Solvation free energies of amino acid side chain analogs for common molecular mechanics water models

Michael R. Shirts and Vijay S. Pande
Department of Chemistry, Stanford University, Stanford, California 94305-5080

(Received 23 November 2004; accepted 28 January 2005; published online 7 April 2005)

Quantitative free energy computation involves both using a model that is sufficiently faithful to the experimental system under study (accuracy) and establishing statistically meaningful measures of the uncertainties resulting from finite sampling (precision). In order to examine the accuracy of a range of common water models used for protein simulation for their solute/solvent properties, we calculate the free energy of hydration of 15 amino acid side chain analogs derived from the OPLS-AA parameter set with the TIP3P, TIP4P, SPC, SPC/E, TIP3P-MOD, and TIP4P-Ew water models. We achieve a high degree of statistical precision in our simulations, obtaining uncertainties for the free energy of hydration of 0.02–0.06 kcal/mol, equivalent to that obtained in experimental hydration free energy measurements of the same molecules. We find that TIP3P-MOD, a model designed to give improved free energy of hydration for methane, gives uniformly the closest match to experiment; we also find that the ability to accurately model pure water properties does not necessarily predict ability to predict solute/solvent behavior. We also evaluate the free energies of a number of novel modifications of TIP3P designed as a proof of concept that it is possible to obtain much better solute/solvent free energetic behavior without substantially negatively affecting pure water properties. We decrease the average error to zero while reducing the root mean square error below that of any of the published water models, with measured liquid water properties remaining almost constant with respect to our perturbations. This demonstrates there is still both room for improvement within current fixed-charge biomolecular force fields and significant parameter flexibility to make these improvements. Recent research in computational efficiency of free energy methods allows us to perform simulations on a local cluster that previously required large scale distributed computing, performing four times as much computational work in approximately a tenth of the computer time as a similar study a year ago. © *2005 American Institute of Physics*. [DOI: 10.1063/1.1877132]

I. INTRODUCTION

An important goal of computational chemistry is the accurate prediction of free energies in molecular systems. The calculation of free energies is a computationally intensive task, even for small systems, because it requires sampling of all thermally relevant configurations of the system. Extensive work over the last 20 years has gone into developing the theoretical and computational apparatus to perform free energy computations, and these calculations can now be performed for many systems with moderate levels of accuracy.[1–3] However, current calculations have still not progressed to be generally useful in applications. For example, reliably predicting experimental observables such as drug binding affinities, which require accuracies of 0.5–1.0 kcal/mol, is still not possible.

There are two main problems in the quantitative prediction of interaction free energies. First, there are the discrepancies between the models used for simulation and the experimentally measured reality. This lack of *accuracy* means the model may not adequately represent the experimental system under study. Second, the phase space of the model must be sufficiently sampled in order to capture all thermally relevant contributions to the ensemble average of the observables of interest. Otherwise, the results will lack the necessary *precision*, and independent calculations will most likely lead to different answers. Only if sufficient precision is obtained in a statistically well-defined manner is it possible to design models that are accurate enough for the application at hand.

In a previous study of neutral amino acid side chain analogs,[4] we demonstrated that precision equivalent to the experimental precision, 0.02–0.05 kcal/mol, is possible to achieve for small molecule energies of solvation. This was an order of magnitude greater precision than for any other set of multiple free energies of solvation. We demonstrated that current levels of computational resources (albeit using several thousand computers under the Folding@Home distributed computing environment[5]) could achieve the time scales necessary to sample all of the relevant degrees of freedom for free energy of hydration. We found that current biomolecular force fields (specifically OPLS-AA,[6] CHARMM,[7] and AMBER[8]) did not quantitatively match experimental measurements. The hydration free energies obtained from simulations were uniformly too high compared to experiment, with a root mean square error of $\approx$1 kcal/mol.

Purposes of present study. In this paper, we extend this previous study by examining the free energy of solvation of the neutral amino acids side chain analogs as a function of

TABLE I. Correspondence between amino acids and the amino acid side chain analogs used in this study, and the naturally occurring frequency of the amino acids in proteins (Ref. 41).

Ala	methane	8.3%
Val	propane	6.6%
Ile	*n*-butane	5.2%
Leu	*iso*-butane	9.0%
Phe	toluene	3.9%
Ser	methanol	6.9%
Thr	ethanol	5.8%
Tyr	*p*-cresol	5.8%
Cys	methanethiol	1.7%
Met	methyl ethyl sulfide	2.4%
Asn	acetamide	4.4%
Gln	propionamide	4.0%
Hid	4-methylimidazole	2.2%
Hie	4-methylimidazole	2.2%
Trp	3-methylindole	1.3%

the water model used for solvation. Amino acid side chain analogs represent a natural test case for biomolecular interaction. First, experimental data are available for direct comparison.[9–14] Second, the conformation space of these molecules is simple enough that we expect to be able to obtain good statistical sampling. Our previous study[4] demonstrated that we were indeed able to sample these molecules to an extent sufficient to obtain experimental precision. Third, a variety of well-defined, extensively tested, and widely used potential sets are available for these molecules—in this paper, we use OPLS-AA, which had the best performance of free energy relative to experiment in our previous study.[4,6,15] If we succeed in obtaining sufficient sampling of these systems, and thus obtain high precision values, remaining deviations from experiment must arise from inaccuracies in these models—where "models" refers both to the force field parameters and to the choice of simulation protocols. The correspondence between amino acids and amino acid side chain analogs is shown in Table I. In our previous paper, we analyzed other possible test systems, and discussed why amino acid side chain analogs are the most preferable of possible test cases.

When evaluating models, one is torn between using more recent simulation methods with the best proven scientific validity, the methods used for the actual parametrization of the model, and the most commonly used methods in the literature—which may all be somewhat different. For example, most force fields (for both proteins and water) were parametrized using simulations that employed finite-ranged nonbonded interactions, which suggests that one should use the corresponding truncation protocol for nonbonded interactions in order to be faithful to the parametrization of the model. However, truncation schemes for electrostatic interactions have been shown to lead to qualitatively incorrect results,[4,16,17] as compared with more sophisticated and increasingly more common methods such as Ewald summation or the use of a reaction field.

In the previous amino acid side chain analog paper, we demonstrated how the group-based tapered cutoffs, even group-based smoothed ones that do not have many of the irregularities noted earlier,[17] gave highly variable results depending on the cutoff distance. This is a clear problem for comparison between studies carried out by different researchers.[4] In this study, we use methodology that meet two separate criteria as much as possible. First, we choose methods as physical as possible, for example, which rigorously implement the desired ensemble, are ergodic, and conserve energy and other relevant quantities. This removes several sources of error and irreproducibility. Second, we choose methods that are as free as possible of arbitrary, nonphysical parameters, such as cutoffs for treatment of long-range interactions. Specifically, we use Ewald methods for evaluation of the long-range Coulombic interactions, and a long-range correction for the Lennard-Jones cutoff. A greater discussion of these and other methodological choices is included in the Methods.

For sampling, we chose to use molecular dynamics as the most direct method for isobaric-isothermal Boltzmann sampling of condensed phase molecular systems. We realize that this method may not be the most computationally efficient in many cases. But it is sufficient for the problem at hand, as well as being simple to implement, frequently utilized, well studied, and well understood.

These results can serve as one possible test of the applicability of explicit fixed-charge water models with current classical protein models, and therefore, the present work could guide the development of more accurate force fields. This work demonstrates that it is now easier to incorporate very important yet computationally intensive observables such as hydration free energy into force field parametrization efforts. Free energy simulations thus serve as useful methods for parameter verification as they quantify the effects of both the enthalpy and entropy of molecular interactions.

II. METHODS

A. Computational methods

The simulations were performed on the Stanford Bio-X cluster of 2.8 GHz Intel Xeon Pentium IV processors. The simulations were performed using a modified version of the molecular dynamics suite GROMACS (v3.1.4).[18,19] Our modifications included adding Andersen pressure and temperature control,[20] and velocity Verlet (which is required for correct pressure control implementation).[21] We also implemented SETTLE[22] for four-point water models, and a variety of modifications for free energy computation, mostly bookkeeping, to allow for decoupling free energies to be computed through the Bennett acceptance ratio.[23,24] We also implemented a modified "soft-core" λ dependence, described later in this paper. Arrangements are being made to include versions of these modifications in the next version of GROMACS. GROMACS was compiled in double precision, instead of the usual single precision, as this was determined to be necessary for converging the bond constraints to the desired relative precision. Although GROMACS is optimized for single precision, on the Pentium IV processors in our cluster with SSE2 vector instructions enabled, the additional time

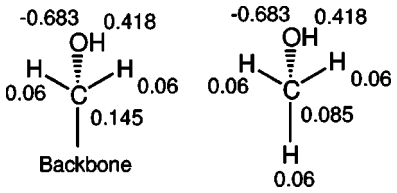

FIG. 1. Differences in partial charges between (left) the OPLS-AA parameter set serine side chain and (right) the OPLS-AA-derived parameters used for the side chain analog in this study. Only the β carbon changes in charge.

taken for double versus single precision was only about 10% greater.

Parameters for the amino acid side chain analogs were generated from the OPLS-AA[6] force fields, and were identical to those used in our previous amino acid side chain study, with the exception of the charge and Lennard-Jones parameters for sulfur, which were obtained from the OPLS-AA/L modifications.[15] We studied all the neutral amino acid side chain analogs except for those of proline and glycine. In addition, we examined analogs of both neutral protonation states of His (referred to as Hid and Hie analogs, where the proton is attached to the δ and ε nitrogens, respectively).

We have omitted the side chain analogs of amino acids carrying a formal charge at neutral *p*H in our computations. Ewald methods provide a much better treatment of charged species than the cutoffs used in our previous paper, as they rigorously treat the long-range Coulombic interactions in the periodic system. However, to simulate a charged molecule, we must either add counterions or add a uniform neutralizing background charge distribution. The effect of either of these on the free energy of hydration becomes complicated to evaluate. Additionally, although experimental data does exist for the free energy of hydration of these charged amino acid side chain analogs, it is more complicated to compute and leads to larger uncertainties, as it is the neutral species that enters the vapor state, and the energy of proton association must be included as well.[9,25]

Some modifications to the side chain β carbon parameters were necessary in order to use the protein force field parameters for the side chain analogs. For the nonbonded parameters for these atoms, the Lennard-Jones term was taken from the standard aliphatic carbons (either CH_3 or CH_2, depending on the amino acid) of each parameter set. The partial charge of this atom was determined for each amino acid side chain analog by reducing the partial charge on the β carbon by the amount of the charge of one hydrogen atom (see Fig. 1). If bonded parameters (bonds, angles, and torsions) were already defined in the parameter set for atoms of the new resulting type and connectivity, these were used instead of the bonded parameters for the original atom type. We discuss in our previous paper how these assumptions are reasonable for OPLS-AA, and are in fact approximately the reverse process used for building the amino acids from small molecules, as can be seen by comparison to the original OPLS-AA amino acid parameters. Input topology and coordinate files for all the amino acid side chain analogs in all parameter sets is provided in the Supplementary Materials.[26]

Six different published water models were compared. The original TIP3P water model[27] was used, which is rigid and includes no Lennard-Jones term on the hydrogens. This is slightly different than the TIP3P-like water model used to develop the CHARMM parameter set, which does have Lennard-Jones terms on the hydrogens.[7] TIP3P (including slight variants, such as the CHARMM TIP3P mentioned above) is probably the most common water model used in explicit water simulations of biomolecules. TIP4P[27] was developed at the same time as TIP3P, and most of the OPLS-AA parameters were developed in conjunction with TIP4P. It is less common than TIP3P in general use, however, partly because the treatment of the dummy site is somewhat complicated in molecular dynamics and many code bases did not support it for some time, delaying adoption. SPC[28] is an important early model still used by many simulations groups. SPC/E is an improved version of SPC,[28] designed to correct for the self-polarization in water. TIP4P-Ew[29] was very recently designed specifically for use with Ewald summation methods, and has demonstrated an extremely good fit to experimental data over a range of temperatures. TIP3P-MOD[30] is not in common usage, and was designed by modifying the Lennard-Jones ϵ and σ parameters in order to try to improve the solute-solvent interaction between methane and water without adversely affecting the pure liquid properties of water. The various parameters and pure liquid properties of these water models are compared in Table II. There are, of course, many other less common effective pair potential water models that have been developed, for both biological and other chemical systems, but a complete discussion of the subject is beyond the scope of this paper.

The hydration free energies were computed using the Bennett acceptance ratio method,[24] with molecular dynamics used to sample the potential energy differences between a number of neighboring intermediate states between the coupled and uncoupled end states. For each intermediate state, simulations using a 2.0 fs time step were equilibrated for 1.0 ns, followed by 5.0 ns data collection. Potential energy differences between intermediates were output every 50 steps, or 0.1 ps. Integration of the equations of motion were performed using the velocity Verlet algorithm,[21] and all bonds were constrained using RATTLE[31] with a relative accuracy of 10^{-6} for the small molecules and SETTLE[22,29] for the water molecules.

Andersen pressure control and Andersen temperature control[20] were used to rigorously implement an isobaric-isothermal (NPT) ensemble. The pressure was calculated using an atom-center based virial. Isotropic scaling of the cubic volume at each step was performed by scaling the atomic positions, and then correcting the intramolecular distances with RATTLE or SETTLE.

The Anderson piston mass was chosen to correspond with a compressibility with 4.5×10^{-5} bar^{-1} and a time constant of 1.67 ps. The thermodynamic observables are formally independent of the choice of the piston mass. Therefore, as long as the piston is heavy enough that the simulation is stable with the choice of time step, and light enough so that that the time scale for the volume fluctuations (which were of order 1 ps for these simulations) is much shorter than the simulation length, the value of the piston

TABLE II. Model input parameters and pure liquid properties for SPC, SPC/E, TIP3P, TIP4P, TIP3P-MOD, and TIP4P-Ew water models. Uncertainties for the density are all $\leqslant 0.0002$ g cm^{-3}, for the heat of vaporization are $\leqslant 0.005$ kcal/mol, for the free energies are $\leqslant 0.03$ kcal/mol, and for the self-diffusion constant are $\leqslant 0.04 \times 10^{-5}$ cm^2 s^{-1}. Computation of the heat of vaporization, including the polarization correction, was done according to previously published methods (Ref. 29). Water experimental data of −6.33 and −6.32 come from Refs. 13 and 10, respectively.

	SPC	SPC/E	TIP3P	TIP4P	TIP3P-MOD	TIP4P-Ew	Expt.
ϵ(kcal/mol)	0.155 50	0.155 50	0.152 10	0.155 50	0.190 00	0.162 75	
σ(Å)	3.165 57	3.165 57	3.150 61	3.153 65	3.121 71	3.164 35	
q	0.410 00	0.423 80	0.417 00	0.520 00	0.417 00	0.524 22	
dOD (Å)	N/A	N/A	N/A	1.50	N/A	1.25	
dOH (Å)	1.0000	1.0000	0.9572	0.9572	0.9572	0.9572	
∠HOH	109.47	109.47	104.52	104.52	104.52	104.52	
Dipole (Debye)	2.27	2.35	2.35	2.177	2.35	2.321	
Density	0.9775	0.9988	0.9859	0.9997	0.9945	0.9954	0.9972
Diffusion	4.37	2.49	5.56	3.71	6.26	2.61	2.23,2.30
ΔH_{vap}	10.474	11.686	10.091	10.412	10.109	11.610	10.518
ΔH_{vap}[pol]	9.645	10.506	8.911	9.913	8.929	10.565	10.518
ΔG	−6.16	−7.05	−6.10	−6.11	−6.17	−6.98	−6.33, −6.32

mass is unimportant. Periodic boundary conditions were used and the simulation cell was constrained to a cubic shape.

All simulations were carried out at 298 K. Andersen temperature control was implemented by reassignment of all velocities from the Maxwell–Boltzmann distribution at periodic intervals, which in the limit of long time is rigorously equivalent to an isothermal ensemble.[20] In the 5.0 ns simulations presented here, velocities were reassigned every 100 steps (0.2 ps), resulting in 25 000 separate reassignments. The average kinetic energy of the simulations was checked to verify that it was in agreement with the control temperature of 298 K.

Perhaps the most important methodological choices in molecular simulations are those used to evaluate the long-range interactions. In this study, a neighbor list of 9.0 Å updated every ten steps was used for the short-range interactions. Particle Mesh Ewald[32] (PME) was used to evaluate the Coulombic interactions, with a real space cutoff of 7.5 Å and a PME order of 6. The Fourier spacing was chosen to be as close to 1.0 Å as possible given the box size and the need for integer numbers of grid points. The relative tolerance between the long- and short-range Coulomb energy was set to 10^{-6}, yielding a Gaussian width of 2.168 31 Å. The Lennard-Jones energy was computed with a switched, atomic-based cutoff between 7.0 Å and 7.5 Å. A long-range correction was added to the energy and pressure to eliminate the effect of this finite-range cutoff of the Lennard-Jones, including the effect of the switch.[4,33]

These cutoffs are significantly shorter than used in many other simulations. They were chosen as the cutoff parameters that made the simulations as fast as possible while still changing the instantaneous potential energy by less than 0.05% and the instantaneous pressure by less than 1% from much longer (up to 14 Å) short-range cutoffs. Simulation methodologies that are independent of cutoff parameters provide a distinct advantage, as they remove arbitrary, nonphysical variables that can greatly affect observables from the comparisons between the results of different published simulations, and thus affect the validity of scientific conclusions and comparisons.

The switching function $S(r)$ is the standard GROMACS switching function, chosen so that the function and its derivative are continuous. We note that in order to strictly conserve energy, $S(r)$ would also have to satisfy the conditions that $d^2S(r_{\text{off}})/dr^2 = d^2S(r_{\text{switch}})/dr^2 = 0$. However, trial constant energy simulations with 900 TIP3P molecules using 2 fs time steps exhibited no statistically meaningful drift in total energy over hundreds of picoseconds. Short trajectories from the same starting point were run with varying time step size, and the standard deviation of the total energy (or enthalpy, in the case of the isobaric-isoenthalpic ensemble) was quadratic in the time step size (as checked down to 0.125 fs), consistent with the inherent accuracy of the velocity Verlet algorithm.[21] We therefore conclude that this switching function is sufficiently accurate.

To prepare the simulation cell, a previously equilibrated cubic box of 900 water molecules was taken, and the amino acid side chain analogs were then solvated by placing them in the center of this box and removing all water molecules with atomic centers within 2.0 Å of any atomic center of the solute. An extremely short (200–400 steps) isochoric-isothermal (NVT) simulation with very frequent velocity reassignments was then used to remove highly unfavorable interactions. A further 1.0 ns NPT equilibration simulation was performed for each intermediate state prior to the 5.0 ns of data collection.

A single simulation is not statistically relevant; it is usually necessary to run multiple independent simulations to verify that the computed uncertainty estimates from single simulations are reasonable. For each free energy between two given intermediate states, we estimated the uncertainty by dividing the full 5.0 ns simulations into five 1.0 blocks, and computed the variance of the average over these five simulations, by the standard formula $\Sigma(\langle X\rangle - X)^2/(n-1)$. The

total variance for each molecular solvation free energy is computed as the sum of the variance for each interval between intermediates, with the overall uncertainty equal to the square root of the variance. This analysis is not as sophisticated as in our previous study of amino acid side chain analogs,[4] and in general, this level of uncertainty analysis would not be sufficient. However, in the previous study we demonstrated that these time scales are sufficient to sample over the relevant degrees of freedom,[34] and thus, making the reasonable approximation that the changes in simulation conditions do not significantly alter these correlations, this simpler uncertainty analysis is justified.

For the pure water simulations, 900 water molecules were used, and the averages were collected over a single 5.0 ns run. Statistical uncertainties δA for the ensemble observables A in Table II were calculated by using the variance $\langle A^2\rangle_c$ and correlation time of the observable τ_A over the $T=5.0$ ns simulation. The uncertainty in the mean for the observable A is then given by the formula $\delta A = \sqrt{\langle A^2\rangle_c}\sqrt{2\tau_A/T}$.[21,33] δA corresponds to one standard deviation σ and all results are reported as $\langle A\rangle \pm 1\sigma$. Correlation times were computed using GROMACS analysis tools.

Diffusion constants were computed by running a 5.0 ns constant enthalpy simulation, with initial kinetic energies sampled from a 298 K ensemble, divided into 100 50.0 ps blocks. For each 50.0 block, the mean square distance traveled by the oxygen atoms was computed, and the result was fitted to a straight line with granularity of 1.0 ps. The slope of the line is six times the self-diffusion constant, and results from the 100 samples are averaged to give the final value, with uncertainty computed from the sample variance. All simulations of pure water properties were run with 900 molecules and with pressure controls as described above.

B. Free energy calculations

Experimental free energies of hydration for weakly soluble solutes are determined from concentration measurements made on two phase systems, where one phase consists of a vapor with a partial pressure P_s of some solute molecule of type s, and the other phase consists of an aqueous solution with a number density concentration for solute molecules of ρ_s^l. When such a two phase system is at equilibrium with respect to transfer of molecules of type s between the phases, the solvation free energy is given by[9,10]

$$\Delta G_{solv} = kT \ln(P_s/\rho_s^l kT). \tag{1}$$

In these simulations, H is the parametrized Hamiltonian, where the coupled state ($\lambda=1$) corresponds to a simulation where the solute is fully interacting with the solvent and the uncoupled state ($\lambda=0$) corresponds to a simulation where the solute does not interact with the solvent. Free energies of hydration are computed by simulation by decoupling a molecule of the solute (here, amino acid side chain analogs) from the solvent using the Bennett acceptance ratio. We have found that the Bennett acceptance ratio can be much more efficient than TI for classical force field simulations in condensed phase.[35]

We analyzed the correction needed to relate the actual solvation free energy ΔG_{solv} and the decoupling free energy ΔG_{sim}:

$$\Delta G_{solv} = \Delta G_{sim} - kT \ln(V^*/V_1), \tag{2}$$

where V_1 is the mean volume at full coupling ($\lambda=1$) and V^* is the mean volume of a box of pure solvent, with the same number of solvent molecules as in the coupled simulation.[4] In all cases, this correction was between 0.005 and 0.001 kcal/mol (roughly proportional to molecular size), an order of magnitude less than the uncertainty, and so was not included.

We parametrized only the nonbonded interactions between the solute and solvent molecules by λ, i.e., the nonbonded potential function is of the form:

$$U(\lambda_C,\lambda_{LJ}) = U_s + U_w + U_{w-w} + U_{s-w}(\lambda_C,\lambda_{LJ}), \tag{3}$$

where U_s is the intramolecular potential energy of the solute, U_w is the intramolecular potential energy of the waters (zero for the classical models used in this study), U_{w-w} is the intermolecular energy of the water-water interactions, and U_{s-w} is the intermolecular energy of the solute-water interactions. It was necessary to modify the GROMACS code to make it possible to neglect the intramolecular terms when changing parameters with respect to λ.

There are a wide variety of choices for parametrization in λ from the initial to the final Hamiltonian, but the choice of a good path for a given system is still an unsolved problem.[1,36–38] As discussed in our previous paper,[4] many of the common choices, such as linear interpolation between Hamiltonians, can be extremely computationally inefficient, and have poor convergence properties at the end states. We avoided many of these problems by using the following expression for the λ-dependent nonbonded interaction energy:

$$U_{s-w}(\lambda_C,\lambda_{LJ}) = \sum_{i,j} \lambda_C \frac{q_i q_j}{r_{ij}} + \lambda_{LJ} 4\epsilon_{ij} \times \left(\frac{1}{[\alpha_{LJ}(1-\lambda_{LJ}) + (r_{ij}/\sigma_{ij})^6]^2} - \frac{1}{\alpha_{LJ}(1-\lambda_{LJ}) + (r_{ij}/\sigma_{ij})^6} \right), \tag{4}$$

where the sum i is over all solute atoms and the sum j is over all solvent atoms. Equation (4) includes the standard Coulombic term with a linear dependence on λ, but also incorporates a soft-core parametrization, where the infinity in the Lennard-Jones interaction is smoothed to zero as a function of λ. The parameter α_{LJ} is a positive constant which controls this transition, and is equal to 0.5 for all simulations in this study. In our previous study, we used the function of Beutler *et al.*,[36] but some unpublished experimentation suggests that the form used here, where the leading λ^4 term is replaced by at λ^1 term and $\alpha_{LJ}(1-\lambda_{LJ})^2$ is replaced by $\alpha_{LJ}(1-\lambda_{LJ})$, is more efficient, by perhaps 30%. In applying the transformation from solute fully coupled with the solvent ($\lambda_C=1, \lambda_{LJ}=1$) to the solute fully decoupled from the solvent ($\lambda_C=0, \lambda_{LJ}=0$), we first decouple solvent-solute Coulombic interactions and then the solvent-solute Lennard-Jones terms.

TABLE III. Free energy of solvation of amino acid side chain analogs in published water models in Table II. Average deviations from experiment are presented in Table V. All results in kcal/mol. Uncertainties of the individual free energy solvation used in the average are 0.02–0.06 kcal/mol, roughly proportional to molecular size. TIP4P, SPC/E, and TIP4P-Ew have longer diffusion times, hence longer water decorrelation times, and hence slightly larger uncertainties within that range for each amino acid side chain analog.

	Experiment	TIP3P-MOD	TIP3P	SPC	TIP4P	SPC/E	TIP4P-Ew
Ala	1.94	2.18	2.24	2.17	2.25	2.27	2.27
Val	1.99	2.15	2.34	2.41	2.55	2.56	2.63
Ile	2.15	2.10	2.43	2.48	2.68	2.76	2.82
Leu	2.28	2.08	2.27	2.50	2.71	2.75	2.64
Phe	−0.76	−1.31	−0.86	−0.79	−0.65	−0.54	−0.42
Ser	−5.06	−4.58	−4.51	−4.62	−4.71	−4.55	−4.65
Thr	−4.88	−4.42	−4.22	−4.43	−4.47	−4.37	−4.44
Tyr	−6.11	−5.96	−5.46	−5.39	−5.06	−5.07	−4.80
Cys	−1.24	−0.83	−0.55	−0.50	−0.43	−0.33	−0.34
Met	−1.48	−0.78	−0.35	−0.42	−0.07	−0.05	−0.08
Asn	−9.68	−8.88	−8.51	−8.53	−8.52	−8.32	−8.51
Gln	−9.38	−9.02	−8.63	−8.59	−8.49	−8.42	−8.63
Hid	−10.27	−9.36	−8.88	−8.71	−8.66	−8.55	−8.68
Hie	−10.27	−9.46	−9.08	−9.14	−9.10	−8.99	−8.96
Trp	−5.88	−5.55	−4.88	−4.70	−4.14	−4.30	−4.02

The Coulombic interactions were decoupled with λ_C values of 0.0, 0.5, and 1.0. For the Lennard-Jones terms, preliminary investigations showed that uneven spacing in λ_{LJ} could give better results than even spacing, as closer spacing is desired when the curvature of $\langle dH/d\lambda \rangle$ is larger. The λ_{LJ} values chosen were therefore 0.0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 1.0. We chose these values after testing showed that they were sufficiently closely spaced to converge the free energy of the largest molecule, the tryptophan analog, to within the desired uncertainty.

At each nth λ value, the potential energy using the Hamiltonian with the $n-1$ and $n+1$ λ values is computed, and the difference between these energies and original potential energy is output. The Bennett acceptance algorithm is applied to the distributions of potential energy differences between the $n+1$ and n states, sampled over both the n and $n+1$ states.[23,24] This process extracts the free energy difference that best corresponds to these potential energy difference distributions, and is further explained in the cited references. Since we are computing the difference between energies in the isobaric-isothermal ensemble, the relevant free energy is the Gibbs free energy or the free enthalpy.

We also computed long-range Lennard-Jones corrections to the solvation free energy, an estimate for total attractive energy from the Lennard-Jones term neglected from outside the cutoff regions.[4,33,39] These long-range corrections to the solute-solvent interaction were computed after the simulation. This is in addition to the long-range dispersion correction performed during the simulation, which serves only to correct the potential energy and pressure for the effect of the finite-range Lennard-Jones cutoff, and was not included in the potential energy difference used for the free energy computation.

For each solute the long-range contribution to the free energy can be calculated by numerically integrating

$$E_{LRC} = \sum_i 16\pi\rho\epsilon_{ij} \int_{r=r_{\text{switch}}}^{\infty} \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^{6} \right] S(r) r^2 dr, \tag{5}$$

where i runs over the solute atoms, r is the distance from solute atom i, ρ is the number density of the solvent molecules, ϵ_{ij} and σ_{ij} are the standard Lennard-Jones parameters between solute atom i and the solvent molecule Lennard-Jones site of type j, and $S(r)$ is the switching function used in GROMACS. In this study, there is only one Lennard-Jones site per solvent molecule, so we can average over j. This derivation of the correction assumes that the solvent radial distribution function $g(r_{ij})=1$ in the tapering region and beyond, which is an adequate assumption for the water models used here, and for the size of the ligands. With this free energy correction, the results from cutoffs starting at distances as short as 6 Å were tested in a previous study, and shown to be statistically indistinguishable from those obtained with longer cutoffs.[4]

III. RESULTS

The main results of this study are a set of highly precise hydration free energy values for the 15 neutral amino acid side chain analogs solvated in the six previously published water models. Additionally, we present free energies of hydration of the side chain analogs with a series of modified water models that demonstrate that it is possible to greatly improve solvent-solute free energies without significantly adversely affecting the pure liquid properties of a water model. These free energies are presented in Tables III and IV. We include a further breakdown of the results, including charging and uncharged solvation free energies, long-range corrections, and the individual uncertainty estimates in further charts in the Supplementary Materials.[26]

TABLE IV. Free energy of solvation of amino acid side chain analogs in TIP3P, TIP3P-MOD, and the Lennard-Jones modified water parameters presented in Table VII. Average deviations from experiment are presented in Table VI. All result in kcal/mol. Uncertainties of the individual free energy solvation used in the average are 0.02–0.05 kcal/mol, roughly proportional to molecular size.

	Experiment	TIP3P	TIP3P-MOD	M20	M21	M22	M23	M24	M25
Ala	1.94	2.24	2.18	2.13	2.19	2.19	2.14	2.21	2.22
Val	1.99	2.34	2.15	2.08	2.10	2.05	2.05	2.00	2.07
Ile	2.15	2.43	2.10	2.14	2.12	2.10	2.07	2.00	2.03
Leu	2.28	2.27	2.08	2.13	2.11	1.96	1.90	1.99	2.01
Phe	−0.76	−0.86	−1.31	−1.32	−1.45	−1.55	−1.60	−1.65	−1.64
Ser	−5.06	−4.51	−4.58	−4.63	−4.57	−4.66	−4.56	−4.65	−4.62
Thr	−4.88	−4.22	−4.42	−4.45	−4.51	−4.53	−4.58	−4.58	−4.57
Tyr	−6.11	−5.46	−5.96	−6.20	−6.31	−6.40	−6.45	−6.63	−6.69
Cys	−1.24	−0.55	−0.83	−0.86	−0.96	−1.09	−1.06	−1.09	−1.10
Met	−1.48	−0.35	−0.78	−0.86	−0.90	−1.10	−1.04	−1.09	−1.11
Asn	−9.68	−8.51	−8.88	−9.05	−9.02	−9.21	−9.25	−9.36	−9.32
Gln	−9.38	−8.63	−9.02	−9.26	−9.24	−9.41	−9.39	−9.54	−9.59
Hid	−10.27	−8.88	−9.36	−9.50	−9.48	−9.58	−9.74	−9.86	−9.87
Hie	−10.27	−9.08	−9.46	−9.60	−9.73	−9.89	−9.91	−10.02	−9.97
Trp	−5.88	−4.88	−5.55	−5.63	−5.76	−5.93	−6.02	−6.17	−6.19

In Table II, we present the model definitions as well as our data collected on pure water using the simulation parameters used for the free energies of hydration as defined in this paper. We see that all models, with the simulation conditions described in the Methods, are relatively accurate in predicting water properties at 298 K. However, some models, such as TIP3P and SPC, do not give as accurate predictions under the current conditions as in previously published papers.[27,29,40] This is most likely because of differences of simulation conditions between this and the previous studies; for example, Ewald methods generally result in slightly lower densities than cutoffs.[29]

The published values for TIP4P-Ew, which was developed using different conditions as ours but with the same philosophy of maximum transferability of simulation parameters, are encouragingly close to our results. Our density of 0.9954 ± 0.0002 g cm^{-3} agrees exactly with their density of 0.9954 ± 0.0003 g cm^{-3}, and our (polarization-corrected) enthalpy of solvation of 10.565 ± 0.005 kcal/mol is only very slightly outside the extremely strict precision of the published value of 10.583 ± 0.004.[29] These simulations were performed on completely independent codes, with many differences in input parameters. Some of these differences, with the previously published TIP4P-Ew study first and the current study second, are 512 vs 900 water molecules, 1 vs 2 fs time steps, velocity reassignment every 2.0 vs 0.2 ps, 0.35 Å vs 2.16 Å screening Gaussian, and Lennard-Jones switched cutoffs of 9.0 Å–9.5 Å vs 7.0 Å–7.5 Å. The diffusion constants were somewhat different (2.335 ± 0.004 vs 2.61 ± 0.02 10^{-5} cm^2 s^{-1}), but still within 10%. This may be related to the shorter cutoffs used in this study, as correlated motions may not be as transferable as the thermodynamics are. Since both simulations used methods that are relatively insensitive to the input parameters (as described earlier in this paper), we believe this represents a small success of the philosophy of parametrization for simulation conditions that are not dependent on arbitrary parameters. As such methodologies become more widely adapted, comparisons between simulation studies may become significantly easier.

An analysis of the free energy of solvation amino acids side chain analogs in the six published water models is presented in Table V. The average error and rms error are computed in two ways; first, as the flat average over all 15 side chain analogs, and second, as an average weighted by the naturally occurring frequency of the corresponding side chains in proteins,[41] with frequencies listed in Table I. This averaging method is more relevant as a validation for protein simulations because data relating to the accuracy of a less

TABLE V. Free energy of solvation of amino acid side chain analogs in published water models. The values for "Slope" and "Constant" are from the linear fit from the flat averages over the 15 side chain analogs. All averages and root mean square (rms) errors in kcal/mol. Uncertainties of the individual free energy solvation used in the average are 0.02–0.06 kcal/mol.

	Flat		Weighted			
	Average error	rms Error	Average error	rms Error	Slope	Constant
TIP3P-MOD	0.33	0.51	0.23	0.42	0.943	0.119
TIP3P	0.67	0.79	0.50	0.64	0.948	0.389
SPC	0.69	0.82	0.53	0.64	0.929	0.427
TIP4P	0.84	0.97	0.71	0.82	0.936	0.596
SPC/E	0.90	1.01	0.71	0.82	0.932	0.642
TIP4P-Ew	0.90	1.02	0.65	0.77	0.939	0.668

common side chain, such as cystine, should not be as important as the data relating to leucine, more than five times as common.

In both the weighted and unweighted averages in Table V, all models underestimate solvation free energy to some extent. Indeed, all but TIP3P-MOD (which, although published, is not a water model in common usage) have solvation free energies that are worse, on average, than TIP3P in comparison to experiment. Interestingly, models that better capture pure liquid properties do not necessarily have improved solute-solvent free energies. TIP4P-Ew is perhaps the most accurate three or four point fixed-charge model for representing water over a range of temperatures, thanks to rigorous quantitative fitting to a range of experimental data.[29] However, it has one of the worst fits of free energies of solvation to experiment of the models in this study, as seen in Table V.

It is useful to compare the free energies of hydration in TIP3P between the current paper and our previous amino acid side chain paper. Because the sulfur parameters have changed between the data sets, it is only valid to compare the other 13 molecules. Any other statistically significant differences must be a result of the simulation conditions. We would expect there to be some small differences in the results because of the difference in the treatment of the long-range interactions. We find an unweighted average difference of −0.20 kcal/mol (with the new results being more favorable towards solvation), with a rms difference of 0.24 kcal/mol, and variance 0.14 kcal/mol. The differences ranged between 0.01 and −0.44 kcal/mol, and interestingly four of the five largest differences in magnitude belonged to the four hydrophobic side chain analogs of Val, Ile, Leu, and Phe. Most likely, since the charges on these molecules are relatively small, the differences are due to changes in water structure in PME vs group-based tapered cutoffs. In general these differences are rather low, in that different simulation methods yield fairly close free energies, which is somewhat encouraging. However, it also demonstrates the problems inherent in comparing results of studies run with different simulation methodologies.

Perhaps the most important reason that current water and protein models do not accurately predict free energies of solvation is simply that they were never parametrized to predict such observables. Free energies of solvation are computationally expensive; and computational limitations at the time of parametrization usually did not allow parameters to be further tuned to also reproduce accurate free energies of solvation in water, a thermodynamic measure that is likely to be very relevant for biological processes. Current generation biomolecular force fields such as AMBER(*ff94*),[8] CHARMM22,[7] GROMOS,[42] and OPLS-AA[6] are in many cases successful in capturing the qualitative behavior of protein structure and dynamics.[43–45] However, parameters were usually tuned to give accurate fits to quantum mechanical energy barriers, as well as to reproduce enthalpies of vaporization and densities for pure liquids.[6–8,42] Other molecular force fields such as TraPPE,[46–48] NERD,[49] and GIBBS99[50] have incorporated phase equilibria data that implicitly include free energies of transfer, but as yet do not include the molecular diversity necessary for biomolecular simulation. Only very recently, since our previous publication on this subject,[4] has at least one such parametrization to free energies has been attempted.[51]

Given that most current molecular force fields are not designed for predicting solvation free energies, then it may be the case that there is sufficient flexibility within the parameters to better predict the free energy of solvation while still not greatly altering the pure liquid properties. So as a proof of concept, we decided to try to improve the average solvent-solute free energies with respect to experiment while maintaining the solvent-solvent and solute-solute properties.

We decided to adopt the philosophy that was used in the development of TIP3P-MOD.[30] First, adjusting the Lennard-Jones parameters for the solute is not an option, because the physical properties of the pure liquid solutes depend almost entirely on these Lennard-Jones parameters. Adjusting the charges for the solutes was not feasible either, as making the charges sufficiently large to improve the solvation free energy would lead to entirely unrealistic dipole moments for the hydrophobic species.

Of course, because of the combining rules used, the solute-solvent behavior depends just as much on the solvent parameters as the solute properties. The precise partial charges on the water are vital to maintain the correct water properties. The Lennard-Jones parameters, however, are less significant. Because of the strength of the Coulombic attractions, the average position of the oxygen atoms of the water molecules is within the r^{-12} Lennard-Jones repulsive core. Therefore, the depth of the well is not nearly as important as the location of the repulsive wall. It should therefore be possible to simultaneously increase ϵ, decreasing the depth of the well, increasing the solute-solvent attraction, and decrease σ, keeping the repulsive core in approximately the same place, and thus generally maintaining pure liquid properties.

TIP3P has $\epsilon = 0.152\,10$ kcal/mol and $\sigma = 3.150\,61$ Å, while TIP3P-MOD has exactly the same geometry and partial charges, but a drastically different $\epsilon = 0.1900$ kcal/mol and $\sigma = 3.121\,71$ Å (see Table VII). Despite this difference, TIP3P-MOD has an almost identical enthalpy of vaporization (10.109 kcal/mol for TIP3P-MOD vs 10.091 kcal/mol for TIP3P), and a density that is actually closer to the experimental value of 0.9972 (0.9998 g cm^{-3} for TIP3P-MOD vs 0.9859 g cm^{-3} for TIP3P) under the current simulation conditions.

We therefore define a series of six models, M20, M21, M22, M23, M24, and M25, identified by the significant digits of their ϵ value: M20 has $\epsilon = 0.20$ kcal/mol, M21 has $\epsilon = 0.21$ kcal/mol, and so on. We adjust the σ for each of these to fit as closely as possible to the experimental density of water at 298 K, with a granularity of 0.001 Å. The final parameters and pure liquid properties are shown in Table VII, in comparison to TIP3P and TIP3P-MOD. This choice of density as the primary variable for fitting to pure liquid properties is somewhat arbitrary. This was intended as a proof of concept to demonstrate the parameter flexibility in current models, not a production-level water model for extensive simulations.

TABLE VI. Free energy of solvation of amino acid side chain analogs in TIP3P, TIP3P-MOD, and the Lennard-Jones modified water parameters presented in Table VII. The values for slope and constant are from the linear fit from the flat averages over the 15 side chain analogs. All averages and root mean square (rms) errors in kcal/mol. Uncertainties of the individual free energy solvation used in the average are 0.02–0.06 kcal/mol.

	Flat		Weighted			
	Average error	rms Error	Average error	rms Error	Slope	Constant
TIP3P	0.67	0.79	0.50	0.64	0.927	0.389
TIP3P-MOD	0.33	0.51	0.23	0.42	0.943	0.119
M20	0.25	0.43	0.18	0.37	0.957	0.090
M21	0.22	0.43	0.15	0.37	0.959	0.060
M22	0.11	0.38	0.06	0.35	0.965	−0.026
M23	0.08	0.38	0.04	0.36	0.968	−0.041
M24	0.01	0.38	0.00	0.36	0.979	−0.066
M25	0.02	0.39	0.01	0.37	0.981	−0.052

Looking at the pure liquid properties, we see that with all models M20-M25, we are able to get a density that is within 0.009 g cm^{-3} of experiment for all models, and are actually much closer to experiment than TIP3P, at 0.1% error vs 1.1% error. The heat of vaporization is further from experiment (not surprisingly, since we were not attempting to constrain it), but no model is further than 1.6% from the heat of vaporization of TIP3P, the original model that all parameters but σ and ϵ are taken from. The free energy of solvation is slightly further away from the experimental value of −6.32 kcal/mol[10] or −6.33 kcal/mol,[13] but are still almost within the uncertainty of the TIP3P and TIP3P-MOD. So we have a class of models with very similar (and in some respects, slightly improved, considering the density) pure liquid properties compared to TIP3P.

How do these models compare to the other models with respect to the free energy of solvation of the amino acid side chain analogs? This data is presented in Table VI. We see that the increased well depth greatly increases the affinity of the solute molecules for the solvent, such that for the M24 and M25 models, the average error goes almost exactly to zero. Taking the amino acid frequency-weighted average, this is an average improvement of −0.50 kcal/mol, with almost no affect on the pure liquid properties. This is accomplished while at the same time reducing the frequency-weighted rms error from 0.51 kcal/mol to 0.31 kcal/mol, and improving the slope of the fit between experiment and simulation.

IV. DISCUSSION

In the preceding section, we presented the results of our simulations and analyzed the precision of the results and the agreement with experimental measurements. In this section, we discuss the utility and obtainability of high precision results, as well as what these results can tell us about the models and simulations using these models.

There are many studies of free energies of solvation of models of small molecules, and it is important to compare these results with the previous studies where possible. Unfortunately, comparisons between these simulations and the current simulations can prove extremely difficult. A variety of modified force fields and simulation parameters have been used in the literature, which can lead to results that are significantly different. In our previous study of amino acid side chain analogs,[4] we reviewed a number of studies, and we refer the reader to this paper for a discussion of those comparisons. In this paper we will review several comparisons of free energies of similar molecules that have been published in the last year, since the publication of the previous study.

One important study is the development of a new version of the GROMOS force field, parametrized to fit the free energy of solvation to within 0.2 kcal/mol error,[51] which is more accurate than OPLS-AA with any of the original published water models. These parameters were developed by first fitting the charges and Lennard-Jones parameters to capture pure liquid properties and solvation free energies in cy-

TABLE VII. ϵ, σ, and pure liquid properties for TIP3P, TIP3P-MOD, and novel Lennard-Jones-modified waters M20-M25. All model parameters except for ϵ and σ are the same as TIP3P, as shown in Table II. Uncertainties in density are $\leqslant$0.0002 g cm^{-3}, for the heat of vaporization are $\leqslant$0.003 kcal/mol, for the free energies are $\leqslant$0.02 kcal/mol, and for the self-diffusion constants are $\leqslant 0.04\times10^{-5}$ cm^2 s^{-1}. Computation of the heat of vaporization was done according to previously published methods (Ref. 29), but without the polarization correction.

	TIP3P	TIP3P-MOD	M20	M21	M22	M23	M24	M25
ϵ(kcal/mol)	0.1521	0.1900	0.20	0.21	0.22	0.23	0.24	0.25
σ(Å)	3.15061	3.12171	3.120	3.118	3.115	3.113	3.111	3.110
Density	0.9859	0.9998	0.9976	0.9963	0.9976	0.9970	0.9976	0.9969
Diffusion	5.56	6.26	6.15	6.57	6.70	6.74	6.64	6.88
ΔH_{vap}(kcal/mol)	10.091	10.109	10.044	9.998	9.983	9.961	9.948	9.928
ΔG(kcal/mol)	−6.16	−6.17	−6.12	−6.14	−6.12	−6.13	−6.12	−6.13

clohexane, and then fitting new charges to the solvation free energies in water. The fact that reparametrizing the solutes instead of the solvent, as in this study, achieved similarly improved results reinforces the idea that there is sufficient flexibility in the parameters to fit to new data. However, the fact that the researchers were unable to use the same charges for both water and cyclohexane solvation also demonstrates the limitations of fixed-charge models and may point to the need for future polarizable force fields. Changing the solvent parameters at the same time will add some flexibility as demonstrated in this study, but is not at all clear if it would be sufficient.

Another recent study evaluated the free energies of solvation for OPLS-AA in SPC and TIP4P water.[52] We found that most energies presented in this study were different from ours by between 0.0 and 2.0 standard deviations of the mutual uncertainty, with a tendency to be slightly less favorable for solvation. In general, this is good agreement, given the difference in simulation conditions. The uncertainties of this study were approximately five times the uncertainty of the current study, so this is most likely a result of differing conditions, and perhaps a slight underestimation of the uncertainties. One other study, of Deng and Roux,[53] found agreement to our previous study of amino acid side chain analogs[4] in almost all cases to within 0.8 kcal/mol on average for both the CHARMM and OPLS-AA parameter sets. However, this methods were not entirely comparable, as a limited number of solvent molecules were simulated, with an effective solvent boundary potential to account for the bulk water, and are expected to give rise to differences of the magnitude noted in the study.

One important question is whether the experimental free energy solvation is actually the best number for comparison to computational free energy in order to asses the model validity. For example, we do not necessarily expect classical fixed-charge models to be able to correctly model both the condensed phase and vapor phase behavior, because of the change in polarization that we would expect to occur. Therefore, it might be preferable to compare the simulation results to the experimental results with a correction added to account for the change in polarization when transferring to the vapor state, as the repolarization in the vapor phase is not important to the solute-solvent interactions. This decision was made in the parametrization of SPC/E and TIP4P-Ew.[29,54] As can be seen in Table II, these models yield polarization-corrected enthalpies of hydration that are much closer to experiment than the uncorrected enthalpy of hydration.

This correction for the enthalpy could also be applied to the free energy. Unfortunately, for arbitrary small molecules, these gas phase dipole moments have not been well studied. For the relatively nonpolar molecules, this change is very small, and the dipole readjustment terms are negligible. However, for the larger and more polar molecules in this study, experimental numbers do not exist, or are extremely imprecise, and a correction is therefore not possible. If this method were pursued, it might therefore be best to calculate the gas dipole moments and polarizabilities from high level quantum calculations of a few representative configurations. It is still unclear if the direct or polarization-corrected energies are the proper ones to use, as it depends somewhat on the philosophy of the parametrization, so this subject does bear closer scrutiny in the future.

How accurate do we expect the hydration free energies computed using these parameter sets to be? OPLS-AA and other current force fields (with the exception of GROMOS 53A5 and 53A6,[51] published in mid-2004) were not parametrized to reproduce free energies of hydration because, until recently, it was too computationally intensive to so do in an iterative manner with sufficient precision. Instead, Lennard-Jones terms were generally derived to reproduce bulk properties using condensed phase simulations of pure liquids, and charges were generated from quantum mechanical calculations in isolated (gas phase) molecules, with some constraints and scaling factors used to account for the solvent environment.[6–8] In this light, there is no *a priori* reason to expect that the free energies correspond particularly well to experimental results.

Given the high precision free energy values obtained in this study, we must ask the unusual question if the experimental data have sufficient precision to provide accurate comparison. As described in the previous study,[4] a metaanalysis of a number of experimental results indicated that the experimental values are between 0.03 and 0.07 kcal/mol for each molecule. The level of uncertainty in this paper is then approximately equal to the experimental uncertainty.

A companion problem to ask is whether simulations this precise are even necessary. Certainly the high level of precision obtained in this paper will not be necessary for all uses of free energy calculations. For many predictive purposes, such as ligand binding, we only are interested in free energies of binding correct to 0.5–1.0 kcal/mol. However, we note that these results presented are for small molecules, and the error increases with size and molecular complexity. For larger molecules, many fewer uncorrelated measurements are accessible, and more efficient methods and extensive computational power as presented in this paper will be necessary to achieve precision levels necessary for ligand binding.

Additionally, higher precisions are much more important for the parametrization of models than for other predictions or experiments. Assuming Gaussian error distribution, reporting a standard deviation implies that $\approx$68% of the measurements would fall within one standard deviation of the answer. However, for parametrization, as the results would be used over and over again, we would prefer a much greater confidence level when at all possible. Not all observables have the same dependence on the parametrizations; some might be more sensitive to changes in parameters that the ones chosen for parametrization. In the process of parametrization, we would therefore suggest that the computational experiments be as precise as the experimental numbers where at all possible or plausible. True differences of the models from experiment sometimes become masked by large uncertainties, making model improvement difficult. Uncertainties of this magnitude may rapidly accumulate for simulations of larger numbers of atoms.

Most prior simulations computed only relative free ener-

gies of hydration, or computed absolute free energies by summing relative free energies of transformation relative to a single absolute energy calculation. We emphasize that the calculated free energies in this paper are all *absolute* free energies, using simulations in which entire molecules are completely decoupled from the solvent. Complete solvent-solute decoupling, rather than mutating between two solutes, entails a significantly larger change in the Hamiltonian, and is thus harder to do with small error. It requires greater care to ensure that the nonbonded interaction terms change in a well-behaved manner with respect to λ, and requires more intermediate steps from starting to final Hamiltonian. However, it is a significantly more flexible method, not requiring a series of structurally similar molecules.

The pathway that is chosen is vital to ensure the best overlap possible between intermediate states. This paper is not intended as a comprehensive guide to such parametrizations. However, we do note that the procedure of decoupling of electrostatic and Van der Waals interactions and the use of the soft-core potential in atomic site introduction/removal presented in Eq. (4) results in relatively small variance of the intermediate free energies, making possible the level of precision presented here.

It is also very desirable to run multiple copies of each simulation, starting from initial conditions that are as uncorrelated as possible. If the variation between the multiple runs does not agree with the statistical uncertainty estimates, then insufficient sampling has been performed. This establishes a useful necessary but not sufficient condition on the sampling required for a computed free energy to converge. In order to strictly guarantee that sufficient sampling has taken place, one must monitor all the relevant degrees of freedom and ensure they are extensively sampling their allowed phase space. For small molecules, this is relatively straightforward, as detailed in our previous amino acid side chain study.[4] However, for even small proteins, this becomes very difficult, and it is as yet unclear how this can be done in a general case.

In our previous study, we described how such free energy computations, suitable for parametrization of force fields, are now accessible to most researchers. The advances in free energy methodology presented in this paper make such computations much more possible. For example, in the previous paper, we estimated the total computer time required was ≈140 CPU years on a 1 GHz processor benchmark computer, or roughly equivalent to 50 CPU years on the 2.8 GHz processors used in this study. Here, however, we were able to perform approximately four times as many free energy computations on the same size systems in only 8.4 CPU years, performed entirely on a local cluster, for an increased efficiency ratio of more than 20 times. Part of this speedup is due to using an efficient code like GROMACS. But the majority of this speedup is from the use of a smoother pathway with respect to λ, and using the Bennett acceptance ratio to extract energies between states with relatively low overlap much more efficiently. Even greater efficiency could be obtain for a small decrease in precision by running somewhat decreased length of simulations, combined with slightly fewer intermediate states of λ, or a slightly smaller water box.

We have not attempted to compute the entire matrix of water models vs force fields, as it is sufficiently clear that current generations of fixed-charge force fields in general do not fit solute-solvent interactions, and it would be preferable to update them to get this data closer. GROMOS 53a5 and 53a6 represent a promising start.[51] But it has been noted that the best effective pair-potentials for some situations are not necessarily the best for other situations. For example, alkane parameters that have been optimized for gas-phase and neat liquid simulations can be worse than earlier models in the ability to compute accurate free energies of hydration of solutes,[55,56] and in the parametrization of GROMOS it proved impossible to fit accurately to free energies of solvation in both water and cyclohexane.[51] It seems likely that more complicated force fields may be necessary for transferability to multiple chemical environments. For example, polarizable force fields that have the ability to incorporate some many-body effects and to respond to different environments are being developed by many different groups.[57–62]

We do not intend to propose the M20-M25 water models as ideal water models for broad adoption by the simulation community. The data shown here demonstrate that these new models would likely serve as well as TIP3P or SPC for many purposes. However, we note that they have only been specifically parametrized to get the density and average free energies of hydration correct, which might limit their applicability. There are a range of other important observables to calculate correctly, such as the heat capacity, isothermal compressibility, self-diffusion constant, and radial and angular distribution functions, and ideally the parametrization would be done over a range of temperatures that are relevant to biological simulations. For example, the O–O radial distribution functions obtained in this study of TIP3P-MOD and M20-25 have a shallow minimum close to 4.5 Å, significantly further out from experiment than TIP3P, at 3.5 Å. TIP4P-Ew, on the other hand, which compares extremely well with experimental x-ray scattering data,[29] has a somewhat deeper minimum at 3.3 Å, closer to experimental observation. On the other hand, these new models also appear to give a closer agreement with water-amino acid pair correlation functions in simulations of proteins.[63] These models could therefore still be useful for free energy calculations in solvent-solute systems until water models more comprehensively parametrized using solvent-solute data are developed.

A more preferable strategy would be to follow the philosophy used in the development of TIP4P-Ew, and to quantitatively parameterize various pure liquid properties and solvent/solute properties simultaneously, using either free energy calculations or simulations of mixtures. This study and the other recent studies[4,43,51,64] have demonstrated that current fixed-charge biological force fields work much better than they have any right to, but still have many flaws. A clear next step is the inclusion of polarizability,[61,64] but improving current fixed-charge force fields may solve many problems that are currently extant before needing to invoke polarizability.

Our results have some implications for simulations using

standard water models such as the ones studied in this paper in estimating free energies. We cannot expect that calculations performed on more complicated systems, such as those used to compute ligand-protein binding free energies, will be any more accurate than the hydration free energies (or at least the relative hydration free energies) of the respective small constituents. Calculations may be more accurate than this limit, but that is dependent on fortuitous compensating errors between ligand-water and ligand-protein interactions. Our results may also have implications for the utility of these force fields for predictions of protein structure, stability, and dynamics. However, it is also true that many computed observables may be relatively insensitive to the details of the potentials that produced the differences from experiment noted in this paper.

V. CONCLUSIONS

We have examined the accuracy of a range of common water models used for protein simulation for their solute/solvent properties, calculating the free energy of hydration of 15 amino acid side chain analogs derived from the OPLS-AA parameter set in the TIP3P, TIP4P, SPC, SPC/E, TIP3P-MOD, and TIP4P-Ew water models. We achieve a high degree of statistical precision in our simulations, obtaining uncertainties for the free energy of hydration of 0.02–0.06 kcal/mol, equivalent to our previous study and to that obtained in experimental hydration free energy measurements of the same molecules. This degree of precision is at least an order of magnitude better than most previous studies, and if the model under study is close enough to experiment, this degree of precision is sufficient for virtually all practical uses.

All previously published models evaluated had free energies of hydration that were less favorable to hydration than experiment. TIP3P-MOD, a rarely used model designed to give improved free energy of hydration for methane, gave uniformly the closest match to experiment. All other models were worse than TIP3P, the model used in our previous high precision study.[4] Importantly, the ability to accurately model pure water properties does not necessarily predict ability to predict solute/solvent behavior.

We also evaluated the free energies of a number of new modifications of TIP3P designed as a proof of concept to show that it is possible to obtain much better solute/solvent free energy behavior without significantly changing pure liquid properties. We are able to decrease the average error in the free energy of solvation to zero while reducing the root mean square error below that of any of the published water models, with liquid water properties remaining almost constant. This demonstrates that there is still both significant room for improvement within current fixed-charge biomolecular force fields and significant flexibility within current functional forms for such improvements to be made.

Because of recent developments we have made in computational efficiency of free energy calculations, we were able to perform simulations that previously required large scale distributed computing on smaller local cluster. We performed four times as much computational work in a little more than a fifth of the computer time as a similar study a year ago, making these sorts of free energy computations amenable to running on a local cluster of PCs. This study demonstrates that the pathway to accurate and reliable methods to compute the free energies of interaction of ligand/macromolecular models may be clearer than previously thought.

ACKNOWLEDGMENTS

M.R.S. acknowledges the support of the Fannie and John Hertz Foundation. This work was supported in part by the Center on Polymer Interfaces and Macromolecular Assemblies (CPIMA) as part of the NSF Materials Science and Engineering Center program under Grant No. DMR 9808677. The authors would like to thank Bill Swope, Jed Pitera, and Hans Horn at IBM for comments, insights, and discussions, Erik Lindhal for help in modifying GROMACS, Teresa Head-Gordon and Margaret Johnson for providing scattering data for water and interpreting the computational structural data, and Guha Jayachandran for help in running on the Bio-X cluster.

[1] C. Chipot and D. A. Pearlman, Mol. Simul. **28**, 1 (2002), see also the references in this review.
[2] P. A. Kollman, Chem. Rev. (Washington, D.C.) **7**, 2395 (93).
[3] W. L. Jorgensen, Science **303**, 1813 (2004).
[4] M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, J. Chem. Phys. **119**, 5740 (2003).
[5] M. Shirts and V. S. Pande, Science **290**, 1903 (2000).
[6] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).
[7] J. A. D. MacKerell, D. Bashford, M. Bellot *et al.*, J. Phys. Chem. B **102**, 3586 (1998).
[8] W. D. Cornell, P. Cieplak, C. I. Bayly *et al.*, J. Am. Chem. Soc. **117**, 5179 (1995).
[9] R. Wolfenden, L. Andersson, P. M. Cullis, and C. C. B. Southgate, Biochemistry **20**, 849 (1981).
[10] A. Ben-Naim and Y. Marcus, J. Chem. Phys. **81**, 2016 (1984).
[11] S. Cabani, P. Gianni, V. Mollica, and L. Lepori, J. Solution Chem. **10**, 563 (1981).
[12] J. Hine and P. K. Mookerjee, J. Org. Chem. **40**, 292 (1975).
[13] M. H. Abraham, G. S. Whiting, R. Fuchs, and E. J. Chambers, J. Chem. Soc., Perkin Trans. 2 **2**, 291 (1990).
[14] V. N. Viswanadhan, A. K. Ghose, U. C. Singh, and J. J. Wendoloski, J. Chem. Inf. Comput. Sci. **39**, 405 (1999).
[15] G. A. Kaminski, R. A. Friesner, J. Rives, and W. L. Jorgensen, J. Phys. Chem. B **105**, 6474 (2001).
[16] J. S. Bader and D. Chandler, J. Phys. Chem. **96**, 6424 (1992).
[17] P. J. Steinbach and B. R. Brooks, J. Comput. Chem. **15**, 667 (1994).
[18] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).
[19] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Model. **7**, 306 (2001).
[20] H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).
[21] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, J. Chem. Phys. **76**, 637 (1982).
[22] S. Miyamoto and P. A. Kollman, J. Comput. Chem. **13**, 952 (1992).
[23] C. H. Bennett, J. Comp. Physiol. **22**, 245 (1976).
[24] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. **91**, 140601 (2003).
[25] A. Radzicka and R. Wolfenden, Biochemistry **27**, 1664 (1988).
[26] See EPAPS Document No. E-JCPSA6-122-521515 for supplementary files described in the text. A direct link to the document may be found in the online article's reference section. The document may also be reached via the EPAPS homepage (http://www.aip.org/pubservers/epaps.html) or from ftp.aip.org in the directory /epaps/. See the EPAPS homepage for more information.
[27] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[28]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, P. A. K. J. Hermans, R. Dixon, W. Cornell, T. Fox, C. Chipot, and A. Pohorille, in *Intermolecular Forces*, edited by A. Wilkinson, P. Weiner, and W. F. van Gunsteren (Reidel, Dordrecht, 1981), Vol. 3, p. 331.
[29]H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, J. Chem. Phys. **120**, 9665 (2004).
[30]Y. Sun and P. A. Kollman, J. Comput. Chem. **16**, 1164 (1995).
[31]H. C. Andersen, J. Comp. Physiol. **52**, 24 (1983).
[32]U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[33]M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).
[34]In the previous paper, we found that the water degrees of freedom were all sampled 50–100 times over the course of 1.0 ns simulations. The torsional degrees of freedom were not always well sampled in 1.0 ns—the isoleucine side chain analog, the worse case, only made one transition in this time. However, both configurations had approximately the same free energy, implying that this lack of torsional exploration does not affect the free energy computations presented here, although it implies computing free energies of molecules with more complex torsional barriers may be much harder.
[35]M. R. Shirts and V. S. Pande, J. Chem. Phys. (in press).
[36]T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, Chem. Phys. Lett. **222**, 529 (1994).
[37]A. Dejaegere and M. Karplus, J. Phys. Chem. **100**, 11148 (1996).
[38]J. W. Pitera and W. F. Van Gunsteren, Mol. Simul. **28**, 45 (2002).
[39]J. T. Wescott, L. R. Fisher, and S. Hanna, J. Chem. Phys. **116**, 2361 (2002).
[40]W. L. Jorgensen and C. Jenson, J. Comput. Chem. **19**, 1179 (1998).
[41]P. McCaldon and P. Argos, Proteins **4**, 99 (1988).
[42]W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996).
[43]D. J. Price and C. L. Brooks, J. Comput. Chem. **23**, 1045 (2002).
[44]M. Karplus and J. A. McCammon, Nat. Struct. Biol. **9**, 646 (2002).
[45]T. Hansson, C. Oostenbrink, and W. F. Van Gunsteren, Curr. Opin. Struct. Biol. **12**, 190 (2002).
[46]B. Chen and J. I. Siepmann, J. Phys. Chem. B **103**, 5370 (1999).
[47]B. Chen, J. J. Potoff, and J. I. Siepmann, J. Phys. Chem. B **105**, 3093 (2001).
[48]C. D. Wick, M. G. Martin, and J. I. Siepmann, J. Phys. Chem. B **104**, 8008 (2000).
[49]S. K. Nath, F. A. Escobedo, and J. J. de Pablo, J. Chem. Phys. **108**, 9905 (1998).
[50]J. R. Errington and A. Z. Panagiotopoulos, J. Phys. Chem. B **103**, 6314 (1999).
[51]C. Oostenbrink, A. Villa, A. E. Mark, and W. F. V. Gunsteren, J. Comput. Chem. **25**, 1656 (2004).
[52]J. L. Maccallum and D. P. Tieleman, J. Comput. Chem. **24**, 1930 (2003).
[53]Y. Deng and B. Roux, J. Chem. Phys. **108**, 16567 (2004).
[54]H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[55]D. Yin and A. D. MacKerell, J. Comput. Chem. **19**, 334 (1998).
[56]G. Kaminski, E. M. Duffy, T. Matsui, and W. L. Jorgensen, J. Phys. Chem. **98**, 13077 (1994).
[57]A. Toukmaji, C. Sagui, J. Board, and T. Darden, J. Chem. Phys. **113**, 10913 (2000).
[58]J. L. Banks, G. A. Kaminski, R. H. Zhou, D. T. Mainz, B. J. Berne, and R. A. Friesner, J. Chem. Phys. **110**, 741 (1999).
[59]G. A. Kaminski, H. A. Stern, B. J. Berne, and R. A. Friesner, J. Phys. Chem. A **108**, 621 (2004).
[60]S. Patel and C. L. Brooks, J. Comput. Chem. **25**, 1 (2004).
[61]J. W. Ponder and D. A. Case, in Adv. Protein Chem. **66**, 27 (2003).
[62]P. Y. Ren and J. W. Ponder, J. Comput. Chem. **23**, 1497 (2002).
[63]Y. M. Rhee and V. S. Pande, Chem. Phys. (submitted)
[64]A. D. Mackerell, J. Comput. Chem. **25**, 1584 (2004).