THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

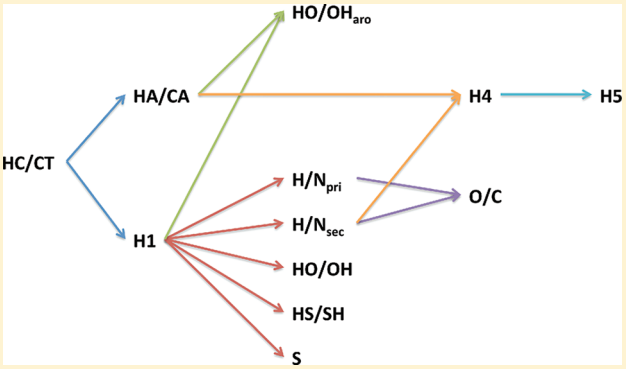
Optimizing Solute–Water van der Waals Interactions To Reproduce Solvation Free Energies

Paul S. Nerenberg,[†,§] Brian Jo,[‡] Clare So,[‡] Ajay Tripathy,[‡] and Teresa Head-Gordon*[,†,‡]

[†]California Institute of Quantitative Biosciences and [‡]Department of Bioengineering, University of California, Berkeley, Berkeley, California 94720-3220, United States

Supporting Information

ABSTRACT: An accurate representation of solute–water interactions is necessary for molecular dynamics simulations of biomolecules that reside in aqueous environments. Modern force fields and advanced water models describe solute–solute and water–water interactions reasonably accurately but have known shortcomings in describing solute–water interactions, demonstrated by the large differences between calculated and experimental solvation free energies across a range of peptide and drug chemistries. In this work, we introduce a method for optimizing solute–water van der Waals interactions to reproduce experimental solvation free energy data and apply it to the optimization of a fixed charge force field (AMBER ff99SB/GAFF) and advanced water model (TIP4P-Ew). We show that, with these optimizations, the combination of AMBER ff99SB/GAFF and TIP4P-Ew is able to reproduce the solvation free energies of a variety of biologically relevant small molecules to within 1.0 k_BT. We further validate these optimizations by examining the aggregation propensities of dipeptide–water solutions, the conformational preferences of short disordered peptides, and the native state stability and dynamics of a folded protein.



INTRODUCTION

Over the last three decades, molecular dynamics (MD) simulations have emerged as an important tool for investigating most aspects of biomolecular structure and function and can be used to both directly calculate experimental observables and analyze properties that are otherwise inaccessible to experiment. Currently, MD simulations offer the ability to observe the behavior of biomolecules at an atomistic level of detail over time scales ranging from femtoseconds to microseconds, thereby contributing valuable insights into such diverse areas as protein folding and stability, ligand binding, protein–protein and protein–nucleic acid interactions, and enzyme catalysis.[1−3]

The vast majority of MD simulation studies performed to date have used classical pairwise fixed-charge force fields. Many of the fixed-charge force fields currently in use were first developed more than 15 years ago,[4−6] and their continued refinements have yielded simulation results that are in qualitative and sometimes quantitative agreement with a wide range of reference data.[7−13] However, continued improvements in both simulation methodologies and computer hardware have allowed for more complete sampling to reveal discrepancies in backbone conformational preferences,[14,15] side-chain conformational preferences,[10,16] and solvation free energies[17−19] when compared to quantitative experimental data. While developing more sophisticated electrostatic models is undoubtedly a necessary component for improving empirical biomolecular force fields,[19,20] several studies have suggested that more advanced fixed-charge, polarizable, or even hybrid electrostatic models are insufficient on their own to systematically remedy solvation free energies for small organic molecules whose chemistries are representative of those found in larger biomolecules.[19−22]

Historically, the commonly used van der Waals parameters for the atom types of organic molecules were derived by matching the densities and enthalpies of vaporization of various neat liquids.[4,6] However, when these parameters are used to calculate the solvation free energy of methane in liquid water, for example, they result in solvation free energies that are 0.3–0.6 kcal/mol less favorable than experiment – a significant amount given that methane is such a simple molecule.[19,20,23] Calculations with more complex small molecules (e.g., amides or alcohols) demonstrate even larger errors on the order of 1–2 kcal/mol.[17−19,24] (One notable exception to this trend is the GROMOS 53A6 force field, which was optimized specifically to reproduce solvation free energies at the expense of neat organic liquid properties.[25]) Summed over an entire protein with hundreds of such moieties, it becomes apparent how these sorts of errors could lead to the observed quantitative discrepancies. Although this was probably the best strategy for parametrizing van der Waals interactions at a time when there was more limited parametrization and validation data, newer experimental

Received: December 8, 2011
Revised: March 12, 2012
Published: March 23, 2012

 © 2012 American Chemical Society dx.doi.org/10.1021/jp2118373 | J. Phys. Chem. B 2012, 116, 4524–4534

data collected on biomolecules in the presence of aqueous solvent is now available with the potential to yield a significant quantitative advance in a new generation fixed charge force field. Such an "alternative" philosophy of force field parametrization appears to be gaining traction within the biophysical community.[11−13,26,27]

This study is part of a larger goal to create a next generation fixed-charge force field based on an advanced water model that could remedy the quantitative disagreement between simulations and experimental data (e.g., NMR scalar couplings and fractional helicities, solvation free energies, and high temperature unfolding data). Such a force field would be more accurate in the simulation of stable biomolecular structure and dynamics in aqueous solution and more broadly capable of describing unfolded protein ensembles and intrinsically disordered peptides and proteins. Any improvements in quantitative prediction via simulation would serve to enable a truly fruitful interplay between simulation and experiment. Here we develop new solute−water van der Waals interactions between the popular AMBER ff99SB protein force field[9] and the robust TIP4P-Ew water model[28] which shows excellent agreement with many reference experimental data over a large range of temperatures and pressures. We demonstrate that the combination of AMBER ff99SB and TIP4P-Ew with these optimizations to the solute−water interactions is able to reproduce the solvation free energies of a variety of biologically relevant small molecules to within 1.0 k_BT. We further validate these optimizations by examining the aggregation propensities of dipeptide−water solutions, the conformational preferences of short disordered peptides, and the native state stability and dynamics of ubiquitin.

METHODS

Charge Derivation for Small Molecules. Semiempirical Merck-Frosst AM1-BCC partial atomic charges[29,30] for all 47 small molecules were obtained from the Supporting Information of Mobley et al.[18] AMBER-compatible ab initio partial atomic charges for these molecules were derived by optimizing their geometries and calculating electrostatic potentials at the HF/6-31G* level of theory and then fitting charges to these potentials using the RESP method,[31] as implemented with RED Tools III-4.[32] We have included mol2 files with the HF/6-31G* charges in the Supporting Information. All ab initio calculations were performed using GAMESS-US.[33]

Dipole−Dipole Polarization Cost Calculations. Recent work by Swope et al. has demonstrated a straightforward method for calculating the energetic cost of polarizing (fixed) charge distributions from the gas phase to the condensed phase.[34] Because of known deficiencies in representing higher order multipoles with only atom-centered partial charges,[19] we approximate this cost using just the dipole−dipole polarization cost

$$W_{\mathrm{pol}} \approx W_{\mathrm{pol}}^{\mathrm{d-d}} = \frac{1}{2}(\vec{\mu}_{\mathrm{aq}} - \vec{\mu}_{\mathrm{gas}})^T(\underline{\underline{\alpha}}^{-1})^T(\vec{\mu}_{\mathrm{aq}} - \vec{\mu}_{\mathrm{gas}}) \quad (1)$$

where $\vec{\mu}_{\mathrm{gas}}$ is the gas phase dipole, $\vec{\mu}_{\mathrm{aq}}$ is the aqueous phase dipole (note that eq 1 is applicable for any condensed phase system although this work focuses on the aqueous phase), and $\underline{\underline{\alpha}}$ is the molecular dipole−dipole polarizability tensor. Aqueous-phase dipole moments were computed directly from the AM1-BCC or HF/6-31G*-derived charge distributions. Gas-phase dipole moments and dipole−dipole polarizability tensors were computed from the geometry-optimized structures (described above) at the MP2/aug-cc-pV(T+d)Z level of theory.[34]

Solvation Free Energy Calculations. Molecules were solvated in truncated octahedral boxes with approximately 500−800 TIP3P[35] or TIP4P-Ew water molecules (depending on the size of the molecule) using the tleap module of AmberTools 1.4. Solvation free energies were calculated by a simulated decoupling of the molecules from solvent using thermodynamic integration with softcore potentials for both van der Waals and electrostatic interactions as implemented in AMBER 11.[36] Differences in the potential energies were calculated at 12 λ values, spaced according to a Gaussian−Legendre quadrature between 0 and 1, and integrated to find the solvation free energy. For each λ value, the system first underwent 1000 steps of conjugate gradient (CG) minimization. The system was then equilibrated in the NVT ensemble for 50 ps at 298 K using a Langevin thermostat with a coupling constant of 1.0 ps^{-1} followed by an equilibration in the NPT ensemble for 100 ps at 298 K at 1.0 bar with a Langevin thermostat coupling constant of 2.0 ps^{-1} and a Berendsen barostat coupling constant of 2.0 ps^{-1}. Production runs in the NPT ensemble of 2 ns were then carried out using the same thermostat and barostat settings. All simulations utilized periodic boundary conditions with a 9-Å nonbonded cutoff for direct-space nonbonded interactions. Long-range electrostatics were calculated using particle mesh Ewald (PME) with default parameters for grid spacing and spline interpolation, and an analytic correction was employed for the van der Waals interactions beyond the cutoff. Dynamics were conducted with a 2-fs time step, and all bonds involving hydrogen atoms were constrained with SHAKE. Representative $dU/d\lambda$ curves for these calculations are shown in Figure 1 of Supporting Information to verify that the number of λ values was adequate.

Simulations of Dipeptide Solutions. Solutions of 1.5 M *N*-acetyl-glycine-methyl-amide (NAGMA or glycine dipeptide) and 1.0 M *N*-acetyl-leucine-methyl-amide (NALMA or leucine dipeptide) were constructed by building cubic simulation boxes with 48 or 32 dipeptide molecules solvated by 1760 TIP4P-Ew water molecules. Each system first underwent 1000 steps of CG minimization with harmonic restraints (10.0 kcal mol^{-1} $Å^{-2}$) on the NAGMA or NALMA molecules, followed by another 1000 steps of CG minimization with no restraints. The system was then equilibrated in the NVT ensemble for 50 ps at 298 K using a Langevin thermostat with a coupling constant of 1.0 ps^{-1}, followed by an equilibration in the NPT ensemble for 200 ps at 298 K at 1.0 bar with a Langevin thermostat coupling constant of 2.0 ps^{-1} and a Berendsen barostat coupling constant of 2.0 ps^{-1}. During these two equilibration simulations, the NAGMA or NALMA molecules were harmonically restrained with force constants of 10.0 and 2.0 kcal mol^{-1} $Å^{-2}$, respectively. A final equilibration in the NPT ensemble with no restraints was performed for 2 ns, followed by a production simulation of 60 ns using the same thermostat and barostat settings. Snapshots were saved every 1 ps for radial distribution function analysis. The nonbonded interactions, time step, and SHAKE settings were the same as the solvation free energy calculations previously described.

Simulations of Short Peptides. Four independent 150-ns MD simulations were performed of the Gly_3 peptide in TIP4P-Ew water. Similarly, two independent 50 ns replica exchange MD simulations were performed of the Val_3 and Ala_5 peptides in TIP4P-Ew water. For all three peptides, simulations were performed using the optimized solute−water van der Waals

dx.doi.org/10.1021/jp2118373 | J. Phys. Chem. B 2012, 116, 4524−4534

parameters developed in this work. The exact methodology for both types of simulations is described in our previous work regarding optimized backbone dihedral angle potentials for AMBER ff99SB with TIP4P-Ew water.[11] NMR scalar (or J) couplings were then calculated from the simulation snapshots, and an overall χ^2 value for the resulting conformational ensemble was determined using

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{(\langle J_i\rangle_{sim} - J_{i,expt})^2}{\sigma_i^2} \quad (2)$$

where N is the number of J coupling constants measured, $\langle J_i\rangle_{sim}$ is the ith calculated J coupling constant averaged over all structures in the simulated ensemble, $J_{i,expt}$ is the ith experimental J coupling constant, and σ_i is the uncertainty in $\langle J_i\rangle_{sim}$, which we expect to be dominated by the uncertainty in the Karplus parameters in addition to the inherent experimental uncertainty in $J_{i,expt}$ and sampling uncertainty in $\langle J_i\rangle_{sim}$, which we neglect.[15]

Simulations of Ubiquitin. A 100-ns MD simulation of the folded ubiquitin protein (PDB ID: 1UBQ) in TIP4P-Ew water was performed using the optimized solute−water van der Waals parameters. The exact methodology for this simulation is described in our previous work.[11] Similar simulations were also performed using a 10−12 potential for backbone hydrogen bonds within the protein as described further in Results.

RESULTS

Parameterization Approach. We developed a training set and a validation set based on high-quality experimental solvation free energies of 47 small molecules that incorporate all of the chemical functionalities of standard protein side chains and backbone groups. The 47 small molecules were further separated into 12 categories based on their functional groups. Each category contained four molecules with the exception of the imidazoles, for which we had experimental solvation free energies for only three molecules. Two molecules from each category were chosen as the training set for solute−water van der Waals (vdW) parameter optimization detailed below. Finally, the optimized parameters were validated by calculating the solvation free energies of the two molecules in the given category not used for optimization (one molecule in the case of imidazoles).

While we would ideally optimize all of the vdW parameters simultaneously, this is currently beyond the capabilities of our computational resources. As such, we elected to optimize the parameters in a stepwise fashion, beginning with the simplest organic molecules alkanes (atom types HC and CT) and progressing to more complex molecules containing other moieties (Figure 1). Additionally, given limited experimental data, whenever two atom types were involved for a given chemical moiety (e.g., HC and CT for alkanes) we elected to use the same optimization constants (see below) for both atom types. Although in principle better agreement should be achievable with unique constants for each atom type, our approach reduces the chance of overfitting the parameters to the available experimental data and limits the total number of new parameters introduced into the force field. Moreover, as we demonstrate below, this approach is sufficient to achieve k_BT accuracy in calculated solvation free energies.

Recently Baker et al. used an approximate thermodynamic cycle to derive optimal pair-specific Lennard-Jones parameters for the CHARMM polarizable force field.[21] In this work we employ a similar methodology and briefly review the formalism used. In the thermodynamic cycle shown in Figure 2, the solvation free energy using a new set of van der Waals parameters, ΔG^{new}_{solv}, can be related to the solvation free energy using the original parameters, ΔG^{orig}_{solv}, and the changes in free energy associated with the change in parameters in both the gas and aqueous phases, ΔG^{Δ}_{gas} and ΔG^{Δ}_{aq}, respectively. As this is a closed thermodynamic cycle, the following is true

$$-\Delta G^{orig}_{solv} + \Delta G^{\Delta}_{gas} + \Delta G^{new}_{solv} - \Delta G^{\Delta}_{aq} = 0 \quad (3)$$

Because the new van der Waals parameters apply only to solute−water interactions

$$\Delta G^{\Delta}_{gas} = 0 \Rightarrow \Delta G^{new}_{solv} = \Delta G^{orig}_{solv} + \Delta G^{\Delta}_{aq} \quad (4)$$

From the polarization cost and solvation free energy calculations described above, we have

$$\Delta G^{orig}_{solv} = \Delta G^{MD(orig)}_{solv} + W_{pol} \quad (5)$$

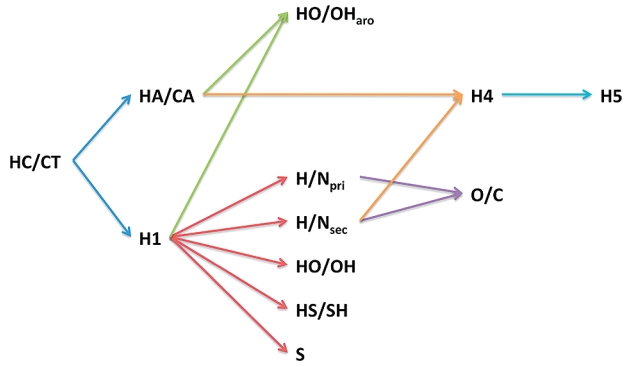


Figure 1. Schematic order of parameter optimization by atom type. van der Waals parameters were first optimized for alkanes (HC/CT) and then followed by hydrocarbon aromatics (HA/CA) and polar molecules (H1). With optimized parameters for HC/CT and H1, it was then possible to optimize the primary amines (H/N$_{pri}$), secondary amines (H/N$_{sec}$), alcohols (HO/OH), thiols (HS/SH), and thioethers (S). Together with the HC/CT and HA/CA parameters, it was then possible to optimize atom types for heterocyclic aromatics (H4), imidazoles (H5), and amides/ketones (O/C). Similarly, with optimized parameters for HC/CT, HA/CA, and H1, it was possible to optimize the oxygen atoms in phenols (HO/OH$_{aro}$).

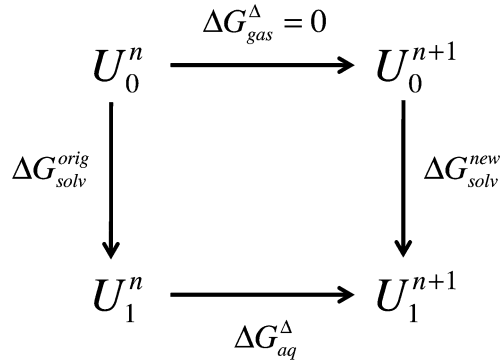


Figure 2. Thermodynamic cycle used to calculate ΔG^{new}_{solv}. U_1^n corresponds to the solvated system using nth iteration solute−water vdW parameters, U_0^n corresponds to desolvated system at using nth iteration parameters, etc. From this cycle, we can conclude that $\Delta G^{new}_{solv} = \Delta G^{orig}_{solv} + \Delta G^{\Delta}_{aq}$, since $\Delta G^{\Delta}_{gas} = 0$.

dx.doi.org/10.1021/jp2118373 | J. Phys. Chem. B 2012, 116, 4524−4534

where $\Delta G_{solv}^{MD(orig)}$ is the solvation free energy calculated from MD simulation using the original vdW parameters, and W_{pol} is the polarization cost. We can then *estimate* ΔG_{aq}^{Δ} by evaluating the energies of MD trajectory snapshots using the new vdW parameters. In the limit of small perturbations to the solute−water vdW parameters, this is a valid method of estimating ΔG_{aq}^{Δ} and therefore ΔG_{solv}^{new}. The primary benefit of this formalism is that it enables the estimate of ΔG_{solv}^{new} for many different possible vdW parameters without having to run a costly solvation free energy calculation for each one. The energy analysis used to calculate ΔG_{aq}^{Δ} (and therefore ΔG_{solv}^{new}) requires roughly 1/1000th of the CPU time necessary for a solvation free energy calculation.

For each training set molecule, we first performed a solvation free energy calculation, as well as MD of the solvated system, using the initial AMBER/GAFF vdW parameters for the interactions between the solute atoms and the oxygen atom of TIP4P-Ew water. (It is important to note that GAFF vdW parameters are identical to those of AMBER ff99SB *for the atom types found in proteins*. We will therefore refer to AMBER ff99SB and GAFF interchangeably in this work when discussing our small molecule results.) The AMBER force fields use a Lennard-Jones potential to describe van der Waals interactions

$$U_{vdW}(r_{ij}) = \sum_{i \neq j} \varepsilon_{ij}[(R_{ij}/r_{ij})^{12} - 2(R_{ij}/r_{ij})^{6}] \quad (6)$$

where the potential well depth, ε_{ij}, and excluded volume, R_{ij}, are defined by the Lorentz−Berthelot mixing rules:

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \qquad R_{ij} = R_i + R_j \quad (7)$$

We then introduced optimization constants to

$$\varepsilon_{ij}^{s-Ow} = c_n^{opt} \sqrt{\varepsilon_i \varepsilon_j} \qquad R_{ij}^{s-Ow} = d_n^{opt}(R_i + R_j) \quad (8)$$

where the superscripts s and Ow denote the solute atom and TIP4P-EW oxygen atom, respectively, and the subscript n indicates the nth iteration of the optimization process (starting from $n = 1$). We chose an initial set of values for the optimization constants, with c_1^{opt} varying from 0.8 to 1.2 and d_1^{opt} varying from 0.9 to 1.1, respectively, leading to a set of 25 c_1^{opt},d_1^{opt} pairs—or 25 different solute−water vdW interactions—"centered on" the original solute−water vdW parameters. We then conducted a free energy analysis of the MD trajectory snapshots of the solvated system, using each one of these pairs to estimate $\Delta G_{aq}^{\Delta,n=1}$ and therefore $\Delta G_{solv}^{new,n=1} = \Delta G_{solv}^{orig} + \Delta G_{aq}^{\Delta,n=1}$. The c_1^{opt}, d_1^{opt} pair that minimized the root-mean-square error (RMSE) between $\Delta G_{solv}^{new,n=1}$ and the experimental solvation free energies for the two "training" molecules was then selected as the starting point for the next iteration of optimization, and the process of finding optimal parameters was repeated. In other words, we generated a new set of 25 c_2^{opt},d_2^{opt} pairs (now centered on c_1^{opt},d_1^{opt}), conducted a new solvation free energy calculation to accurately determine $\Delta G_{solv}^{new,n=1}$, used free energy analysis to estimate $\Delta G_{aq}^{\Delta,n=2}$ and therefore $\Delta G_{solv}^{new,n=2} = \Delta G_{solv}^{new,n=1} + \Delta G_{aq}^{\Delta,n=2}$, and then selected the c_2^{opt},d_2^{opt} pair that minimized the RMSE to the experimental solvation free energies. We repeated this process of finding more optimal c_n^{opt}, d_n^{opt} pairs until we reached convergence to a RMSE of less than 0.15 kcal/mol ($^1/_4$ k_BT), which generally required 3−5 iterations of "new" solute−water vdW parameters. In some cases, it was not possible to reduce the RMSE to 0.15 kcal/mol, and we instead terminated the optimization procedure when changes in RMSE from one iteration to the next were on the order of the statistical uncertainties in the solvation free energies (i.e., there was no statistically significant improvement with further iteration of the parameters). The solute−water vdW parameter optimization procedure was applied to the various AMBER protein atom types in the order described by Figure 1.

Benchmark Calculations. To validate the use of softcore potentials for both electrostatic and van der Waals interactions, as well as the single-step decoupling of solute molecules from solvent, we first calculated the solvation free energies of the 47 small molecules in TIP3P water using the Merck-Frosst AM1-BCC charges used by Mobley et al.[18] When comparing our results to their previous free energy calculations we achieve excellent agreement, with a root mean square (rms) difference of 0.13 kcal/mol, in spite of the different MD software packages, different solvent box geometries and numbers of solvent molecules, single-step decoupling vs the more traditional three-step transformation, and the use of TI vs BAR. Relative to experiment, we achieve a RMSE of 1.43 kcal/mol when using TIP3P, as compared to Mobley et al.'s 1.49 kcal/mol (Figure 3, Table 1 of Supporting Information). RMSEs are unsigned, however, and therefore do not reflect if the error is due to solute−solvent interactions being too favorable or too unfavorable. Thus it is worth noting that the calculated solvation free energies of these molecules are all too *unfavorable* relative to experiment—the mean signed errors are 1.21 (our work) and 1.28 kcal/mol (Mobley et al.)—with the exceptions of naphthalene and 3-methylindole (Figure 3). Because of the shorter simulation lengths used, our sampling errors are 2−3× greater than that of Mobley et al. but are still less than 0.06 kcal/mol ($^1/_{10}$ k_BT). These data suggest that the protocol we use to calculate solvation free energies is sufficiently robust and accurate for the purposes of parametrization.

To compare the performance of semiempirical vs ab initio-derived charge models, we next calculated the solvation free energies for the same molecules using 6-31G*/RESP charges (i.e., the same charge model used in the AMBER ff99SB force field). Overall, the RMSE to experiment is 1.53 kcal/mol, and in most cases, the solvation free energies are again generally unfavorable relative to experiment (mean signed error of 1.24 kcal/mol) (Figure 3 and Table 1 of Supporting Information). While there is little difference in the solvation free energies computed for the alkanes, there is an immediate difference for purely hydrocarbon aromatic compounds, which become less favorable than the AM1-BCC benchmark. Another interesting difference between the two charge models is that primary amine compounds become more favorable with 6-31G*/RESP relative to AM1-BCC, but secondary amine compounds become less favorable. Simple alcohols become more favorable with 6-31G*/RESP, whereas phenols—like simple aromatics—become less favorable. Of particular interest among the phenols, the solvation free energies for *o*-cresol and *p*-cresol become "reversed" relative to the experimental and AM1-BCC results. This observation has been corroborated in previous studies (with more sophisticated ab initio methods) and suggests a possible deficiency in the use of gas-phase ab initio calculations for deriving partial atomic charges for such molecules.[20] Thiols and thioethers are generally similar, although it worth noting the anomalously high solvation free energy of dimethyl disulfide with the Merck-Frosst AM1-BCC charges; interestingly, AM1-BCC charges computed using antechamber do not show the same anomaly (data not shown).

dx.doi.org/10.1021/jp2118373 | J. Phys. Chem. B 2012, 116, 4524−4534

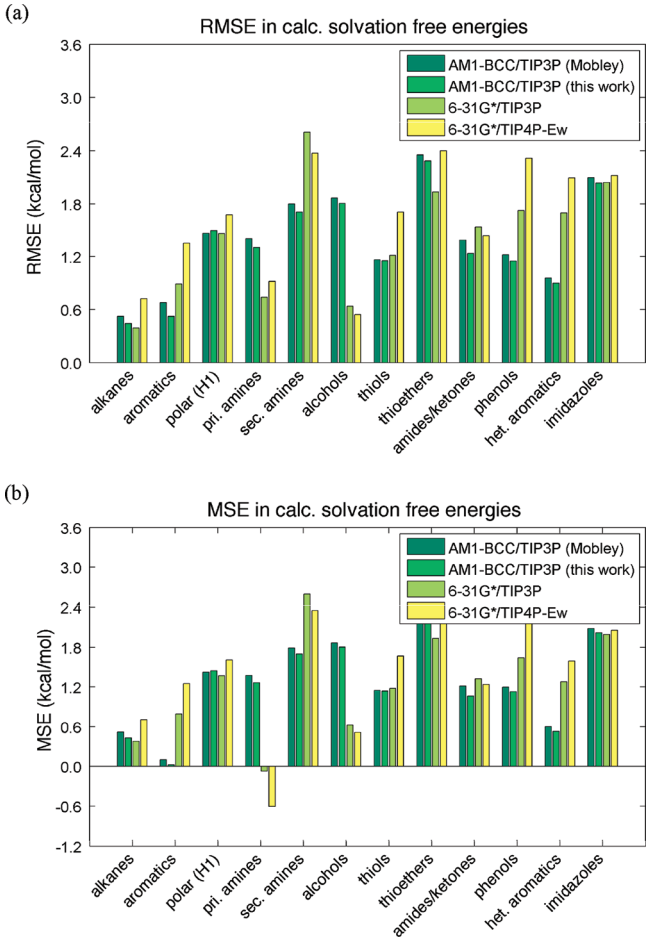


Figure 3. (a) Root mean square and (b) mean signed errors in calculated solvation free energies for the 12 categories of small molecules used for benchmarking. Results are shown for the AM1-BCC charge model with TIP3P water (green), HF/6-31G* charge model with TIP3P water (light green), and HF/6-31G* charge model with TIP4P-Ew water (yellow), along with AM1-BCC/TIP3P results from Mobley et al. (dark green).

While there is little difference between the charge models for ketone/amide and imidazole compounds, other nitrogen-containing aromatic molecules (e.g., 3-methylindole or pyrrole) show large changes in solvation free energy. In particular, three of these compounds are much less favorable with 6-31G*/RESP charges, with the noted exception of pyrrole, which becomes significantly more favorable. Further studies with other five-membered heterocyclic aromatic rings (e.g., thiophene) confirm this specific trend (data not shown).

Finally, we compare the performance of the TIP3P and TIP4P-Ew water model for this set of molecules using the 6-31G*/RESP charges (Figure 3 and Table 1 of Supporting Information), as our ultimate goal is to optimize the combination of the AMBER ff99SB force field and TIP4P-Ew water. While TIP4P-Ew has been optimized to reproduce many properties of liquid water more accurately than TIP3P, it of course has not been optimized to combine with any standard protein force field. Overall, the RMSE to experiment increases to 1.75 kcal/mol with most of the molecules being more unfavorable in TIP4P-Ew water relative to TIP3P. Specifically, the alkanes are 0.1−0.4 kcal/mol less favorable and the hydrocarbon aromatics are 0.3−0.6 kcal/mol less favorable in TIP4P-Ew water relative to TIP3P. Interestingly, the primary amine compounds are more favorable in TIP4P-Ew by roughly 0.6 kcal/mol, while secondary amine compounds are only slightly more favorable. The alcohols are only slightly more favorable (0.1−0.2 kcal/mol), whereas the phenols are strongly less favorable (0.5−0.7 kcal/mol)—consistent with the trends observed with the aromatic compounds. Thiols and thioethers are both less favorable in TIP4P-Ew water (0.3−0.7 kcal/mol and 0.2−0.6 kcal/mol, respectively). As before, we do not observe any differences for the imidazole and ketone/amide compounds, with the noted exception of ethanamide, which contains a primary amine group and is 0.5 kcal/mol more favorable in TIP4P-Ew. Finally, the heterocyclic aromatic compounds are also less favorable in TIP4P-Ew (0.2−0.7 kcal/mol), with the exception of pyrrole, which becomes slightly more favorable.

Optimization of Solute−Water van der Waals Interactions. Our benchmark calculations suggest that the solvation of nearly every chemical moiety encountered in proteins is too unfavorable using the combination of AMBER ff99SB and TIP4P-Ew water, with the exception of primary amines. As previously mentioned, this has far-reaching consequences for solvation free energy calculations, simulations of intrinsically disordered peptides and proteins, as well as protein folding. By use of the methodology described above, we proceeded to optimize the van der Waals (vdW) interaction of each chemical moiety with TIP4P-Ew water by iteratively determining the combined vdW radius and well depth that result in the lowest RMSE in the solvation free energies for two different molecules containing that moiety (e.g., methane and *n*-butane for alkanes). We then validated these parameters by calculating the solvation free energies for a separate set of two different molecules containing the same moiety (e.g., propane and isobutane for alkanes). These results are summarized in Figure 4 and Table 2 of Supporting Information but are analyzed in more detail below.

As shown in Figure 1, we began our optimization with the nonpolar alkanes—the simplest building block of organic molecules. Our benchmark calculations had a RMSE in solvation free energies of 0.72 kcal/mol, but after optimizing the parameters for the HC and CT atom types, this error was reduced to 0.12 kcal/mol. The mean signed error was also reduced accordingly. Even greater improvements were obtained for the hydrocarbon aromatics (atom types HA and CA), which saw a decrease in RMSE from 1.35 to 0.28 kcal/mol.

In the AMBER/GAFF force field, there is a special atom type (H1) for hydrogen atoms bound to carbons with other electron-withdrawing neighbors (e.g., nitrogen). This presented a challenge for our parametrization because H1 is necessary for describing molecules that contain anything other than carbon and hydrogen, but by virtue of its very nature, it is impossible to find molecules that contain only H1 and the HC and CT (or HA and CA) atom types. As such, we chose four molecules in which there are a large number of H1 atoms and only one noncarbon (i.e., non-CT) electron-withdrawing atom: dimethyl ether, fluoromethane, dimethyl sulfide, and chloromethane. For the two molecules used for optimization, we chose the two which had the lowest discrepancy in experimental vs calculated solvation free energy—dimethyl ether and fluoromethane. We then assumed that the H1 atoms were responsible for most ($^{6}/_{7}$ and $^{4}/_{5}$, respectively), but not all of the error in the solvation free energies, as the oxygen and fluorine atom vdW parameters are also likely nonoptimal. By use of this approach, we were able to reduce the RMSE in solvation free energies from 1.68 to 0.54 kcal/mol, again recognizing that the noncarbon heavy

dx.doi.org/10.1021/jp2118373 | J. Phys. Chem. B 2012, 116, 4524−4534

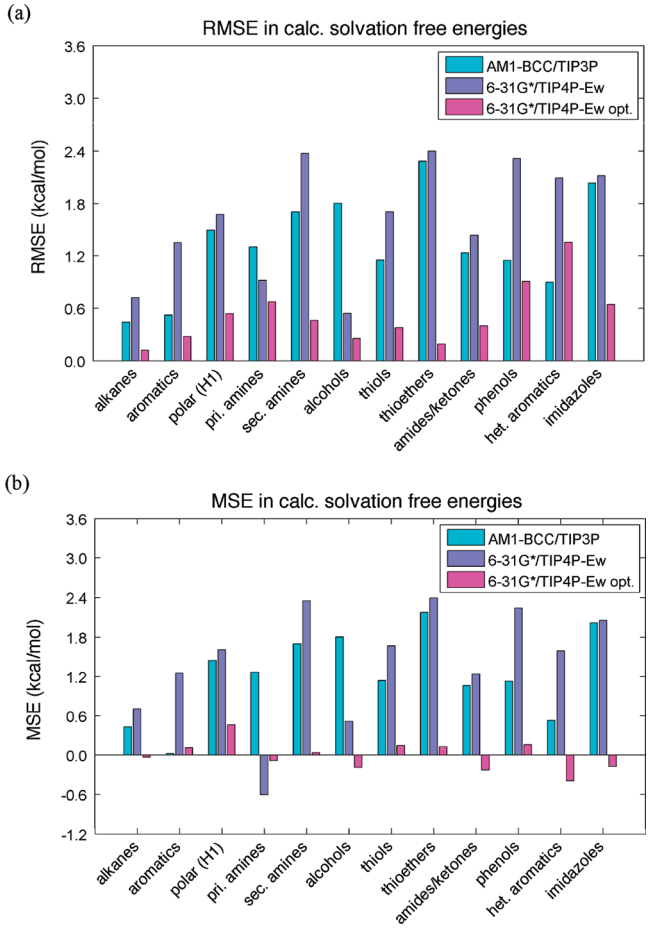


Figure 4. (a) Root mean square and (b) mean signed errors in calculated solvation free energies for the 12 categories of small molecules used for optimization and validation. Results using the HF/6-31G* charge model with optimized solute−water (TIP4P-Ew) van der Waals parameters (magneta) are compared with benchmark results for the AM1-BCC charge model with TIP3P water (light blue) and HF/6-31G* charge model with TIP4P-Ew water (purple).

atoms were no doubt contributing significantly to the overall error.

During our benchmark calculations, we observed that primary and secondary amines had different solvation properties, with the solvation of primary amines being significantly more favorable. As such, we elected to create a separate nitrogen and hydrogen atom type for the primary amines and treat primary and secondary amines separately in our optimization. We were only able to improve the primary amines slightly—a reduction in RMSE from 0.92 to 0.68 kcal/mol. By contrast, we were able to greatly improve the solvation of the secondary amines from a RMSE of 2.37 to 0.46 kcal/mol. We noticed a similar trend with alcohols and phenols—alcohols were much more favorable than phenols—and elected to again create a separate atom type (for the oxygen atoms of phenols) and optimize the related oxygen atoms for these moieties separately as well. Doing so enabled a small reduction in RMSE for alcohols (0.54 to 0.26 kcal/mol) and a larger one for phenols (2.32 to 0.91 kcal/mol). It is worth noting that the hydroxyl group hydrogen atom typically has zero vdW radius and well depth in the AMBER force fields. We experimented with nonzero vdW parameters for the interaction between this atom and TIP4P-Ew water but were unable to significantly improve the solvation free energy results. Therefore, we elected to keep the zero parameters for this atom type.

Sulfur-containing compounds have long been known to have poor agreement between computed and experimentally measured solvation free energies.[7,18,19,37,38] Our TIP4P-Ew benchmark data indicated RMSEs of 1.71 and 2.40 kcal/mol for thiols and thioethers, respectively. Optimizing these categories separately, we were able to obtain excellent agreement with experiment with RMSEs of 0.38 and 0.19 kcal/mol. We had similar success with the amides/ketones (1.44 to 0.40 kcal/mol) and imidazoles (2.12 to 0.65 kcal/mol). Unfortunately, our optimization of the H4 atom type for heterocyclic aromatics (e.g., tryptophan analogues) was not as successful—the RMSE decreased from 2.09 to 1.35 kcal/mol—but it is worth noting that the majority of the error is due to a single molecule (pyrrole) being overly soluble. As briefly described in the benchmark results, we investigated this further by performing solvation free energy calculations on other five-membered ring heterocyclic compounds (thiophene, 1-methylpyrrole, and 2-methylthiophene) and found that their solvation was also more favorable than expected (data not shown), pointing to a possible deficiency in the fixed charge model for these compounds or the need for explicit polarizability.

Overall, the results of the solute−solvent vdW parameter optimization process were quite encouraging. We were able to reduce the RMSE in solvation free energies of the 47 molecule set from 1.75 kcal/mol to only 0.61 kcal/mol (i.e., on the order of 1.0 k_BT at 300 K), equivalent to the accuracy of current water models and polarizable force fields such as AMOEBA.[39] Moreover, the mean signed error of the solvation free energies—an indicator of systematic under- or oversolvation—was reduced from 1.40 to 0.00 kcal/mol, suggesting that the optimizations result in a balanced solvation of these various moieties by TIP4P-Ew water. The optimized c^{opt} and d^{opt} coefficients for the radius (R_{ij}) and well depth (ε_{ij}) of the vdW interaction between each AMBER ff99SB protein atom type and the oxygen atom of TIP4P-Ew water are included in Table 3 of Supporting Information. In addition, we have included the corresponding van der Waals A and B coefficients (where $A = \varepsilon_{ij}R_{ij}^{12}$ and $B = 2\varepsilon_{ij}R_{ij}^{6}$, respectively) for these protein−water interactions, as these are the numbers found in AMBER prmtop (parameter/topology) files (Table 3 of Supporting Information).

Validation: Dipeptide Solutions. Previous work by Johnson et al. examined the properties of concentrated (>1.0 M) solutions of glycine and leucine dipeptides (also called NAGMA and NALMA, respectively).[40] Using the AMBER ff03 force field,[41] which employs a somewhat different charge model but the same van der Waals parameters as ff99SB, and TIP4P-Ew water, they observed significant aggregation of the dipeptides at concentrations for which these molecules are experimentally known to be soluble.[42] Conversely, using the AMOEBA force field and water model, Johnson et al. found that the dipeptide molecules remained soluble, with no discernible aggregation.[40]

We performed similar calculations using AMBER ff99SB (instead of ff03) and TIP4P-Ew water and likewise observed significant aggregation for both 1.5 M NAGMA and 1.0 M NALMA solutions. This can be demonstrated by calculating a carbon−carbon radial distribution function and observing that $g(r)$ is significantly greater than 1.0 at short distances (e.g., less than 10 Å), as opposed to being roughly 1.0 for all distances, which would indicate a roughly even distribution of the molecules throughout the solution (Figure 5). By contrast, when we

dx.doi.org/10.1021/jp2118373 | J. Phys. Chem. B 2012, 116, 4524−4534

simulated these solutions using the optimized solute−water van der Waals parameters, we found that $g(r)$ stays nicely around 1.0 for all distances (Figure 5); moreover, these data are nearly

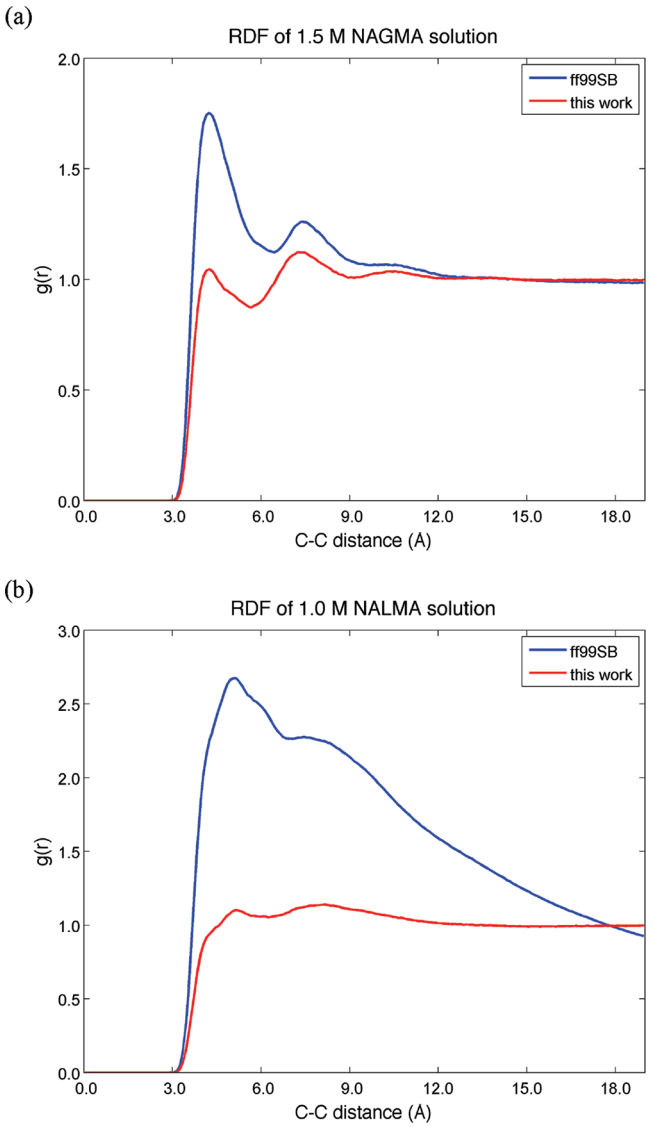


Figure 5. Carbon−carbon radial distribution functions for (a) 1.5 M NAGMA and (b) 1.0 M NALMA solutions. AMBER ff99SB/TIP4P-Ew results are shown in blue; results using the optimized solute−water van der Waals parameters are shown in red. For both cases, there is a marked decrease in aggregation using the optimized vdW parameters.

identical to the AMOEBA results obtained by Johnson et al. for the same solutions. These data suggest that the optimized van der Waals parameters not only enable the accurate solvation of small molecules in TIP4P-Ew, but of short peptides as well.

Validation: Short Disordered Peptides. To follow up on our dipeptide solution results and further test the applicability of these parameters to larger biomolecules, we turned to simulations of short (3−5 residue) peptides. In particular, we simulated the Gly_3, Val_3, and Ala_5 peptides, which have previously been characterized extensively using NMR spectroscopy by Graf et al.[43] and which we utilized in optimizing and validating a change to the ϕ' backbone dihedral angle potential in ff99SB.[11] (In this previous work, we modified just the $n = 2$ term of the ϕ' backbone dihedral angle potential in AMBER ff99SB to increase sampling of the PPII conformation and consequently obtain better agreement with experimental NMR scalar coupling data for a variety of short peptides.) Similar to our previous work, we assessed the effect of the new solute−water van der Waals parameters by calculating NMR scalar couplings (using the Karplus equation) and comparing them with experimentally measured couplings via the χ^2 metric (see Methods). If χ^2 is less than or equal to 1.0, this indicates that the conformational ensemble of the simulation is essentially the same as the experimental ensemble, within the limitations of the Karplus equation parameters.[15]

For Gly_3 we obtain χ^2 values of 2.93 ± 0.04 with ff99SB and 2.79 ± 0.03 with the new solute−water van der Waals parameters (Table 1). However, in our previous work we noted

Table 1. χ^2 Values for Calculated Scalar Couplings of Gly_3 at 300 K in TIP4P-Ew Water Using Unmodified ff99SB and ff99SB with Optimized Solute−Water van der Waals Parameters[a]

force field	all couplings	no $^2J(N',C_\alpha)$ or $^3J(C,C')$ coupling
ff99SB	2.93 (0.04)	0.47 (0.05)
optimized vdW	2.79 (0.03)	0.48 (0.04)

[a]Values are given as the means over four independent MD simulations, with the standard error of the means given in parentheses. Only the original Karplus equation parameters are used for these calculations.[11,15]

that the Karplus equations for at least one of the couplings, $^2J(N',C_\alpha)$, has parameters that are likely insufficient for describing glycine accurately.[11] In addition, another coupling, $^3J(C,C')$, is subject to considerable measurement uncertainty due to spectral crowding.[43] If we exclude these two couplings, we obtain χ^2 values of 0.47 ± 0.05 and 0.48 ± 0.04, respectively, indicating that both ff99SB and ff99SB with the optimized van der Waals parameters are sufficiently accurate to capture the conformational ensemble of Gly_3 (Table 1). In addition, these data suggest that there are no substantial differences between the two conformational ensembles and therefore that the simulated conformational ensembles of very short peptides do not depend strongly on the solvent−solute van der Waals interactions, at least within the regimes sampled by our work.

For the Val_3 peptide, we compared our results with ff99SB, ff99SB with the optimized ϕ' potential (which applies to all nonglycine residues), and ff99SB with both the optimized ϕ' potential and solute−water van der Waals parameters. For these three different simulations, we obtain χ^2 values of 1.98 ± 0.10, 1.63 ± 0.02, and 1.63 ± 0.04 (Table 2). Once again these data demonstrate no major differences in simulations with the optimized solute−water vdW parameters, as well as a small improvement using the optimized dihedral potential. If we exclude the $^3J(C,C')$ coupling, we obtain χ^2 values of 1.30 ± 0.06, 1.19 ± 0.18, and 1.24 ± 0.06, again echoing the first result (Table 2).

Lastly, we performed simulations of the Ala_5 peptide using the same three simulation conditions. For this peptide we obtain χ^2 values of 2.44 ± 0.10, 1.33 ± 0.04, and 1.15 ± 0.03, or if we exclude the $^3J(C,C')$ coupling, 1.73 ± 0.09, 0.86 ± 0.05, and 0.71 ± 0.04 (Table 3). These data suggest that the optimized backbone potential is indeed a significant improvement for the Ala_5 peptide but also that the optimized solute−water vdW parameters have a greater effect as the length of the peptide chain increases. For Ala_5, these parameters actually

dx.doi.org/10.1021/jp2118373 | J. Phys. Chem. B 2012, 116, 4524−4534

Table 2. χ^2 Values for Calculated Scalar Couplings of Val_3 at 300 K in TIP4P-Ew Water χ^2 Using Unmodified ff99SB, ff99SB with the Optimized ϕ' Backbone Dihedral Angle Potential, and ff99SB with Both Optimized Dihedral Potential and Solute–Water van der Waals Parameters[a]

force field	all couplings	no $^3J(C,C')$ coupling
ff99SB	1.98 (0.10)	1.30 (0.06)
optimized dihedral	1.63 (0.02)	1.19 (0.18)
optimized dihedral + vdW	1.63 (0.04)	1.24 (0.06)

[a]Values are given as the means over two independent REMD simulations, with the differences between the two simulations given in parentheses. Only the original Karplus equation parameters are used for these calculations.[11,15]

Table 3. χ^2 Values for Calculated Scalar Couplings of Ala_5 at 300 K in TIP4P-Ew Water Using Unmodified ff99SB, ff99SB with the Optimized ϕ' Backbone Dihedral Angle Potential, and ff99SB with Both Optimized Dihedral Potential and Solute–Water van der Waals Parameters[a]

	all couplings		No $^3J(C,C')$ coupling	
force field	original	DFT2	original	DFT2
ff99SB	2.44 (0.10)	2.14 (0.11)	1.73 (0.09)	1.37 (0.08)
optimized dihedral	1.33 (0.04)	1.26 (0.02)	0.86 (0.05)	0.86 (0.02)
optimized dihedral + vdW	1.15 (0.03)	1.12 (0.01)	0.71 (0.04)	0.78 (0.02)

[a]Values are given as the means over two independent REMD simulations, with the differences between the two simulations given in parentheses. Both the original and DFT2 Karplus equation parameters are used for these calculations.[11,12,15]

result in a lower χ^2 value, suggesting a more accurate conformational ensemble.

Validation: Stability and Dynamics of the Folded Ubiquitin Protein. All of our previous validation studies examined relatively small biomolecules that are highly, if not completely, solvent-exposed. We were therefore motivated to investigate how our optimized solute–water van der Waals parameters might apply to a larger folded protein with a hydrophobic core. Several MD force fields, including AMBER ff99SB, have been validated (in part) by simulating ubiquitin, and we likewise previously used this protein in validating the aforementioned ϕ' backbone dihedral angle potential.[9,11,12]

We found that within the first 20 ns of a 100-ns simulation, ubiquitin began to unfold when simulated with the optimized vdW parameters (parts a and b of Figure 6). We observed, however, that the unfolding was not the result of an exposure of the hydrophobic core but rather the breaking of certain hydrogen bonds between secondary structure elements. Moreover, replica exchange MD simulations of two folded peptides, GB1 hairpin and $(AAQAA)_3$, using these new parameters yield conformational ensembles with noticeably smaller folded populations than the experimental data suggest (data not shown). These observations make sense in light of the fact that with the new solute–water vdW parameters, protein–water hydrogen bonds (H-bonds)—specifically those involving the backbone amide group—might become more energetically favorable than protein–protein hydrogen bonds. Previous studies have also suggested that the protein–protein backbone H-bonds in AMBER ff99SB might not be sufficiently energetically favorable[12] and therefore that the more favorable solvation properties of the backbone might further exacerbate the issue.

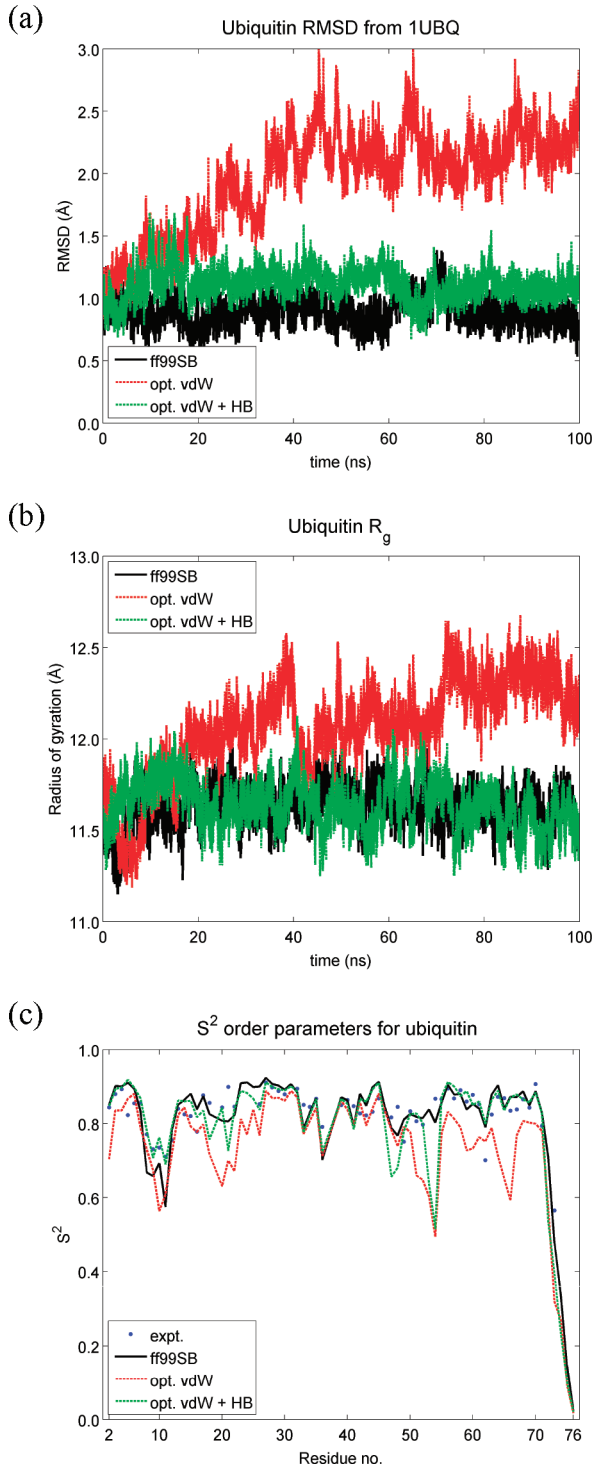


Figure 6. Native state stability and dynamics of ubiquitin for the unmodified AMBER ff99SB force field (black), the optimized solute–water van der Waals parameters (red), and the optimized solute–water vdW parameters with an added 10–12 potential for protein–protein backbone hydrogen bonds (green). (a) The root-mean-square deviation (rmsd) for backbone heavy atoms of residues 1–71 with respect to the crystal structure 1UBQ. (b) Radius of gyration of backbone C_α atoms. (c) Lipari–Szabo (S^2) order parameters with the experimentally derived isotropic values shown in blue. For all three simulations, the TIP4P-Ew water model was employed.

dx.doi.org/10.1021/jp2118373 | J. Phys. Chem. B 2012, 116, 4524–4534

To test if this was indeed the case, we added a 10−12 potential (i.e., $A/r^{12} - B/r^{10}$) to the force field to model the nonelectrostatic component of H-bonds between the backbone carbonyl oxygen and amide protons of the protein.[44,45] We note that this potential *replaces* the regular vdW potential for the atoms in question and can be considered as a modified vdW interaction rather than an altogether new interaction. Such potentials were commonly used in MD force fields of ∼25 years ago[46] but were phased out because it was believed that the electrostatic and regular (i.e., 6−12 or Lennard-Jones) van der Waals potentials were sufficient to accurately model these interactions.[6] By use of the A and B coefficients from the AMBER ff86 force field, which yield a well depth of 0.5 kcal/mol per hydrogen bond, we observed some unfolding of ubiquitin after ∼50 ns (data not shown). We then proceeded to double the well depth of the potential (i.e., doubled the A and B coefficients) so as to further stabilize protein−protein H-bonds relative to protein−water H-bonds. With this modification to the potential, we observed a near total avoidance of the unfolding observed with the optimized solute−water vdW parameters (Figure 6b). Moreover, we found through multiple metrics—rmsd, radius of gyration, and S^2 order parameters—that the H-bond potential appears to yield results that are significantly closer to the unmodified AMBER ff99SB result or even closer agreement with the NMR experiment (Figure 6c). Nonetheless, this is not the final say on anisotropic short-ranged interactions. We found that residues 54 (an arginine), 48 (a lysine), and 23 (an isoleucine) are more disordered with respect to experimental S^2 values (Figure 6c) and were more solvent-exposed (data not shown). These residues were found to have a significant reduction in the percentage of trajectory frames in which backbone hydrogen bonds were formed. In fact, statistically, the overall 10−12 potential is just a little too weak overall, but it is remarkable that pulling the hydrogen bond potential "off the shelf" with no optimization worked as well as it did.

DISCUSSION AND CONCLUSION

While most previous biomolecular force field development efforts have focused on improving the agreement between gas-phase ab initio and molecular mechanics calculations for substituent amino acid chemistry,[7,9,10] the resulting parameters are often "transferred" to describe these same molecules in aqueous solution, where success is (not surprisingly) more limited. In fact, it is generally well appreciated that most protein−water force fields are undersolvated, meaning that solute−solute and water−water interactions are systematically more favorable than the corresponding solute−water interactions.[13,18,19] This in turn limits the predictive capacity of simulations for properties such as small molecule solvation free energies, NMR observables that are collected in solution, and conformational equilibria involving secondary structure or even folding/unfolding populations. For example, Best and Mittal recently analyzed the folding of several small peptides and found that their unfolded states are typically too collapsed,[13,47] which in turn gives poor qualitative agreement with FRET efficiencies at higher temperatures.[47] Limitations on accuracy are exacerbated further by the use of newer nonbonded Ewald interaction schemes for solvated biomolecules,[48] which are different from the energy and force truncation schemes used in the original parametrization of the water models themselves.[35]

Given the current acceptance in the biomolecular simulation community of treating long-ranged electrostatics under Ewald conditions, we felt it was time to consider a new standard default water model such as TIP4P-Ew, which was parametrized specifically for use with Ewald as well as long-range van der Waals corrections.[28] Nonetheless, the more current ff99SB protein force field and TIP4P-Ew water model required further optimization to be well utilized in predicting condensed phase properties. We have used a hybrid strategy of relying on the "original" parametrization that was more focused on either neat organic liquid or gas-phase ab initio data (which has been central to subsequent refinements of both the bonded interactions and the electrostatic models) and further fine-tuning of *only* the solute−water parameters against condensed phase experimental data. Although there are now additional parameters in the new protein−water force field we have developed here, the potential problem of overfitting is reduced by our relatively restrained optimization approach and the fact that there are more abundant solution phase data than in the past that make this parametrization strategy possible. This is a credit to the much appreciated engagement from the experimental small molecule, protein folding, and NMR communities that have contributed data and analyses of small but complex biomolecular systems that operate on time scales that are accessible to MD simulation, allowing for a one-to-one comparison of simulated and experimental observables.

Finally, to maintain a force field that works not only for solvent-exposed small molecules or peptides but also for folded globular proteins we found a need to restore some balance by strengthening the protein−protein interactions given that the favorable enhancement of solute−water interactions resulted in the partial unfolding of ubiquitin. One of the important vindications of the parametrization approach taken here was that the hydrophobic core of ubiquitin remained intact under the new force field, suggesting that the protein−protein van der Waals interactions are still relatively well-balanced with the protein−water van der Waals interactions. We did, however, observe a degradation of ubiquitin's secondary structure that may be caused by a relative weakening of protein−protein hydrogen bonds with respect to protein−water hydrogen bonds. By reintroducing the 10−12 hydrogen bonding potential for protein−protein backbone H-bonds, we were able to show that ubiquitin remained stable over a 100-ns trajectory and significantly improved its fidelity to experimentally measured Lipari−Szabo (S^2) order parameters relative to ff99SB with only the modified protein−water vdW interactions. However, we do not believe this is the final say on hydrogen bonding potentials or even relevant functional forms that are trying to capture short-ranged directionality and cooperativity. It is our view that force fields such as AMOEBA,[39] although often classified as a polarizable force field (which is true), derive much of their accuracy from more robust representation of short-ranged electrostatics through the use of permanent multipoles.[49,50] This elegant approach, requiring a factor of 3−4 in additional computational cost, would need to be weighed against a simpler geometric definition of hydrogen bonding, which would be cheaper to evaluate but may be inherently limited in accuracy. We hope to report on a final parametrization phase that incorporates short-ranged directional protein−protein interactions to complete a force field that operates across a diversity of biomolecules. At present, the force field presented here is certainly reliable for aqueous solution studies involving small molecules and peptides.

■ ASSOCIATED CONTENT

Supporting Information

Documents containing the results of the benchmark solvation free energy calculations, solvation free energy calculations with optimized solute−water vdW parameters, the optimization coefficients c^{opt} and d^{opt} (and corresponding optimized A and B coefficients) for each AMBER ff99SB atom type, and a set of representative $dU/d\lambda$ curves for the solvation free energy calculations. In addition, mol2 files of all 47 small molecules used in the benchmark and optimization process with HF/6-31G*-derived partial atomic charges. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: tlhead-gordon@lbl.gov.

Present Address

§W. M. Keck Science Department, Claremont McKenna, Pitzer, and Scripps Colleges, Claremont, California 91711−5716.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work reported here is supported by resources of UC Berkeley CITRIS and the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We thank Prof. David A. Case at Rutgers University for help with resurrecting the hydrogen bonding code in pmemd.

■ REFERENCES

(1) Schlick, T.; Collepardo-Guevara, R.; Halvorsen, L. A.; Jung, S.; Xiao, X. *Q. Rev. Biophys.* **2011**, *44*, 191−228.

(2) van Gunsteren, W. F.; Dolenc, J. *Biochem. Soc. Trans.* **2008**, *36*, 11−15.

(3) van der Kamp, M. W.; Shaw, K. E.; Woods, C. J.; Mulholland, A. J. *J. R. Soc. Interface* **2008**, *5* (Suppl 3), S173−190.

(4) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(5) MacKerell, A. D. Jr.; Bashford, D.; Dunbrack, R. L. Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(6) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(7) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(8) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−1415.

(9) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712−725.

(10) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950−1958.

(11) Nerenberg, P. S.; Head-Gordon, T. *J. Chem. Theory Comput.* **2011**, *7*, 1220−1230.

(12) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.

(13) Best, R. B.; Mittal, J. *J. Phys. Chem. B* **2010**, *114*, 14916−14923.

(14) Wickstrom, L.; Okur, A.; Simmerling, C. *Biophys. J.* **2009**, *97*, 853−856.

(15) Best, R. B.; Buchete, N.-V.; Hummer, G. *Biophys. J.* **2008**, *95*, L07−L09.

(16) Jiang, F.; Han, W. *J. Phys. Chem. B* **2010**, *114*, 5840−5850.

(17) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.

(18) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.

(19) Swope, W. C.; Horn, H. W.; Rice, J. E. *J. Phys. Chem. B* **2010**, *114*, 8631−8645.

(20) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007**, *111*, 2242−2254.

(21) Baker, C. M.; Lopes, P. E. M.; Zhu, X.; Roux, B.; Mackerell, A. D. *J. Chem. Theory Comput.* **2010**, *6*, 1181−1198.

(22) Dyer, P. J.; Docherty, H.; Cummings, P. T. *J. Chem. Phys.* **2008**, *129*, 024508.

(23) Sun, Y.; Kollman, P. A. *J. Comput. Chem.* **1995**, *30*, 1164−1169.

(24) Klimovich, P. V.; Mobley, D. L. *J. Comput. Aided Mol. Des.* **2010**, *24*, 307−316.

(25) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656−1676.

(26) Ashbaugh, H. S.; Collett, N. J.; Hatch, H. W.; Staton, J. A. *J. Chem. Phys.* **2010**, *132*, 124504.

(27) Horta, B. A. C.; Fuchs, P. F. J.; van Gunsteren, W. F.; Hünenberger, P. H. *J. Chem. Theory Comput.* **2011**, *7*, 1016−1031.

(28) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665−9678.

(29) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132−146.

(30) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623−1641.

(31) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(32) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821−7839.

(33) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(34) Swope, W. C.; Horn, H. W.; Rice, J. E. *J. Phys. Chem. B* **2010**, *114*, 8621−8630.

(35) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(36) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER*, *11*, 2010.

(37) Price, D. J.; Brooks, C. L. *J. Comput. Chem.* **2005**, *26*, 1529−1541.

(38) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Dill, K. *J. Phys. Chem. B* **2009**, *113*, 4533−4537.

(39) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.

(40) Johnson, M. E.; Malardier-Jugroot, C.; Murarka, R. K.; Head-Gordon, T. *J. Phys. Chem. B* **2009**, *113*, 4082−4092.

(41) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999−2012.

(42) Hura, G.; Sorenson, J. M.; Glaeser, R. M.; Head-Gordon, T. *Perspect. Drug Disc. Des.* **1999**, *17*, 97−118.

(43) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J. Am. Chem. Soc.* **2007**, *129*, 1179−1189.

(44) McGuire, R. F.; Momany, F. A.; Scheraga, H. A. *J. Phys. Chem.* **1972**, *76*, 375−393.
(45) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361−2381.
(46) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230−252.
(47) Best, R. B.; Mittal, J. *Proteins* **2011**, *79*, 1318−1328.
(48) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(49) Shi, Y.; Wu, C.; Ponder, J. W.; Ren, P. *J. Comput. Chem.* **2010**, *32*, 967−977.
(50) Rasmussen, T. D.; Ren, P.; Ponder, J. W.; Jensen, F. *Int. J. Quantum Chem.* **2006**, *107*, 1390−1395.