Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins

Michael R. Shirts
Department of Chemistry, Stanford University, Stanford, California 94305-5080

Jed W. Pitera and William C. Swope
IBM Almaden Research Center, San Jose, California 95120-6099

Vijay S. Pande
Department of Chemistry, Stanford University, Stanford, California 94305-5080

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). We use large-scale distributed computing to access sufficient computational resources to extensively sample molecular systems and thus reduce statistical uncertainty of measured free energies. In order to examine the accuracy of a range of common models used for protein simulation, we calculate the free energy of hydration of 15 amino acid side chain analogs derived from recent versions of the OPLS-AA, CHARMM, and AMBER parameter sets in TIP3P water using thermodynamic integration. We achieve a high degree of statistical precision in our simulations, obtaining uncertainties for the free energy of hydration of 0.02–0.05 kcal/mol, which are in general an order of magnitude smaller than those found in other studies. Notably, this level of precision is comparable to that obtained in experimental hydration free energy measurements of the same molecules. Root mean square differences from experiment over the set of molecules examined using AMBER-, CHARMM-, and OPLS-AA-derived parameters were 1.35 kcal/mol, 1.31 kcal/mol, and 0.85 kcal/mol, respectively. Under the simulation conditions used, these force fields tend to uniformly underestimate solubility of all the side chain analogs. The relative free energies of hydration between amino acid side chain analogs were closer to experiment but still exhibited significant deviations. Although extensive computational resources may be needed for large numbers of molecules, sufficient computational resources to calculate precise free energy calculations for small molecules are accessible to most researchers. © *2003 American Institute of Physics.* [DOI: 10.1063/1.1587119]

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] However, the accuracy of current calculations greatly limits their applicability, especially for predictions of experimental observables such as drug binding affinities. For example, the threshold for pharmaceutically useful predictions is an accuracy of approximately 0.5–1.0 kcal/mol, implying a binding constant correct to within an order of magnitude of the experimental value. Obtaining such precision in a repeatable and consistent manner is an unsolved problem.[1] Also, it is unclear to what extent uncertainties presented in many published studies truly capture the statistical uncertainty inherent in free energy simulations.[2]

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 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 differing answers. In general, insufficient attention has been given to separating the differences between free energies computed from simulation and those derived experimentally into contributions from inaccuracies in the model and the lack of precision in the measurement of ensemble averages. Only if sufficient precision is obtained in a statistically well-defined manner is it possible to design models that are sufficiently accurate for the application at hand.

A. Difficulties in obtaining high-precision measurements

In order for the free energy of a physical or chemical process to be correctly computed by simulation techniques, thermally relevant configurations of the system must be thor-

TABLE I. Correspondence between amino acids and the amino acid side chain analogs used in this study.

Ala	methane	Val	propane	Leu	iso-butane
Ile	*n*-butane	Ser	methanol	Thr	ethanol
Phe	toluene	Tyr	*p*-cresol	Cys	methanethiol
Met	methyl ethyl sulfide	Asn	acetamide	Gln	propionamide
Trp	3-methylindole	Hid	4-methylimidazole	Hie	4-methylimidazole

oughly explored. This is a difficult challenge for the simulation of large biological systems with many degrees of freedom and long dynamical correlation times. Long, "brute-force" molecular dynamics simulations, typically on the nanosecond time scale, are currently insufficient to capture the long time scales of typical biological events, which are frequently on the microsecond or millisecond time scale. Without proper statistical uncertainty analysis, reported uncertainties can be significantly smaller than the spread in values that would likely be generated under repeated independent executions of the simulation. One common practice in reporting computed free energies is to assign as an uncertainty the hysteresis in the free energy as the "mutation" is performed forward and backward. This is not actually a valid uncertainty estimate, as it is related to systematic error, not the statistical uncertainty. Moreover, it is frequently smaller than the actual statistical uncertainty of the measurement.[3,4] Additionally, observables from molecular dynamics and Monte Carlo simulations can falsely appear to have converged if these observables are coupled to degrees of freedom with long correlation times.[5] There is a need for rigorous analysis in the determination and reporting of the uncertainty due to limited sampling.

B. Accuracy of models

Current generation biomolecular force fields such as AMBER($ff94$),[6] CHARMM22,[7] GROMOS,[8] and OPLS-AA[9] are in many cases successful in capturing the qualitative behavior of protein structure and dynamics.[10–12] These force fields for proteins were primarily built and parameterized to match small molecule data, both experimental and theoretical. 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.[6–9] However, computational limitations 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. Other molecular force fields such as TraPPE,[13–15] NERD,[16] and GIBBS99[17] 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.

It must also be kept in mind when performing a simulation that simply specifying a set of force field parameters is not sufficient to define a computational model. Anything that affects the Hamiltonian changes the model, and thus will affect the end results. For example, there are a large number of choices of simulation methodology, such as choice of boundary conditions, truncation or tapering of nonbonded interactions, and constraint of bonds or other degrees of freedom, all of which can change the Hamiltonian. Each of the force fields mentioned above was parameterized using certain assumptions about simulation procedures and parameters. Using different simulation conditions than those used to parameterize the force field has consequences for the observed results.

C. Purposes of the present study

In this paper, we have attempted to separate these two sources of deviation from experiment for a simple test system, the hydration of neutral amino acid side chain analogs. The correspondence between amino acids and amino acid side chain analogs is shown in Table I. We chose this system for a variety of reasons. First, experimental data are available for direct comparison.[18–23] Second, these molecules are simple enough that we expect to be able to obtain good statistical sampling, thus allowing us to obtain a lower bound on the amount of the sampling necessary for larger, biological systems. Third, a variety of well-defined, extensively tested, and widely-used potential sets are available for these molecules—in this paper, we use AMBER($ff94$), CHARMM22, and OPLS-AA.[24] 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.

There are other possible choices of a test system, such as partition coefficients of blocked amino acids between water and a hydrophobic solvent or the free energy of hydration of whole or blocked amino acids, but these alternate test systems have serious disadvantages. Partition coefficients are likely better measures of force field utility for proteins than direct hydration, and extensive experimental data is available. Force field parameterizations of nonaqueous solvents (such as 1-octanol, cyclohexane, chloroform, and other solvents for which experiments have been performed) have been proposed and partially explored.[25–28,14] However, models for these solvents are less developed than are the various water models, especially for solvent–solute simulations, and very few simulations have used protein force field parameters in conjunction with these solvents. Since the solvent model is just as important as the solute model in computing the free energy of solvation, we believe that computation of partition coefficients would be more a test of these less developed nonaqueous solvent models than of the more mature solute models.[29] Additionally, published protein/solute parameters have generally been developed for use with high dielectric solvents like water and may not be as accurate in

lower dielectric solvents. In the case of whole amino acid solvation, experimental results for solvation of the entire amino acids are very difficult to obtain, as isolated amino acids are extremely soluble in water in zwitterionic or even blocked forms. As additional support for computing the free energies of hydration as a test of protein force fields, the free energy of hydration of these amino acid side chain analogs is highly correlated to such measures as Chothia's criterion of hydrophobicity[30] (with $R = 0.90$) as measured from the distribution of amino acids within globular proteins.[18,31]

When evaluating models, one is torn between using more recent simulation methods with the best proven scientific validity, the methods used for the actual parameterization 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 parameterized 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 parameterization of the model. However, truncation schemes for electrostatic interactions have been shown to lead to qualitatively incorrect results,[32] as compared with more sophisticated and increasingly more common methods such as Ewald summation or the use of a reaction field. Our goal here is to compare models and we choose to be faithful, where possible, to the methods used in the original parameterizations. Additionally, despite their limitations, cutoffs are still frequently used in simulations of proteins.

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 is not the most computationally efficient in general. But it is sufficient for the problem at hand, as well as being simple to implement, frequently utilized, well-studied, and well-understood.

Some studies have examined the free energy of hydration of a few representative molecules using standard force fields in explicit water simulations.[33–38] However, there has not been a comprehensive attempt to compare common force fields for their accuracy in predicting the hydration free energies for a set of molecules relevant to protein simulations employing explicit representations of water. One reason has been a historical problem of adapting force fields to a diversity of molecular dynamics software packages. The code used in this study is a modified version of TINKER,[39] which supports a variety of common force fields. Another major reason for this lack of extensive comparison is a lack of sufficient computational resources to evaluate free energies of large numbers of molecules with sufficient precision. Using the power of distributed computing,[40] we are able to access sufficient computer time for such large computational tasks.

These results can serve as one possible test of the applicability of current protein models studied with explicit representation of water molecules. The present work could guide the development of more accurate force fields. It demonstrates that it is now easier to incorporate very important yet computationally intensive observables such as hydration free energy into force field parameterization efforts. Free energy simulations are useful parameter verification methods as they quantify the effects of both the enthalpy and entropy of molecular interactions.

II. METHODS

A. Computational methods

The Folding@Home distributed computing infrastructure[40] was used to perform the free energy calculations presented in this paper. Folding@Home (at http://folding.stanford.edu) is a collection of approximately 90 000 volunteered computers across the world running the Folding@Home client program.[41] Volunteered computers automatically download calculations to be performed and return the processed results to servers at Stanford University. Folding@Home has previously been used to perform massively parallel calculations successfully simulating the folding kinetics of small peptides and proteins[42–46] and to design ensembles of new protein sequences which fit to specified structures.[47]

The main data set alone (3 force field parameter sets×15 amino acid side chain analogs×5 trials×1.2 ns×61 λ values) represents approximately 140 CPU years on a 1 GHz Celeron, and all of the simulations together represent closer to 200 CPU years. The majority ($>95\%$) of data was collected in <2 months, approximately equivalent to an 800 processor cluster of similar machines working uninterrupted for the same time period. These calculations used only a small fraction of Folding@Home's total capacity at any given time.

The simulations were performed using a modified version of the molecular dynamics suite TINKER (v3.8).[39] Our modifications included adding Andersen pressure and temperature control[48] and a variety of methods for free energy computation using thermodynamic integration (such as "soft-core" λ dependence[49]), as well as adding some performance optimizations.

Parameters for the amino acid side chain analogs were generated from the versions of AMBER(ff94),[6] CHARMM22,[7] and OPLS-AA(1996),[9] force fields included in the TINKER v3.8 distribution. CHARMM22 is the most recent CHARMM version for proteins (CHARMM27 parameters are only different for nucleic acids and lipids). ff94 and ff96 in Amber differ only in peptide backbone torsion angle parameters,[50] and ff98 did not change any protein parameters from earlier parameter sets, so these differences are not relevant here. OPLS-AA[9] parameters used date from October 1997 and are similar to the most recent OPLS-AA/L,[51] only differing somewhat in the torsion angle and the sulfur Lennard-Jones parameters. In this paper, we will refer to these parameter sets as CHARMM, AMBER, and OPLS-AA, recognizing that there are variants of each. We studied all the neutral amino acid side chains 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 pH in our computations.
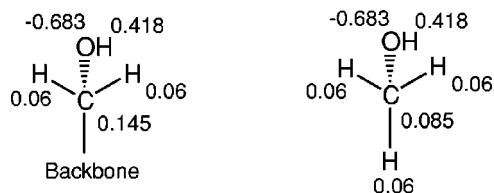
FIG. 1. Differences in partial charges between (left) the OPLS-AA parameter set serine side chain and (right) the OPLS-AA-derived parameters used for the side chain analog in this study. Only the β-carbon changes in charge.

A finite-ranged treatment for nonbonded interactions is ultimately inappropriate for the inherently long-ranged Coulombic interactions of charged species. Complex corrections to free energies from finite-ranged simulations are required to recover approximations to the full electrostatic free energy of charged species, and errors in these approximations are somewhat uncontrolled.[52,53] Experimental data does exist for the free energy of hydration of these charged amino acid side chain analogs at alternate pH values where they are neutral,[18] but not all of the force fields have parameters defined for these neutral species.

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 resulting type and connectivity, these were used. Otherwise, the bonded terms originally assigned that atom were retained. TINKER input parameter and coordinate files for all the amino acid side chain analogs in all parameter sets used are provided as Supplementary Material.[54]

In the case of the AMBER parameter set, additional changes were necessary. CHARMM and OPLS-AA are formulated to have electostatically neutral side chain moieties, while AMBER charges are only neutral for the entire residue. To create a neutral side chain analog with the AMBER parameters, we followed the procedure for adding an additional hydrogen described in the previous paragraph, and then added the charge on the β-carbon necessary to make the overall side chain analog neutral. The average of this last correction to the partial charges on the β-carbon across the AMBER parameter set was -0.0725, with root-mean-square difference 0.021, and maximum difference of -0.1197 (for the leucine analog). These changes result in a slightly different dipole moment for these side chain analog atoms than might occur if the same atoms were in an intact residue, or alternatively if the charges were rederived with *ab initio* calculations, and will therefore slightly change the free energy of solvation. To understand the effect in hydration free energy that this dipole moment difference would create, we performed a separate simulation on the Leu analog in which the correction to the charges to achieve neutrality was per-

formed by adjusting the charges on the three hydrogens attached to the β-carbon instead of the charges on the β-carbon itself, with the assumption that most other choices would fall between these extremes. The difference in hydration free energies obtained was of the size of the uncertainty (as discussed at length in the Results). This demonstrates that differences between this simulation and simulations with other possible variants of the AMBER parameters are likely to be statistically insignificant.

The original TIP3P water model[55] 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, like the CHARMM TIP3P mentioned above) is probably the most common water model used in explicit water simulations, and was selected to best capture typical current methods, even though TIP4P and TIP5P have been shown to reproduce some properties of water more effectively than TIP3P.[56]

Thermodynamic integration was used to compute the hydration free energy (full details below), with molecular dynamics used to compute the ensemble average of $dH/d\lambda$ at a number of values of λ. For each value of λ, simulations using a 2 fs time step were equilibrated for 200 ps, followed by 1.0 ns data collection. Since each simulation was carried out in fivefold replica, a total of 5 ns of simulation time was collected at each value of λ. Integration of the equations of motion were performed using the velocity Verlet algorithm,[57] and all bonds were constrained using RATTLE (Ref. 58) with a distance constraint of 10^{-6} Å. To allow for better load balancing in the Folding@Home distributed computing environment, simulations were divided up into blocks of 100 or 200 ps. The only side effect of this windowing is on the computation of the correlation times of the system, as we discuss below.

Andersen pressure control[48] was used to produce controlled pressure simulations. Andersen pressure control (in conjunction with Andersen temperature control) rigorously implements an isobaric–isothermal (NPT) ensemble. The pressure was calculated using a molecule-based virial, with the molecular centers defined as the molecular center-of-mass. Isotropic scaling of the cubic volume at each step was performed by scaling the coordinates of molecular centers, not the atomic positions.[59–61] A piston mass of 0.0004 g mol^{-1} Å^{-4} was used. 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 mass is unimportant. Periodic boundary conditions were used and the simulation cell was constrained to a cubic shape. At all times, the edge length of the simulation cell was more than twice the range of the finite-ranged potentials used.

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.[48] In the 1.0 ns simulations presented here, velocities were reassigned every 500 steps (1.0 ps), resulting in 1000 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.

Statistical uncertainties, δA, for all observables, A, were calculated by using the variance, $\langle A^2 \rangle_c$, and correlation time of the observable, τ_A, over the entire $T = 1.0$ ns simulation. The uncertainty in the mean for the observable A is then given by the formula $\delta A = \sqrt{\langle A^2 \rangle_c} \sqrt{2 \tau_A / T}$.[57,62] δA corresponds to one standard deviation, σ, and all results are reported as $\langle A \rangle \pm 1 \sigma$.

The normalized fluctuation autocorrelation functions were calculated with a resolution of 20 fs within the window of each 100 or 200 ps block sent to each participating Folding@Home machine. Correlation times for the components (Coulombic or van der Waals) of $dH/d\lambda$ were computed by numerically integrating the autocorrelation functions of each of these observables by the trapezoid rule.[63] It is possible there are correlations comparable to or longer than 100 or 200 ps in the components of $dH/d\lambda$, thus making this windowing inaccurate. To ensure the validity of the windowing procedure and to identify the possible presence of long-time scale correlations, values were sampled every 0.2 ps for the entire 1.0 ns run (for a total of 5000 samples) for 20 of the 45 molecule/force field combinations, and autocorrelation times were computed for these entire 1.0 ns runs for comparison with the windowed correlation times. These times were found to deviate little, as discussed below.

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 combination of force field and amino acid side chain analog, we ran five separate simulations and examined the deviations of the individual results from their mutual average to decide if the simulations had truly converged.

Charge-neutral group-centered tapered cutoffs were used for the nonbonded interactions of both solvent and solute. In all cases, the smallest (in number of atomic sites) possible groups of contiguous atom centers which had zero total partial charge were selected, which resulted in different group definitions depending on the parameter set used. The center of each group was chosen as the atom closest to the geometric center of the group—if two atoms are equally close, the atom with the largest absolute value charge was chosen as the group center. We used combining rules for Lennard-Jones terms and exclusion rules for 1,4 terms appropriate for each force field.

Interactions between atoms using group switching are defined as follows. We first compute the standard Lennard-Jones term (representing the van der Waals interactions) and Coulombic term for each atom pair i and j located at $\mathbf{r}_i$ and $\mathbf{r}_j$,

$$U(r_{ij}) = \frac{q_i q_j}{r_{ij}} + 4 \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \qquad (1)$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$, q_i is the partial charge on atom i, and ϵ_{ij} and σ_{ij} are the standard Lennard-Jones parameters. If the beginning of the switching region (moving outwards from $r_{ij} = 0$) is defined as r_{switch} and the end of the switch region is defined as r_{off}, then the interaction energy between neutral group A, with n_A atomic sites and group center located at r_A, and neutral group B, with n_B atomic sites and group center located at r_B, is defined as

$$U_{AB} = S(r_{AB}) \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} U(r_{ij}), \qquad (2)$$

where $r_{AB} = |\mathbf{r}_A - \mathbf{r}_B|$, and $S(r_{AB})$ is defined as[64]

$$S(r_{AB}) = \begin{cases} 1 & \text{for } 0 < r_{AB} < r_{\text{switch}} \\ \dfrac{(r_{\text{off}}^2 - r_{AB}^2)^2 (r_{\text{off}}^2 + 2 r_{AB}^2 - 3 r_{\text{switch}}^2)}{(r_{\text{off}}^2 - r_{\text{switch}}^2)^3} & \text{for } r_{\text{switch}} < r_{AB} < r_{\text{off}} \\ 0 & \text{for } r_{AB} > r_{\text{off}}, \end{cases} \qquad (3)$$

where A and B are the centers of the groups to which atoms i and atoms j belong, respectively.

This function of r^2 satisfies the conditions $S(r_{\text{off}}) = 0$, $S(r_{\text{switch}}) = 1$, and $dS(r_{\text{off}})/dr = dS(r_{\text{switch}})/dr = 0$. We note than in order to strictly conserve energy, $S(r)$ would also have to satisfy the conditions that $d^2 S(r_{\text{off}})/dr^2 = d^2 S(r_{\text{switch}})/dr^2 = 0$. However, trial constant energy simulations with 900 TIP3P molecules using 2 fs time steps exhibited no statistically meaningful drift in total energy, even for simulations as long as 200 ps. 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 (NPH) 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.[57] For the main data set, all solvent–solute and solute–solute interactions were tapered with $r_{\text{switch}} = 10$ Å and $r_{\text{off}} = 12$ Å, while the solvent–solvent interactions were tapered from $r_{\text{switch}} = 8$ Å to $r_{\text{off}} = 10$ Å.[65]

To prepare the simulation cell, a cubic box of 900 TIP3P water molecules was taken from a larger, previously equilibrated water box and was further equilibrated under NPT conditions described above for an additional 200 ps. The average volume over this equilibration run was 2.700

$\pm 0.003 \times 10^4$ Å^3. We obtained a density of 0.997 ± 0.001 g cm^{-3} for pure TIP3P water, in good agreement with the experimental density of water at 298 K.[66] Our computed density is only slightly different than the 1.001 ± 0.001 g cm^{-3} previously reported.[67] We believe this difference is most likely attributable to different truncation schemes for long-range interactions, as well as an apparent slight system-size dependence of densities and other observables, under periodic boundary conditions.[68] However, the pair distribution functions of the O–O, O–H, and H–H atoms pairs agree well with previously published results,[67] as does the heat capacity (19.9$\pm$0.2 in this study vs 20.0$\pm$0.6 kcal/mol K). 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 1.5 Å of any atomic center of the solute. An extremely short (200 steps) NVT (isochoric–isothermal) simulation with very frequent velocity reassignments was then used to remove highly unfavorable interactions. A further 200 ps NPT equilibration simulation was performed for each value of λ prior to the 1.0 ns of data collection.

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[18,19]

$$\Delta G_{\text{solv}} = kT \ln(P_s / \rho_s^l kT). \tag{4}$$

Free energies of hydration are computed by simulation by decoupling a molecule of the solute (here, amino acid side chain analogs) from the solvent by thermodynamic integration (TI), using the identity,

$$\Delta G_{\text{sim}} = \int_0^1 d\lambda \left\langle \frac{dH(\lambda)}{d\lambda} \right\rangle, \tag{5}$$

where H is parameterized Hamiltonian, and 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. We demonstrate in the Appendix that the relationship between ΔG_{solv} and ΔG_{sim} is given by

$$\Delta G_{\text{solv}} = \Delta G_{\text{sim}} - kT \ln(V^*/V_1), \tag{6}$$

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. We will show below that the term that depends on the ratio of these volumes is negligible with respect to the statistical uncertainty in our calculations of ΔG_{sim}, and can thus be safely ignored. We feel that the rigorous derivation of this

relationship that is covered in the Appendix is particularly useful and general. For example, there are no assumptions that the solution must be dilute or ideal.

We have chosen to use the ''multiconfiguration'' TI formulation,[69] where equilibrium runs at a variety of fixed λ values are performed and then integrated numerically, as opposed to a so-called ''slow growth'' method, with λ changing continuously throughout the simulation, as the latter methodology introduces uncontrolled systematic error into the calculation.[3]

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

$$U(\lambda_C, \lambda_{\text{LJ}}) = U_s + U_w + U_{w-w} + U_{s-w}(\lambda_C, \lambda_{\text{LJ}}), \tag{7}$$

where U_s is the intramolecular potential energy of the solute, U_w is the intramolecular potential energy of the waters (zero for TIP3P water), 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. There are a wide variety of choices for parameterization 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.[1,49,70,71] Simply multiplying the difference between the initial and final Hamiltonians by λ, although analytically correct, has the disadvantage of producing discontinuities in $\langle dH/d\lambda \rangle$ at $r = 0$ and $\lambda = 0$ (and thus extremely high variances in $dH/d\lambda$ anywhere near $r = 0$ and $\lambda = 0$) due to the $1/r$ term, as well as other unphysical behavior like the ''fusion'' of bare charges. This can be avoided by first decoupling the Coulombic interaction while the Lennard-Jones term is still present. However, we are then left with a similar discontinuity in the $1/r^{12}$ term. Although these terms are only repulsive, they still cause such problems such as the ''fence post'' effect, with a near-delta function excluding all other molecules from the region occupied by the ''mutating'' ones for all $\lambda > 0$. Volumes near these atoms are never sampled at any value of $\lambda > 0$, and thus numerical integration of Eq. (5) will not give good results. The discontinuities are also present in the potential, leading to instabilities in the dynamical integration for small values of λ.

We avoid these singularities by using the following expression for the λ-dependent nonbonded interaction energy,

$$U_{s-w}(\lambda_C, \lambda_{\text{LJ}}) = \sum_i \sum_j \lambda_C \frac{q_i q_j}{r_{ij}}$$
$$+ \lambda_{\text{LJ}}^4 4 \epsilon_{ij} \left(\frac{1}{[\alpha_{\text{LJ}}(1-\lambda_{\text{LJ}})^2 + (r_{ij}/\sigma_{ij})^6]^2} \right.$$
$$\left. - \frac{1}{\alpha_{\text{LJ}}(1-\lambda_{\text{LJ}})^2 + (r_{ij}/\sigma_{ij})^6} \right), \tag{8}$$

where the sum i is over all solute atoms, and the sum j is over all solvent atoms. Equation (8) includes the standard Coulombic term with a linear dependence on λ, but also incorporates the ''soft-core'' parameterization of Beutler et al.,[49,71] in which the Lennard-Jones term goes to zero in a well-behaved manner with λ. The additional parameter α_{LJ} is a positive constant which controls this transition, and is equal to 0.5 for all simulations in the study. 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. This procedure avoids the singularities in the potential discussed earlier. It is essential to use this or a similarly well-behaved function in order to get accurate results when atomic sites are being "grown" into or removed from the solvent.

It is tempting to assign the free energy resulting from the turning off of the Coulombic potential as the "electrostatic contribution to" the free energy, and the turning off of the Lennard-Jones term as the "van der Waals contribution to" the free energy of solvation. However, we note that it is impossible to directly assign a Coulombic or van der Waals partitioning of the total free energy, as these are both path-dependent terms.[72] Instead, the free energy of the Coulombic decoupling can more accurately be identified as the "charging" or "discharging" free energy of the solute (though this is not completely correct, as the intramolecular Coulombic interactions are not turned off), and the van der Waals contribution as the free energy of "insertion" of the bonded assembly of uncharged Lennard-Jones spheres that make up the molecule into the simulation cell.

The Coulombic interactions were decoupled by reducing λ_C in steps of 0.05. For the van der Waals 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. Therefore, the following 41 values of λ_{LJ} were used: 0.00, 0.10, 0.20, 0.22, 0.24, 0.26, 0.28, 0.30, 0.32, 0.34, 0.36, 0.38, 0.40, 0.41, 0.42, 0.43, 0.44, 0.45, 0.46, 0.47, 0.48, 0.49, 0.50, 0.52, 0.54, 0.56, 0.58, 0.61, 0.64, 0.67, 0.70, 0.73, 0.76, 0.79, 0.82, 0.85, 0.88, 0.91, 0.94, 0.97, and 1.00. Larger λ spacing gave hydration free energies that deviated from those free energies obtained with the given λ spacing by a significant fraction of the statistical uncertainty (as discussed below). Numeric integration was carried out by the trapezoid rule. Representative curves of $\langle dH/d\lambda \rangle$ versus λ are shown in Fig. 2.

Each set of 61 total λ values for each of the 15 molecules was simulated in fivefold replica. For each the five replicas, the same initial coordinates were used, but each replicated simulation was started with different random number seeds for the velocity reassignment used in the thermal control protocol. Ideally, we would start with something approximating a Boltzmann-weighted, uncorrelated ensemble of starting states, but the correlation times for $dH/d\lambda$ are sufficiently short that the five replicas are essentially uncorrelated after the 200 ps equilibration, with the possible exception of some torsional degrees of freedom discussed later in this study.

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.[62,73] These long range corrections to the solute–solvent interaction were computed after the simulation, and were not included in the calculation of the pressure or the dynamics. Long-range corrections were not applied at any point to the solvent–solvent interactions. The corrections
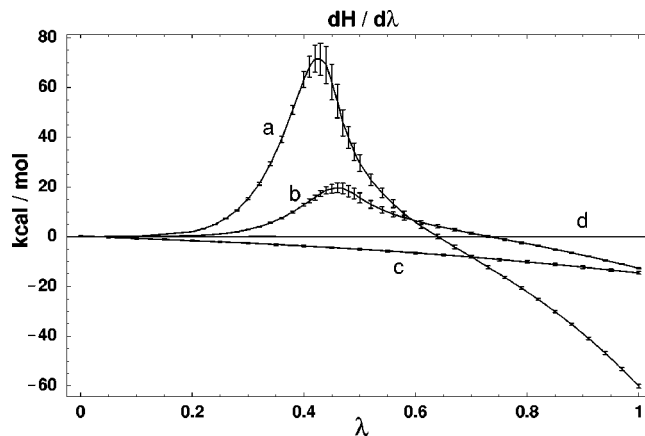

FIG. 2. $\langle dH/d\lambda_C \rangle$ and $\langle dH/d\lambda_{LJ} \rangle$ as a function of λ, for the amino acid side chain analogs of Trp and Ala using the OPLS-AA force field. These two side chain analogs were chosen as they are extrema of the behaviors of the other analogs. The van der Waals component of Trp is curve **a**, the van der Waals component of Ala is curve **b**, the Coulombic component of Trp is curve **c**, and the Coulombic component of Ala is curve **d** (essentially coincident with the *x*-axis). The curves for all of the other amino acid side chain analogs and force field combinations resemble the Trp and Ala OPLS-AA curves for both components, differing mainly in magnitude. The location of the maximum in the van der Waals component differs between the amino acid side chain analogs, but this maximum occurs in the range $\lambda=0.35–0.5$ for all of the molecule/force field sets under study. Uncertainties are multiplied by 5.0 in order to guide the eye, as they would otherwise be difficult to distinguish from the lines themselves.

included both the tapered and group-based nature of the intermolecular interactions. If we assume a TIP3P-like model of water with only one Lennard-Jones term per molecule, then for each solute the long range contribution can be calculated by numerically integrating

$$E_{LRC} = \sum_i 8\pi\rho\epsilon_{ij}$$

$$\times \int_{\theta=0}^{\pi} \int_{r=r_{switch}}^{\infty} \left[\left(\frac{\sigma_{ij}}{\sqrt{r^2 \sin^2\theta + (r\cos\theta - r_0)^2}} \right)^{12} - \left(\frac{\sigma_{ij}}{\sqrt{r^2 \sin^2\theta + (r\cos\theta - r_0)^2}} \right)^6 \right] S(r) r^2 \sin\theta\, d\theta\, dr,$$

(9)

where i runs over the solute atoms, r is the distance from solute atom i, r_{switch} is the radius at which tapering begins, 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 described in Eq. (3). r_0 is the distance from the atom i to the center of the group to which the atom belongs, and is determined by each solute-force field combination. It is treated as a constant for the purposes of computing this correction, as for virtually all of the neutral groups, these distances are fixed by the bond constraints. 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 TIP3P. The pressure would be slightly changed if a similar correction was also included during the simulation, but since

TABLE II. Free energies of hydration for 15 neutral amino acid side chain analogs, for the AMBER(ff94), CHARMM22, and OPLS-AA force fields, in kcal/mol. Each reported value is the average of five trials. LRC is the long range van der Waals correction as computed by Eq. (9). Uncertainties are the averaged uncertainty of the five runs, and represent one standard deviation from the mean ($\pm 1\sigma$). Experimental values are from Wolfenden *et al.* (Ref. 18).

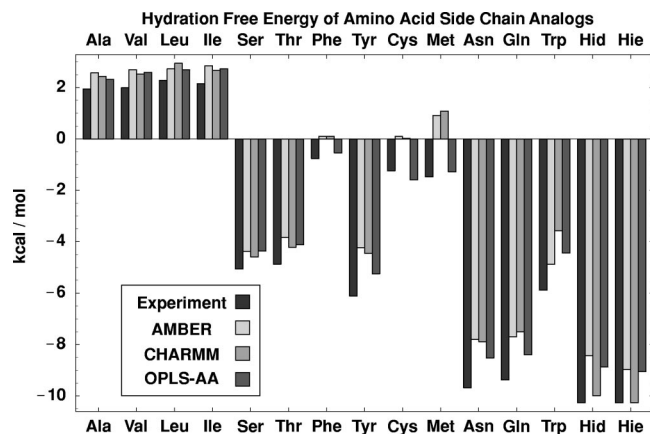| Side chain analog	Coulombic	van der Waals	Total	LRC	Total (w/LRC)	Expt.
			AMBER			
Ala	$-0.00\pm<0.001$	2.69 ± 0.02	2.68 ± 0.02	-0.12	2.57 ± 0.02	1.94
Val	-0.02 ± 0.001	3.02 ± 0.03	3.00 ± 0.03	-0.31	2.69 ± 0.03	1.99
Leu	-0.05 ± 0.001	3.19 ± 0.03	3.13 ± 0.03	-0.41	2.72 ± 0.03	2.28
Ile	-0.00 ± 0.001	3.27 ± 0.03	3.27 ± 0.03	-0.42	2.84 ± 0.03	2.15
Ser	-6.13 ± 0.01	1.93 ± 0.02	-4.20 ± 0.02	-0.17	-4.37 ± 0.02	-5.06
Thr	-5.77 ± 0.01	2.21 ± 0.03	-3.56 ± 0.03	-0.27	-3.83 ± 0.03	-4.88
Phe	-2.35 ± 0.01	3.03 ± 0.04	0.68 ± 0.04	-0.58	0.10 ± 0.04	-0.76
Tyr	-6.19 ± 0.02	2.65 ± 0.04	-3.54 ± 0.04	-0.69	-4.23 ± 0.04	-6.11
Cys	-2.04 ± 0.01	2.37 ± 0.02	0.33 ± 0.02	-0.22	0.11 ± 0.02	-1.24
Met	-1.61 ± 0.01	2.96 ± 0.03	1.35 ± 0.03	-0.44	0.91 ± 0.03	-1.48
Asn	-9.27 ± 0.02	1.78 ± 0.03	-7.49 ± 0.03	-0.31	-7.80 ± 0.03	-9.68
Gln	-9.35 ± 0.02	2.08 ± 0.03	-7.27 ± 0.04	-0.42	-7.69 ± 0.04	-9.38
Trp	-6.39 ± 0.02	2.35 ± 0.04	-4.05 ± 0.05	-0.83	-4.88 ± 0.05	-5.88
Hid	-9.51 ± 0.02	1.56 ± 0.03	-7.95 ± 0.04	-0.48	-8.43 ± 0.04	-10.27
Hie	-10.07 ± 0.02	1.57 ± 0.03	-8.50 ± 0.04	-0.48	-8.98 ± 0.04	-10.27
		Average absolute error w/o LRC: 1.64			rms error with w/o LRC: 1.76	
		Average absolute error with LRC: 1.22			rms error with LRC: 1.35	
			CHARMM			
Ala	$-0.03\pm<0.001$	2.57 ± 0.02	2.55 ± 0.02	-0.11	2.44 ± 0.02	1.94
Val	-0.05 ± 0.001	2.87 ± 0.03	2.82 ± 0.03	-0.29	2.52 ± 0.03	1.99
Leu	-0.04 ± 0.001	3.35 ± 0.03	3.31 ± 0.03	-0.36	2.94 ± 0.03	2.28
Ile	-0.06 ± 0.001	3.12 ± 0.03	3.05 ± 0.03	-0.38	2.67 ± 0.03	2.15
Ser	-6.66 ± 0.01	2.25 ± 0.02	-4.42 ± 0.02	-0.17	-4.59 ± 0.02	-5.06
Thr	-6.49 ± 0.01	2.53 ± 0.03	-3.96 ± 0.03	-0.26	-4.22 ± 0.03	-4.88
Phe	-1.69 ± 0.01	2.31 ± 0.04	0.62 ± 0.04	-0.53	0.09 ± 0.04	-0.76
Tyr	-6.10 ± 0.02	2.22 ± 0.04	-3.87 ± 0.04	-0.58	-4.46 ± 0.04	-6.11
Cys	-1.12 ± 0.004	1.41 ± 0.02	0.30 ± 0.02	-0.27	0.02 ± 0.02	-1.24
Met	-0.59 ± 0.004	2.14 ± 0.03	1.55 ± 0.03	-0.47	1.08 ± 0.03	-1.48
Asn	-9.63 ± 0.01	2.04 ± 0.03	-7.59 ± 0.03	-0.30	-7.89 ± 0.03	-9.68
Gln	-9.48 ± 0.02	2.36 ± 0.03	-7.12 ± 0.03	-0.39	-7.51 ± 0.03	-9.38
Trp	-4.60 ± 0.01	1.77 ± 0.04	-2.84 ± 0.05	-0.74	-3.57 ± 0.05	-5.88
Hid	-11.70 ± 0.02	2.11 ± 0.03	-9.59 ± 0.04	-0.41	-10.00 ± 0.04	-10.27
Hie	-11.94 ± 0.02	2.08 ± 0.03	-9.86 ± 0.04	-0.41	-10.27 ± 0.04	-10.27
		Average absolute error w/o LRC: 1.44			rms error with w/o LRC: 1.67	
		Average absolute error with LRC: 1.06			rms error with LRC: 1.31	
			OPLS-AA			
Ala	$-0.00\pm<0.001$	2.43 ± 0.02	2.42 ± 0.02	-0.11	2.31 ± 0.02	1.94
Val	-0.00 ± 0.001	2.88 ± 0.03	2.87 ± 0.03	-0.28	2.59 ± 0.03	1.99
Leu	0.00 ± 0.001	3.06 ± 0.03	3.06 ± 0.03	-0.37	2.69 ± 0.03	2.28
Ile	-0.01 ± 0.001	3.10 ± 0.03	3.10 ± 0.03	-0.37	2.73 ± 0.03	2.15
Ser	-6.42 ± 0.01	2.23 ± 0.02	-4.19 ± 0.02	-0.17	-4.36 ± 0.02	-5.06
Thr	-6.32 ± 0.01	2.46 ± 0.03	-3.86 ± 0.03	-0.25	-4.11 ± 0.03	-4.88
Phe	-2.51 ± 0.01	2.50 ± 0.04	-0.01 ± 0.04	-0.53	-0.54 ± 0.04	-0.76
Tyr	-7.01 ± 0.02	2.34 ± 0.04	-4.67 ± 0.04	-0.58	-5.25 ± 0.04	-6.11
Cys	-3.88 ± 0.01	2.51 ± 0.02	-1.37 ± 0.03	-0.22	-1.59 ± 0.03	-1.24
Met	-3.88 ± 0.01	3.02 ± 0.03	-0.86 ± 0.03	-0.42	-1.27 ± 0.03	-1.48
Asn	-9.70 ± 0.02	1.50 ± 0.03	-8.20 ± 0.03	-0.32	-8.53 ± 0.03	-9.68
Gln	-9.79 ± 0.02	1.80 ± 0.03	-7.99 ± 0.04	-0.41	-8.40 ± 0.04	-9.38
Trp	-5.77 ± 0.02	2.08 ± 0.04	-3.69 ± 0.05	-0.74	-4.44 ± 0.05	-5.88
Hid	-9.99 ± 0.02	1.61 ± 0.03	-8.38 ± 0.04	-0.49	-8.87 ± 0.04	-10.27
Hie	-10.16 ± 0.02	1.60 ± 0.03	-8.56 ± 0.04	-0.49	-9.05 ± 0.04	-10.27
		Average absolute error w/o LRC: 1.11			rms error w/o LRC: 1.23	
		Average absolute error with LRC: 0.75			rms error with LRC: 0.85	

FIG. 3. Free energies of solvation for 15 neutral amino acid side chain analogs, for the AMBER(ff94), CHARMM22, and OPLS-AA force fields, in kcal/mol. Each reported value is the average of five independent trials, and includes the long range van der Waals correction as computed by Eq. (9). Statistical uncertainties are not shown, as they would be too small to read relative to the resolution of the plot, but are listed in Table II.

this correction is only applied to the solute–solvent interactions, the effect of this pressure change on the density or on the hydration free energy is expected to be negligible.

III. RESULTS

The main results of this study are a set of highly precise hydration free energy values of the fifteen neutral amino acid side chain analogs modeled by three different parameter sets. The averages over the five replicas are shown in Table II and Fig. 3. The results of all five simulations for each amino acid analog are included in the Supplementary Material.[74] Additionally, we report on a series of calculations that examine the effect of changes in certain simulation parameters and protocols on the accuracy and precision of the results.

From Fig. 4, one of the first things we notice is that virtually all side chain analogs are not sufficiently soluble relative to experiment. Large, polar side chain analogs are particularly inaccurate, off by 1–2 kcal/mol. After adding the long range Lennard-Jones correction, the average absolute
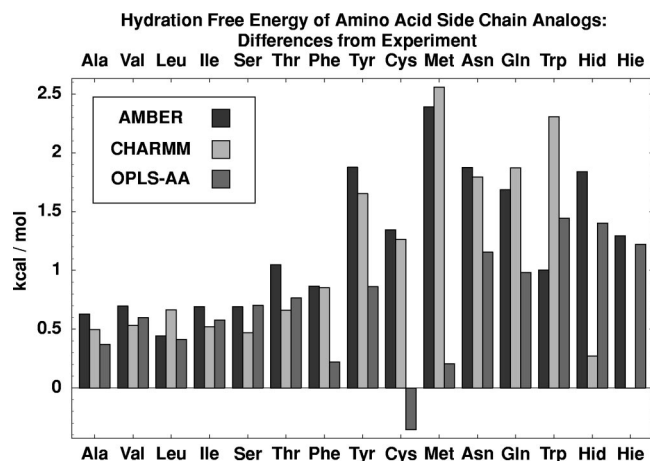

FIG. 4. Difference of free energy of solvation of the 15 amino acid side chain analogs from experiment, as listed in Table II, and shown in Fig. 3. All values in kcal/mol.

error from experiment for AMBER, CHARMM, and OPLS-AA were 1.22 kcal/mol, 1.06 kcal/mol, and 0.75 kcal/mol and the corresponding root-mean-square (rms) errors from experiments were 1.35 kcal/mol, 1.31 kcal/mol, and 0.85 kcal/mol (from Table II).

For some purposes, the *relative* hydration free energies are more relevant for simulation than the *absolute* hydration free energies. For example, one may be concerned with the relative preference for different amino acids to be solvent-exposed on a protein surface, or one may try to calculate the change of free energy of binding of an inhibitor when there is a mutation in the protein. In this case, the errors are somewhat smaller than are the errors in absolute hydration free energy. For the $\Delta\Delta G$ of Ala→X (where X is any other side chain analog), AMBER has an absolute error of 0.67, and a rms error of 0.85, while CHARMM has an absolute/rms error of 0.71/0.98, and OPLS-AA has absolute/rms errors of 0.51/0.60 (all values in kcal/mol), with the Lennard-Jones correction included. These relative free energy results are shown in Figs. 5 and 6. They can be computed from Table II, and are also included as a separate table in the Supplementary Material.[74] For $\Delta\Delta G$'s of all 105 mutations of X→Y (where X and Y are any two of the amino acid side chain analogs), we get absolute/rms errors for AMBER, CHARMM, and OPLS-AA of 0.68/0.89, 0.90/1.20, 0.58/0.75, with the long range Lennard-Jones correction included. We note that the long range correction improves the values of relative as well as the values of the absolute hydration free energy for all three force fields.

We also computed the free energy of hydration of a single molecule of TIP3P water solvated in a box of TIP3P water. This calculation used 8/10 Å tapered cutoffs for all water molecules (including the "solute"). The calculated free energy was -6.15 ± 0.02 kcal/mol. With the long range Lennard-Jones correction included, it becomes -6.21 ± 0.02 kcal/mol. This compares to experimental values of -6.33 kcal/mol[22] and -6.32 kcal/mol.[19] Interestingly, this value is closer to experiment than most of the amino acid side chain analogs, demonstrating that parameter sets may be optimized for some interactions (solvent–solvent) but not for others (solvent–solute).

In this study, we use parameters that are derived from protein force fields, and compare them with small molecule experimental data. It is important to address both whether the modified parameters are significantly different from the original protein parameters, and whether it is reasonable to compare these protein force field-derived parameters to the small molecule data.

The explicit changes in the published parameter sets were the truncations of the amino acids to the side chain analogs, and the changes in the β-carbon charge needed to produce molecular electrostatic neutrality for AMBER. The side chain parameters in these force fields were derived primarily from condensed phase and quantum mechanical data of the same or very similar small molecules as are studied in this paper as side chain analogs.[6–9] Essentially, our method of truncation (see Fig. 1) is the opposite of the method used to construct the side chain parameters in the first place (the exception being the case of AMBER partial charges). Com-

parison to small molecule experimental free energy data is reasonable as it is consistent with the methodology used to construct the force fields.

The effect of the additional charge changes for AMBER can be estimated. We examined the effect of different charge reassignment schemes in the Leu analog, which had the largest β-carbon partial charge correction. We found that the first charge distribution of the Leu analog (the one used in Table II) has a $\Delta G_{solv}=3.13\pm0.03$ kcal/mol, whereas the alternate-charge Leu analog, where the charges of the hydrogens of the β-carbon were changed instead of the charge of the β-carbon itself, has $\Delta G_{solv}=3.07\pm0.03$ kcal/mol (both of them before the van der Waals correction, which is identical for both). Since the difference is small, we believe that these changes do not affect the conclusions reached in this paper.

We must take into account in our calculations differences between the end states of the simulated system and the experimental system. This was derived in Eq. (6), and involves only the average volume V_1 of the fully coupled system, and average volume V^* of a box of pure solvent containing the same number of solvent molecules as the coupled system. We estimate V^* by finding the average volume of a box of 900 TIP3P waters, and scaling linearly to the actual number of waters in each solute simulation. The largest value for the $kT\ln(V^*/V_1)$ correction in all three force fields was for the Leu analog, at 0.011 kcal/mol, approximately 40% of the uncertainty in the measurement. In all cases, this correction was between 0.011 and 0.003 kcal/mol, and between 10% and 30% of the uncertainties reported in Table II. The uncertainties in these correction terms are negligible, approximately 10^{-4} kcal/mol. Because of the magnitude of this term, we can therefore safely neglect it in our calculations.

Another consideration related to the reproducibility and transferability of these results is their robustness to changes in the nonbonded interaction truncation protocol. Ideally, we would like to perform the calculations under the same conditions as those used to develop the parameters and/or the conditions under which they are typically used, sets of conditions that are not necessarily the same. For example, OPLS-AA was developed using quadratically tapered atomic spherical cutoffs,[9] and AMBER was developed with residue-based abrupt cutoffs,[6] which result in discontinuous interaction energies and forces at the cutoff boundary. We adopted the tapered cutoffs for solvent/solute interactions which are similar to those implemented in CHARMM,[64] though we chose 10/12 Å tapered truncations, longer than the range used in the parameterization of CHARMM.[7]

However, it is important to note that the methods used to make nonbonded interactions finite ranged can significantly affect the free energy or other simulation results.[75,76] In order to examine how different nonbonded truncation schemes can affect free energy calculations, we first examined the difference between group-based and residue-based tapering schemes in the case of five-side chain analogs (Ile, Thr, Tyr, Gln, and Trp). Atom-based schemes were not examined as they can produce anomalous simulation results unless they employ extremely long truncation or tapering distances.[57,64,77] We compared the OPLS-AA results (from
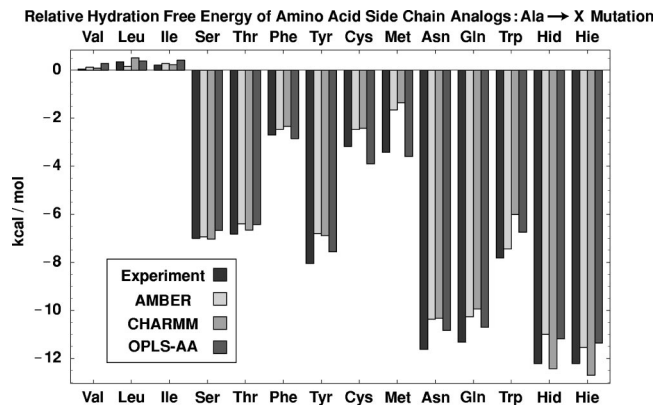

FIG. 5. $\Delta\Delta G$ of the mutation of Ala$\rightarrow$X side chain analogs, computed by subtracting the ΔG of hydration for the two amino acids. Note that the $\Delta\Delta G$s' differences from experiment are smaller than for the ΔGs' shown in Table II and Fig. 3.

the simulations shown in Table II) produced using the group-based finite ranged potentials to another set of results produced using "residue"-based cutoffs (where the entire side chain analog was treated as one group). For this test, we used only every other λ value, for a total of 31, as the extremely high precision is not as vital in this comparison and the errors in numerical integration will mostly cancel out, but in all other respects these simulations were run using the same parameters and protocols as those in the main data set. In Table III, we see that group and residue-based truncation schemes generally have differences that are statistically significant, but are still very small (about 0.1 kcal/mol). The exception is the Gln analog, which has a difference in the Coulombic energy of 0.42 ± 0.03 kcal/mol. The difference between group-based and residue-based truncation schemes may be expected to be the greatest in the Gln analog, as it contains several neutral groups with large partial charges inside each group. We thus see that there may be issues with comparing results produced with residue and group-based truncation schemes.

We also examined the effect of using different interaction truncation ranges, again using only 31 values of λ. In Table V, we see that different ranges of tapered cutoffs can
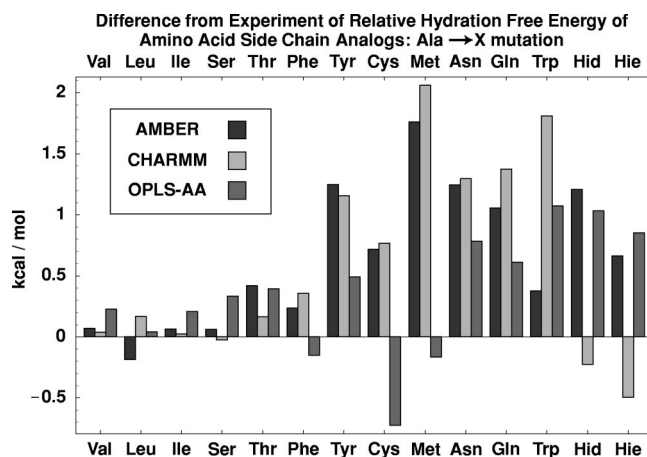

FIG. 6. Differences in $\Delta\Delta G$ of the mutation of Ala$\rightarrow$X side chain analogs from experiment. Note that the $\Delta\Delta G$s' mean differences from experiment are closer than for the ΔGs' shown in Table II and Fig. 3.

TABLE III. The effect of using group-based or residue-based tapered cutoffs on calculated free energies of hydration for selected side chain analogs using the OPLS-AA force field. Differences are somewhat greater than the statistical uncertainty, especially for the Gln analog. These results do not include the long range correction for the van der Waals terms. All energies are in kcal/mol.

Side chain analog	Residue			Group		
	Coulombic	van der Waals	Total	Coulombic	van der Waals	Total
Ile	-0.01 ± 0.001	3.25 ± 0.05	3.24 ± 0.05	-0.01 ± 0.001	3.17 ± 0.05	3.16 ± 0.05
Thr	-6.42 ± 0.02	2.52 ± 0.04	-3.90 ± 0.04	-6.32 ± 0.02	2.51 ± 0.04	-3.81 ± 0.04
Tyr	-7.07 ± 0.02	2.64 ± 0.06	-4.43 ± 0.06	-7.01 ± 0.02	2.47 ± 0.05	-4.54 ± 0.06
Gln	-9.39 ± 0.02	1.91 ± 0.05	-7.48 ± 0.05	-9.81 ± 0.02	1.90 ± 0.04	-7.90 ± 0.05
Trp	-5.79 ± 0.02	2.28 ± 0.06	-3.51 ± 0.07	-5.78 ± 0.02	2.24 ± 0.06	-3.54 ± 0.07

have a drastic difference on the computed free energies, up to 1.56 ± 0.07 kcal/mol in the case of the Trp analog. However, the difference in free energy between different cutoff ranges is due primarily to the attractive Lennard-Jones terms at long range. If we add the Lennard-Jones correction term, adjusted for the different truncation range, the difference in the van der Waals interactions becomes essentially negligible relative to the uncertainties of the simulations. The long range Lennard-Jones correction thus becomes useful not only in improving the match to experiment for the hydration free energies, but in eliminating almost all of the effect of the truncation of the Lennard-Jones term at different cutoff lengths. However, the effect of changing cutoff lengths has a nonmonotonic effect on the Coulombic component of the free energy of hydration. For the Ile analog, it is negligible, and for the Trp analog, it is relatively small, but for the Gln analog, with large partial charges and a large dipole moment, it can be up to 0.46 ± 0.07 kcal/mol. Many published simulation results use cutoff schemes shorter than our 10/12 Å tapered cutoff, which will most likely result in values even less favorable for solvation and thus further from the experimental results than the ones described in this paper.

To make reasonable comparisons to experiment and to other simulations, we also must understand the precision of these results. One source of numerical error is from the numerical integration with respect to λ of $\langle dH/d\lambda\rangle$ to produce the overall free energy change. We can examine this by looking at how much the calculated free energy changes when we omit some values of $\langle dH/d\lambda\rangle$. First, we examine the convergence of the Coulombic contribution with increasing sampling of λ_C. For the results in Table II, we used a λ_C spacing of 0.05. If we instead were to use a λ_C spacing of 0.1, over the 15 side chain analogs of OPLS-AA data set, we get a rms difference of 0.010 kcal/mol for the Coulombic contribution to the hydration free energy from the $\Delta\lambda=0.05$ results, with a maximum absolute difference of 0.025 kcal/mol for the Asn analog. For the Asn analog, this is approximately the same as the calculated uncertainty in the Coulombic component, while for the other analogs it is less, suggesting that λ spacing of 0.1 is slightly too large given the computed uncertainty, while 0.05 is probably sufficient, as the integration error is quadratic in $\Delta\lambda$. If we increase to a spacing of 0.2, the rms difference in the Coulombic component increases to 0.045 kcal/mol, with a maximum of 0.096 kcal/mol, again for the Asn analog, indicating that this spacing is clearly insufficient. We can see the shape of the Coulombic compo-

nents of $\langle dH/d\lambda\rangle$ for the Trp and Ala analogs in Fig. 2. Because of the smooth shape, it may be possible to achieve a higher level of accuracy in the Coulombic part with fewer values of λ using a higher order integration scheme.[78] However, higher order integration schemes can be problematic due to the difficulty of determining precise higher derivatives in λ.[79]

Next, we look at the convergence with respect to the number of sampled $\langle dH/d\lambda\rangle$ values of the van der Waals contributions to the free energy. When we use the data at only every other value of λ_{LJ} (of the initial 41 values specified above), we obtain a rms difference of 0.09 kcal/mol, with the maximum of 0.160 kcal/mol for the Trp analog. This is approximately three times greater than the uncertainty in the computed value of the van der Waals component of solvation, suggesting that this spacing over the entire λ_{LJ} range is not sufficient for the precision that is possible. Since using simple trapezoidal integration yields an error that converges with the square of the spacing, the original (41 λ_{LJ}) spacing scheme is most likely sufficiently within the uncertainty of the simulation. Using only every fourth λ_{LJ} value yields even worse results, with a rms difference of 0.337 kcal/mol, and maximum error of 0.58 kcal/mol, again for the Trp analog. The shape of the curve of $\langle dH/d\lambda\rangle$ of the van der Waals component is also shown in Fig. 2 for the Trp and Ala analogs. We see from the curve of the $\langle dH/d\lambda\rangle$ that there is a peak that requires such close spacing in λ for adequate integration. For other molecules, such as the Ala analog, the curvature is much less, so that such close spacing is not necessarily required.

It also is important in free energy simulations to compute the uncertainty due to limited sampling in a systematic way. In this study, we have done this by computing the autocorrelation function and correlation time of the components of $dH/d\lambda$ and using this information to compute statistical uncertainty estimates for $\langle dH/d\lambda\rangle$. We looked at the behavior of the correlation times in the simulations which included both windowed and full 1.0 ns correlation functions, which were performed for 20 out of the total 45 molecule/force field combinations. These latter simulations included some of the larger molecules, like the Trp and Tyr side chain analogs. The average (over all runs and λ values) of the windowed correlation times of the Coulombic component of $dH/d\lambda$ was 0.16 ps with a 0.10 ps standard deviation, and all were less than 0.55 ps. For the Coulombic component correlation times from the full 1.0 ns simulations, the average was 0.23

ps, with a 0.11 ps standard deviation, with all being less than 0.86 ps. For the windowed correlation times of the van der Waals terms, the average was 0.35 ps, with a standard deviation of 0.26 ps. All were less than 1.5 ps. For the full 1.0 ns van der Waals component correlation times, the average was 0.42 ps with a standard deviation of 0.28 ps, and all were less than 2.3 ps. These correlation times were of course λ and molecule dependent, so these numbers cannot be used as predictions for future studies. But it is clear that we are running, at minimum, hundreds of correlation times in order to collect $\langle dH/d\lambda \rangle$ at each λ, so the statistical analysis is indeed meaningful—if, of course, all the relevant degrees of freedom are sampled.

It was found that in all cases, the uncertainties determined from the 1.0 ns correlation function were between 10% and 20% larger than the 100 ps window uncertainties. We note the 200 fs resolution of the autocorrelation function from the entire 1.0 ns of data overestimates the contribution to the correlation time from the part of the correlation function before 200 fs, and thus may account for a portion of the somewhat longer estimates for τ_A. Since the correlation time contributes to the overall uncertainty in the square root, the effects on correlation time estimation from the windowing are negligible in the small molecule systems under study.

If these uncertainty estimates results are valid, we would expect that deviations from the average over the five separate simulations roughly obeys a normal distribution with a variance given by our uncertainty estimate. There are 45×5 each Coulombic and van der Waals values determined in the main data set, and we can compare them from the 45 averages for each combination of molecule and force field. We find that for the Coulombic values, 180 (80.0%) are within one standard deviation from the average, 220 (97.7%) within two standard deviations, and all but two (99.1%) within three standard deviations. If anything, this seems to suggest that our estimated uncertainties are slightly too large, as the percentages found within the first three standard deviations of the normal distribution are 68.3%, 95.5%, and 99.7%, respectively. For the van der Waals values, 155 (68.9%) are within one standard deviation, 215 (95.6%) are within two standard deviations, and all but one (99.6%) are within three standard deviations. We can also look at the distribution of all $\langle dH/d\lambda \rangle$ values at all values of λ used to calculate ΔG_{solv}. Here, we see that for the Coulombic component the breakdown in terms of percentage within (1σ, 2σ, 3σ) is (72.0%, 96.8%, 99.9%), (using $45 \times 21 \times 5 = 4725$ values) and for the van der Waals component (using $45 \times 40 \times 5 = 9000$ values, excluding $\lambda = 0$ where $\langle dH/d\lambda \rangle$ is constrained to be zero), we get (71.2%, 96.3%, 99.7%). In general, we seem to be fairly close to a normal distribution, suggesting that our calculated uncertainty, based on the correlation times of our observables, is really what it purports to be.

Even if our observations seem to behave as expected statistically, we must also ensure that we are sampling all the relevant degrees of freedom. In general, this is a difficult problem, but for smaller molecules like those examined here it is not insurmountable, even by the brute force methods used in this study. To test the extent of conformational sampling in these systems, we ran separate 1.0 ns simulations of the Ala, Ile, Phe, Ser, and Gln side chain analogs using the OPLS-AA force field, under the same conditions as the Folding@Home simulations. These simulations were run fully coupled, i.e., with both the Coulombic and van der Waals components at $\lambda = 1$. At smaller values of λ, most barriers to motion should be smaller, and thus sampling should in general occur even more rapidly. There are some exceptions, for example, intramolecular hydrogen bonding in vapor phase, but there are no intramolecular hydrogen bonds or other similar intramolecular attractions possible within the molecules in this study. We anticipate that these fully coupled simulations are therefore either representative of or slower in sampling these degrees of freedom than simulations run at other values of λ.

The degrees of freedom that contribute most to $\langle dH/d\lambda \rangle$ are those associated with the rearrangement of the solvent. To evaluate the sampling of these degrees of freedom, we computed the residence time of water molecules in the hydration shell for the Phe and Ala analogs. We defined a water molecule as being "in the hydration shell" if the distance between the oxygen atom of the water and any of the solute carbon atoms is <4.5 Å. We found a mean time of residence in the shell of 0.56 ± 0.68 ps for the Ala analog, and 0.69 ± 1.05 ps for the Phe analog, which are over three orders of magnitude less than the overall simulation time of 1.0 ns. We also computed the correlation function of occupation of the hydration shell, averaging over all water molecules which are in the hydrophobic shell at some point during the simulation, and integrating this function to obtain the correlation time. This correlation time also captures returns to the hydration shell after the first departure. For the Ala, this time was 7.4 ps, and for Phe, it was 13.0 ps. During the Ala analog simulation, 767 of 897 water molecules spent some time in the solvation shell, while during the Phe analog simulation, 803 of 892 did, indicating extensive sampling, as the hydration shell itself covers only about 5% of the simulation cell volume. It is clear that these simulations are indeed running tens and hundreds of times the correlation times of any of the observables that are correlated to the free energy components that we are trying to measure.

For the Ser analog, we examined the lifetime of the hydrogen bonds between the water molecules and the OH moiety. The hydrogen bonds are defined as pairs of molecules with O–O distance less than 3.5 Å and $\angle$O–H–O<110°, and with this definition, the residence time is 0.22 ± 0.54 ps, also orders of magnitude smaller than the simulation time. The correlation time of hydrogen-bonding molecules (computed in the same manner as the correlation time of hydration shell occupation) was longer, 2.3 ps, and 610 of the 897 water molecules hydrogen bonded with the solute over the course of the simulation. In all cases, it appears we can have confidence that these solvent degrees of freedom are well sampled in our simulations.

In general, the slowest degrees of freedom in molecular dynamics simulations are those associated with torsional motion. To explore the sampling of these degrees of freedom in a long "brute-force" calculation, we looked at selected heavy-atom torsional angles, specifically the C–C–C–C angle in the Ile analog, the C–C–C–O angle in the Gln
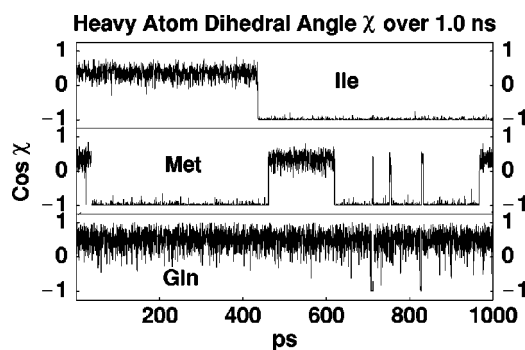
Heavy Atom Dihedral Angle χ over 1.0 ns

FIG. 7. Cosine of the heavy atom dihedral angle of the three amino acid side chain analogs studied in 1.0 ns simulations, with λ=1 (i.e., fully coupled to the solvent). The Ile analog (C–C–C–C dihedral) is on top, the Met analog (C–C–S–C dihedral) is in the middle, and the Gln analog (C–C–C–O dihedral) is on the bottom. We see that the Met analog dihedral angle is not sampled well, and the Ile analog dihedral angle is very poorly sampled.

analog, and the C–C–S–C angle in the Met analog. We computed the value of cos χ (where χ is one of the three torsional angles described above) as well as its correlation function. The 1.0 ns time courses of cos χ of the three angles are shown in Fig. 7.

For the Gln analog, we find that there is fairly good sampling for this torsional angle. For the Met analog, the *anti* and the *gauche* states are only moderately sampled—we note approximately 12 transitions back and forth during the run, which may not be enough to adequately sample the true thermodynamic probability of the different states. More worrisome, for the Ile analog, there is only one transition between the *gauche* and *anti* conformations, and this lack of sampling could very well result in biases in the results.

Previous data suggests that the difference in hydration free energy of *n*-butane (the Ile side chain analog) fixed in the *gauche* and *anti* conformations is around 0.3 kcal/mol,[35] which is certainly larger than the precision of our calculations here, though difficult to compare to as no uncertainties are provided in that study. In order to assess the size of errors in free energy due to a lack of sufficient sampling of the dihedral angle, we examined the difference in hydration free energy of the Ile conformations separately. One way to do this is to rigidly restrain *anti* and *gauche* conformations and evaluate the free energy of hydration of both, but the distribution of the torsional angle for Ile shown in Fig. 7 shows a fairly broad angular distribution, so relative free energies of fixed conformations are not the most relevant observable. We instead constructed a restraining potential of the form,

$$H(\chi) = \exp\left(\frac{0.01}{(0.5 + \cos \chi)^2}\right) - 1 \qquad (10)$$

and then added it to the torsional term. This serves to produce an impenetrable barrier between the *anti* and *gauche* configurations, as it goes to infinity at ±60° from the *anti* configuration, while adding <20% of kT to any state outside of the range of ±(60±20)°. We can see from Fig. 7 that this excluded region is essentially never occupied even without the restraining potential.

We performed a set of simulations with the OPLS-AA Ile side chain analog which differed from the simulations

whose results are presented in Table II only in the presence of the restraining potential. We examined only the van der Waals contribution to the free energy of hydration, as the Coulombic contribution was only about 0.01 kcal/mol, and any changes in this value with respect to the torsional angle are expected to be smaller than even this amount.

We found that the *anti* conformation had a van der Waals component of the free energy of solvation of 3.12±0.03 kcal/mol, while the corresponding value for the *gauche* conformation is 3.07±0.03 kcal/mol. These values compare with the computed value in Table II of 3.10±0.03 kcal/mol. These free energy results demonstrate that for *n*-butane, the free energy of hydration is almost completely independent of the conformation. It is possible, however, that molecules with large dipole differences between torsional conformations could have a larger hydration free energy conformational dependence. Additionally, the fact that some dihedral angles have such long correlation times implies that accurate calculations for even moderately more complex systems will be challenging, and require more complicated sampling algorithms.[80–85]

There are many studies of free energies of solvation of models of small molecules, and it is important to compare to these previous studies where possible. Unfortunately, comparisons between these simulations and the current simulations can prove extremely tricky. The majority of previous studies were performed using earlier parameter sets (such as CHARMM19 or ''united-atom'' OPLS), or different solvent models (such as TIP4P or SPC) and thus are not expected to be directly comparable to the present study. There are also many simulation parameters (such as the number of water molecules, methods of truncation of long range interactions, and inclusion of long range corrections to cutoffs) that can drastically affect the results. A complete review of the wealth of solvation free energy simulations is outside the scope of the present paper, so we will focus in our comparisons instead on a few studies performed under similar conditions, and that should be expected to give similar results.

One recent study of free energies of solvation using OPLS-AA is closely comparable to the present study. Wescott *et al.* reported a thorough study of small, straight-chain hydrocarbons, using the OPLS-AA force field and TIP3P water.[73] These results include a long-range correction, so are expected to be very comparable to the study described in this paper. They obtain values of 2.35 kcal/mol for methane (the Ala side chain analog), 2.40 kcal/mol for propane (the Val analog), and 2.59 kcal/mol for *n*-butane (the Ile analog), compared to our results of 2.31, 2.59, and 2.73, respectively. Unfortunately, no uncertainty estimates are given, so it is impossible to judge if these results are within the mutual uncertainty of the two studies, but they are relatively close, within 0.2 kcal/mol for all molecules. Wescott *et al.* also agree qualitatively that the models' free energies of solvation are uniformly too high. Inspection of Fig. 4(a) in this study also indicates that the area of greatest positive curvature in $\langle dH/d\lambda \rangle$ is not well sampled in λ compared to the present study (see Fig. 2), meaning their reported results are most likely somewhat too negative.

There are significantly fewer discussions in the literature

TABLE IV. Six different experimental data sets of free energy of hydration of amino acid side chain analogs. There is some variation between different studies, in most cases of the same order of magnitude as the uncertainties in the computed hydration free energies (Ref. 88).

Side chain analog	Wolfenden[a]	Cabani[b]	Hine[c]	Abraham[d]	Plyasunov[e]	Viswanadhan[f]	Ben-Naim[g]
Ala	1.94	2.00	1.95	1.98	1.98	2.00	1.93
Val	1.99	1.96	1.99	1.96	2.02	1.95	1.98
Leu	2.28	2.32	2.29	2.32	2.32	2.25	2.28
Ile	2.15	2.08	2.15	2.07	2.18	2.07	2.15
Phe	−0.76	−0.89	−0.76	−0.79	−0.77	−0.80	−0.77
Ser	−5.06	−5.11	−5.07	−5.10	−5.14	−5.05	···
Thr	−4.88	−5.01	−4.90	−5.00	−4.96	−5.00	···
Cys	−1.24	−1.24	−1.24	−1.36	−1.26	···	···
Tyr	−6.11	−6.14	−6.12	−6.14	···	···	···
Asn	−9.68	−9.71	···	···	···	···	···

[a]Reference 18.
[b]Reference 20.
[c]Reference 21.
[d]Reference 22.
[e]Reference 89.
[f]Reference 23.
[g]Reference 19.

of hydration free energy calculations for the amino acid side chain analogs with CHARMM and AMBER. The only directly comparable result (i.e., using the same charge and Lennard-Jones parameters) found using the CHARMM22 force field was for methane (Ala analog), with a value of 2.10 ± 0.25 kcal/mol.[86] Our calculation was slightly outside this uncertainty range (2.44 ± 0.02 kcal/mol), but again, a variety of differences in simulation parameters make exact comparison difficult.

Although the GROMOS96 potential set was not examined in this study, a recent publication makes it possible to draw some comparisons.[34] This work used SPC water, rather than TIP3P water, but as GROMOS96 was designed with SPC water,[8] the GROMOS96/SPC combination represents the fairest test. From the data in Table II of Villa and Mark,[34] we obtain an average absolute difference from experiment of 2.2 kcal/mol, and a rms difference from experiment of 2.7 kcal/mol, significantly larger than the errors for the parameter sets studied above. Reported uncertainties for these simulations spanned the range from 0.15 kcal/mol (for Ala) to 0.5 kcal/mol (for Tyr), approximately 5–10 times those of the present study. These results were even more positive relative to experiment that the force fields used in this study.

IV. DISCUSSION

In the previous 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 about simulations using these models.

How accurate do we expect the hydration free energies computed using these parameter sets to be? The parameter sets used in this study were not parameterized to reproduce free energies of hydration because, until recently, it was too computationally intensive to so do in an iterative manner with sufficient precision.[87] 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.[6,7,9] In this light, there is no *a priori* reason to expect the free energies correspond particularly well to experimental results. Indeed, it is encouraging that all three parameter sets are qualitatively close to the experimental values.

These uniformly high values are not unexpected. As noted above, Wescott *et al.* reports free energies for OPLS-AA/TIP3P that are significantly higher than experiment. Other simulations using OPLS-AA and TIP4P water have given free energies that were uniformly higher than experiment,[35,38] though slightly less so. This is reasonable, as TIP4P has a slightly deeper Lennard-Jones minimum, which should result in greater attraction, and thus free energies more favorable to solvation.

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. A number of studies have reported free energies of hydration for many of the amino acid side chain analogs studied here,[18–23] as shown in Table IV. We note good agreement among these experimental studies, and we also note the variation among experimental measurements is comparable to the uncertainties in our simulated results.[88] In addition to the variation shown in Table IV, Plyasunov *et al.* examined a large number of literature values of the free energy of solvation for several of these molecules.[89] For toluene (Phe), 46 literature values were found, with a standard deviation of 0.06 kcal/mol. For methanol (Ser), 14 values were found, with a standard deviation of 0.05 kcal/mol, and for ethanol (Thr), 17 values were found with a standard deviation of 0.04 kcal/mol. In light of this variation, it appears for these molecules that the uncertainty in the computed free energies approaches the experimental uncertainty. However, it is important to note that for more complex systems of interest, such as the free energy of binding in ligand-protein com-

TABLE V. The effect of using different nonbonded interaction truncation ranges on the calculated hydration free energy for selected amino acid side chain analogs using the OPLS-AA force field. We note drastic differences in the simulated van der Waals energies, but these can mostly by accounted for by adding the analytic Lennard-Jones long range correction (LRC) described in Eq. (9) in the text. Differences in the Coulombic effect are lower, but occasionally significant and nonmonotonic. Uncertainties are not listed for space reasons, but are all very close to the uncertainties for the corresponding molecules in Table II. All energies are in kcal/mol.

Side chain analog	Cutoffs	Coulombic	vdW	Total	LRC	With LRC vdW	With LRC Total
	6/10	−0.01	3.77	3.67	−0.97	2.81	2.80
	7/10	−0.01	3.62	3.61	−0.80	2.81	2.80
Ile	8/10	−0.01	3.52	3.51	−0.68	2.84	2.83
	10/12	−0.01	3.17	3.16	−0.37	2.80	2.80
	12/14	−0.00	3.02	3.02	−0.22	2.80	2.79
	6/10	−5.77	3.45	−2.32	−2.01	1.44	−4.33
	7/10	−5.86	3.10	−2.75	−1.66	1.45	−4.41
Trp	8/10	−5.85	2.82	−3.03	−1.39	1.44	−4.42
	10/12	−5.78	2.24	−3.54	−0.74	1.49	−4.28
	12/14	−5.66	1.86	−3.80	−0.45	1.42	−4.24
	6/10	−9.92	2.55	−7.38	−1.07	1.48	−8.45
	7/10	−10.00	2.34	−7.66	−0.89	1.45	−8.54
Gln	8/10	−10.08	2.20	−7.89	−0.75	1.45	−8.63
	10/12	−9.81	1.90	−7.90	−0.41	1.50	−8.31
	12/14	−9.62	1.70	−7.91	−0.25	1.46	−8.16

plexes, both calculations and experiments are very likely to be less precise.

This study also draws attention to the sensitivity of results to changes in the conditions of the simulation. In this paper, we studied the effect on computed free energies of changes in parameter sets, inclusion or exclusion of long range Lennard-Jones corrections, and different methods of implementing finite-ranged interactions. As has also been noted by other researchers,[53,75,76] many of these effects can result in significant differences in calculated values. It is only with high precision results that these effects can emerge from the statistical noise inherent in free energy simulations.

We call particular attention to the variation in results caused by the differences in methods to make nonbonded interactions finite-ranged. Although the Lennard-Jones terms are quite small outside the cutoff ranges, they are attractive everywhere, and thus can contribute significantly to the solvation free energy.[90] It has been previously suggested that this correction can significantly affect the results of a simulation.[73,91] As nonbonded potential sets are *effective* pair potentials, force field designers are of course free to include or exclude a long range correction in their parameterization. The Lennard-Jones functional form is an approximation to van der Waals forces in any case, and consistency between the development and application of the force fields is the important point. The OPLS-AA parameter set was parameterized including a long range correction,[9,90] whereas CHARMM and AMBER were not. However, we find that including this long range correction improves hydration free energy results for all three parameter sets, as shown in Table II. Perhaps most importantly, as noted in Table V, applying a proper long range correction to the van der Waals term eliminates much of the difference in hydration free energies that results from truncation schemes with different distance ranges.

The differences that the same finite-ranged schemes cause in the electrostatic terms can presumably be eliminated by moving to methods like Ewald summation or reaction field schemes for computing the electrostatic interactions. Other studies have pointed out additional reasons for moving away from finite-ranged treatments of electrostatic interactions.[32,75] As cutoffs for van der Waals terms are still necessary when using methods that eliminate the need for finite-ranged Coulombic forces, like Ewald summation, the use of a correction term can remove the van der Waals truncation scheme as a parameter which arbitrarily affects the solvation free energy. Moving to treatments of long ranged interactions that require fewer user-adjusted arbitrary parameters will greatly improve the ability to compare simulations from different simulation studies and to develop parameters for force fields that are more easily transferable.

In general, previous simulations with a variety of models reproduced experimental values of free energy within or near the uncertainty of the studies performed, though this is not always true for all molecules and all sets. However, the uncertainties that the studies list (usually 0.2–0.5 kcal/mol) are sufficiently large as to make comparisons to experiments difficult. True differences of the models from experiment are masked by such 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 energies 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.

We would like to emphasize several steps that are required in order to obtain high precision, apart from simply utilizing more computational power. A proper statistical uncertainty analysis should always be conducted in order to quantify reproducibility. Computation of correlation times and variances for the $dH/d\lambda$ terms are required, and this information is relatively straightforward to collect from a simulation. To obtain highly precise results, methods should be chosen that give small variance in $\langle dH/d\lambda \rangle$. The choice of λ parameterization greatly affects this variance. This paper is not intended as a comprehensive guide to such parameterizations. However, the decoupling of electrostatic and van der Waals interactions and the use of a soft-core potential in atomic site introduction/removal results in magnitudes of variance of $\langle dH/d\lambda \rangle$ that make possible the level of preci-

sion 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 above. However, for even small proteins, this becomes very difficult, and it is as yet unclear how this can be done in a general case.

The high level of precision obtained in this paper will not be necessary for all uses of free energy calculations. However, high precision measurements are necessary in order to understand the magnitude and source of deviations from experiment, and to understand the subtle effects of changing simulation conditions on the free energy results. High precision measurements are also necessary to study the effectiveness of new methodologies, and to judge small differences between molecules.

Although determining free energies for large numbers of small molecule systems may require special computational resources like Folding@Home, obtaining the equivalent degree of precision for a few molecules is within the reach of most academic and industrial research groups. The computational power required to produce the precision in these studies for one molecule with one force field would be equivalent to about 10 dual processor 1.5 GHz Athlons running for <2 days (about a week for the fivefold replication performed in this study). In retrospect, there are also many ways to improve the computational efficiency of the simulations, aside from hardware and software optimizations. For example, the simulation cells were conservatively large, and a speedup factor of at least two could be obtained with a smaller cell. Sixty-one λ points were used, but analysis afterward showed that at least 10 of the Coulombic λ points were unnecessary for obtaining the desired precision for these molecules. For the smaller molecules, the number of intermediate Coulombic and van der Waals λ points could be reduced even further. If a slightly larger uncertainty is acceptable (say, 0.05–0.1 kcal/mol), both the total simulation time required and the number of λ values needed could be reduced even more. In this case, and given the correlation times of these molecules, it would be possible to reduce the total collection time to 200–300 ps and still include enough uncorrelated data points necessary for good statistical validation. Under these conditions, it would be possible to evaluate several molecules per day with the cluster described above. For absolute solvation free energies of small molecules with few torsional degrees of freedom in explicit water, uncertainties <0.1 kcal/mol are easily obtainable by most researchers. It is important to note that these should not be taken as general guidelines for all free energy simulations. The appropriate spacing in λ and the necessary simulation time required need to be determined individually for each system.

It is still to be determined how the required computa-

tional effort to obtain sufficiently precise free energies scales with solute size and complexity, especially for those solutes with many torsional barriers and many stable conformations. The current study suggests that simply using straightforward long-running molecular dynamics simulations may not be sufficient, as torsional degrees of freedom are not well-sampled. This means that more sophisticated sampling schemes, such as replica exchange and other multicanonical methods,[81–85,92,93] umbrella sampling,[94,95] or locally enhanced sampling[80] may be necessary for more complex solutes like biomolecules.

Although the free energies of hydration of amino acid side chain analogs are one measure of fitness of a force field, it should be emphasized that there are other measures of fitness that are also important. For example, Price and Brooks[10] have run explicit water protein simulations with these same force fields, monitoring such observables as the rms deviation from experimental structures, the radius of gyration, and the solvent-accessible surface area, finding no significant difference between AMBER, CHARMM, and OPLS-AA. Other studies have compared conformational and interaction energies of a variety of force fields.[96,97]

This study leaves some important questions unanswered. One is the sensitivity of these calculated free energies of hydration to the choice of other very commonly used water models, such as SPC, TIP4P, and TIP5P. Each model has enough differences that discrepancies in solute–solvent energetics and kinetics are likely.[67,98] Another is whether improved results are possible with improved treatments of Coulombic forces, such as reaction field and Ewald summation methods. Comparisons of solvation free energies with simulations using different water models and Ewald summation are currently underway.

Our results have some implications for simulations using parameter sets such as AMBER, CHARMM, and OPLS-AA to estimate 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. It may be possible for sufficiently precise calculations to 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.

This study also reveals the power that distributed computing can provide. Thanks to the rapid increase in available computer power in the last decade, we can afford somewhat more costly algorithms when we are striving to obtain precise results with rigorous approaches, we can apply these algorithms to large numbers of molecules, and we can collect more comprehensive data. We also note that distributed resources such as Folding@Home may be of great use in the development of potential sets. First, it becomes much easier to explore multidimensional parameter space in a systematic

way, and second, it becomes possible to use a wider variety of computationally intensive observables (such as free energies of solvation or partition coefficients) in the actual development of parameters.

V. CONCLUSIONS

We have demonstrated that free energies of hydration for popular computational models can be computed with a precision of three significant figures in a reasonable amount of time, at least in the case of solutes with low numbers of degrees of torsional freedom. This degree of precision is at least an order of magnitude better than most previous studies, and the uncertainties reported have been rigorously computed and verified statistically. The computational precision we have obtained is approximately equal to the experimental precision for the systems examined, and can be obtained for the absolute free energy of hydration, not simply for relative free energies of hydration. If the model under study is close enough to experiment, this degree of precision is sufficient for virtually all practical uses.

Current generation parameter sets can be off as much as 2 kcal/mol per side chain analog, and therefore may be insufficient to compute free energies that accurately compare to experimental results for many applications. The three studied potential sets are fairly similar, and deviate from experimental hydration free energies in similar ways, though to somewhat different degrees. However, we have found that these results can be highly dependent on simulation protocols such as the choice of implementation of finite-range interactions, making sweeping generalizations about suitability of potential sets impossible.

It is clear from the merely adequate sampling of the torsional degrees of freedom in these simple systems that more complicated systems, with many correlated torsional terms producing large barriers between physically relevant states, will present difficulties and may be largely impossible to study adequately with simple molecular dynamics. In order to sufficiently sample these systems it will require increased computational power, smarter sampling schemes, and a clear separation and understanding of the errors arising from insufficient sampling versus improperly represented energetics.

This paper also 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. This is especially true in the light of new distributed computing techniques, which provide the greatly increased computational power needed for both the development of improved parameter sets and the sufficient sampling of complicated models. We expect that this will be enormously useful to both the scientific study of macromolecular interaction and to the commercial development of useful drugs.

APPENDIX: DERIVATION OF KEY RELATIONSHIPS IN THE COMPUTATION OF SOLVATION FREE ENERGIES IN ISOBARIC–ISOTHERMAL SIMULATIONS

This Appendix introduces notation used in other parts of this manuscript, conveys our definition for the free energy of solvation, develops relationships between the free energy of solvation and chemical potentials, and describes some aspects of the computation of absolute solvation free energies. The classical theory of liquids and solutions is well developed[19,99] and is included in most graduate statistical mechanics courses. However, we present here some aspects of an isobaric–isothermal treatment that are not usually discussed.

Most of the experiments with which we make contact determine solvation free energies by measuring concentrations of some molecular solute, s, in two phases, a gas and a liquid, that are in equilibrium with each other and both subject to a pressure P and a temperature T. (In the experiments of interest, s are the amino acid side chain analogs and the solvent in the liquid phase is water.) If the partial pressure of s in the gas phase is P_s and the concentration of s in the liquid phase is ρ_s^l, the experimental results are reported as *solvation free energies*, computed from these measured quantities by use of the following:

$$\Delta G_{\text{solv}} = kT \ln(P_s / \rho_s^l kT). \tag{A1}$$

The usual interpretation of ΔG_{solv} is that it is the free energy associated with *coupling* a molecule of solute into the rest of the solution.[100] In this Appendix we will derive a relationship between ΔG_{solv} and quantities computable by simulation techniques, thereby allowing comparison of the simulation and experimental results.

First consider the gas phase. The Gibbs free energy of a gas phase system of N_s^g molecules of s at a pressure P_s and a temperature T is

$$G(N_s^g, P_s, T) = -kT \ln \Delta(N_s^g, P_s, T), \tag{A2}$$

where k is the Boltzmann constant, and Δ is the isobaric–isothermal partition function for such a system,

$$\Delta(N_s^g, P_s, T) = \int_0^\infty d(V/V') \exp(-\beta P_s V) Q(N_s^g, V, T). \tag{A3}$$

Here, Q is the canonical partition function for a system of N_s^g molecules in a container of volume V at a temperature T and $\beta = 1/kT$. The factor of V' is an arbitrary constant with units of volume that serves to make the partition function Δ dimensionless.[101]

If one assumes the gas phase can be treated as ideal, the canonical partition function is

$$Q(N_s^g, V, T) = \frac{1}{N_s^g!} (\Lambda^{-3} V q_s(T))^{N_s^g}, \tag{A4}$$

where $\Lambda = (h^2/2\pi M_s kT)^{1/2}$ is the thermal de Broglie wavelength, that arises from the translational partition function for a molecule of mass M_s, h is Planck's constant, and $q_s(T)$ is a canonical *internal* partition function for a single molecule whose center-of-mass is constrained to be at some fixed point and whose center-of-mass momentum is constrained to be zero. $q_s(T)$ would treat rotational, vibrational and electronic degrees of freedom and could be computed classically or quantum mechanically, but either way it does not affect the essential outcome here. The feature of q_s we exploit is that it is independent of volume.

Combining the relationships in Eqs. (A2), (A3), and (A4) one can obtain

$$G(N_s^g, P_s, T) = -kT N_s^g \ln\{\Lambda^{-3} q_s(T) kT/P_s\}. \tag{A5}$$

The chemical potential of s in the gas phase is given as

$$\mu_s^g(N_s^g, P_s, T) = G(N_s^g, P_s, T) - G(N_s^g - 1, P_s, T)$$
$$= -kT \ln\{\Lambda^{-3} q_s(T) kT/P_s\}. \tag{A6}$$

Second, consider the liquid phase. The Gibbs free energy of a system of N_s^l molecules of s, and N_w molecules of solvent at pressure P, and a temperature T is given by

$$G(N_s^l, N_w, P, T) = -kT \ln \Delta(N_s^l, N_w, P, T). \tag{A7}$$

Several points should be noted. In general, the pressure P (*total* pressure exerted on the liquid) will be different from P_s (*partial* pressure of s in the gas phase). Also, the notation suggests that the solvent is water, but the formulation is general at this point, and the solvent may itself be a mixture. Furthermore, the formulation does not assume that the concentration of s in the liquid phase is dilute.

The chemical potential of s in the liquid phase is given by the following:

$$\mu_s^l(N_s^l, N_w, P, T) = G(N_s^l, N_w, P, T) - G(N_s^l - 1, N_w, P, T)$$
$$= -kT \ln\left\{ \frac{\Delta(N_s^l, N_w, P, T)}{\Delta(N_s^l - 1, N_w, P, T)} \right\}. \tag{A8}$$

The Hamiltonian to be used for evaluating the numerator will contain the kinetic and all (intramolecular and intermolecular) potential energy for the $N_s^l + N_w$ molecules of the system. Symbolically, for the types of Hamiltonians usually used in classical simulations, where the interaction energies are of a separable form, we could write the following:

$$H(N_s^l, N_w) = T_s(N_s^l) + T_w(N_w) + U_s(N_s^l) + U_w(N_w)$$
$$+ U_{w-w}(N_w) + U_{s-s}(N_s^l) + U_{s-w}(N_s^l, N_w),$$

where T_i indicates the kinetic energy from all molecules of type i, U_i represents intramolecular potential energy for molecules of type i, and U_{i-j} indicates intermolecular potential energy for molecules of type i interacting with other molecules of type j.

Next a special molecule of s is chosen and a coupling parameter, λ, is introduced to the $(N_s^l + N_w)$-molecule Hamiltonian such that when $\lambda = 0$ the special molecule does not interact with *any* of the other material, and when $\lambda = 1$ the Hamiltonian is the same as the one described above for the fully interacting $(N_s^l + N_w)$-molecule system. This coupling parameter turns off only the intermolecular interactions between the one special molecule and the N_w solvent molecules (affecting U_{s-w}), as well as between the special molecule and the remaining $N_s^l - 1$ other molecules of type s, (affecting U_{s-s}). However, the kinetic and intramolecular components of the energy of the special molecule are *not* affected, and are included for all values of λ. Note that this formulation requires treating intramolecular electrostatic interactions for the special molecule of type s (not λ-coupled) differently from intermolecular electrostatic interactions between the special molecule of type s and other material (λ-coupled).

In practice, the intermolecular potential between molecules consists of van der Waals and Coulombic interactions, and there are many ways to couple these interactions. The only formal requirement is that the Hamiltonian have the above properties at the $\lambda = 0$ and $\lambda = 1$ endpoints. The parameterization used in this work, along with its rationale, is described in the Methods.

Note that even though the fully coupled Hamiltonian, $H(\lambda = 1)$, is identical to the original Hamiltonian, the act of selecting a special molecule of type s for the coupling process results in a slightly different partition function,

$$Q(N_s^l, N_w, V, T, \lambda = 1) = N_s^l Q(N_s^l, N_w, V, T). \tag{A9}$$

This is because, once the choice of a special molecule is made, instead of N_s^l indistinguishable molecules, there are now only $N_s^l - 1$, resulting in a different combinatorial factor. The factor of N_s^l corresponds to the number of ways one could have chosen the special molecule.

Note that $H(\lambda = 0)$ is *not* the Hamiltonian used to evaluate the denominator in Eq. (A8). This Hamiltonian is for a $(N_s^l + N_w - 1)$-molecule system, whereas the parameterized Hamiltonian, $H(\lambda)$, is actually for a $(N_s^l + N_w)$-molecule system. However, consider the following version of the canonical partition function, developed in terms of $H(\lambda)$ and evaluated at $\lambda = 0$,

$$Q(N_s^l, N_w, V, T, \lambda = 0)$$
$$= Q(N_s^l - 1, N_w, V, T) Q(N_s^l = 1, N_w = 0, V, T). \tag{A10}$$

This factorization is possible because, by the design of $H(\lambda)$, when $\lambda = 0$ the special molecule of type s is uncoupled from the rest. The last factor is the partition function for an isolated gas phase molecule of type s in a volume V at temperature T. Using the relationships in Eqs. (A10) and (A4), and multiplying and dividing by the denominator in Eq. (A8), we see that the isobaric–isothermal partition function becomes when $\lambda = 0$,

$$\Delta(N_s^l,N_w,P,T,\lambda=0)=\int_0^\infty d(V/V')\exp(-\beta PV)Q(N_s^l-1,N_w,V,T)Q(N_s^l=1,N_w=0,V,T)$$

$$=\int_0^\infty d(V/V')\exp(-\beta PV)Q(N_s^l-1,N_w,V,T)(\Lambda^{-3}Vq_s(T))$$

$$=(\Lambda^{-3}q_s(T))\Delta(N_s^l-1,N_w,P,T)\frac{\int_0^\infty d(V/V')V\exp(-\beta PV)Q(N_s^l-1,N_w,V,T)}{\Delta(N_s^l-1,N_w,P,T)}$$

$$=\{\Lambda^{-3}V^*q_s(T)\}\Delta(N_s^l-1,N_w,P,T). \tag{A11}$$

In the last equality, V^* is the mean volume of the $(N_s^l+N_w-1)$-molecule system at pressure P and temperature T. The term in braces is recognizable as $Q(N_s=1,V^*,T)$. We may take advantage of this final relationship in Eq. (A8) to obtain

$$\mu_s^l(N_s^l,N_w,P,T)=-kT\ln\left\{\frac{\Delta(N_s^l,N_w,P,T)}{\Delta(N_s^l-1,N_w,P,T)}\right\}$$

$$=-kT\ln\left\{\frac{\Delta(N_s^l,N_w,P,T)}{\Delta(N_s^l,N_w,P,T,\lambda=1)}\right.$$

$$\times\frac{\Delta(N_s^l,N_w,P,T,\lambda=1)}{\Delta(N_s^l,N_w,P,T,\lambda=0)}$$

$$\left.\times\frac{\Delta(N_s^l,N_w,P,T,\lambda=0)}{\Delta(N_s^l-1,N_w,P,T)}\right\}$$

$$=-kT\ln(1/N_s^l)+\Delta G_{\text{sim}}$$

$$-kT\ln\{\Lambda^{-3}V^*q_s(T)\}, \tag{A12}$$

where we have introduced ΔG_{sim}, defined by the following:

$$\Delta G_{\text{sim}}\equiv-kT\ln\left\{\frac{\Delta(N_s^l,N_w,P,T,\lambda=1)}{\Delta(N_s^l,N_w,P,T,\lambda=0)}\right\}. \tag{A13}$$

The first term on the right-hand side of Eq. (A12) provides the free energy associated with making the special molecule indistinguishable from the other molecules of type s. ΔG_{sim} provides the free energy that comes from *coupling* the interactions of the special molecule of s with the $N_s^l-1+N_w$ molecules of the rest of the system. By comparison with Eq. (A6) one sees that the third term can be recognized as the free energy associated with introducing a noninteracting molecule into a system to establish a mean number density of $1/V^*$, since such a gas would have a "partial pressure" consistent with $P^*/kT=1/V^*$.

A useful interpretation is suggested by Eq. (A12). The introduction of a molecule of type s can be thought of as happening in four steps. The first step consists of producing from the $(N_s^l-1+N_w)$-molecule system, a system that consists of two totally independent subsystems. One of these two subsystems is an $(N_s^l-1+N_w)$-molecule system at a pressure P, and the other is a system that consists of a single molecule of type s at a pressure of $P^*=kT/V^*$, the pressure exerted by a single ideal gas particle if it were confined to a volume of V^*, the mean volume of the $(N_s^l-1+N_w)$-molecule system. The isobaric–isothermal partition function for

this compound system is simply the product of the two-partition functions for the subsystems. The free energy change for this process is $-kT\ln\{\Lambda^{-3}q_s(T)kT/P^*\}$, equivalent to that given by the third term on the right-hand side of Eq. (A12).

The second step consists of producing from the system of two independent subsystems a new system of $N_s^l+N_w$ molecules where the new molecule of type s is included, but does not interact with the rest of the system. The partition function for this latter system is $\Delta(\lambda=0)$. Since the newly included particle increases the internal pressure, this process step causes the mean volume of the system to change from V^* to some new volume, say, $V_0>V^*$, against the external pressure, P. However, by virtue of the relationship given by Eq. (A11), the partition functions of the two end states of this process step are the same, and so the Gibbs free energy change associated with it is zero, even though the mean volume has increased..[102]

The third step consists in taking the system from uncoupled ($\lambda=0$) to fully coupled ($\lambda=1$), resulting in a free energy change of ΔG_{sim}, given by the second term on the right-hand side of Eq. (A12). The fourth step consists of making the special particle indistinguishable from the remaining N_s^l-1, with a consequent free energy change given by the first term on the right-hand side of Eq. (A12).

Although Eq. (A12) is formally correct, in order to be useful one needs a relationship for the chemical potential of the solute that is in terms of its concentration at full coupling. In order to obtain such a relationship we add and subtract $kT\ln V_1$ on the right-hand side to obtain

$$\mu_s^l(N_s^l,N_w,P,T)=\Delta G_{\text{sim}}-kT\ln(V^*/V_1)$$

$$-kT\ln\{\Lambda^{-3}V_1q_s(T)/N_s^l\}, \tag{A14}$$

where we let V_1 be the mean volume of the $(N_s^l+N_w)$-molecule system at full coupling. Like ΔG_{sim}, both V_1 and V^* can be computed from simulations. The difference between V_1 and V^* is, of course, related to the partial molar volume of the solute.

ΔG_{sim} can be computed using any one of the various methods available for determining free energy changes. Our calculations use thermodynamic integration,

$$\Delta G_{sim} = -kT \ln\left\{\frac{\Delta(N_s^l, N_w, P, T, \lambda=1)}{\Delta(N_s^l, N_w, P, T, \lambda=0)}\right\}$$

$$= -kT \int_0^1 d\lambda \frac{d\ln\Delta(\lambda)}{d\lambda}$$

$$= \int_0^1 d\lambda \left\langle \frac{dH(\lambda)}{d\lambda}\right\rangle_{N_s^l, N_w, P, T, \lambda}, \qquad (A15)$$

where the angle brackets indicate that an isobaric–isothermal ensemble average is to be performed at the conditions indicated by the subscript.

The Gibbs free energy change, $\Delta G_{g\to l}$, for the process of transferring a molecule of solute from a (ideal) gas phase having a partial pressure of P_s into the solution to produce a concentration of ρ_s^l in the liquid phase is given by the difference of the chemical potentials in the gas and the liquid phase. We combine Eqs. (A14) and (A6), to obtain

$$\Delta G_{g\to l} = -\mu_s^g + \mu_s^l = \Delta G_{sim} - kT\ln(V^*/V_1)$$

$$- kT\ln\{\Lambda^{-3} V_1 q_s(T)/N_s^l\}$$

$$+ kT\ln\{\Lambda^{-3} q_s(T)kT/P_s\}$$

$$= \Delta G_{sim} - kT\ln(V^*/V_1) - kT\ln(P_s/\rho_s^l kT). \qquad (A16)$$

At *equilibrium*, the concentrations of s in the gas and liquid phases adjust until $\Delta G_{g\to l}$ is zero, yielding a relationship between experimentally measurable and computable quantities,

$$\Delta G_{sim} - kT\ln(V^*/V_1) = kT\ln(P_s/\rho_s^l kT). \qquad (A17)$$

In fact, using the relationship in Eq. (A1), we see immediately the relationship between the experimentally reported solvation free energies, ΔG_{solv}, and quantities computed by simulation,

$$\Delta G_{solv} = \Delta G_{sim} - kT\ln(V^*/V_1). \qquad (A18)$$

This relationship is the key result of this Appendix. With Eq. (A1), ΔG_{solv} is determined by experiments that measure equilibrium concentrations of a solute in the liquid and gas phases. And with Eq. (A18) above, it is possible to compute ΔG_{solv} and compare this computed result with that from experiment.

For free energy changes due to similar transfer processes but that are relative to other reference states, such as ones where the gas phase has a partial pressure of s of one atmosphere, or where the concentration of s in the solution is one molar, we can simply substitute the relevant concentration properties of the reference states into Eq. (A16).

The formulation described in this section has a number of desirable aspects. It develops a rigorous procedure for the computation of chemical potentials and absolute solvation free energies that is applicable even when the solutions are nonideal. The relationships between these chemical potentials and solvation free energies relative to the same at different reference conditions is readily apparent. During the simulations implied by this formulation, the molecule being solvated is kept intact by virtue of the fact that the intramolecular potential of the solute molecule is not affected by the coupling process. In this formulation, there is no requirement for the solute molecule to be rigid, or otherwise conformationally constrained. The solute responds fully to the solvent, and consequently, any possible effects on the solvation free energy due to solvent-induced changes in the energetics of solute conformations are properly treated.

Finally, we note that the above formulation is applicable even if $N_w = 0$, in which case one obtains relationships between the vapor pressure of a pure liquid in terms of the coupling energy, ΔG_{sim}, for the addition of a new molecule into the pure liquid. For hydration free energies of amino acid side chain analogs, where the equilibrium solute concentrations are small (dilute) and the solute molecules are not expected to aggregate, the appropriate simulations involving the above formulation use $N_s = 1$ and, of course, pure water as the solvent.

[1] C. Chipot and D. A. Pearlman, Mol. Simul. **28**, 1 (2002), and references therein.

[2] D. A. Pearlman and P. R. Connelly, J. Mol. Biol. **248**, 696 (1995).

[3] D. A. Pearlman and P. A. Kollman, J. Chem. Phys. **91**, 7831 (1989).

[4] C. Chipot, P. A. Kollman, and D. A. Pearlman, J. Comput. Chem. **17**, 1112 (1996).

[5] H. J. C. Berendsen, in *Proteins: Structure, Dynamics, and Design*, edited by V. Renugopalakrishnan, P. R. Carey, I. C. P. Smith, S. G. Huang, and A. C. Storer (ESCOM, Leiden, 1991), pp. 384–392.

[6] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Mertz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).

[7] J. A. D. MacKerell, D. Bashford, M. Bellot *et al.*, J. Phys. Chem. B **102**, 3586 (1998).

[8] 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).

[9] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).

[10] D. J. Price and C. L. Brooks, J. Comput. Chem. **23**, 1045 (2002).

[11] M. Karplus and J. A. McCammon, Nat. Struct. Biol. **9**, 646 (2002).

[12] T. Hansson, C. Oostenbrink, and W. F. Van Gunsteren, Curr. Opin. Struct. Biol. **12**, 190 (2002).

[13] B. Chen and J. I. Siepmann, J. Phys. Chem. B **103**, 5370 (1999).

[14] B. Chen, J. J. Potoff, and J. I. Siepmann, J. Phys. Chem. B **105**, 3093 (2001).

[15] C. D. Wick, M. G. Martin, and J. I. Siepmann, J. Phys. Chem. B **104**, 8008 (2000).

[16] S. K. Nath, F. A. Escobedo, and J. J. de Pablo, J. Chem. Phys. **108**, 9905 (1998).

[17] J. R. Errington and A. Z. Panagiotopoulos, J. Phys. Chem. B **103**, 6314 (1999).

[18] R. Wolfenden, L. Andersson, P. M. Cullis, and C. C. B. Southgate, Biochemistry **20**, 849 (1981).

[19] A. Ben-Naim and Y. Marcus, J. Chem. Phys. **81**, 2016 (1984).

[20] S. Cabani, P. Gianni, V. Mollica, and L. Lepori, J. Solution Chem. **10**, 563 (1981).

[21] J. Hine and P. K. Mookerjee, J. Org. Chem. **40**, 292 (1975).

[22] M. H. Abraham, G. S. Whiting, R. Fuchs, and E. J. Chambers, J. Chem. Soc., Perkin Trans. 2 **2**, 291 (1990).

[23] V. N. Viswanadhan, A. K. Ghose, U. C. Singh, and J. J. Wendoloski, J. Chem. Inf. Comput. Sci. **39**, 405 (1999).

[24] In this paper, the force fields are listed alphabetically.

[25] S. E. DeBolt and P. A. Kollman, J. Am. Chem. Soc. **117**, 5316 (1995).

[26] T. Fox and P. A. Kollman, J. Phys. Chem. B **102**, 8070 (1998).

[27] I. G. Tironi and W. F. Van Gunsteren, Mol. Phys. **83**, 381 (1994).

[28] J. E. Eksterowicz, J. L. Miller, and P. A. Kollman, J. Phys. Chem. B **101**, 10971 (1997).

[29] Additionally, the time scales important for equilibration and sampling can be longer than for water. For example, our initial experiments with blocked amino acids in a 1-octanol/water mixture suggested correlation times for some degrees of freedom of at least 100s of picoseconds.

[30] C. Chothia, J. Mol. Biol. **105**, 1 (1976).

[31] A. Radzicka and R. Wolfenden, Biochemistry **27**, 1664 (1988).

[32] J. S. Bader and D. Chandler, J. Phys. Chem. **96**, 6424 (1992).

[33] H. A. Carlson, T. B. Nguyen, M. Orozco, and W. L. Jorgensen, J. Comput. Chem. **14**, 1240 (1993).

[34] A. Villa and A. E. Mark, J. Comput. Chem. **23**, 548 (2002).

[35] G. Kaminski, E. M. Duffy, T. Matsui, and W. L. Jorgensen, J. Phys. Chem. **98**, 13077 (1994).

[36] W. L. Jorgensen and T. B. Nguyen, J. Comput. Chem. **14**, 195 (1993).

[37] V. Helms and R. C. Wade, J. Am. Chem. Soc. **120**, 2710 (1998).

[38] B. Chen and J. I. Siepmann, J. Am. Chem. Soc. **122**, 6464 (2000).

[39] Produced and distributed by the Professor Jay Ponder's research group at Washington University. Documentation and programs for TINKER are available at http://dasher.wustl.edu/tinker/.

[40] M. Shirts and V. S. Pande, Science **290**, 1903 (2000).

[41] 90 000 is the number of processors running Folding@Home worldwide at the time the paper was written. At the time the simulations were run, the number was closer to 30 000.

[42] C. D. Snow, N. Nguyen, V. S. Pande, and M. Gruebele, Nature (London) **420**, 102 (2002).

[43] B. Zagrovic, E. J. Sorin, and V. S. Pande, J. Mol. Biol. **313**, 151 (2001).

[44] V. S. Pande, I. Baker, J. Chapman, S. P. Elmer, S. Khaliq, S. M. Larson, Y. M. Rhee, M. R. Shirts, C. D. Snow, E. J. Sorin, and B. Zagrovic, Biopolymers **68**, 91 (2003).

[45] B. Zagrovic, C. D. Snow, S. Khaliq, M. R. Shirts, and V. S. Pande, J. Mol. Biol. **323**, 153 (2002).

[46] S. M. Larson, C. D. Snow, M. R. Shirts, and V. S. Pande, in *Computational Genomics*, edited by R. P. Grant (Horizon Scientific, Wymondham, Norfolk, U.K., 2002) (in press).

[47] S. M. Larson, J. L. England, J. R. Dejarlais, and V. S. Pande, Protein Sci. **11**, 2804 (2002).

[48] H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).

[49] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, Chem. Phys. Lett. **222**, 529 (1994).

[50] P. A. Kollman, R. Dixon, W. Cornell, T. Fox, C. Chipot, and A. Pohorille, in *Computer Simulation of Biomolecular Systems*, edited by A. Wilkinson, P. Weiner, and W. F. van Gunsteren (Elsevier, Amsterdam, The Netherlands, 1997), Vol. 3, pp. 83–96.

[51] G. A. Kaminski, R. A. Friesner, J. Rives, and W. L. Jorgensen, J. Phys. Chem. B **105**, 6474 (2001).

[52] H. Resat and J. A. McCammon, J. Chem. Phys. **108**, 9617 (1998).

[53] T. P. Straatsma and H. J. C. Berendsen, J. Chem. Phys. **89**, 5876 (1988).

[54] See EPAPS Document No. E-JCPSA6-119-521329 for parameter files and a brief description. A direct link to this document may be found in the online article's HTML reference section. The document may also be reached via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html) or from ftp.aip.org in the directory /epaps/. See the EPAPS homepage for more information.

[55] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[56] M. L. Mahoney and W. L. Jorgensen, J. Chem. Phys. **112**, 8910 (2000).

[57] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, J. Chem. Phys. **76**, 637 (1982).

[58] H. C. Andersen, J. Comput. Phys. **52**, 24 (1983).

[59] J. P. Ryckaert and G. Ciccotti, Mol. Phys. **58**, 1125 (1986).

[60] J. P. Ryckaert and G. Ciccotti, J. Chem. Phys. **78**, 7368 (1983).

[61] G. Ciccotti and J. P. Ryckaert, Comput. Phys. Rep. **4**, 345 (1986).

[62] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).

[63] To reduce the effect of the noise that occurs when the correlation function approaches 0, the tail of the function was replaced by a linear fit (for the purposes of integration) between the first time the normalized correlation function reaches 0.1 and the first time it reaches 0.0 by connecting these time points. Inspection indicated that a linear approximation is more accurate than an exponential fit. This process is simplistic (and fails if the correlation function is too far from monotonically decreasing in the tail), but is of sufficient accuracy here, since we ultimately require only one significant digit in the determination of uncertainty.

[64] P. J. Steinbach and B. R. Brooks, J. Comput. Chem. **15**, 667 (1994).

[65] In some simulations, the electrostatic potential is shifted before the switch is applied in order to minimize forces arising from sharply tapered potentials (Ref. 64). However, since all groups are neutral in these simulations, a shift actually has no net effect on intergroup interactions.

[66] *CRC Handbook of Chemistry and Physics*, edited by D. R. Lide (CRC, Boston, 2001).

[67] W. L. Jorgensen and C. Jenson, J. Comput. Chem. **19**, 1179 (1998).

[68] An important but rarely commented upon artifact of periodic boundary conditions in that there is a system size dependence for any observable that has nonisotropic components (Refs. 103–105). The density of water depends on the nonisotropic dipole–dipole interaction between water molecules, especially in conjunction with finite-range treatments of electrostatic interactions. A recent study of TIP5P water (Ref. 75), as well as our unpublished results have found this size-dependence of density with finite-ranged Coulombic potentials, but more work is necessary to understand this phenomenon.

[69] T. P. Straatsma and J. A. McCammon, J. Chem. Phys. **95**, 1175 (1991).

[70] A. Dejaegere and M. Karplus, J. Phys. Chem. **100**, 11148 (1996).

[71] J. W. Pitera and W. F. Van Gunsteren, Mol. Simul. **28**, 45 (2002).

[72] A. E. Mark and W. F. Van Gunsteren, J. Mol. Biol. **240**, 167 (1994).

[73] J. T. Wescott, L. R. Fisher, and S. Hanna, J. Chem. Phys. **116**, 2361 (2002).

[74] See Ref. 54.

[75] M. Lisal, J. Kolafa, and I. Nezbeda, J. Chem. Phys. **117**, 8892 (2002).

[76] P. Mark and L. Nilsson, J. Comput. Chem. **23**, 1211 (2002).

[77] A. R. Leach, *Molecular Modelling: Principles and Applications* (Addison–Wesley Longman Limited, Harlow, Essex, England, 1996).

[78] H. Resat and M. Mezei, J. Chem. Phys. **99**, 6052 (1993).

[79] P. E. Smith and W. F. Van Gunsteren, J. Chem. Phys. **100**, 577 (1994).

[80] C. Simmerling, T. Fox, and P. A. Kollman, J. Am. Chem. Soc. **120**, 5771 (1998).

[81] C. D. Wick and J. I. Siepmann, Macromolecules **33**, 7207 (2000).

[82] Z. Chen and F. A. Escobedo, J. Chem. Phys. **113**, 11382 (2000).

[83] N. Rathore, T. A. Knotts, and J. J. de Pablo, J. Chem. Phys. **118**, 4285 (2003).

[84] V. G. Mavrantzas, T. D. Boone, E. Zervopoulou, and D. N. Theodorou, Macromolecules **32**, 5072 (1999).

[85] J. P. Ulmschneider and W. L. Jorgensen, J. Chem. Phys. **118**, 4261 (2003).

[86] D. Yin and J. Alexander D. Mackerell, J. Comput. Chem. **19**, 334 (1998).

[87] OPLS-AA and AMBER used some free energy simulations to gauge the general applicability of the force field for free energy calculations, but force fields were not explicitly tuned to give correct free energies—nor could they, with the precision available at the time.

[88] Some recasting of the experimental data is necessary in order to convert the experimental results to the form used in this paper. In Abraham *et al.* (Ref. 22), and in Hine and Mookerjee (Ref. 21) the desired hydration free energy can be obtained by multiplying the Oswalt coefficient by $-kT \ln 10$. With the data from Plyasunov and Shock (Ref. 89) we can convert from $\Delta_h G^o$ to a hydration free energy by subtracting $RT \ln(V_g/V_l)$, which can be approximated by neglecting the volume in liquid state and assuming ideality the gas (Ref. 19) with an error of less than 0.01 kcal/mol.

[89] A. V. Plyasunov and E. L. Shock, Geochim. Cosmochim. Acta **64**, 439 (2000).

[90] W. L. Jorgensen, J. D. Madura, and C. J. Swenson, J. Am. Chem. Soc. **106**, 6638 (1984).

[91] C. L. Lin and R. H. Wood, J. Phys. Chem. **100**, 16399 (1996).

[92] Y. Sugita, A. Kitao, and Y. Okamoto, J. Chem. Phys. **113**, 6042 (2000).

[93] A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers **60**, 96 (2001).

[94] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).

[95] C. Bartels, M. Schaefer, and M. Karplus, J. Chem. Phys. **111**, 8048 (1999).

[96] M. D. Beachy, D. Chasman, R. B. Murphy, T. A. Halgren, and R. A. Friesner, J. Am. Chem. Soc. **119**, 5908 (1997).

[97] T. A. Halgren, J. Comput. Chem. **20**, 730 (1999).

[98] P. M. King, Mol. Phys. **94**, 717 (1998).

[99] D. A. McQuarrie, *Statistical Mechanics* (Harper and Row, New York, 1976).

[100] ΔG_{solv} is referred to as $W(s|l)$ in the work of Ben-Naim and Marcus (Ref. 19). They refer to this as the coupling work, or the Gibbs free energy of interaction of s with the rest of the solution.

[101] See discussions in Refs. 106 and 107. In general, all observable thermodynamic quantities are expressed as logarithms of *ratios* of partition functions, so volume scale factors cancel and the particular choice has no consequence. A particularly useful value of V' is kT/P_s, since this choice gives the purely extensive expression for the Gibbs free energy of an ideal gas in Eq. (15). However, other choices give identical results for chemical potentials and solvation free energies.

[102] The work described in Wescott *et al.* (Ref. 73) employed a different

coupling scheme, whereby the contribution to the internal pressure from the "internal energy" of the special molecule of type s was also coupled. Using the terminology of the current study, their coupling scheme results in a mean volume at $\lambda = 0$ of V^*, whereas in our coupling scheme the mean volume at $\lambda = 0$ is V_0. We note first that the contribution to the internal pressure from the *relative* kinetic energy and the intramolecular contributions to the virial from the special molecule have a time average of 0 at all values of λ, and thus need not be considered. Second, we also couple the contributions to the internal pressure from the intermolecular interactions between the special molecule and the rest of the material simply by virtue of the fact that those energy and force components and consequently also their contribution to the virial are coupled. The difference between their treatment and ours is in the handling of the contribu-

tion to the internal pressure from the center-of-mass kinetic energy of the special molecule. However, since the effect of coupling this contribution corresponds to the second step in our four-step process, its contribution to the free energy of solvation will be zero. Thus, the contribution to the internal pressure from the molecular internal energy of the coupled molecule does not actually need to be included.

[103] L. R. Pratt and S. W. Haan, J. Chem. Phys. **74**, 1864 (1981).

[104] L. R. Pratt and S. W. Haan, J. Chem. Phys. **74**, 1873 (1981).

[105] M. W. Evans, Comput. Phys. Commun. **59**, 495 (1990).

[106] T. L. Hill, *Statistical Mechanics: Principles and Selected Applications* (McGraw–Hill, New York, 1956).

[107] T. L. Hill, *An Introduction to Statistical Thermodynamics* (Addison–Wesley, Reading, Massachusetts, 1960).