems in this study, a neighbor list 0.2 nm larger than the maximum of the Coulombic and Lennard-Jones cutoffs is sufficient and does not significantly increase the computation time. This largest cutoff +0.2 nm rule for rlist is used for all simulations in this study.

Switching Functions. In both the Lennard-Jones and Coulomb cases, as described above, there is a discontinuity in the short-range potential and the force at the cutoff. This abrupt cutoff can lead to errors in the numerical integration, as the force is discontinuous for a particle crossing the cutoff. A switching function in the potential can be introduced at the cutoff for both the Lennard-Jones and Coulomb potentials. A spline of at least third order in the potential is required so that the force is differentiable. In GROMACS, this switching function is quintic:

$$
\begin{aligned}
U_{\text{nonbonded}} &= S(r)U(r) \\
S(r) &= 1 && r < r_{\text{sw}} \\
&= 1 - 10\frac{(r - r_{\text{sw}})^3}{(r_c - r_{\text{sw}})^3} + 15\frac{(r - r_{\text{sw}})^4}{(r_c - r_{\text{sw}})^4} && r_{\text{sw}} < r \le r_c \\
&\quad - 6\frac{(r - r_{\text{sw}})^5}{(r_c - r_{\text{sw}})^5} \\
&= 0 && r > r_c
\end{aligned}
\tag{4}
$$

In GROMACS, the switch begins at r_{sw} and ends at r_c, and is specified by rvdw_switch and rcoul_switch depending on the potential type. We will occasionally refer to a switching distance $r_{\text{dsw}} = r_c - r_{\text{sw}}$, a switching range.

In the case of Lennard-Jones particles, a complementary switch function can be added to the dispersion correction, meaning no nonbonded energy is lost from the full interaction as long as the radial distribution function is constant outside the start of the switch. In the case of the Coulomb interaction, a short-range real space potential switch cannot be directly matched by a compensating switch in the Fourier space potential, so non-negligible switching distances will distort the true $1/r$ potential in the cutoff region. Because etol is usually quite low near the cutoff, this energy loss should be fairly small, but it has not been quantified before. A smaller switch width will minimize this distortion of the potential. However, small switch widths also cause moderate spiking in the force at the cutoff, which can potentially affect the kinetics of the system

dx.doi.org/10.1021/ct4005068 | J. Chem. Theory Comput. 2013, 9, 4700−4717

and potentially the stability of the integration. To achieve a balance between the loss in energy without introducing a substantial force spike, we must tune the switching distance both for LJ and Coulomb potentials.

It is also possible to eliminate cutoffs using a shifting function, where the potential surface is shifted down to zero at the cutoff, or a combination of the shifts and switches.[45,46] In this study, we only examine the switching function. A shift affects the entire potential energy surface and thus causes large effects on ensemble averages such as the enthalpy or the free energy examined in this paper. On the other hand, properly balanced shift functions eliminate the force spike, and where conformational properties are of most interest, rather than free energies, they may be more appropriate. Additionally, group-based shifts for electrostatics have no effect since the shift components cancel for net-neutral groups. In any case, we will find that properly tuned switch functions do not significantly affect free energies or ensemble averages, and at least in this study, we will not explore any other the trade-offs between switching and shifting functions.

The space of all the possible nonbonded simulation parameters is clearly multidimensional. Specifically, we have the order of B-spline interpolation (order), tolerance in the real space cutoff (etol), the Fourier spacing (FS), Coulomb cutoff ($r_{c,coul}$), Coulomb switching range ($r_{sw,coul}$), Lennard-Jones cutoff ($r_{c,LJ}$), and Lennard-Jones switching range ($r_{sw,LJ}$). Many of the parameters cannot be set independently for high-quality simulation results, and the PME parameters in particular are interdependent; a good Fourier spacing for one choice of etol may not be good for another. Each choice of a set of nonbonded simulation parameter affects the computational cost and accuracy of the thermodynamic sampling differently. Certain choices of parameters, such as such as large cutoffs or very fine Fourier spacing significantly increase the simulation time. However, beyond a certain point, the increased cost gives limited returns in accuracy, and we seek to identify the speed required for a specified level of accuracy.

The "expensive" reference parameter set we examine includes an Ewald tolerance (etol) of 10^{-10} (smaller is more accurate), Fourier grid spacing (FS) of 0.04 nm (smaller is more accurate), and a B-spline order of 6 (higher is more accurate). The expensive Coulomb real-space cutoff ($r_{c,coul}$) and Lennard-Jones real-space cutoff ($r_{c,LJ}$) varied with the system and were both 1.3 nm for methane solvation, 1.5 nm for dipole inversion, and 1.4 nm for anthracene solvation, chosen to be the longest possible cutoffs for the respective simulation boxes. Switch distances for the converged PME parameters were chosen to be zero for Coulomb and 0.001 nm for Lennard-Jones; switches only improve the stability of the numerical integration, not the energies themselves. The Lennard-Jones switch distance for comparison was chosen as 0.001 nm, effectively zero, because of implementation-specific constraints on parameter choices in GROMACS.

The order of interpolation range tested was between 3 (the minimum value in GROMACS) and 6, beyond which the potential energy change was negligible for systems of this size for any choice of Coulomb parameters examined. The Fourier spacing range was between 0.04 nm (any smaller became computationally untenable because of expense, but only negligibly changed the energy) and 0.20 nm, which is longer scale than many molecular details. etol ranged between 10^{-10} and 10^{-2}, equivalent to a β between approximately 5 and 2 nm^{-1}, thus extending beyond the range of typical values for β. The largest cutoffs possible were limited by the box size for the benchmark molecular set. The lowest cutoff of 0.6 nm for all systems was set to be smaller than for all standard MD simulations.

In the first phase of Coulomb parameter optimization, LJ parameters fixed to those used in our previous benchmark study,[14] which had been selected based on anecdotal experience from previous large scale studies.[7,8,47] The selected values used to search through the multidimensional parameter space are given in Table 1.

Table 1. Nonbonded Interaction Parameter Values Examined in the Optimization Process

param.	value
order of interpolation (order)	3, 4, 5, 6
Ewald tolerance (etol)	10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}
Fourier spacing (FS) (nm)	0.04, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18, 0.20
Coulomb cutoff ($r_{c,coul}$) (nm)	0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5
Coulomb switch width (nm) $r_{c,coul}-r_{sw,coul}$	0.2, 0.18, 0.16, 0.14, 0.12, 0.10, 0.08, 0.06, 0.04, 0.02, 0.01, 0.001
LJ cutoff ($r_{c,LJ}$) (nm)	0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5
LJ switch width (nm) $r_{c,LJ}-r_{sw,LJ}$	0.2, 0.18, 0.16, 0.14, 0.12, 0.10, 0.08, 0.06, 0.04, 0.02, 0.01, 0.001

2.2. Phase one: Optimization of PME Parameters. We first perform the search through Coulomb parameter space, specifically examining the values of the order of B-spline interpolation, the Ewald tolerance, the Fourier spacing, and the Coulomb cutoff. The first portion of search was performed over 180 parameters per $r_{c,coul}$ distance, giving 1440 combinations for methane solvation, 1800 combinations for dipole inversion and 1620 combinations for anthracene solvation, since each molecule had different box size and therefore a different maximum allowable cutoff.

2.2.1. Difference in Free Energy Estimates Due to Approximations Introduced by Simulation Parameters. Samples from equilibrium isobaric−isothermal simulations corresponding to just one parameter set are used to reevaluate energies for different parameter sets. MBAR reweights the sampled energies using the reevaluated energies to estimate the observables at the unsampled parameter set. The variance-minimizing free energy estimating equation for MBAR is

$$f_i = -\ln \sum_{k=1}^{K} \sum_{n=1}^{N_k} \frac{\exp[-u_i(x_{kn})]}{\sum_{k'=1}^{K} N_{k'} \exp[f_{k'} - u_{k'}(x_{kn})]} \quad (5)$$

where K is the total number of states and N_k is the total number of uncorrelated samples available from an equilibrium simulation at state k. Each free energy is thus a weighted sum of all samples from all k states. $f_i = \beta G_i$, where G_i is the Gibbs free energy associated with state i. $u_i(x_{kn})$ is the reduced potential energy of the n^{th} sample belonging to equilibrium simulation of state k but evaluated at state i. For the NPT simulations carried out here, the reduced potential $u_i(x_{kn}) = \beta(U_i(x_{kn}) + PV_{kn})$, where P is the applied pressure and V_{kn} is the simulation volume of sample kn.

Equilibrium samples $\vec{x}_B$ were obtained from benchmark simulations from our previous study, with nonbonded simulation parameters P_B listed in Table 2. Samples from

dx.doi.org/10.1021/ct4005068 | J. Chem. Theory Comput. 2013, 9, 4700−4717

Table 2. Benchmark, Expensive, and Optimized Parameters[a]

		expensive		
param.	benchmark	optimization	validation	optimized
order of interpolation	4	6	6	4
Ewald tolerance	10^{-8}	10^{-10}	10^{-10}	10^{-4} or 10^{-5}
Fourier spacing (nm)	0.12	0.04	0.06	0.12
Coulomb cutoff (nm)	0.9	1.3[a]	1.2[a]	0.9 or 1.0
		1.5[b]	1.4[b]	
		1.4[c]	1.3[c]	
Coulomb switch (nm)	0.88	1.299[a]	1.15[a]	0.85 or 0.95
		1.499[b]	1.35[b]	
		1.399[c]	1.25[c]	
LJ cutoff (nm)	0.9	1.3[a]	1.2[a]	0.9 or 1.0
		1.5[b]	1.4[b]	
		1.4[c]	1.3[c]	
LJ switch (nm)	0.8	1.299[a]	1.15[a]	0.85 or 0.95
		1.499[b]	1.35[b]	
		1.399[c]	1.25[c]	

[a]Parameters marked with *a*, *b*, *c* are parameters for methane solvation, dipole inversion, and anthracene solvation, respectively; otherwise, the parameters are common to all three molecular sets. rlist is always 0.2 nm greater than the largest of the two nonbonded cutoff, which is a change from the benchmark. All validations are done for the first optimized parameter set.

other simulation parameter sets i are denoted as $\vec{x}_i$. Samples at the benchmark parameters were obtained from 100 5 ns uncorrelated simulation runs for free energy calculations for the systems in our benchmark study,[14] with simulation parameters noted there. The samples were stored as full precision coordinate sets 50 ps apart, which is longer than the maximum autocorrelation time of 30 ps for the potential energies found in our previous study of the same system. The samples can thus be considered effectively independent. The first 10% (0.5 ns) of data from each run was discarded, leaving at total of 9000 samples at each of the intermediate states from the free energy calculation.

We recalculate potential energies using these samples at a range of parameters sets P_i, with P_E being the most expensive nonbonded simulation parameters discussed above, corresponding to "gold standard" reference parameters. Energies were calculated with GROMACS 4.5.3 with bug fixes later incorporated into the 4.5.7 and 4.6 versions.[43] $U_i(x_B)$ is therefore the set of potential energies calculated using parameter set P_i with the samples $\vec{x}_B$, with $U_E(x_B)$ denoting energies evaluated with the expensive parameter set on samples from the simulations done with benchmark parameter set. Benchmark energies were recalculated along with all other energies to maximize consistency of the energies.

To calculate free energy differences and the uncertainties in these free energy differences between P_E and a given choice of P_i using MBAR, we first construct a 3-dimensional matrix **U** of size $(3K,3K,N_k)$, where K is the number of states in the free energy simulation performed in the benchmark state. When we re-evaluate energies for expensive and i^{th} trial parameter sets, we introduce $2K$ new thermodynamic states, K states for fully converged, "expensive" parameters, and K states for i^{th} trial parameter set. Thus, we have a total of $3K$ thermodynamic states defined. In the optimization procedure, only the first K intermediate states simulated using the original benchmark parameters have samples, though more generally for multistate reweighting, all states might include samples. There are $K = 8$ states spanning the initial and final states for methane solvation, 11 for dipole inversion, and 15 for anthracene solvation, with the choice of K for each system discussed in the previous study.[14]

We write **U** as the 3D matrix:

$$\mathbf{U}(3K \times 3K \times N_k) = \begin{bmatrix} U_0(x_{0,1\cdots N_0}) & \cdots & U_{3K-1}(x_{0,1\cdots N_0}) \\ \vdots & \ddots & \vdots \\ U_0(x_{3K-1,1\cdots N_{3K-1}}) & \cdots & U_{3K-1}(x_{3K-11\cdots N_{3K-1}}) \end{bmatrix} = \begin{bmatrix} \mathbf{U_B}(\mathbf{x_B}) & \mathbf{U_E}(\mathbf{x_B}) & \mathbf{U_i}(\mathbf{x_B}) \\ \mathbf{U_B}(\mathbf{x_E}) & \mathbf{U_E}(\mathbf{x_E}) & \mathbf{U_i}(\mathbf{x_E}) \\ \mathbf{U_B}(\mathbf{x_i}) & \mathbf{U_E}(\mathbf{x_i}) & \mathbf{U_i}(\mathbf{x_i}) \end{bmatrix} \quad (6)$$

Each element of the above matrix is a $K \times K \times N_k$ matrix as it extends out to N_k samples in the third dimension. In these submatrices, the subscript to U indicates the parameter set used for energy evaluation. The subscript to X indicates the parameter set with which the sample configurations were generated. The vector **N** in eq 7 contains the number of uncorrelated equilibrium samples collected at each state k.

$$\mathbf{N} = [N_0, ..., N_{3K-1}] \quad (7)$$

For the optimization procedure, $N_k = 0$ for the $2K$ total states of the free energy calculations carried out with parameter sets P_E and P_i. This means that $\mathbf{U}_i(x_E)$ and $\mathbf{U}_i(x_i)$ for any choice of state i are both undefined and unused in the initial phases of this study. More generally, we could use this formalism to compute free energies using samples from multiple states simultaneously. In this paper, we initially restrict ourselves to the case of sampling from a single set of parameters, which we will find will be sufficient to predict thermodynamic properties over the range of simulation parameters of interest. We can then validate the final results with data collected from multiple states.

We also note that there is no specific reason to restrict the property prediction to handling three states (reference, converged and trial) at a time; the multistate reweighting formalism is general enough to perform computations between any number of sampled and unsampled states simultaneously. For example, we could predict the properties of *all* of the unsampled states simultaneously instead of one at a time. However, the memory requirement for the entire matrix of all states become prohibitive, and it provides no extra statistical efficiency; the results only change when we change the sampled data, not the order in which the properties at unsampled parameter sets are computed.

Instead, we focus on the most flexible and straightforward approach, using the thermodynamic states with energies sampled (in this case, the K benchmark states), the states being used as a reference (in this case, the K expensive parameter states), and the set of states at one other trial parameter set of interest. The process can be made even more efficient if the free energies and weights are saved from the first round of self-consistent iteration, as generating weights for new states with no samples does not require any additional iterations, but for this study, such optimization was not used.

For this optimization process, we can rewrite the matrix as

dx.doi.org/10.1021/ct4005068 | J. Chem. Theory Comput. 2013, 9, 4700−4717

$$U(3K \times 3K \times N_k) = \begin{bmatrix} U_0(x_{0,1\cdots N_0}) & \cdots & U_{3K-1}(x_{01\cdots N_0}) \\ \vdots & \ddots & \vdots \\ U_0(x_{K-1,1\cdots N_{K-1}}) & \cdots & U_{3K-1}(x_{K-1,1\cdots N_{K-1}}) \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{U_B}(\mathbf{X_B}) & \mathbf{U_E}(\mathbf{X_B}) & \mathbf{U_i}(\mathbf{X_B}) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (8)$$

$$\mathbf{N} = [N_0, ..., N_{K-1}, 0, 0, ..., 0, 0] \quad (9)$$

The **U** matrix and the **N** vector in eqs 8 and 9 are the inputs to MBAR (eq 5) which is solved using standard optimization techniques[17] for G_i for all $3K$ states. The result is a matrix $\Delta\mathbf{G}$ with all pairwise free energy difference estimates

$$\Delta\mathbf{G}(3K \times 3K) = \begin{bmatrix} G_0 - G_0 = 0 & \cdots & G_{3K-1} - G_0 \\ \vdots & \ddots & \vdots \\ G_0 - G_{3K-1} & \cdots & G_{3K-1} - G_{3K-1} = 0 \end{bmatrix} \quad (10)$$

and also a matrix of an estimate of the statistical error $\delta(\Delta G)$ for each of these free energy differences. The uncertainties in the pairwise difference estimates ΔG are estimated using eq 12 in Shirts and Chodera.[17]

$$\delta(\Delta\mathbf{G})(3K \times 3K) = \begin{bmatrix} 0 & \cdots & \delta(G_{3K-1} - G_0) \\ \vdots & \ddots & \vdots \\ \delta(G_0 - G_{3K-1}) & \cdots & 0 \end{bmatrix} \quad (11)$$

where the diagonal is identically zero as there is no uncertainty in the value of zero for the free energy between a state and itself. For standard molecular transformation calculations, these error estimates are extremely reliable estimates of the true sample standard deviation.[14] The free energy of molecular transformation in the benchmark molecular set, solvation or dipole inversion, is equal to the difference in free energies of the end states, in this case $\Delta G_B = G_{K-1} - G_0$, $\Delta G_E = G_{2K-1} - G_K$, and $\Delta G_i = G_{3K-1} - G_{2K}$.

However, we are most interested in the free energy difference $\Delta\Delta G_{Ei} = \Delta G_i - \Delta G_E$ rather than any of the free energies individually. At first, it may seem that we cannot estimate this difference with any significant degree of precision, because of the uncertainties in the individual free energy estimations. However, the variance estimate of $\Delta\Delta G_{Ei} = \mathrm{var}(\Delta\Delta G_{Ei})$ is not simply the sum of the variances $\mathrm{var}(\Delta G_E)$ and $\mathrm{var}(\Delta G_i)$, because the two free energy estimates are computed from the same samples and are thus are highly correlated. We must instead derive the variance estimate for $\Delta\Delta G_{Ei}$ using the properties of the covariance:

$$\begin{aligned} \mathrm{var}(\Delta\Delta G_{Ei}) &= \mathrm{var}((G_{2K-1} - G_K) - (G_{3K-1} - G_{2K})) \\ &= \mathrm{cov}(G_{2K-1} - G_K, G_{2K-1} - G_K) \\ &+ \mathrm{cov}(G_{3K-1} - G_{2K}, G_{3K-1} - G_{2K}) \\ &- 2\mathrm{cov}(G_{2K-1} - G_K, G_{3K-1} - G_{2K}) \end{aligned} \quad (12)$$

where $\mathrm{cov}(x,y)$ is the covariance of the two variables x and y. With repeated application, the above equation yields the following:

$$\begin{aligned} \mathrm{var}(\Delta\Delta G_{Ei}) &= \mathrm{var}(G_{2K-1}) + \mathrm{var}(G_K) + \mathrm{var}(G_{3K-1}) \\ &+ \mathrm{var}(G_{2K}) - 2[\mathrm{cov}(G_{2K-1}, G_K) \\ &+ \mathrm{cov}(G_{3K-1}, G_{2K}) + \mathrm{cov}(G_{2K-1}, G_{3K-1}) \\ &- \mathrm{cov}(G_{2K-1}, G_K) \\ &- \mathrm{cov}(G_K, G_{3K-1}) + \mathrm{cov}(G_K, G_{2K})] \end{aligned} \quad (13)$$

Each term in the above equation can be straightforwardly extracted from the asymptotic covariance matrix, Θ_{ij}, estimated using eq 8 of Shirts and Chodera.[17]

We emphasize that this equation, and the resulting ability to calculate the covariances between two observables calculated from the data, is the absolute core of the success of this reweighting approach to estimate *changes* in thermodynamic parameters. With it, we can drastically reduce the variance in differences of thermodynamic variables differences. Once we have all the $\Delta\Delta G_{Ei}$, provided the statistical uncertainties $\delta(\Delta\Delta G_{Ei}) = (\mathrm{var}(\Delta\Delta G_{Ei}))^{1/2}$ are low enough, we can find the parameter set for which $\Delta\Delta G_{Ei}$ is less than the desired level of bias in free energy estimate and then choose the parameter set that are most computationally efficient for that level of bias.

2.2.2. Difference in Enthalpy of Transformation. We also calculate the enthalpy of molecular transformation, the difference between the initial and the final state enthalpies $(\langle H\rangle_{\lambda=1} - \langle H\rangle_{\lambda=0}) \pm \delta(\langle H\rangle_{\lambda=1} - \langle H\rangle_{\lambda=0})$. For methane solvation and anthracene solvation, the enthalpy of molecular transformation is the solvation enthalpy.

The expectation value of $H = U_{\mathrm{int}} + PV$ is the enthalpy of the simulated state. Each energy term in the **U** matrix, eq 8, is equal to total internal energy U_{int} plus the pressure times volume term PV. We can therefore use this expression **U** as the input matrix of observables to eqs 15 and 16 of Shirts and Chodera for expectations of observables.[17] Solving these equations (which requires the weights generated solving for the free energies) yields the equilibrium expectation values of enthalpies at all states and corresponding uncertainties in the expectation values.

$$\langle\mathbf{H}\rangle = \begin{bmatrix} \langle H\rangle_0 \\ \vdots \\ \langle H\rangle_{3K-1} \end{bmatrix}, \quad \delta(\langle\mathbf{H}\rangle) = \begin{bmatrix} \delta(\langle H\rangle_0) \\ \vdots \\ \delta(\langle H\rangle_{3K-1}) \end{bmatrix} \quad (14)$$

Where these two vectors are size $(3K \times 1)$, we can also obtain the pairwise difference in expectation values between two states and associated uncertainties in these differences, which are size $(3K \times 3K)$ arrays.

dx.doi.org/10.1021/ct4005068 | J. Chem. Theory Comput. 2013, 9, 4700−4717

$$\Delta\langle \mathbf{H} \rangle = \begin{bmatrix} \langle H \rangle_0 - \langle H \rangle_0 = 0 & \cdots & \langle H \rangle_{3K-1} - \langle H \rangle_0 \\ \vdots & \ddots & \vdots \\ \langle H \rangle_0 - \langle H \rangle_{3K-1} & \cdots & \langle H \rangle_{3K-1} - \langle H \rangle_{3K-1} = 0 \end{bmatrix} \quad (15)$$

$$\delta(\Delta\langle \mathbf{H} \rangle) = \begin{bmatrix} 0 & \cdots & \delta(\langle H \rangle_{3K-1} - \langle H \rangle_0) \\ \vdots & \ddots & \vdots \\ \delta(\langle H \rangle_0 - \langle H \rangle_{3K-1}) & \cdots & 0 \end{bmatrix} \quad (16)$$

We obtain $\Delta\Delta H$ from eq 15.

$$\Delta\Delta H_{Ei} = (\langle H \rangle_{2K-1} - \langle H \rangle_K) - (\langle H \rangle_{3K-1} - \langle H \rangle_{2K}) \quad (17)$$

We derive the error estimate $\delta(\Delta\Delta H_{Ei})$ using the covariance formula, similar to the one used for the error estimate of $\Delta\Delta G_{Ei}$.

$$\begin{aligned} \mathrm{var}(\Delta\Delta H_{Ei}) &= \mathrm{cov}(H_{2K-1} - H_K) - (H_{3K-1} - H_{2K}), \\ &\quad (H_{2K-1} - H_K) - (H_{3K-1} - H_{2K})) \\ &= \mathrm{var}(H_{2K-1} - H_K) + \mathrm{var}(H_{3K-1} - H_{2K}) \\ &\quad - 2[\mathrm{cov}(H_{2K-1}, H_{3K-1}) - \mathrm{cov}(H_{2K-1}, H_{2K}) \\ &\quad - \mathrm{cov}(H_K, H_{3K-1}) + \mathrm{cov}(H_K, H_{2K})] \end{aligned} \quad (18)$$

The first two variance terms in eq 18 can be directly read from the matrix in eq 16. Unlike in the case of $\delta(\Delta\Delta G_{Ei})$, the covariance terms cannot be read directly from the covariance matrix Θ corresponding to the expectation values. The estimator of the uncertainty for equilibrium expectation values is given by eq 16 in Shirts and Chodera:[17]

$$\delta^2\hat{A} \equiv \mathrm{cov}(\hat{c}_A/\hat{c}_a, \hat{c}_A/\hat{c}_a) = \hat{A}^2(\hat{\Theta}_{AA} + \hat{\Theta}_{aa} - 2\hat{\Theta}_{Aa}) \quad (19)$$

Using eq 10 of Shirts and Chodera[17] we can derive an expression for the covariance terms of the form $\mathrm{cov}(\hat{c}_A/\hat{c}_a, \hat{c}_B/\hat{c}_b)$:

$$\mathrm{cov}(\hat{c}_A/\hat{c}_a, \hat{c}_B/\hat{c}_b) = \hat{A}\hat{B}(\hat{\Theta}_{AB} - \hat{\Theta}_{Ab} - \hat{\Theta}_{aB} + \hat{\Theta}_{ab}) \quad (20)$$

The four covariance terms in eq 18 can be calculated using eq 20. This allows us to compute $\Delta\Delta H_{Ei} \pm \delta(\Delta\Delta H_{Ei})$ between the expensive parameters and any other set of simulation parameters taking into account the correlation between the two measurements. For notational clarity, we will sometimes use $\Delta\Delta G$ and $\Delta\Delta H$ to refer to $\Delta\Delta G_{Ei}$ and $\Delta\Delta H_{Ei}$ in the rest of the paper.

2.2.3. Difference in Enthalpy of Vaporization of Water. We calculate the enthalpy of vaporization of water using the estimate of enthalpy of the last alchemical state of methane solvation, which includes water with only a single non-interacting dummy united atom methane molecule. We assume that the enthalpy difference in vapor is negligible between parameter sets, which is exact in the ideal gas limit. The enthalpy of the simulation box after subtracting the kinetic energy of the dummy methane molecule is the enthalpy of liquid TIP3P water. The difference in the enthalpies of vaporization of water evaluated at two different simulation parameter sets, P_E and P_i, will be equal to the difference in the enthalpies of the last state of methane solvation of the two sets, as the PV terms and kinetic energies of the dummy methane molecule cancel out. Thus, the difference in the enthalpy of vaporization of water ΔH_{vap} can also be read directly from eq 15. Similarly, the uncertainty in ΔH_{vap} can be directly read from eq 16.

$$\begin{aligned} \Delta H_{vap} &= H_{vap,E} - H_{vap,i} \\ &\approx \langle H_{liq,i)} \rangle - \langle H_{liq,E)} \rangle \\ &= \langle H \rangle_{3K-1} - \langle H \rangle_{2K-1} \\ \delta(\Delta H_{vap}) &= \delta(\langle H \rangle_{3K-1} - \langle H \rangle_{2K-1}) \end{aligned} \quad (21)$$

2.3. Phase Two: Optimization of the Coulomb Switch. Once we have the optimized Coulomb parameters from the first phase of our search, we can investigate how much of a difference Coulomb switch parameters make in the thermodynamic estimates $\Delta\Delta G$, $\Delta\Delta H$, and ΔH_{vap}. Although the switching parameters could have been optimized simultaneously with the other electrostatic parameters, the Coulomb switch was optimized independently to make it easier to identify the effect of this switch on the thermodynamic properties, which is an open question. We construct the input matrix $\mathbf{U}$ in a similar way as described in previous section. The search proceeds identically, except now varying only the $r_{sw,coul}$. Because the optimized Coulomb cutoff is chosen as 0.9 nm (see the Results section for selection criteria) we calculate $\Delta\Delta G$, $\Delta\Delta H$, and ΔH_{vap} for Coulomb switches starting at 0.7, 0.72, 0.74, 0.76, 0.78, 0.8, 0.82, 0.84, 0.86, 0.88, 0.89, and 0.899 nm, terminating at 0.9 nm.

2.4. Phase Three: Optimization of the Lennard-Jones Parameters. After understanding the behavior of the Coulomb switch, we optimize the Lennard-Jones cutoff $r_{c,LJ}$ and Lennard-Jones switching distance $r_{sw,LJ}$ with fixed Coulomb parameters. Here, we simultaneously optimize both the LJ cutoff and the LJ switching distances. The treatment of the Lennard-Jones switch is somewhat different than in the Coulomb switch case, as in the Lennard-Jones the missing potential energy can be properly treated in the dispersion correction by analytically including the switch function in a way that it cannot be treated in the Fourier space portion of the Coulomb calculation. The maximum cutoffs, set by the box sizes of 1.3 nm for methane solvation, 1.5 nm for dipole inversion and 1.4 nm for anthracene solvation along with optimized PME parameters are chosen as the 'expensive' LJ parameter reference. Switch width varies from 0.001 to 0.2 nm as shown in Table 1.

An important factor in our search is the choice of criteria to by which we decide whether optimized parameters are sufficiently close to the more expensive parameters for general use. We consider two possible choices:

1. A set of parameters is sufficiently close to the converged parameters if $\Delta\Delta G_{Ei} \leq \delta(\Delta G_E)$; in words, the difference in free energies between the two parameter sets is smaller than the uncertainty in a standard calculation, in our case the 10 ns (with 1 ns used for equilibration) calculation used for validation.
2. A set of parameters is sufficiently close to the converged parameters if $\Delta\Delta G_{Ei} \leq \delta(\Delta\Delta G_{Ei})$; in words, the difference in free energies is smaller than the uncertainty in the calculation of this difference.

If calculations are being used for a large scale simulation testing, then likely the first criteria is sufficient; it says that any errors due to the parameters will be statistically indistinguishable from statistical errors. However, to be fully converged, and

dx.doi.org/10.1021/ct4005068 | J. Chem. Theory Comput. 2013, 9, 4700−4717

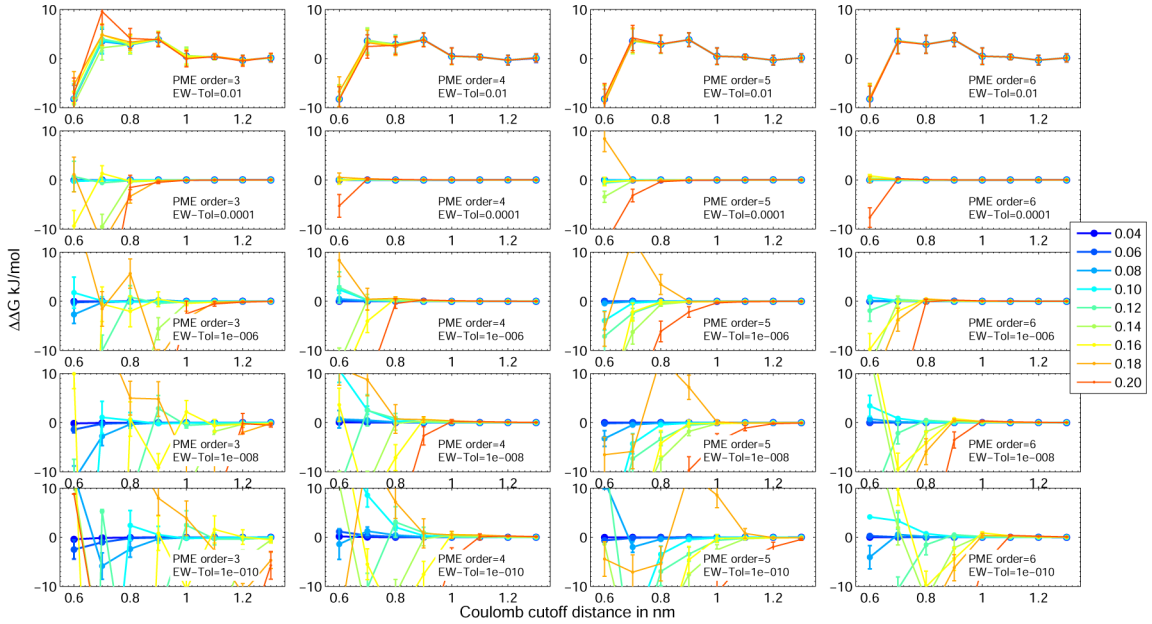


Figure 1. $\Delta\Delta G_{Ei}$ for methane solvation between reference expensive PME parameters and a choice of each of the four PME parameters: Coulomb cutoff (*x*-axis of each graph), Fourier spacing (color), PME order (graphs arranged left to right), and Ewald tolerance (graphs arranged top to bottom). The difference in free energy differences goes to zero as the reference parameters are approached.

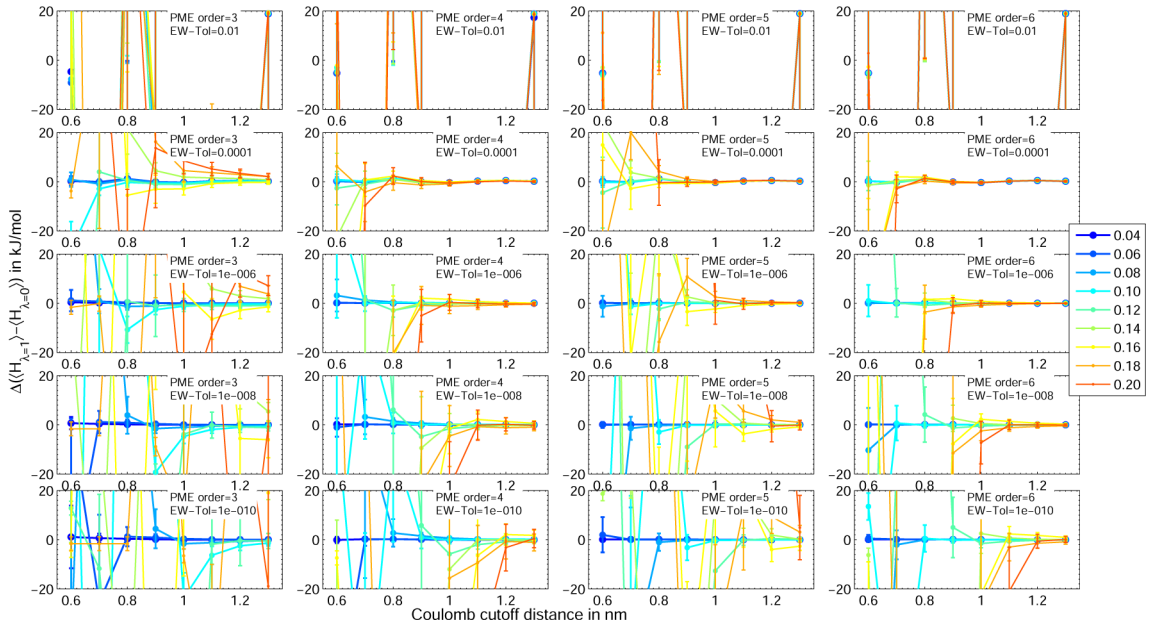


Figure 2. Change in enthalpy of methane solvation ($\langle H\rangle_{\lambda=1} - \langle H\rangle_{\lambda=0}$) between expensive reference PME parameters and a choice of each of the four PME parameters: Coulomb cutoff (*x*-axis of each graph), Fourier spacing (color), PME order (graphs arranged left to right), and Ewald tolerance (graphs arranged top to bottom). Although the differences go to zero in the limit of the expensive reference parameters, because of the statistical noise in the enthalpy of solvation calculation, the level of bias for any given choice of parameters is difficult to identify.

for the highest precision calculations, the second criteria would be required. The second criteria is much more stringent, because using reweighting we are able to calculate differences of differences of free energies with extremely high precision.

For the first and second phase of the search, over the Coulomb parameters, we choose the first criteria. For the third, van der Waals phase, we find that all choices already match the first criteria, and then turn to the second criteria, in some cases requiring us to return and perform additional refinement of the Coulombic calculations in order to reach this more stringent second criteria.

Once we have determined which sets of parameters are sufficiently accurate, we identify the parameters that are the most computationally efficient. In all cases, computational expense for a given set of parameters is reported in ns/h, estimated using 1000 steps of simulation at a single intermediate state run in parallel on 8 cores on the same node. Experimentation showed that 1000 steps of MD is sufficient to eliminate error due to time used in initializing and finalizing simulation runs. Running on different cores may change the trade-off of expense slightly, but examining all

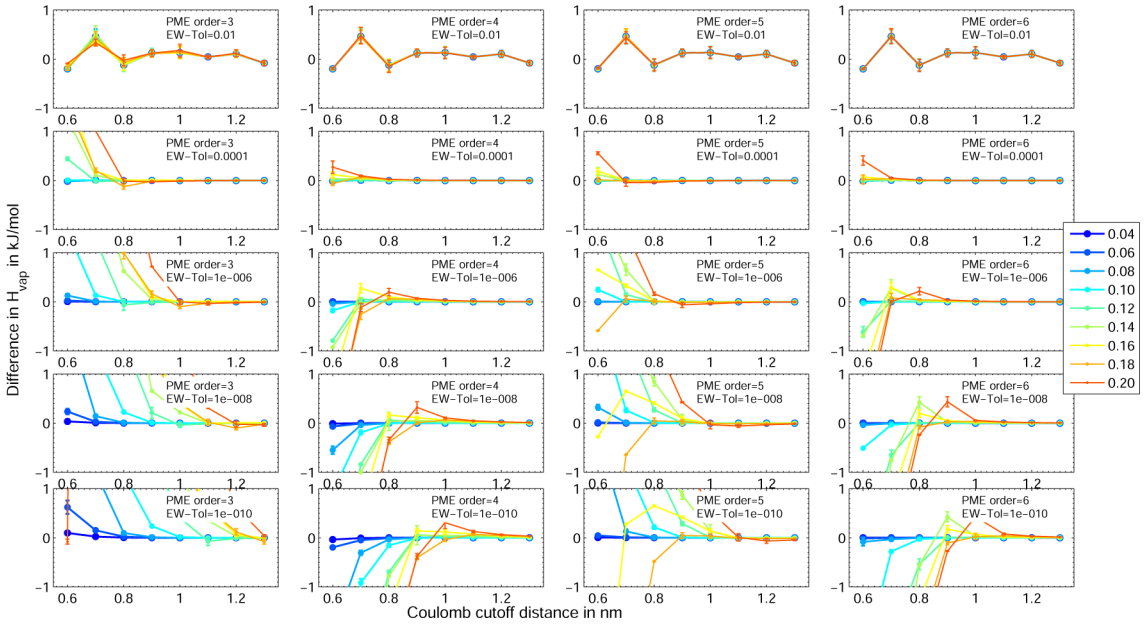


Figure 3. Change in enthalpy of vaporization ΔH_{vap} for TIP3P water between expensive reference PME parameters and a choice of each of the four PME parameters: Coulomb cutoff (x-axis of each graph), Fourier spacing (color), PME order (graphs arranged left to right), and Ewald tolerance (graphs arranged top to bottom) for methane solvation, showing clear convergence to zero for more expensive parameters.

possible combinations of hardware configurations for timings is beyond the scope of this study.

3. RESULTS

3.1. Optimized PME Parameters. The difference in thermodynamic quantities between simulations run with converged PME parameters and the 1440 other parameters examined for methane solvation over the entire multidimensional space is plotted in Figures 1−3. The $\Delta\Delta G$ for dipole inversion and anthracene solvation, which are qualitatively the same, are shown in Figures S-2 and S-4 respectively. Figure 1 shows $\Delta\Delta G$ plotted as a function of four PME parameters for methane solvation. The order of interpolation of B-spline varies by columns from 3 on left to 6 on right. The Ewald tolerance varies by rows from 10^{-2} on the top row to 10^{-10} at the bottom row. At a fixed order of interpolation and Ewald tolerance, $\Delta\Delta G$ is plotted in each subplot as a function of Fourier spacing and Coulomb cutoff.

From Figure 1, $\Delta\Delta G$ indeed approaches zero for converged PME parameters such as high Coulomb cutoff and small Fourier spacing, providing a sanity check on the procedure. Similar plots for the difference in enthalpy of molecular transformation $\Delta\Delta H$, Figure 2, the difference in enthalpy of vaporization of water ΔH_{vap}, Figure 3, show expected convergence for expensive parameters.

Of the three quantities, $\Delta\Delta G$, $\Delta\Delta H$, and ΔH_{vap}, we use only $\Delta\Delta G$ and ΔH_{vap} in our parameter search. The errors $\delta(\Delta\Delta H)$ are larger than the errors $\delta(\Delta\Delta G)$ and $\delta(\Delta H_{vap})$ by more than 2 orders of magnitude, and thus they are not nearly as useful in detecting small changes due to parametrizations. We note that the predictions of $\Delta H_E - \Delta H_O$ are statistically indistinguishable from the validated direct calculations for even the starting benchmark parameters, as can be seen in Tables S-1−S-3 (Supporting Information). However, since the uncertainties in ΔH are of the order 1−100 kJ/mol, and the uncertainties $\delta(\Delta\Delta H)$ themselves are of the order 1−70 kJ/mol, it is impossible to optimize the fine details of the simulation parameters when including this value in the criteria. This difference in the amount of uncertainty for these observables is a consequence of enthalpies and entropies being statistically much harder to compute than free energies of the same processes.

The computational speed in ns/h is plotted as a function of the four PME parameters in Figures S-1, S-3, and S-5 (Supporting Information) for methane solvation, dipole inversion, and anthracene solvation, respectively. Several patterns in the computational time required as a function of parameters can be noted. First, the computation time required is essentially independent of etol. The computational time increases with increasing order of interpolation for large Fourier spacings 0.12−0.2 nm, but smaller Fourier spacings show no change in computational expense with respect to change in order of interpolation. However, with decreasing Fourier spacing and increasing Coulomb cutoff the computational expense increases significantly. Speed of simulation was essentially identical for all three molecules, as it is determined primarily by the system size in number of atoms, which was roughly equal. Since we can easily determine the difference in thermodynamic properties from the converged estimates as well as computational speed as a function of the four PME parameters, we can easily find which set of parameters yield a given error $\Delta\Delta G$, $\Delta\Delta H$, and ΔH_{vap} in the least amount of computational time.

Although the proper choice of parameters depends on the degree of accuracy desired for an application, we propose optimized parameters for which deviations in equilibrium observables satisfy the criteria described in the Methods section in minimum time. Using the first criteria, a parameter set i is statistically indistinguishable from the expensive parameter set if $\Delta\Delta G_{Ei} \leq \delta(\Delta G_E)$. $\delta(\Delta G_E)$ is 0.07 kJ/mol for methane solvation, 0.16 kJ/mol for dipole inversion, and 0.13 kJ/mol for anthracene solvation. $\delta(H_{vap,E})$ is 0.002 kJ/mol. Over the collection of all parameter sets i which have $\Delta\Delta G_{Ei}$ less than corresponding $\delta(\Delta G_E)$, the fastest results can be achieved with

dx.doi.org/10.1021/ct4005068 | J. Chem. Theory Comput. 2013, 9, 4700−4717

order of interpolation 4, relative Ewald tolerance 10^{-4}, Fourier spacing 0.12 nm, and Coulomb cutoff 0.9 nm for all three benchmark systems.

Coulomb cutoffs larger than 0.9 nm or Fourier spacings smaller than 0.12 nm give statistically indistinguishable results from for standard free energy calculations, that is, within 0.05 kJ/mol for methane solvation and within 0.1 kJ/mol for dipole inversion and anthracene solvation. There are no significant gains in accuracy if we increase Coulomb cutoff or decrease Fourier spacing beyond this point. However, the predicted $\Delta\Delta G_{Ei}$ increases by an order of magnitude when rcoul is decreased to 0.8 nm, or for any Fourier spacing larger than 0.12 nm, so we cannot improve the efficiency of the simulation using those parameters without drastically worsening the results.

3.2. Optimized Coulomb Switch. In Figure 4, we plot $\Delta\Delta G_{Ei}$ for methane solvation and ΔH_{vap} for TIP3P water for different Coulomb switching distances. Figure S-6 (Supporting Information) shows $\Delta\Delta G_{Ei}$ for dipole inversion and anthracene solvation for different Coulomb switching distances. ΔH_{vap} for switching starting points beyond 0.85 nm (for a Coulomb cutoff of 0.9 nm) converge to the switchless values and are within one or two standard deviations of each other. The difference $\Delta\Delta G$ is also within one or two standard deviations for switching distances less than 0.1 nm. $\Delta\Delta G$ for methane solvation and dipole inversion are lower than the first error criteria $\delta(\Delta G_E)$ for all three molecules. ΔH_{vap} is also less than or equal to 0.002 kJ/mol for all Coulomb switches starting at any further than 0.84 nm.

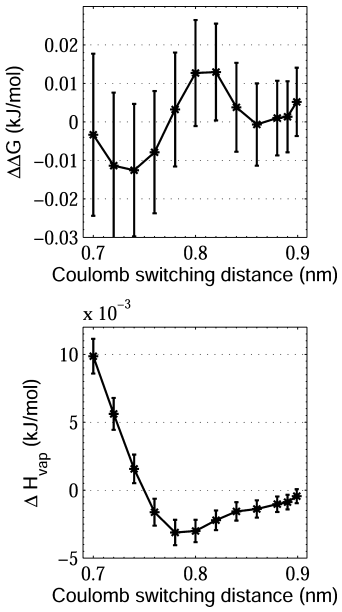


Figure 4. $\Delta\Delta G_{Ei}$ for methane solvation and ΔH_{vap} for TIP3P water between reference expensive PME parameters and a simulation with specified Coulomb switch width, for an optimized Coulomb cutoff of 0.9 nm. Coulomb switch widths of less than 0.1 nm at this cutoff distance and optimized PME parameters are statistically indistinguishable from the reference PME potential. Error bars are one standard deviation.

We should also consider the effect of a switch on the force near the cutoff due to short switching regions. Figure S-7 (Supporting Information) shows how the short-range nonbonded potentials go to zero at cutoff of 0.9 nm with different choice of switches. A very short switch between 0.899 and 0.9 nm clearly will cause the least distortion of the potential energy. However, the force spiking becomes larger, as seen in Figure S-8 (Supporting Information). We need a rational criterion to decide how large a force spike from a single nonbonded interaction is acceptable. For example, the force for passing through the shift region should not be greater than the force experienced due to a thermal collision. The force F experienced by an atom with three degrees of freedom during a collision can be estimated by dividing the energy imparted during a collision at temperature T, which equals $(3/2)k_BT$, where k_B is Boltzmann constant, by a characteristic length, which we choose as the Lennard-Jones radius of oxygen σ_O.

$$F = \frac{3k_BT}{2\sigma_O} \approx 12 \text{ kJ/mol·nm} \tag{22}$$

We note that the force spiking calculated here is due to a single nonbonded interaction pair, and that particles will experience forces in all directions. If the solvent is homogeneous, then these radial forces will cancel out on average, so this 12 kJ/mol·nm value might be an overestimate, but serves as a useful baseline motivated by physical criteria.

We see in Figure S-8(a) (Supporting Information) that the force spike for all examined Coulomb switches are below 12 kJ/mol·nm; indeed, for any switch width larger than 0.01 nm, the magnitude of any force spike is an order of magnitude below this level. The $\Delta\Delta G$ versus Coulomb switch plots for dipole inversion and anthracene solvation (Figure S-6 (Supporting Information)) also show that the $\Delta\Delta G$ for switches between 0.8 and 0.89 are within one to two standard deviations of each other. We therefore choose our optimized Coulomb switch as 0.85 nm, midway between 0.8 and 0.89 nm to balance the effect of force spikes and loss of electrostatic energy, though any switch in that range is statistically identical in thermodynamic properties to free energy calculations carried out without a switch.

3.3. Optimized LJ Cutoff and Switch. In the last phase of the search, we examine the effect of varying the LJ cutoff and the LJ switching function using our optimized Coulomb parameters to calculate $\Delta\Delta G$ for all three transformations in the benchmark set and ΔH_{vap} for TIP3P water. As noted, the initial benchmark parameter choices already satisfy the first criteria. We thus see if we can now obtain the second more stringent criteria, pushing the difference due to deviations in the parameters from converged results below the uncertainty in this calculation of $\Delta\Delta G_{Ei}$ itself, which is 0.01 kJ/mol for methane solvation, 0.02 kJ/mol for dipole inversion, 0.015 kJ/mol for anthracene solvation, and 0.001 kJ/mol for the enthalpy of vaporization (from Tables 3−6). Figure 5 shows $\Delta\Delta G$ for methane solvation, dipole inversion and anthracene solvation as a function of different LJ cutoffs and LJ switches. Figure 6 shows ΔH_{vap} for TIP3P water as a function of different LJ cutoffs and LJ switches.

In all cases, we see that $\Delta\Delta G$ and ΔH_{vap} approaches zero as the LJ cutoff is increased from 0.8 nm. Methane solvation and dipole inversion approach the stricter condition of indistinguishability of parameters at a LJ cutoff of 0.9 nm and switch starting from 0.84 nm, when $\Delta\Delta G$ is equal to or less than 0.01 kJ/mol. $\Delta\Delta G$ for anthracene solvation does not quite reach this level of accuracy until reaching a LJ cutoff of 1.0 nm; at 0.9 nm cutoff $\Delta\Delta G$ anthracene solvation is closer to 0.03 kJ/mol. The predicted $\Delta\Delta G$ for methane solvation and dipole inversion are nearly equal to the error and for anthracene solvation $\Delta\Delta G$ is

dx.doi.org/10.1021/ct4005068 | J. Chem. Theory Comput. 2013, 9, 4700−4717

Table 3. Predictions and Validation $\Delta\Delta G_{EO}$ Results for Methane Solvation Match within One and Two Standard Deviations[a]

	prediction	validation using: one parameter set	validation using: two parameter sets simultaneously		
	B	B or E or O	B and E	B and O	E and O
		ΔG (kJ/mol) for Methane Solvation			
benchmark (B)	8.619 ± 0.069	8.619 ± 0.069	8.701 ± 0.049	8.668 ± 0.049	N/A
expensive (E)	8.687 ± 0.071	8.868 ± 0.069	8.811 ± 0.049	N/A	8.851 ± 0.049
optimized (O)	8.675 ± 0.072	8.863 ± 0.069	N/A	8.764 ± 0.049	8.857 ± 0.049
		$\Delta\Delta G$ (kJ/mol) for Methane Solvation			
$\Delta G_B - \Delta G_E$	−0.068 ± 0.019	−0.249 ± 0.098	−0.109 ± 0.012	N/A	N/A
$\Delta G_B - \Delta G_O$	−0.057 ± 0.022	−0.244 ± 0.098	N/A	−0.096 ± 0.013	N/A
$\Delta G_E - \Delta G_O$	0.011 ± 0.011	0.005 ± 0.098	N/A	N/A	−0.006 ± 0.006

[a]In column two ΔG and $\Delta\Delta G$ are estimated using samples from benchmark parameter set and re-evaluations done at expensive and optimized parameter sets. In column three we use samples from only one parameter set: either only the benchmark parameter set (B) or only the expensive (E) or only the optimized (O) parameter set. For columns 4−6 we use samples from a pair of parameter sets. Re-evaluation is not done for the remaining parameter sets, so we have no estimate for the unsampled parameter set (labeled as N/A).

Table 4. Predictions and Validation ΔH_{vap} Results Match for All the Three Parameter Sets within Two Standard Deviations

	prediction	validation using: one parameter set	validation using: two parameter sets simultaneously		
	B	B or E or O	B and E	B and O	E and O
		H_{vap} kJ/mol for TIP3P Water			
benchmark (B)	42.402 ± 0.002	42.402 ± 0.002	42.398 ± 0.001	42.401 ± 0.001	N/A
expensive (E)	42.408 ± 0.002	42.404 ± 0.002	42.403 ± 0.001	N/A	42.406 ± 0.001
optimized (O)	42.410 ± 0.002	42.412 ± 0.002	N/A	42.409 ± 0.001	42.407 ± 0.001
		ΔH_{vap} kJ/mol for TIP3P Water			
$H_{vap\ B} - H_{vap\ E}$	−0.006 ± 0.001	−0.002 ± 0.003	−0.005 ± 0.001	N/A	N/A
$H_{vap\ B} - H_{vap\ O}$	−0.008 ± 0.001	−0.010 ± 0.003	N/A	−0.008 ± 0.001	N/A
$H_{vap\ E} - H_{vap\ O}$	−0.002 ± 0.001	−0.008 ± 0.003	N/A	N/A	−0.001 ± 0.000

Table 5. Predictions and Validation $\Delta\Delta G_{EO}$ Results for Discharging, Charging, and Complete Dipole Inversion Also Match within One and Two Standard Deviations

	prediction	validation using: one parameter set	validation using: two parameter sets simultaneously		
	B	B or E or O	B and E	B and O	E and O
		ΔG (kJ/mol) for Complete Dipole Inversion +e/−e to −e/+e			
benchmark (B)	−0.079 ± 0.157	−0.079 ± 0.157	−0.019 ± 0.111	0.003 ± 0.112	N/A
expensive (E)	−0.120 ± 0.160	−0.056 ± 0.158	−0.021 ± 0.112	N/A	0.030 ± 0.111
optimized (O)	−0.092 ± 0.162	0.088 ± 0.157	N/A	0.008 ± 0.112	0.021 ± 0.111
		$\Delta\Delta G$ (kJ/mol) for Complete Dipole Inversion +e/−e to −e/+e			
$\Delta G_B - \Delta G_E$	0.040 ± 0.035	−0.023 ± 0.223	0.002 ± 0.021	N/A	N/A
$\Delta G_B - \Delta G_O$	0.013 ± 0.041	−0.167 ± 0.222	N/A	−0.005 ± 0.023	N/A
$\Delta G_E - \Delta G_O$	−0.028 ± 0.021	−0.144 ± 0.223	N/A	N/A	0.009 ± 0.010
		ΔG (kJ/mol) for First Half of Dipole Inversion Transformation, Discharging the Dipole +e/−e to 0e/0e			
benchmark (B)	93.241 ± 0.108	93.241 ± 0.108	93.259 ± 0.077	93.282 ± 0.078	N/A
expensive (E)	93.400 ± 0.112	93.442 ± 0.109	93.433 ± 0.077	N/A	93.483 ± 0.077
optimized (O)	93.352 ± 0.114	93.464 ± 0.110	N/A	93.393 ± 0.078	93.418 ± 0.077
		$\Delta\Delta G$ (kJ/mol) for First Half of Dipole Inversion Transformation, Discharging the Dipole +e/−e to 0e/0e			
$\Delta G_B - \Delta G_E$	−0.159 ± 0.030	−0.201 ± 0.153	−0.175 ± 0.018	N/A	N/A
$\Delta G_B - \Delta G_O$	−0.110 ± 0.035	−0.223 ± 0.154	N/A	−0.112 ± 0.020	N/A
$\Delta G_E - \Delta G_O$	0.048 ± 0.018	−0.022 ± 0.155	N/A	N/A	0.066 ± 0.009
		ΔG (kJ/mol) for Second Half of Dipole Inversion Transformation, Charging the Dipole 0e/0e to −e/+e			
benchmark (B)	−93.320 ± 0.109	−93.320 ± 0.109	−93.277 ± 0.078	−93.278 ± 0.077	N/A
expensive (E)	−93.520 ± 0.113	−93.498 ± 0.110	−93.454 ± 0.078	N/A	−93.453 ± 0.077
optimized (O)	−93.444 ± 0.114	−93.375 ± 0.109	N/A	−93.385 ± 0.077	−93.397 ± 0.077
		$\Delta\Delta G$ for Second Half of Dipole Inversion Transformation, Charging the Dipole 0e/0e to −e/+e			
$\Delta G_B - \Delta G_E$	0.199 ± 0.031	0.178 ± 0.155	0.177 ± 0.018	N/A	N/A
$\Delta G_B - \Delta G_O$	0.123 ± 0.036	0.055 ± 0.154	N/A	0.106 ± 0.020	N/A
$\Delta G_E - \Delta G_O$	−0.076 ± 0.018	−0.123 ± 0.155	N/A	N/A	−0.057 ± 0.009

dx.doi.org/10.1021/ct4005068 | J. Chem. Theory Comput. 2013, 9, 4700−4717

Table 6. Predictions and Validation Results for Anthracene Solvation Match within One and Two Standard Deviation for $\Delta\Delta G_{EO}$[a]

	prediction	validation using: one parameter set	validation using: two parameter sets simultaneously		
	B	B or E or O	B and E	B and O	E and O
		ΔG (kJ/mol) for Anthracene Solvation			
benchmark (B)	−9.624 ± 0.121	−9.624 ± 0.121	−9.853 ± 0.086	−9.823 ± 0.086	N/A
expensive (E)	−9.023 ± 0.125	−9.261 ± 0.123	−9.267 ± 0.087	N/A	−9.267 ± 0.086
optimized (O)	−8.996 ± 0.125	−9.205 ± 0.121	N/A	−9.214 ± 0.086	−9.237 ± 0.086
		$\Delta\Delta G$ (kJ/mol) for Anthracene Solvation			
$\Delta G_B - \Delta G_E$	−0.602 ± 0.029	−0.363 ± 0.173	−0.585 ± 0.017	N/A	N/A
$\Delta G_B - \Delta G_O$	−0.628 ± 0.031	−0.419 ± 0.171	N/A	−0.609 ± 0.018	N/A
$\Delta G_E - \Delta G_O$	−0.027 ± 0.014	−0.056 ± 0.173	N/A	N/A	−0.030 ± 0.007

[a]The predictions that ΔG for expensive and optimized parameter sets are greater than ΔG for benchmark parameter set is correct.

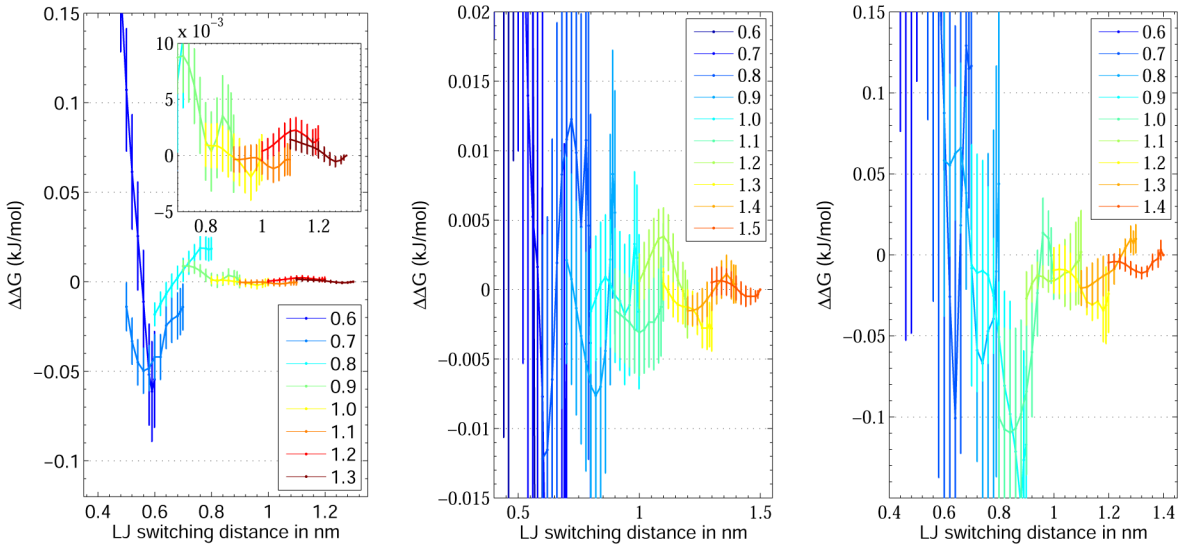


Figure 5. $\Delta\Delta G_{solv}$ for methane solvation (left), dipole inversion (center) and anthracene solvation (right) as a function of different Lennard-Jones switching distances for Lennard-Jones cutoffs in increments of 0.1 nm from 0.6 to the maximum allowed by the simulation box. Each color represents a different cutoff distance, and each line within the color represents the value at a different switching distance, with statistical uncertainty indicated by the vertical bar. Note differences in *y*-axis energy scale, with the dipole inversion scale being approximately 10× smaller than the other two. The inset for methane solvation demonstrates how $\Delta\Delta G$ converges to essentially zero at the largest Lennard-Jones cutoff of 1.3 nm, and below 0.007 kJ/mol for cutoffs of 0.9 nm or greater for any switch width.

within twice the error $\delta(\Delta\Delta G_{Ei})$ if we use the optimized parameter set.

For most cases, an error of 0.03 kJ/mol in a few larger molecules, and thus, a 0.9 nm cutoff might be sufficient. However, we can perform a limited search in the parameter space around this initially optimized parameter set to see if even higher accuracies can be achieved. We discuss this final optimization in the validation section.

$\Delta\Delta G$ and ΔH_{vap} for entire range of LJ switches are all within one and two standard deviations of the unswitched values. In Figure S-8(b) (Supporting Information), we see that a long LJ switch, 0.7 to 0.8 nm, distorts the potential and a relatively small switch like 0.899 nm will result in large force spikes at the cutoff. Unlike the Coulomb switch, the force spike due to the LJ switch is greater than the 12 kJ/mol·nm mark for switching regions narrower than 0.04 nm, as the magnitude of the potential is larger in the switching region for the Lennard-Jones potential. We therefore select 0.05 nm for the switching width. We note particularly that the free energy and enthalpy results are essentially statistically independent for any value switch distance between 0.1 and 0.001 nm for appropriate cutoff values.

3.4. Validation of the Optimization Procedure. To validate our search process, we must actually perform the simulations at the new parameter values to see if the free energy and enthalpy differences between the optimized parameters and the fully converged reference parameters were correctly predicted. We therefore compare results from our optimized parameter set with results from simulations performed at expensive, converged Coulomb and LJ parameters. We directly validate all stages of optimization in a single comparison, as we only need to validate the individual stages if the final results are incorrect. Validation simulations were performed with GROMACS 4.6.1, which gave identical energies to the bug-fixed GROMACS 4.5.3 but was better optimized. All versions were compiled in double precision. Simulations were run using velocity Verlet with a time step of 2 fs and SETTLE constraints for water,[48] with SHAKE/RATTLE[49] with relative tolerance 10^{-12} used to constrain nonwater molecules. Temperature and pressure control were obtained using the algorithm of Martyna et al.,[50] with τ_t = 5.0 ps at 300 K, and τ_p = 5.0 ps at 1 atm.

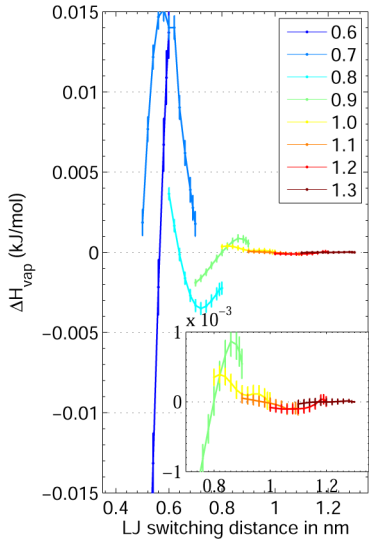


Figure 6. ΔH_{vap} for TIP3P water as a function of different Lennard-Jones switching distances for Lennard-Jones cutoffs 0.6−1.3 nm. ΔH_{vap} at and beyond 0.9 nm Lennard-Jones cutoff is less than 0.005 kJ/mol for any switch distance.

The optimized and reference parameters are summarized in the Table 2. We find that to perform the actual simulations at the expensive parameters, rather than just reevaluating the energies of the configurations, we needed to make slight modifications. When the LJ and Coulomb cutoffs used are very near to the twice the box length, the estimated free energies are inconsistent, especially for anthracene solvation. This appears to be an artifact of the improper potential energy contribution of periodic images at the box boundaries under box size fluctuations. We therefore chose the "expensive" nonbonded cutoffs to be 0.1 nm less than the most expensive cutoffs, which was determined to have a negligible effect on potential energies of representative configurations. Similarly, a Fourier spacing of 0.06 nm is used instead of 0.04 nm to accelerate the otherwise very slow reference parameter set simulations. The predicted $\Delta\Delta G$ between expensive parameter set used in validation simulations and most expensive parameter set use in reweighting is $4 \times 10^{-5} \pm 5 \times 10^{-5}$ kJ/mol, and thus, this approximation is acceptable to any reasonable limit.

There are two types of computations we can perform to validate these predictions. The simplest is to use the samples from a simulation corresponding to just one parameter set in our calculations of differences in thermodynamic variables, and compute properties from this simulation alone. In this case, we use sampled energies from simulations performed using the optimized parameter set, a $[K \times K \times N_k]$ matrix of energy differences $\mathbf{U}_O(\mathbf{X}_O)$ to obtain free energy estimates for the optimized parameter set ΔG_O. Similarly, we can use samples from simulation performed using the expensive parameter set, the $[K \times K \times N_k]$ energies $\mathbf{U}_E(\mathbf{X}_E)$, to obtain free energy estimates for the expensive parameter set ΔG_E, and then compute the difference $\Delta\Delta G_{EO} = \Delta G_E - \Delta G_O$. We can easily get the differences in the free energy estimates $\Delta\Delta G$ for other pairs of parameter sets in the same way. The error estimates of $\Delta\Delta G$ will be the square root of the sum of square of errors in individual ΔG being compared, since the two ΔG belong to completely different simulations whose data is uncorrelated. The third column of Tables 3−6 show $\Delta\Delta G$ calculated using samples corresponding to just one parameter set.

However, it is clear that this standard process for calculating differences in average quantities does not allow us to differentiate very well between the results generated with similar sets of parameters, as the statistical noise is simply too high. We would generally have to run tens of times longer to sufficiently converge our statistics. The much more powerful and statistically efficient way to calculate $\Delta\Delta G$ is by using samples from *both* parameter sets being compared to compute the differences in observables using multistate reweighting.

To compare the expensive and optimized parameter sets, we consider the $[2K \times 2K \times N_k]$ matrix of energy differences $\mathbf{U} = [\mathbf{U}_E(\mathbf{X}_E),\mathbf{U}_E(\mathbf{X}_O);\mathbf{U}_O(\mathbf{X}_E),\mathbf{U}_O(\mathbf{X}_O)]$, and use multistate reweighting via MBAR. Now, using data generated using both parameter sets $\mathbf{U}_E(\mathbf{X}_E)$ and $\mathbf{U}_O(\mathbf{X}_O)$, including reevaluation of samples from each parameter set at the other parameter set, $\mathbf{U}_E(\mathbf{X}_O)$ and $\mathbf{U}_O(\mathbf{X}_E)$ we can compute $(\Delta G_E - \Delta G_O) \pm \delta(\Delta G_E - \Delta G_O)$, and compare to the estimate of $(\Delta G_E - \Delta G_O) \pm \delta(\Delta G_E - \Delta G_O)$ using data from only the benchmark state obtained during the optimization process. The reevaluation required for this process takes negligible simulation time, since we reprocess only uncorrelated samples. We repeat the same process for the other pairs of simulation data and show the results in Tables 3−6.

In the course of this study, we found that the simulations in the original benchmark study[14] used a value of rlist that truncated the total potential, though it did not affect the findings of that study which examined variances. The benchmark free energy estimates for methane solvation, dipole inversion, and anthracene solvation differ from the results of our previous study because in this study the benchmark parameter set has a different rlist. To be consistent with high precision free energy calculations, we started our search process by performing new calculations with the benchmark study parameters, except using an rlist 0.2 nm longer than the longest cutoff. Because free energies are unaffected by any rlist greater than this value, we did not include this parameter in our optimization routine. Longer rlists have some effect on the total time in GROMACS, and the obvious choice is thus the shortest value of rlist that leaves the potential energies and thus free energies unchanged.

We also found that for anthracene, equipartitioning of kinetic energies between solvent and ligand was broken at low coupling of the ligand to the bulk solvent if a global thermostat was used, disrupting both the energy distribution and volume distribution, as validated by the *checkensemble* tool.[28] This deviation changed the free energy differences by approximately −3 kJ/mol. To obtain proper equipartitioning we used separate thermostats for the solute and the solvent for anthracene solvation simulations only. No error appeared in the methane solvation case, perhaps because of the much smaller size of the UA methane. Such a problem did not occur in the case of the dipole because all states remain coupled to the solvent.

For the validation simulations, to match the amount of sampling in this original benchmark study, we performed 10 ns simulations (9 ns postequilibration) at each of the K intermediate states for the expensive and the optimized parameter sets. Each simulation yielded 9000 samples, taking samples every 1 ps. We further subsampled the data using the calculated autocorrelation times of $dH/d\lambda$.

We can see that we get remarkably high agreement between predictions of ΔG_{EO} using samples only from the benchmark parameter set and results from simulations actually run with optimized and expensive parameter sets. The comparisons to

dx.doi.org/10.1021/ct4005068 | J. Chem. Theory Comput. 2013, 9, 4700−4717

Table 7. Predicted $\Delta\Delta G_{Ei}$ for GROMACS Default Parameters, In the Bottom Right Corner, Give Marginally Better Convergence Compared to the First Optimized Guess for a Slight (10−15%) Performance Cost[a]

		$\Delta\Delta G$ (kJ/mol)		
cutoff	system	10^{-4}	10^{-5}	$\delta(\Delta G_E)$ (kJ/mol)
		high accuracy		
0.9	methane solvation	0.011 ± 0.011	0.027 ± 0.006	0.071
	full dipole inversion	−0.028 ± 0.021	−0.007 ± 0.012	0.160
	discharging	0.048 ± 0.018	0.117 ± 0.010	0.112
	charging	−0.076 ± 0.018	−0.124 ± 0.010	0.113
	anthracene solvation	−0.027 ± 0.014	0.009 ± 0.07	0.125
	H_{vap}	−0.002 ± 0.001	0.004 ± 0.000	0.002
			very high accuracy	
1.0	methane solvation	0.013 ± 0.011	0.004 ± 0.003	0.071
	full dipole inversion	0.005 ± 0.020	−0.001 ± 0.006	0.160
	discharging	0.060 ± 0.018	0.069 ± 0.006	0.112
	charging	−0.054 ± 0.018	−0.070 ± 0.006	0.113
	anthracene solvation	−0.009 ± 0.014	−0.002 ± 0.004	0.125
	H_{vap}	0.002 ± 0.001	0.002 ± 0.000	0.002

[a]Both satisfy the first criteria of statistical indistinguishability compared to ΔG_E calculations (one standard deviation error shown in column 5).

results using samples from two parameter sets at a time are shown in columns 4 (benchmark, expensive), 5 (benchmark, optimized) and 6 (expensive, optimized) in Tables 3−6. For methane solvation in Table 3, dipole inversion in Table 5, and anthracene solvation in Table 6, the predicted $\Delta G_E - \Delta G_O$ in column two agree within one to two standard deviations with the estimates calculating with samples generated using expensive and optimized parameter sets in column 6. The statistical convergence is also good, as $\delta(\Delta G_E - \Delta G_O)$ is very low i.e. approximately 0.01 kJ/mol, far closer than would ever be required for molecular simulations, as experimental results for condensed phase data are not that accurate.

Interestingly, even though the sampling at the benchmark state is quite not good enough to predict free energies at these particular alternate (i.e., expensive and optimized) states, it is remarkably able enough to predict *differences* in free energies between these alternate states with dramatically low bias and uncertainty. For methane solvation in Table 3 the predicted ΔG_E and ΔG_O by reweighting samples from the benchmark parameter set are 8.687 ± 0.071 kJ/mol and 8.675 ± 0.072 kJ/mol, respectively. ΔG_E evaluated using samples generated using the expensive parameter set alone is 8.868 ± 0.069 kJ/mol, and ΔG_O evaluated using samples generated using the optimized parameter set alone is 8.863 ± 0.069 kJ/mol. The difference between the predicted and actual ΔG_E is 0.181 ± 0.099 kJ/mol, and the difference between the predicted and actual ΔG_O is 0.188 ± 0.099 kJ/mol, at the edge of statistical error. However, the predicted $\Delta G_E - \Delta G_O$, 0.011 ± 0.011 kJ/mol, is within a standard deviation of $\Delta G_E - \Delta G_O$ estimated using the direct difference of estimates from the expensive and optimized sets, 0.005 ± 0.098 kJ/mol, though the statistical error is much larger in the second case. When $\Delta G_E - \Delta G_O$ when samples from expensive and optimized sets are used simultaneously, the difference is only −0.006 ± 0.006 kJ/mol, within uncertainty of the predicted value 0.011 ± 0.011 kJ/mol. Importantly, using data from *both* parameter sets allows us to push the uncertainty in differences of ΔG values down an order of magnitude.

Similarly for dipole inversion and anthracene solvation, the predicted $\Delta G_B - \Delta G_E$ and $\Delta G_B - \Delta G_O$ in column 2, evaluated using only samples generated with the benchmark data, agree with the free energy differences evaluated with the data from the pairs of simulation in columns 4, 5, and 6 within error. The difference $\Delta G_E - \Delta G_O$ computed with samples from both the expensive and the optimized simulations agree extremely accurately with the predictions made with samples used from only the benchmark simulations.

We next turn to the agreement of the enthalpy of vaporization. We obtain highly accurate estimates obtained for ΔH_{vap} between the expensive and optimized parameter values, with agreement down to 0.0002 kJ/mol (Table 4). The fact that the magnitude of $\Delta H_{vap,EB}$ is one to 2 orders of magnitude smaller than $\Delta\Delta G_{EB}$ indicates that small changes in simulation parameters may influence small molecule free energies more than bulk properties, raising caution about the applicability of parameters optimized for bulk properties to free energy calculations. Here again, reweighting is much more precise than calculating direct differences.

Because the full dipole inversion process could lead to cancellation of errors, we also examine the effect of simulation parameters on the free energy of charging and discharging separately. Therefore, for dipole inversion, we have also reported ΔG and $\Delta\Delta G$ for both halves of the transformation. We notice that $\Delta\Delta G_{EO}$ for the first and the second half of the transformation using data from both the expensive and optimized parameter sets, 0.066 ± 0.009 kJ/mol and −0.057 ± 0.009 kJ/mol, are not zero but the errors do cancel out for the complete dipole inversion. Even in this case, $\Delta\Delta G_{EO}$ for both the first and the second halves satisfy the first criteria of indistinguishability, that is, $\Delta\Delta G_{Ei} \leq \delta(\Delta G_E)$. The $\Delta\Delta G_{EO}$ for the first and the second halves of the dipole inversion do not satisfy the second criteria even at long cutoffs. This could be due to some small artifact of periodic boundary conditions with a total permanent dipole of the system, as this small but statistically meaningful difference does not taper off even for the case of the putatively converged cutoff.

The optimized parameter set also results in good convergence with respect to expensive parameter set. $\Delta\Delta G_{EO}$ for methane solvation and dipole inversion test systems are essentially zero, showing that the optimal parameter set is statistically indistinguishable from the expensive parameter set. For methane solvation $\Delta\Delta G_{EO}$ = −0.006 ± 0.006 kJ/mol and for dipole inversion $\Delta\Delta G_{EO}$ = 0.009 ± 0.010 kJ/mol. However for anthracene solvation the predicted value of $\Delta\Delta G_{EO}$ = −0.027 ± 0.014 kJ/mol. The validated value (in column 6 of

dx.doi.org/10.1021/ct4005068 | J. Chem. Theory Comput. 2013, 9, 4700−4717

Table 6) of $\Delta\Delta G_{EO} = -0.030 \pm 0.007$ kJ/mol using the optimized LJ cutoff of 0.9 nm, which is statistically significant. For most purposes, this small deviation will likely not be relevant, but for complete consistency, slight improvements in these parameters might be necessary.

In our search for optimized PME parameters we used relatively coarse grid spacings for Ewald tolerance, Fourier spacing, and the Coulomb cutoff, which already allowed us to come very close to fully converged results. We found that the increasing Coulomb cutoff and decreasing Fourier spacing gave better convergence at the cost of higher computational expense. However $\Delta\Delta G_{Ei}$ does not change monotonically with Ewald tolerance. For a given Coulomb cutoff and Fourier spacing $\Delta\Delta G_{Ei}$ decreased as Ewald tolerance was increased from 10^{-10} to 10^{-4} and then $\Delta\Delta G_{Ei}$ increased when Ewald tolerance was increased to 10^{-2}, without affecting computational expense. Thus there exists a possibility that slightly improved results could be obtained with etol = 10^{-5} which was not explored in the initial scan.

Perhaps not coincidentally, the GROMACS default parameters determined by the much simpler approach of converging force calculations are Ewald tolerance = 10^{-5}, Coulomb cutoff = LJ cutoff = 1.0 nm, PME order = 4, Fourier spacing = 0.12 nm. These values are close but not exactly the same as our optimized parameters of Ewald tolerance =10^{-4}, Coulomb cutoff = LJ cutoff =0.9 nm (with PME order and Fourier spacing the same). We therefore searched the PME parameter space around the our first optimized parameter set, including the GROMACS default parameters among others, to see if we could refine the current optimized parameter set and push the error further down for the test systems. We therefore also tested Ewald tolerance = [10^{-4},10^{-5}] and LJ cutoff = Coulomb cutoff = [0.9,1.0], with PME order =4, Fourier spacing =0.12 nm, and switches both of width 0.05 nm.

The results are summarized in the Table 7. We found that if we use an Ewald tolerance of 10^{-5}, rvdw = rcoul = 0.9 nm, rvdw_switch and rcoul_switch = 0.85, predicted $\Delta\Delta G_{Ei}$ for anthracene solvation is reduced from -0.027 ± 0.014 kJ/mol to 0.009 ± 0.007 kJ/mol, predicted $\Delta\Delta G_{Ei}$ for dipole inversion changes from -0.028 ± 0.021 to -0.007 ± 0.012 kJ/mol, $\Delta\Delta G_{Ei}$ for methane solvation changes from 0.011 ± 0.011 to 0.027 ± 0.006 kJ/mol. For all systems $\delta(\Delta\Delta G_{Ei})$ improves by a factor of 2 indicating increased phase space overlap. There was no statistically significant improvement in $\Delta\Delta G_{Ei}$ for methane solvation, full dipole inversion, and anthracene solvation. However, $\Delta\Delta G_{Ei}$ increased with reduced etol of 10^{-5} for a cutoff of 0.9 nm, particularly the half-charging calculations, which is surprising, since the electrostatic treatment is presumably getting better. $\Delta\Delta G_{Ei}$ for methane solvation, dipole inversion and anthracene solvation are still less than corresponding $\delta(\Delta G_E)$ and hence follow the first criteria of indistinguishability. The decrease in $\delta(\Delta\Delta G_{Ei})$ seems to indicate that the parameter set with etol = 10^{-5} has a better phase space overlap with the expensive parameter set, although the accuracy in the overall energy may be a bit worse. If we also increased LJ and Coulombic cutoffs to 1.0 nm, with etol remaining at 10^{-5}, then as seen in the bottom right corner of the Table 7, we obtain even lower error and higher accuracy than the original optimized parameter set for all test systems.

The initial optimized parameter set remains approximately the same speed as the benchmark parameter set (perhaps 3−10% slower) for cutoffs of 0.9 nm, but the accuracy is improved for all molecules. For a slightly higher cost, 10−15% in computational speed depending on the system, we can get slightly more accurate anthracene solvation free energy estimate compared to using a benchmark parameter set.

Our reweighting method, using only samples from the initial free energy calculation, allows us to determine optimized parameters that pass stringent validation tests for methane solvation, dipole inversion and anthracene solvation and predicts the direction of the optimized parameter set correctly for all three test cases. The difference in thermodynamic estimates between the converged and the optimized parameters is of the order of 0.01 kJ/mol for methane solvation, dipole inversion, and anthracene solvation, primarily due to improved choices of PME parameters.

For estimating thermodynamic observables with high accuracy, we suggest the simulations be run with B-spline interpolation order 4, Ewald tolerance of 10^{-4}, Fourier spacing of 0.12 nm, Coulomb and LJ cutoff of 0.9 nm, and Coulomb and LJ switch between 0.85 nm. For even higher accuracies and precision with 10−15% reduced speed, Ewald tolerance of 10^{-5}, Coulomb and LJ cutoff of 1.0 nm, and Coulomb and LJ switch of 0.95 could be used.

We note that we get the same optimized parameter set for three very different test systems. The molecules were chosen to represent the extremes of small molecule free energy calculations typically attempted, with a large (7D) dipole change and 14 heavy atoms being removed, and thus the results should be transferable across a wide range of free energy calculations. However, there are some limitations to the parameters proposed. We did not examine charging calculations because of difficulties in identifying the proper long-range reference when the total charge of the system changes. However, the magnitude of the dipole disappearance (∼90 kJ/mol) approaches that of a single ion disappearing (∼250 kJ/mol), suggesting that magnitude of such free energy changes are supported by these parameters.

Additionally, for nonhomogeneous simulations, such as in the case of ligand binding or membrane simulations, it is clear that short Lennard-Jones cutoffs are not sufficiently accurate.[27] Binding free energies differ by 0.8−1.2 kcal/mol when evaluated with shorter (0.8 nm) vs very long (2.5 nm) cutoffs. However, short cutoffs with some reweighted information from longer cutoffs (only one configuration every 200 ps) can give cutoff-independent results.[27] This suggests that at least for ligand binding, a twin range scheme that only occasionally uses much longer cutoffs may be required. Ewald schemes for dispersion interactions are also possible.[39]

Larger box sizes would not result in loss of accuracy of the thermodynamic observables with the parameters described. However, computational speed would drop because of the increased number of interaction sites, which would perhaps require an increase of real space cutoffs to balance an increasing Fourier space cost, requiring additional optimization. In cases where researchers suspect these parameters may not be transferable, then the detailed presentation of methods in this paper for validating parameters should allow those situations to be tested.

4. CONCLUSIONS

In this paper, we describe the use of multistate reweighting, using the MBAR formalism, to estimate thermodynamic observables for a large range of unsampled states using the configurations from only a few sampled states. We demonstrate the utility of this procedure by performing a multidimensional

search in the space of nonbonded interaction simulation parameters to identify an optimized set of simulation parameters. These optimized parameters are chosen to give statistically indistinguishable results to those generated with fully converged nonbonded parameters, but with lower computation time. Careful ordering of the parameter search reduced the final number of parameter sets to be compared to approximately 5200. For each of these parameter sets, we reevaluate the energies of 9000 uncorrelated samples from each simulation performed for the different alchemical states in the original free energy calculations, a much faster procedure than performing full simulations. Specifically, reevaluating every 500th step takes approximately 0.4% of the simulation time of an equivalent MD simulation. Table 8 compares actual MD

Table 8. Re-evaluation is More than Two Orders of Magnitude Faster than Generating Fresh Samples Using MD[a]

system	parameter set	MD (h/ns)	re-evaluation (h/ns)
methane solvation	benchmark	3.620	0.013
	expensive	9.212	0.028
	optimized	3.721	0.013
dipole inversion	benchmark	6.935	0.026
	expensive	22.292	0.073
	optimized	7.582	0.026
anthracene solvation	benchmark	4.629	0.022
	expensive	13.268	0.055
	optimized	5.123	0.024

[a]The initial optimized parameters are used.

sample generation rate versus the rate at which re-evaluations can be done. Since the GROMACS version 4.5.3 did not support parallel rerun for re-evaluation for velocity Verlet this comparison was performed using a single processor for both molecular dynamics and trajectory reruns.

Performing 10 ns of simulation for approximately 60 000 thermodynamic states (5200 parameter sets × the number of alchemical states) with 3 million observables estimated in the optimization process would have taken over 60 years of CPU time. In this study, the calculations took less than a month of CPU time, resulting in a time savings of almost 3 orders of magnitude.

Estimates of $\Delta\Delta G_{EO}$ calculated using only samples from the initial benchmark states agree with the estimates calculated directly using samples from the optimized and expensive parameter set within statistical error. Remarkably, this agreement is usually of the order of 0.01 kJ/mol, with statistical error equally low both for predictions and validations. Both the level of agreement of predictions with the validation calculations and the high precision of all results are quite surprising and demonstrate the power of this multistate reweighting approach. The final computational cost of the optimized parameters is approximately the same as was used in previous benchmarking studies, but with substantial improvements in the accuracy of free energy calculations with respect to the previous benchmark calculations for anthracene and statistically significant accuracy improvements for methane solvation and dipole inversion.

The fact that the optimal parameters for these thermodynamic calculations were similar to the default parameter previously chosen to minimize errors in the force evaluation is encouraging, showing that very different simulation observables have similar dependence on these cutoff parameters, and suggests general transferability of these parameters. The fact that the benchmark set used in this study included fairly large changes (such as the disappearance of a 7 D dipole or removal of 14 heavy atoms) suggest that the results should be relatively transferable to other molecules, with the caveats for heterogeneous systems discussed above. Errors in free energies of solvation for suboptimal parameters were larger than errors in the bulk enthalpy, suggesting the importance of examining multiple types of observables when performing optimization of force field parameters.

Interestingly, even in the case of lower overlap between the sampled states and unsampled thermodynamic states, we obtained highly accurate $\Delta\Delta G$ results between pairs of unsampled states that did share overlap with each other. The key factor leading to these accurate estimations of the statistical uncertainty in $\Delta\Delta G$ and $\Delta\Delta H$ values is the use of MBAR to correctly propagate the correlated uncertainty in differences between two ΔG values.

This reweighting process using MBAR could easily be further extended to any other type of parameter sensitivity analysis, such as the evaluation of free energies at a different temperatures or for different force field parameters for which samples are not available because no new simulations are done at the new conditions. The only constraint is that the new unsampled thermodynamic states should not be particularly far away in phase space from the sampled states. Determining exactly the trade-off between distance in parameter space and amount of error is a topic for further research. It is somewhat difficult to relate the statistical uncertainty in calculated observables to the phase space overlap for even a single state,[51−54] and interpreting overlap in multistate reweighting is even more complicated. However, the very low errors obtained in the present study indicate that the effects of at least moderate changes in parameters on thermodynamic observables should be easily measurable with at most moderate statistical errors.

■ ASSOCIATED CONTENT

Supporting Information

Additional Figures S-1−S-8, Tables S-1−S-3, submatrices of the matrix U in eq 6. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: michael.shirts@virginia.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge financial support from a Oak Ridge Associated Universities Ralph E. Powe Jr. Faculty Enhancement Award and National Science Foundation grant CHE-1152786. The authors also thank Mark Abraham, Magnus Lundborg, and Berk Hess (Royal Institute of Technology, Sweden) as well as John Chodera (Memorial Sloan−Kettering Cancer Center) and David Mobley (UC-Irvine) for useful comments on drafts of the manuscript.

■ REFERENCES

(1) Wereszczynski, J.; McCammon, J. A. *Q. Rev. Biophys.* **2012**, *45*, 1−25.

(2) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.;

Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. *Angew. Chem., Int. Ed. Engl.* **2006**, *45*, 4064−92.
(3) Lau, A. Y.; Roux, B. *Structure* **2007**, *15*, 1203−14.
(4) Smit, B.; Maesen, T. L. M. *Nature* **2008**, *451*, 671−8.
(5) Durrant, J. D.; McCammon, J. A. *BMC Biol.* **2011**, *9*, 71.
(6) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740−5761.
(7) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 144107.
(8) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007**, *111*, 2242−54.
(9) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.
(10) Zwanzig, R. W. *J. Chem. Phys.* **1955**, *23*, 1915.
(11) Widom, B. *J. Chem. Phys.* **1963**, *39*, 2808−2812.
(12) Oostenbrink, C.; van Gunsteren, W. F. *Chem.—Eur. J.* **2005**, *11*, 4340−4348.
(13) Oostenbrink, C.; van Gunsteren, W. F. *Proc. Nat. Acad. Sci.* **2005**, *102*, 6750−6754.
(14) Paliwal, H.; Shirts, M. *J. Chem. Theory Comput.* **2011**, *7*, 4115−4134.
(15) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(16) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1998**, *61*, 2635−2638.
(17) Shirts, M.; Chodera, J. *J. Chem. Phys.* **2008**, *129*, 124105.
(18) *J. Chem. Phys.* **2013**, *138*, 154108.
(19) Radmer, R. R. J. R. R. J.; Kollman, P. P. P. A. P. P. A. *J. Comput. Chem.* **1997**, *18*, 902−919.
(20) Chipot, C. *Free Energy Calculations*, 1st ed.; Springer: New York, 2007.
(21) Jorge, M.; Garrido, N. M.; Queimada, A. J.; Economou, I. G.; Macedo, E. A. *J. Chem. Theory Comput.* **2010**, *6*, 1018−1027.
(22) Meirovitch, H.; Cheluvaraja, S.; White, R. P. *Curr. Protein Pept. Sci.* **2009**, *10*, 229−243.
(23) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, 1420−1426.
(24) Lelievre, T.; Stoltz, G.; Rousset, M. *Free Energy Computations: A Mathematical Perspective*, 1st ed.; Imperial College Press: London, 2010.
(25) Ytreberg, F. M.; Swendsen, R. H.; Zuckerman, D. M. *J. Chem. Phys.* **2006**, *125*, 184114.
(26) Trzesniak, D.; Kunz, A.; van Gunsteren, W. *ChemPhysChem* **2007**, *8*, 162−169.
(27) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. *J. Phys. Chem. B* **2007**, *111*, 13052−63.
(28) Shirts, M. R. *J. Chem. Theory Comput.* **2013**, *9*, 909−926.
(29) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10089.
(30) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
(31) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451.
(32) Deserno, M.; Holm, C. *J. Chem. Phys.* **1998**, *109*, 7678.
(33) Smith, P. E.; Pettitt, B. M. *J. Chem. Phys.* **1996**, *105*, 4289.
(34) Mark, P.; Nilsson, L. *J. Comput. Chem.* **2002**, *23*, 1211−9.
(35) Petersen, H. G. *J. Chem. Phys.* **1995**, *103*, 3668.
(36) Hünenberger, P. H.; McCammon, J. A. *J. Chem. Phys.* **1999**, *110*, 1856.
(37) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
(38) Lagüe, P.; Pastor, R. W. *J. Phys. Chem. B* **2004**, *108*, 363−368.
(39) in't Veld, P. J.; Ismail, A. E.; Grest, G. S. *J. Chem. Phys.* **2007**, 127.
(40) Abraham, M. J.; Gready, J. E. *J. Comput. Chem.* **2011**, *32*, 2031−40.
(41) Wang, H.; Dommert, F.; Holm, C. *J. Chem. Phys.* Vol. *133*, 034117. DOI: 10.1063/1.3446812.
(42) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(43) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845−854.
(44) Schlick, T. *Molecular Modeling and Simulation*, 1st ed.; Springer: New York, 2002.
(45) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667−683.
(46) van der Spoel, D.; van Maaren, P. J. *J. Chem. Theory Comp.* **2006**, *2*, 1−11.
(47) Klimovich, P. V.; Mobley, D. L. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 307−316.
(48) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.
(49) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24−34.
(50) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Mol. Phys.* **1996**, *87*, 1117−1157.
(51) Lu, N. D.; Kofke, D. A. *J. Chem. Phys.* **2001**, *114*, 7303−7311.
(52) Lu, N. D.; Kofke, D. A. *J. Chem. Phys.* **2001**, *115*, 6866−6875.
(53) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 054103.
(54) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 084109.