Title
Optimizing Protein−Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides

Journal
Journal of Chemical Theory and Computation, 7(4)

ISSN
1549-9618

Authors
Nerenberg, Paul S
Head-Gordon, Teresa

Publication Date
2011-04-12

DOI
10.1021/ct2000183

Peer reviewed

Optimizing Protein-Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides

Paul S. Nerenberg[1] and Teresa Head-Gordon[1,2]

[1]*California Institute of Quantitative Biosciences (QB3),*
[2]*Department of Bioengineering,*
University of California, Berkeley
Berkeley, CA 94720-3220 USA

While most force field efforts in biomolecular simulation have focused on the parameterization of the protein, relatively little attention has been paid to the quality of the accompanying solvent model. These considerations are especially relevant for simulations of intrinsically disordered peptides and proteins, for which energy differences between conformations are small and interactions with water are enhanced. In this work, we investigate the accuracy of the AMBER ff99SB force field when combined with the standard TIP3P model or the more recent TIP4P-Ew water model, to generate conformational ensembles for disordered trialanine (Ala_3), triglycine (Gly_3), and trivaline (Val_3) peptides. We find that the TIP4P-Ew water model yields significantly better agreement with experimentally measured scalar couplings – and therefore more accurate conformational ensembles – for both Ala_3 and Gly_3. For Val_3, however, we find that the TIP3P and TIP4P-Ew ensembles are equivalent in their performance. To further improve the protein-water force field combination and obtain more accurate intrinsic conformational preferences, we derive a straightforward perturbation to the ϕ' backbone dihedral potential that shifts the β–PPII equilibrium. We find that the revised ϕ' backbone dihedral potential yields improved conformational ensembles for a variety of small peptides and maintains the stability of the globular ubiquitin protein in TIP4P-Ew water.

INTRODUCTION

Over the last three decades, classical molecular dynamics (MD) and Monte Carlo (MC) simulations have emerged as an important complement to experimental methods for investigating many aspects of biomolecular structure, dynamics, and function[1]. The predictive quality of biomolecular simulation depends on the accuracy of the potential energy function (or force field), comprised of bonded interactions, van der Waals parameters, and (typically) fixed-charge electrostatics, that were first developed approximately 15 years ago[2-4]. Bonded parameters describing bond stretching, bending, and torsions and partial atomic charges are generally derived using quantum chemical methods, and the van der Waals parameters are empirically derived to match experimental densities and enthalpies of vaporization of neat organic liquids (e.g., liquid hydrocarbons) in an effort to model the intramolecular interactions typical of folded globular proteins[2-4]. Since that time, force field development efforts have utilized more advanced *ab initio* methods and focused on improving the backbone and side chain dihedral angle potentials[5-10] and the fixed charges[11-12] by minimizing the differences between gas phase *ab initio* and molecular mechanics energies, charge distributions, etc., for specified molecular structures. The ongoing refinement of empirical biomolecular force fields has resulted in a corresponding increase in the predictive power of MD simulations[13] in modeling the structures, dynamics, and folding of globular proteins[14-16].

In the last 10 years, however, there has been a growing recognition that much of the human proteome (~30%) consists of proteins that are intrinsically disordered[17-19]. Intrinsically disordered proteins (IDPs) are thought to be vital for carrying out a variety of signaling and regulatory functions in the cell, but they have also been implicated in several common diseases, including various cancers, Alzheimer's disease, Parkinson's disease, type II diabetes, and cardiovascular disease[17-18,20]. As it is known that IDPs rapidly sample multiple conformations (i.e., faster than the ms time scale of NMR experiments), the free energy landscapes of these proteins must be relatively flat, with small energetic barriers between conformations and energy differences that are on the order of k_BT[21]. This differs markedly from the single, deep free energy minimum (folding funnel) that is characteristic of folded proteins[22-23] and presents a potentially serious challenge to the predictive power of MD force fields when applied to such systems. In addition, by virtue of having greater net charge per residue and proportionally fewer hydrophobic residues, IDPs are generally more unfolded and solvent-exposed than their globular counterparts[19,24]. Because of their increased solvent exposure, the conformational ensembles of IDPs likely depend more strongly on sensitive protein-water interactions, and therefore deficiencies of the standard TIP3P or SPC solvent models may become more obvious in simulations of these systems[25].

Recent studies enabled by improvements in both simulation methodologies and computer hardware and the development of experimental methods that yield high quality quantitative data (e.g., NMR scalar couplings) have enabled direct comparisons between simulation results and experiment. These comparisons have revealed discrepancies in the backbone and side chain conformational preferences of short peptides[8,26-27] and solvation free energies of small molecules[12,28-29], lending credence to the notion that current force fields may not be optimal for simulations of more solvent exposed peptides and IDPs. Related to the work presented here, Best and Hummer analyzed a number of different protein force fields and concluded that most are too helical when comparing calculated J-coupling observables to experiment, although the AMBER ff99SB[7] force field proved to be more reliable than most[26]. In subsequent work they developed a

correction to the ψ backbone dihedral angle potential for AMBER ff99SB, deemed ff99SB*, which improved the helix-to-coil transition for longer α-helix-forming peptides[5]. However, the only water model considered in that study was the TIP3P model[30], which continues to dominate the biomolecular simulation literature in spite of demonstrable improvements in condensed phase water descriptions by other (potentially) compatible fixed charge water models[31-34].

Evidence that improved water models can yield more accurate conformational ensembles – particularly of disordered peptides – has been mounting in the past few years. A study of the $A\beta_{21-30}$ peptide by Fawzi et al. found that the combination of AMBER ff99SB and the recently developed TIP4P-Ew model[33] yielded better predictions of NMR ROESY crosspeaks than ff99SB and TIP3P[35]. Using the same force field, Wickstrom et al. demonstrated that ensembles generated with TIP4P-Ew predicted NMR scalar couplings for Ala_3 and Ala_5 more accurately than ensembles generated with TIP3P[27]. More recently, Best and Mittal used a methodology similar to the ff99SB* study to develop a correction to the ψ backbone dihedral angle potential for AMBER ff03[11] with the TIP4P/2005 water model[31]. They found that the optimized force field and solvent model combination (ff03w) yielded a more cooperative helix-to-coil transition and a more realistic collapse of unfoldeded conformers with increasing temperature than an optimized combination of ff03 and TIP3P water (ff03*)[6].

Therefore, we set out to understand how well a modern force field, AMBER ff99SB, and two water models, the default TIP3P model and the newer TIP4P-Ew model, could quantify the conformational ensembles of three disordered tripeptides (Ala_3, Gly_3, and Val_3) for which there are a large number of experimentally measured NMR scalar couplings at 300 K[36]. Additionally, we simulate the Ala_3 system at 275, 325, and 350 K – temperatures for which the same coupling data are available[36] – to evaluate how the temperature-dependent characteristics of the water model (and protein force field) impact the simulated conformational ensembles.

Our results suggest that the use of the TIP4P-Ew water model, which is known to reproduce several characteristics of water better than TIP3P[33], produces superior conformational ensembles for both the Ala_3 and Gly_3 peptides; for the Val_3 peptide, the difference in accuracy between the TIP3P and TIP4P-Ew ensembles is less significant. Because there are few cooperative interactions, e.g., hydrogen bonds and dipole alignments to stabilize secondary structure in these short peptides, we hypothesize that the backbone dihedral potentials likely play the primary role in determining the conformational ensembles. We therefore explore if optimizing the backbone dihedral potentials to reproduce intrinsic conformational preferences of single amino acids – again in TIP4P-Ew water – can improve the conformational ensembles. By optimizing a single term of the ϕ' potential of the ff99SB force field with respect to NMR scalar couplings measured on a series of GXG peptides[37], we find that we can broadly improve intrinsic conformational preferences in disordered peptides (GLG, Ala_5, and Val_3) without disrupting the excellent native state stability of the globular protein ubiquitin. Moreover, these studies suggest several avenues for future improvements to the AMBER ff99SB force field, or fixed-charge force fields in general, when combined with more current water models such as TIP4P-Ew.

METHODS

Charges for protonated C-terminal residues

All of the NMR scalar couplings referenced in this work were obtained at pH 2[36-37], which necessitates the use of protonated C-terminal residues (-COOH) in the MD simulations. Charges

for C-terminal Ala and Val residues were derived by generating three conformations (corresponding to alpha, beta, and PPII), optimizing these conformations at the HF/6-31G* level of theory, and then calculating the electrostatic potentials of these structures at the same level of *ab initio* theory[2]. All *ab initio* calculations were performed using GAMESS-US[38]. Charges were fit to these potentials using the RESP method[39], as implemented in the R.E.D. Server version 1.0[40]. The charge derivation for the C-terminal Gly residue followed the same procedure except that five conformations were used instead of three. Charges for all three C-terminal residues are listed in the Supporting Information.

Simulation protocol

Ala_3, Gly_3, Val_3 and GXG peptides were built in an extended conformation using the tleap program included with AmberTools 1.4[41]. Each peptide was solvated in a truncated octahedron of 665-667 TIP3P[30] or TIP4P-Ew[33] water molecules, and the solvated system was neutralized by the addition of a single Cl^- ion[42]. All simulations were performed using AMBER 11 with either the ff99SB[7] or ff99SB*[5] force fields. Periodic boundary conditions were employed with a 9 Å cutoff for direct-space non-bonded interactions. Long-range electrostatics were calculated using particle mesh Ewald (PME)[43-44] 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[45]. First, each system was minimized with 500 steps of steepest descent minimization, followed by 1000 steps of conjugate gradient minimization. The system was then equilibrated at 300 K for 50 ps using a Langevin thermostat with a coupling constant of 0.5 ps^{-1}. During both minimization and heating, peptide atoms were restrained with a force constant of 10.0 kcal mol^{-1} $Å^{-2}$. The system was then brought to appropriate density by equilibrating at a constant temperature and pressure of 300 K and 1.0 bar, respectively, for 250 ps. This NPT equilibration was performed using a Langevin thermostat coupling constant of 1.0 ps^{-1} and a Berendsen barostat coupling constant of 5.0 ps^{-1} [46]. Four independent starting configurations were generated by simulating the equilibrated system at in the NVT ensemble at 400 K for 10 ns and drawing four conformations evenly from the last 8 ns. Each of these configurations was then equilibrated for 250 ps at 300 K (or 275, 325, or 350 K for Ala_3) and 1.0 bar. Four production simulations were then performed in the NPT ensemble for 100-400 ns, depending on the convergence properties of the simulation (discussed in greater detail in the Results section). Structures were saved every 1 ps. Definitions in terms of ϕ/ψ regions for the α, β, and PPII conformations are taken from Best et al.[26]

Replica exchange MD simulations

To improve convergence of simulation data for the low temperature (275 K) Ala_3 and Val_3 systems, two independent reservoir replica exchange MD (RREMD) simulations[47-48] were performed. For Ala_3, starting configurations were generated by first equilibrating the system at 275 K and then performing a 400 K NVT simulation, as described above. For both Ala_3 and Val_3, starting configurations were drawn from the first and fourth structures obtained by the 400 K simulations. Structure reservoirs of 50000 structures were generated by simulating the system at 380 K (Ala_3) or 400 K (Val_3) for 50 ns and saving conformations every 1 ps. For Ala_3, 16 replicas of the system were equilibrated at exponentially-spaced temperatures ranging from 275.00 to 372.40 K for 250 ps. For Val_3, 14 replicas of the system were equilibrated at exponentially-spaced temperatures ranging from 300.0 to 391.9 K for 250 ps. RREMD simulations were then performed with the 380 K (Ala_3) or 400 K (Val_3) reservoirs for 50 ns, with swaps attempted

between neighboring replicas every 1 ps. This temperature spacing yielded acceptance ratios of approximately 30-45%. An identical simulation protocol was used to simulate the Val_3 peptides with two different sets of modified van der Waals parameters, as well as with the modified backbone potential described in the Results section.

For GXG peptides, 20 replicas of the system were equilibrated in the NVT ensemble at exponentially-spaced temperatures ranging from 298 to 450 K for 250 ps. REMD simulations[48] were then performed for 50 ns, with swaps attempted between neighboring replicas every 1 ps. The temperature spacing yielded exchange probabilities of approximately 28-40%. When optimizing the backbone parameters, this simulation protocol was carried out for four equally spaced values of the n=2 ϕ' backbone dihedral angle potential ranging from a barrier height of 2.00 (ff99SB value) down to 1.55 kcal/mol. Piecewise cubic Hermite polynomials were used to interpolate the resulting data from 0.15 to 0.05 kcal/mol intervals.

To validate the optimized n=2 ϕ' (i.e., C-N-Cα-Cβ) backbone dihedral angle potential, we performed REMD simulations of the GLG and Ala5 peptides with both the unmodified and modified ff99SB force fields. The GLG simulations were carried out as described above, but using two different starting conformations: a fully extended conformation (ϕ = 180°, ψ = 180°) and an α-helical conformation (ϕ = -60°, ψ = -45°). The Ala_5 peptides were simulated with the same basic protocol, but were instead solvated with 902 TIP4P-Ew water molecules and used 24 exponentially spaced replicas instead of 20 to account for the larger number of degrees of freedom. Exchange probabilities for the Ala_5 system ranged from 31-43%.

MD simulations of ubiquitin

A native state structure for ubiquitin was obtained from the PDB crystal structure 1UBQ[49]. Hydrogen atoms were added by tleap, and the sole histidine residue was protonated to be consistent with the NMR relaxation data, which were obtained at pH 4.7[50]. The system was solvated in a truncated octahedron of 3602 TIP4P-Ew water molecules and neutralized with 1 Na^+ and 2 Cl^- ions, consistent with the experimental salt concentration of 10 mM NaCl[50]. Simulations were performed with both the unmodified and modified ff99SB force fields. The solvated system was first minimized with 500 steps of steepest descent minimization, followed by 1500 steps of conjugate gradient minimization, using Cartesian restraints on the protein atoms with a force constant of 10.0 kcal mol^{-1} $Å^{-2}$. The entire system was then minimized again with the same number of steps, except without any restraints on the protein atoms. Next, the system was equilibrated in the NVT ensemble at 298 K for 50 ps using a Langevin thermostat with a coupling constant of 0.5 ps^{-1}. The system was then brought to appropriate density by equilibrating at a constant temperature and pressure of 298 K and 1.0 bar, respectively, for 250 ps. During both NVT and NPT equilibration, the protein atoms were restrained with a force constant of 10.0 and 2.0 kcal mol^{-1} $Å^{-2}$, respectively. The system was then equilibrated in the NPT ensemble for an additional 5 ns without any restraints. Production simulations were run for 60 ns, with structures saved every 1 ps.

Generalized order parameters

Assuming that the slower overall motion of ubiquitin is isotropic and independent of faster internal motions[51], we eliminated rigid body rotations from the ubiquitin trajectories by performing a mass-weighted all-atom RMS fit using the first frames of the trajectories as reference structures. Next, we calculated the time autocorrelation function for the NH bond vectors:

$$C_l(t) = \langle P_2(\hat{\mu}(0) \cdot \hat{\mu}(t)) \rangle \quad (1)$$

where $P_2(x)$ is the second Legendre polynomial and $\hat{\mu}(t)$ is the unit vector of the NH bond. We then fit these correlation functions with the simplest approximation for internal motion[51]:

$$C_l(t) = S^2 + (1 - S^2)e^{-t/\tau_{eff}} \quad (2)$$

to determine the S^2 value for each bond vector.

RESULTS

A variety of NMR scalar couplings were calculated for Ala_3 using the ϕ/ψ backbone dihedral angles measured from the structures generated by the MD and RREMD simulations, and compared to the experimental coupling measurements of Graf et al.[36] that probe the ψ_1, ϕ_2, ψ_2, and ϕ_3 dihedral angles. In particular, we calculated $^3J(H_N,H_\alpha)$, $^3J(H_N,C)$, $^3J(H_\alpha,C')$, $^3J(H_N,C_\beta)$, $^3J(H_N,C_\alpha)$, $^1J(N,C_\alpha)$, and $^2J(N',C_\alpha)$ couplings, using three different sets of Karplus equation parameters ("Orig.", "DFT1", and "DFT2")[26]. For Gly_3 and Val_3, we also calculated the $^3J(C,C')$ coupling, again using the same Karplus equation parameters as Best et al.[26]

Similarly, we calculated the overall error between the calculated and experimental couplings as:

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(\langle J_i \rangle_{calc} - J_{i,expt}\right)}{\sigma_i^2} \quad (3)$$

As with previous studies, we assumed that errors in the calculated couplings due to sampling and errors in the experimentally measured couplings were negligible and therefore that the primary source of error (σ_i) for each coupling were the Karplus equation parameters themselves[5,26-27]. The exact values of the coupling errors are given in the Supporting Information. We conservatively increased the error estimates in the Karplus equation parameters for the $^3J(H_N,H_\alpha)$, $^3J(H_N,C)$, $^3J(H_\alpha,C')$, $^3J(H_N,C_\beta)$, and $^3J(C,C')$ couplings by 10% to account for the fact that the values given in the literature are mean absolute deviations, as opposed to root-mean-square deviations[26]. For comparison, previous studies increased these error estimates by 30% for the same reason[5,26-27]. Thus, we expect our χ^2 values to be generally larger than those found in previous studies due to the use of generally smaller error estimates.

Simulations of Ala_3 at multiple temperatures

We carried out four independent MD simulations at 275, 300, 325, and 350 K and compared the results obtained with ff99SB and either the TIP3P or the TIP4P-Ew water model. In addition, for select temperatures, we also used the newly developed ff99SB* force field of Best and Hummer[5], together with the TIP4P-Ew water model.

At 350 K our data demonstrate that both ff99SB and ff99SB*, together with the TIP4P-Ew water model, yield conformational ensembles more consistent with experimental data than ff99SB with the TIP3P water model (Table 1a). Moreover, this result is independent of the Karplus equation parameters used. A comparison between the ff99SB and ff99SB* results reveals that the use of a different water model brings about a larger change in the ensemble than the use of a modestly different force field (Table 1a). The primary difference between the TIP3P and TIP4P-Ew ensembles is an increase in the extended (β and PPII) conformations of the central Ala residue relative to more compact α-helical conformations (Table 2a). The ff99SB* force field is somewhat more helical than ff99SB, but the change in water model again imparts a larger difference than the change in force field (Table 2a). It is important to note, however, that the central residue's conformation does not account for couplings that measure the ψ_1 or ϕ_3 dihedral angles, which we examine separately below.

The TIP4P-Ew water model also generates more accurate ensembles at both 300 and 325 K (Tables 1b and 1c), again showing a higher propensity to sample β and PPII conformations than TIP3P. Interestingly, the β propensity remains relatively constant with decreasing temperature, while the α propensity decreases and the PPII propensity increases – regardless of the solvent model used (Tables 2b and 2c).

The results at 275 K provide the most sensitive test of the force field-water model combinations, as relative differences in energies more strongly contribute to differences in ensembles (via their Boltzmann weights) at this low temperature. We again found that TIP4P-Ew yielded demonstrably better conformational ensembles at this temperature than TIP3P, with the difference being even more pronounced than at higher temperatures (Table 1d). As before, the primary difference between the TIP3P and TIP4P-Ew ensembles is an increase in the extended (primarily PPII) conformations of the central Ala residue relative to more compact α-helical conformations (Table 2d).

Because convergence at low temperatures is difficult, we also conducted two independent RREMD simulations to corroborate the results of the "conventional" MD simulations. Although the RREMD simulations are performed in the NVT ensemble, as opposed to the NPT ensemble employed in the conventional MD simulations, differences between the two are likely minimal at the target temperature (275 K). The results of the RREMD simulations again confirm that TIP4P-Ew yields better ensembles for Ala_3 than TIP3P (Table 1d). In addition, we note that the ff99SB/TIP4P-Ew combination is significantly more accurate than the ff99SB*/TIP4P-Ew combination at this low temperature (Table 1d). While there is little difference between these combinations in the conformational preferences of the central residue (Table 1d), the behavior of the N- and C-terminal residues differs significantly, with ff99SB* stabilizing α-helical conformations of the N-terminal residue (SI Figure 1a) and turn/α_L conformations of the C-terminal residue (SI Figure 1b).

An examination of individual scalar couplings from the 275 K TIP3P and TIP4P-Ew data suggests that much of the observed improvement is due to a decrease in the sampling of the β conformation by the ϕ angle of the third residue and to a lesser extent the second residue, indicated by a decrease in magnitudes of the $^3J(H_N,H_\alpha)$, $^3J(H_N,C)$, and $^3J(H_\alpha,C')$ couplings and an increase in the magnitude of the $^3J(H_N,C_\beta)$ coupling (SI Tables 1-3). There is also the decrease in sampling of the α conformation by the ψ angles of the first and second residue, which is

indicated by an increase in magnitude of the $^3J(H_N,C_\alpha)$ and $^2J(N',C_\alpha)$ couplings (SI Tables 1-3). These observations are consistent with the central residue data that suggest an increase in the sampling of the PPII conformation in TIP4P-Ew water relative to TIP3P (Table 2d) and that the relative stabilization of the PPII conformation in TIP4P-Ew is primarily responsible for its improved performance relative to TIP3P.

Simulations of Gly_3 at 300 K

We performed four independent simulations of Gly_3 at 300K, using both TIP3P and TIP4P-Ew water, as well as the modified ff99SB* force field with TIP4P-Ew water. For all three force field and solvent model combinations, we observed large discrepancies between the observed and calculated scalar couplings, leading to χ^2 values of 2.93-3.45 (Table 3). Nonetheless, both of the TIP4P-Ew simulations had consistently lower χ^2 values than the TIP3P simulation, correlated with slightly greater sampling of the PPII conformation (SI Table 4), while there was little difference between ff99SB and ff99SB* (Table 3). These data suggest that the TIP4P-Ew water model again results in a more accurate conformational ensemble than TIP3P due to enhanced sampling of the PPII conformation.

Graf et al. have suggested that there may be errors in the Karplus equation parameters due to the limited number of measurements on glycine residues[36]. To better understand the cause of the large discrepancies in the scalar couplings, we first examined which of the calculated couplings were making the largest contribution to the χ^2 values. We found that two couplings (out of a total of 12) – the $^2J(N',C_\alpha)$ and $^3J(C,C')$ couplings of the central residue – approximately equally contributed 60-80% of the total χ^2 value, with the variation due to the different sets of Karplus equation parameters used. In addition, the $^2J(N',C_\alpha)$ coupling of the third residue and the $^3J(H_\alpha,C')$ coupling of the central residue made lesser contributions to the overall error.

The underlying cause of the discrepancies in the $^2J(N',C_\alpha)$ couplings appears to be due to a residue-specific effect that renders the Karplus equation parameters for this coupling inaccurate. More precisely, the maximum possible value of the coupling using these parameters is 8.71 Hz (with an uncertainty of ±0.5 Hz), while the experimental values are 10.45 Hz and 9.05 Hz, respectively, for second and third residues of Gly_3. Thus, even if the simulation-generated ensembles were completely identical to the experimental ensemble, the Karplus equation parameters would result in a discrepancy between the calculated and experimentally measured scalar couplings. Because the Cα spin system in glycine differs profoundly from those of the other amino acids in having no attached Cβ atom, it is plausible that coupling parameters involving the Cα atom that were developed for all residues would be the least applicable to glycine.

The difference between calculation and experiment for the $^3J(C,C')$ coupling appears to be slightly more complicated, in that it is somewhat dependent on which parameterization is used. The experimentally measured value for the $^3J(C,C')$ coupling is 0.26 Hz. For the Orig. parameterization, the lowest possible value of the coupling is 0.44 Hz, while for the DFT1 and DFT2 parameterizations the lowest possible values are 0.10 Hz and 0.13 Hz, respectively. (For comparison, the maximum possible values range from 2.89 to 3.90 Hz.) Thus, matching the experimentally measured value would require the simulated ensemble to almost exclusively sample ϕ angles of ±60-90°, which is not commensurate with glycine's conformational flexibility. In addition, Graf et al. remark that there is often severe overlap in the spectra used to

measure this coupling[36], so it is possible that there is non-negligible error in the experimentally measured values themselves.

Given the above considerations, we recomputed the χ^2 values, this time excluding the $^2J(N',C_\alpha)$ and $^3J(C,C')$ couplings. We found that if these couplings were excluded, the χ^2 values dropped below 1.0 for the Orig. and DFT2 calculated couplings and only slightly above 1.0 for the DFT1 calculated couplings (Table 3). Moreover, the differences between the various force field and solvent model combinations became relatively insignificant, although the TIP4P-Ew simulations still yielded better agreement with experiment (Table 3). These data suggest that the ϕ and ψ dihedral angle potentials of the ff99SB force field are likely adequate for condensed phase simulation in TIP4P-Ew water.

Simulations of Val_3 at 300 K

Simulations of the hydrophobic Val_3 peptide probe a force field's ability to reproduce the conformational preferences of residues with side chains more complicated than the simple methyl group of alanine. This is not a trivial point, as backbone dihedral angle parameters are often developed using alanine di- or tetrapeptides[2,7,9,11,52], with the implicit assumption being that all residues with Cβ atoms (with the exception of proline) will behave in essentially the same way and that any differences in the backbone preferences will be accounted for by additional interactions with the side chains.

As with Ala_3 and Gly_3, we performed four independent simulations of the Val_3 peptide using ff99SB and both water models. We found, however, that it was not possible to converge the simulations to our standard of less than 5% standard deviation in the mean inter-proton distances and the χ^2 values, even after 400 ns of simulation. Thus, we also performed RREMD simulations with these force field and solvent model combinations, as well as ff99SB* with TIP4P-Ew.

The results of these simulations suggest that the various force field and solvent model combinations produce conformational ensembles that are equivalently accurate within the statistical uncertainty, with no significant advantage for the TIP4P-Ew water model (Table 4). We do, however, find that the Karplus equation parameters developed using alanine peptides – particularly the DFT1 parameters – yield significantly larger χ^2 values (Table 4). These data suggest that backbone scalar coupling parameters derived from studies of only a single residue type may be inaccurate when applied to residues with different Cβ configurations and that it may be preferable to use parameters that are "averaged" over all residue types (e.g., the Orig. parameters) for assessing the accuracy of the conformational ensembles.

Because we encountered discrepancies with the $^3J(C,C')$ coupling in our studies of Gly_3 (see above), we examined the behavior of this coupling in our studies of Val_3. As can be seen in Table 4, it is clear that this coupling contributes a significant amount of error (reflected in the total χ^2 value), particularly for the Orig. and DFT2 parameters. Moreover, if this coupling is excluded from the χ^2 calculation, then the simulated Val_3 ensembles generate χ^2 values approximately equal to 1.0 (within the uncertainty bounds), which suggests that the simulated ensembles are equivalent to the experimental ensemble within the error of the Karplus equation parameters (Table 4). Nonetheless, the considerable variances in the χ^2 values, both with and without the $^3J(C,C')$ coupling, make it difficult to conclude which force field and water model combination yields more accurate conformational ensembles for Val_3, despite significant differences in the conformational preferences of the central residue (SI Table 5).

Development of an optimized ϕ' backbone dihedral angle potential

While the χ^2 values for Gly_3 indicate that the combination of the ff99SB force field and the TIP4P-Ew water model accurately simulates glycine residues (after excluding the problematic couplings from the calculation), the data we obtained for Ala_3 and, to a lesser extent, Val_3 clearly suggest that there is room for improving the intrinsic conformational preferences of non-glycine residues. More specifically our Ala_3 data imply that an increase in sampling of the PPII conformation could yield better agreement with the experimental data. Similar observations were made previously by Wickstrom et al., who found that increased sampling of the PPII conformation in Ala_3 and Ala_5 in TIP4P-Ew water yielded improved agreement with the scalar coupling measurements and suggested that further increases in sampling the PPII conformation would be desirable[27].

There are several potential avenues for improving the intrinsic conformational preferences of amino acids given a specific force field and water model combination including modifying partial atomic charges, van der Waals parameters, dihedral angle potentials, and other bonded interaction parameters. Of these, the two choices that would likely involve the least perturbation of the ff99SB force field's already excellent description of native state thermodynamics and dynamics (for folded proteins) would be the van der Waals parameters – as they apply to interactions with water molecules – and the backbone dihedral angle parameters.

In the course of our work, we derived optimized van der Waals (vdW) parameters for the interactions between TIP4P-Ew water molecules and alkane hydrogen and carbon atoms by fitting the vdW radii (R_i) and well depth (ε_i) of these atoms to bring the calculated solvation free energies of methane and *n*-butane into satisfactory agreement with experimentally determined values. Ultimately, we found that while these parameters yielded significantly more accurate solvation free energies for other alkane amino acid side chain analogues (e.g., propane and isobutane), they did not significantly improve the accuracy of the conformational ensemble for Val_3, the peptide that should benefit the most from such modifications due to its large alkane side chains (data not shown). We will further describe our methodology for developing optimized vdW parameters and present these data in future work.

The remaining possibility for improving intrinsic conformational preferences lies in the backbone dihedral angle potentials – the subject of numerous previous studies and force field development efforts[5-7,9-10]. In the AMBER force fields, there are two sets of backbone dihedral angle potentials[7]. One set of these potentials is based on the ϕ and ψ angles, defined as the torsions about the C-N-Cα-C and N-Cα-C-N atoms, respectively. These potentials apply to all residues in AMBER. The second set is based on the ϕ' and ψ' angles, defined as the torsions about the C-N-Cα-Cβ and Cβ-Cα-C-N atoms, respectively. These potentials apply to all *non-glycine* residues in AMBER. Given that we obtained χ^2 values less than 1 for Gly_3 (after excluding the problematic $^2J(N',C_\alpha)$ and $^3J(C,C')$ couplings), we elected to focus on improving the ϕ' and ψ' potentials rather than perturb the ϕ and ψ potentials.

There are two ways to modify these potentials that would increase sampling of the PPII conformation. First, one could shift the ψ' potential to increase sampling of both the β and PPII conformations, thereby decreasing the sampling of the α conformation. Recent studies, however, have suggested that the ff99SB force field may actually benefit from increased rather than decreased sampling of the α conformation[5]. The second possibility is to shift the ϕ' potential to

less frequently sample the β conformation and more frequently sample the PPII conformation. In principle, even large changes to the ϕ' potential would not greatly impact the overall sampling of the α conformation, although it would bias residues away from sampling conformations in the α basin with large negative ϕ values toward sampling those with ϕ angles near -60°. We therefore focused on optimizing the ϕ' potential to increase sampling of the PPII conformation.

Rather than introduce a new term to the ϕ' potential which would require determining an optimal energy magnitude and angular offset, we examined the existing ϕ' potential, which contains three terms, each with a different periodicity. The n=2 term (i.e., the term with 2 maxima/minima over the range of ϕ') affects the height of the two potential energy maxima at ϕ = -60° and 120°, assuming a geometric relationship of ϕ = ϕ' + 120°. By modestly decreasing the magnitude of the n=2 term, we could lower these barriers and increase the sampling of conformations near ϕ = -60°, thereby increasing the sampling of the PPII conformation. (Conformations at ϕ = 120° are sufficiently high in energy due to both steric clash and the n=1 term of the potential to preclude significant sampling.)

We performed REMD simulations of a number of different GXG peptides (where X = A, E, F, S, or V) in TIP4P-Ew water, varying the n=2 potential energy term from 2.00 kcal/mol (the ff99SB value) down to 1.55 kcal/mol, and calculating χ^2 values for predicted couplings. We used GXG peptides for the parameterization process because there are high quality coupling data available for them[37] and they provide a minimally perturbing context in which to examine the intrinsic conformational preferences of the central amino acids. In calculating the χ^2 values, we exclusively used the Orig. Karplus equation parameters because they effectively average over all residue types, unlike the DFT1 and DFT2 parameters, which are optimized for alanine. We found that there was an improvement in the χ^2 values for all five peptides as the term was decreased to 1.85 kcal/mol and continued improvement for three of the five peptides (GAG, GEG, and GFG) to 1.80 kcal/mol (Figure 1a). Moreover, at 1.80 kcal/mol, the simulations of all five peptides yielded χ^2 values less than 1.0, indicating that the conformational preferences were accurate to within the limits of the Karplus equation parameterization (Figure 1a).

To validate this change to the n=2 term of the ϕ' potential (Figure 1b), we performed REMD simulations using the aforementioned Val_3 peptide, as well as GLG and Ala_5 peptides, in TIP4P-Ew water. The Val_3 simulations indicate an overall improvement in the χ^2 value – larger than that observed for the optimized vdW parameters – but much of this improvement is in the predicted $^3J(C,C')$ coupling (Table 5). This result is somewhat unexpected in that the parameterization of the potential did not involve evaluation against any $^3J(C,C')$ couplings, as these couplings were not measured for the GXG peptides[37].

The GLG simulations display a modest improvement, with the χ^2 value (using only the Orig. Karplus equation parameters) decreasing from 0.513 ± 0.004 to 0.483 ± 0.011. This is concomitant with an increase in the probability of sampling the PPII conformation from 0.376 to 0.441, as well as a slight increase in sampling of α conformations (SI Table 6). The modified potential substantially improves the conformational ensembles for Ala_5, resulting in a 40-50% decrease in the χ^2 values computed with the Orig. and DFT2 parameters (Table 6). This improvement is accompanied by the probability of sampling PPII conformation increase from

0.443 to 0.542, with no observed increase in the fraction of α conformations (Figure 2, SI Table 7).

Lastly, while this new parameterization is intended to improve intrinsic conformational preferences, it is important to know that it does not adversely affect native protein stability or dynamics. To verify this, we performed 60 ns simulations of ubiquitin, a well-characterized protein used in previous force field validation efforts. We observed no significant difference between our modified force field and ff99SB in examining either RMSDs from the crystal structure (Figure 3a) or the computed NMR S^2 order parameters (Figure 3b).

DISCUSSION

The majority of previous MD force field development and optimization efforts have focused on improving the agreement between gas phase *ab initio* and molecular mechanics calculations as the primary means of improving MD simulation accuracy[2,4,7-9,11,52]. As simulations of proteins, nucleic acids, and other biomolecules are generally carried out in the condensed phase, however, one of the most pressing questions in the field is how accurately such parameters describe these molecules in the condensed phase – particularly in aqueous solution – and to what extent different solvent models may influence their structural ensembles[25]. Moreover, the accuracy of force fields and solvent models is paramount for simulations of intrinsically disordered proteins, as their manifold conformational states are similar in free energy[21] and interactions with solvent are enhanced relative to folded proteins[25].

We therefore set out to assess the intrinsic conformational preferences of alanine, glycine, and valine in the AMBER ff99SB force field in combination with the TIP3P and TIP4P-Ew water models by simulating the Ala_3, Gly_3, and Val_3 peptides, respectively. In the cases of Ala_3 and Gly_3, the ff99SB/TIP4P-Ew combination yielded significantly more accurate conformational preferences than the ff99SB/TIP3P combination. For Val_3, the TIP4P-Ew simulations were systematically in better agreement with the experimental measurements, but the differences from TIP3P were not statistically significant. Simulations of all three peptides demonstrated that the primary difference between the TIP3P and TIP4P-Ew ensembles is an increase in the extended (primarily PPII) conformations relative to more compact α-helical conformations.

In the case of Ala_3, the increase in sampling of the PPII conformation across the temperature range 275-350 K is unambiguously correlated with an improvement in agreement with NMR scalar coupling data. A similar correlation was noted in a previous study comparing the ensembles of Ala_3 and Ala_5 in TIP3P and TIP4P-Ew water at 300 K[27]. These observations are reinforced by data from another recent force field study, which demonstrated that the AMBER ff03/ff03* force fields yielded better agreement with scalar coupling data for the Ala_5 peptide than ff99SB/ff99SB* (in TIP3P water)[5]. One of the primary differences between the two force field families is the greater sampling of the PPII conformation by ff03/ff03* relative to ff99SB/ff99SB*[5]. Moreover, these data are consistent with a multitude of experimental data regarding alanine in short peptides, which suggest that it primarily samples the PPII conformation[36-37,53-55].

While the Gly_3 results suggested that the intrinsic conformational preferences of glycine were already adequate using the ff99SB/TIP4P-Ew combination, the Ala_3 and Val_3 simulation suggested that further improvements were possible. We first explored the creation of van der

Waals parameters optimized for simulation in TIP4P-Ew water, but found that while they yielded significantly more accurate calculations of solvation free energy, but they were unable to significantly improve the conformational ensemble of Val_3. We then considered a second approach – modifying the backbone dihedral angle potentials, specifically focusing on the ϕ' potential governing the balance between the β and PPII conformations. By lowering the energy scale of one term of this potential by 0.20 kcal/mol to increase sampling of the PPII conformation and performing REMD simulations of a variety of GXG peptides to assess the effects of our changes, we revised the ϕ' potential to yield more accurate intrinsic conformational preferences for a wide range of amino acids in TIP4P-Ew water while also maintaining the excellent native state stability of the ff99SB force field.

It is important to note that this revised potential is designed to increase the accuracy of the *intrinsic* conformational preferences of single amino acids and there is no *a priori* reason that it should improve other types of conformational preferences, such as α-helix or β-sheet formation propensities, which are a combination of intrinsic conformational preferences and cooperative interactions between residues (e.g., hydrogen bonds). This makes our optimization strategy distinct from the ff99SB* modification of Best and Hummer, for example, which aims to improve the description of the helix-to-coil transition and uses fractional helicities of a longer α-helix-forming peptide in the parameterization process[5]. As Best and Hummer have suggested, it is likely that additional physics and/or potentials must be introduced into current force fields to accurately capture such cooperative interactions[5].

Another related issue that is often overlooked in the biomolecular simulation community is the use of solvent models with non-bonded interaction schemes that are different from those used in the parameterization of those solvent models. In particular, the TIP3P model was parameterized using simple truncation (cutoffs) for both electrostatic and van der Waals interactions[30], whereas many current studies – including this work, as well as other force field development and assessment studies[5,7,27] – utilize TIP3P along with particle mesh Ewald (PME) for long-range electrostatics and corrections for van der Waals interactions beyond the direct-space cutoff. By contrast, TIP4P-Ew was derived specifically for use with modern simulation techniques, including PME and long-range van der Waals corrections[33]. It has been shown that the accuracy of the TIP3P model is degraded under these simulation conditions[56], and therefore it may be possible that the accuracy of the conformational ensembles generated with TIP3P could be improved if we were to revert to a simple truncated scheme for non-bonded interactions. Nonetheless, even under ideal simulation conditions, TIP3P does not reproduce experimentally measured characteristics of liquid water as accurately as TIP4P-Ew[33,56]. Given the central role of the solvent model in representing peptide-solvent interactions[25], the optimization of current force fields for use with solvent models more advanced than TIP3P is logically sound and has already been suggested to be a promising avenue for force field development[6]. More critically, while abandoning the use of PME would yield increased accuracy of the TIP3P model, it is known that disregarding long-range electrostatic interactions in MD simulations can lead to unphysical behavior of biomolecules[57-59]. We therefore believe it both appropriate and sensible to have performed our assessments of TIP3P (and TIP4P-Ew) using PME and to have carried out our optimization efforts using TIP4P-Ew.

In addition to assessing the impact of solvent models on intrinsic conformational preferences, our simulations of Ala_3, Gly_3, and Val_3 revealed a number of insights into the calculation of

experimental observables (scalar couplings) from structural ensembles. For Ala_3 (and to a lesser extent, Gly_3) we observed that the ff99SB/TIP4P-Ew combination yielded lower χ^2 values than the ff99SB/TIP3P *regardless* of the Karplus equation parameters used. For Val_3, however, decreases in the χ^2 values obtained with the Orig. and DFT2 parameters were correlated with *increases* in χ^2 values obtained with the DFT1 parameters. While the Orig. parameters were obtained by fitting measured NMR couplings to ϕ/ψ angles measured from high resolution X-ray or NMR structures[60-63], the DFT1 and DFT2 parameters were derived from DFT calculations of the scalar couplings for the Ac-Ala-Nme dipeptide and NH_2-Ala-Ala-NH_2 peptide, respectively[64]. Thus, the Orig. parameters implicitly average over all residue types, while the DFT1 and DFT2 parameters were derived explicitly using alanine residues. This in turn suggests that the Orig. parameters may be applied with more or less equivalent accuracy to all amino acids, whereas the DFT1 and DFT2 parameters may be accurate for alanine residues, but their applicability to other residue types is in question. Moreover, the length of the peptide used for the Karplus equation parameterization may also matter. The modified ϕ' potential yielded improved χ^2 values for the Ala_5 peptide when computed using the Orig. or DFT2 parameters, but higher χ^2 values when using the DFT1 parameters.

The need for residue-specific Karplus equation parameters is further exemplified by the case of Gly_3, in which the parameters for the $^2J(N',C_\alpha)$ coupling – obtained by fitting the Karplus equation with ϕ/ψ angles of a refined NMR structure[60] – are not able to generate predicted couplings large enough to match the experimentally measured couplings. An examination of the original data for this set of Karplus equation parameters reveals that while there is some correlation between predicted and measured couplings, there is also considerable spread between the results, especially for the residues in extended conformations[60]. We observed similar shortcomings using other residue-averaged parameterizations[63] for this coupling (data not shown).

Together these data suggest that deriving residue-specific Karplus equation parameters would significantly improve the calculation of backbone couplings, particularly those involving Cα and/or Cβ atoms, which have very different chemistries across the range of amino acids. It may be sufficient to derive a few sets of parameters (e.g., one set for the aromatic side chain residues, one set for the branched Cβ side chain residues, etc.) rather than a unique set for each amino acid, but further study is needed to investigate the accuracy of such approaches. In addition, it is clear that the length of peptide used in the parameterization process affects the results and that dipeptides may be of insufficient length to yield accurate parameters even for fairly short peptides (e.g., Ala_5). Calculating couplings with modestly larger peptides (e.g., GXG) may be sufficient to minimize the parameterization errors due to finite length. We intend to investigate both residue-specific and length effects in deriving Karplus equation parameters in future work.

CONCLUSIONS

Exciting new frontiers in biology, such as intrinsically disordered proteins, require an unprecedented interplay between simulation and experiment to fully understand the behavior of these biomolecules[35,65]. This work and others demonstrate that current force field and water model combinations still require improvement to accurately describe disordered states[26-27] and

that such improvements may be realized by utilizing condensed phase simulations and experimental data to fine-tune parameters[5-6] rather than relying solely on matching gas phase *ab initio* data, as has often been done in the past. A "hybrid" strategy of using gas phase *ab initio* data for initial parameterization and quantitative experimental data for fine-tuning those parameters may be valuable not only in the optimization of existing force fields, but also in the development of next-generation fixed-charge and polarizable force fields. This force field development strategy, however, hinges upon a simultaneous development of more accurate methods for calculating experimental observables from simulated ensembles. Improvements in both of these areas will be critical for the interplay between simulation and experiment necessary for characterizing IDPs, as well as the overall advance of MD simulations as applied to biomolecules in general.

ACKNOWLEDGEMENTS

The work reported here is supported by the NSF Cyber-Infrastructure Award 0344670, as well as the 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.

SUPPORTING INFORMATION AVAILABLE

Calculated scalar couplings for Ala_3 at 275 K, conformational preferences of the central residues of Gly_3, Val_3, and Ala_5 at 300 K and of GLG at 298 K, conformational preferences of the first/N-terminal and third/C-terminal residues of Ala_3 at 275 K, and partial atomic charges for protonated C-terminal (-COOH) Ala, Gly, and Val residues. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

(1) van Gunsteren, W. F.; Dolenc, J. *Biochem. Soc. Trans.* **2008,** *36*, 11-15.
(2) 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.
(3) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996,** *118*, 11225-11236.
(4) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; 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.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998,** *102*, 3586-3616.
(5) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009,** *113*, 9004-9015.
(6) Best, R. B.; Mittal, J. *J. Phys. Chem. B* **2010,** *114*, 14916-14923.
(7) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006,** *65*, 712-725.

(8) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010,** *78*, 1950-1958.
(9) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd. *J. Comput. Chem.* **2004,** *25*, 1400-1415.
(10) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005,** *88*, 2472-2493.
(11) 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.
(12) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007,** *111*, 2242-2254.
(13) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003,** *66*, 27-85.
(14) Daggett, V. *Chem. Rev.* **2006,** *106*, 1898-1916.
(15) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102*, 6679-6685.
(16) Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu. Rev. Phys. Chem.* **2007,** *58*, 57-83.
(17) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell. Biol.* **2005,** *6*, 197-208.
(18) Fink, A. L. *Curr. Opin. Struct. Biol.* **2005,** *15*, 35-41.
(19) Uversky, V. N. *Protein Sci.* **2002,** *11*, 739-756.
(20) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biophys.* **2008,** *37*, 215-246.
(21) Turoverov, K. K.; Kuznetsova, I. M.; Uversky, V. N. *Prog. Biophys. Mol. Biol.* **2010,** *102*, 73-84.
(22) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. *Annu. Rev. Biophys.* **2008,** *37*, 289-316.
(23) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991,** *254*, 1598-1603.
(24) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107*, 8183-8188.
(25) Florova, P.; Sklenovsky, P.; Banas, P.; Otyepka, M. *J. Chem. Theory Comput.* **2010,** *6*, 3569-3579.
(26) Best, R. B.; Buchete, N. V.; Hummer, G. *Biophys. J.* **2008,** *95*, L07-L09.
(27) Wickstrom, L.; Okur, A.; Simmerling, C. *Biophys. J.* **2009,** *97*, 853-856.
(28) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009,** *5*, 350-358.
(29) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005,** *122*, 134508.
(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983,** *79*, 926-935.
(31) Abascal, J. L.; Vega, C. *J. Chem. Phys.* **2005,** *123*, 234505.
(32) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987,** *91*, 6269-6271.
(33) 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.
(34) Rick, S. W. *J. Chem. Phys.* **2004,** *120*, 6085-6093.
(35) Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T. *J. Am. Chem. Soc.* **2008,** *130*, 6145-6158.
(36) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J. Am. Chem. Soc.* **2007,** *129*, 1179-1189.
(37) Hagarman, A.; Measey, T. J.; Mathieu, D.; Schwalbe, H.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2010,** *132*, 540-551.
(38) 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.

(39) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993,** *115*, 9620-9631.
(40) Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y. *R.E.D. Server: a web service designed to derive RESP and ESP charges and to generate force field libraries for new molecules/molecular fragments*, Université de Picardie Jules Verne and Sanford-Burnham Institute for Medical Research, 2010.
(41) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; A. Roitberg; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; 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*, University of California, San Francisco, 2010.
(42) Joung, I. S.; Cheatham, T. E., 3rd. *J. Phys. Chem. B* **2008,** *112*, 9020-9041.
(43) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993,** *98*, 10089-10092.
(44) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995,** *103*, 8577-8593.
(45) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977,** *23*, 327-341.
(46) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984,** *81*, 3684-3690.
(47) Okur, A.; Roe, D. R.; Cui, G. L.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2007,** *3*, 557-568.
(48) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999,** *314*, 141-151.
(49) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987,** *194*, 531-544.
(50) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. *J. Am. Chem. Soc.* **1995,** *117*, 12562-12566.
(51) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982,** *104*, 4546-4559.
(52) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000,** *21*, 1049-1074.
(53) Schweitzer-Stenner, R. *J. Phys. Chem. B* **2009,** *113*, 2922-2932.
(54) Shi, Z.; Chen, K.; Liu, Z.; Ng, A.; Bracken, W. C.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102*, 17964-17968.
(55) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99*, 9190-9195.
(56) Price, D. J.; Brooks, C. L., 3rd. *J. Chem. Phys.* **2004,** *121*, 10096-10103.
(57) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995,** *117*, 4193-4194.
(58) Norberg, J.; Nilsson, L. *Biophys. J.* **2000,** *79*, 1537-1553.
(59) Zuegg, J.; Gready, J. E. *Biochemistry* **1999,** *38*, 13862-13876.
(60) Ding, K.; Gronenborn, A. M. *J. Am. Chem. Soc.* **2004,** *126*, 6232-6233.
(61) Hennig, M.; Bermel, W.; Schwalbe, H.; Griesinger, C. *J. Am. Chem. Soc.* **2000,** *122*, 6268-6277.
(62) Hu, J. S.; Bax, A. *J. Am. Chem. Soc.* **1997,** *119*, 6360-6368.
(63) Wirmer, J.; Schwalbe, H. *J. Biomol. NMR* **2002,** *23*, 47-55.
(64) Case, D. A.; Scheurer, C.; Bruschweiler, R. *J. Am. Chem. Soc.* **2000,** *122*, 10390-10397.
(65) Sgourakis, N. G.; Merced-Serrano, M.; Boutsidis, C.; Drineas, P.; Du, Z.; Wang, C.; Garcia, A. E. *J. Mol. Biol.* **2011,** *405*, 570-583.

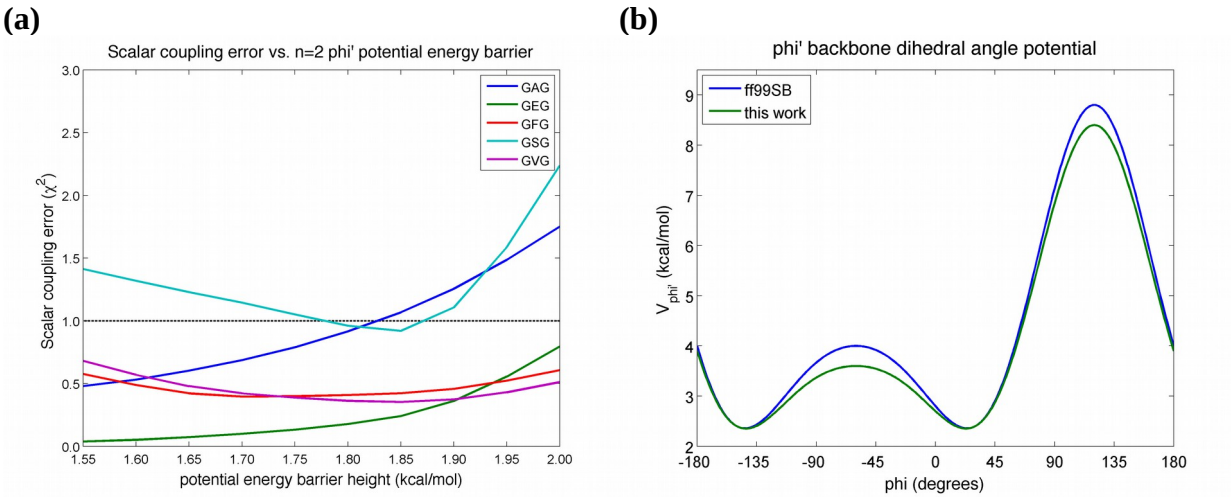

Figure 1. (a) Scalar coupling error (χ^2 value) as a function of the n=2 term of the ϕ' backbone dihedral angle potential energy for GAG (dark blue), GEG (green), GFG (red), GSG (cyan), and GVG (magenta) peptides. (b) Potential energy as a function of ϕ for the ϕ' backbone potential, assuming a geometric relationship of ϕ = ϕ' + 120°, for ff99SB (dark blue) and the optimized potential presented in this work (green).

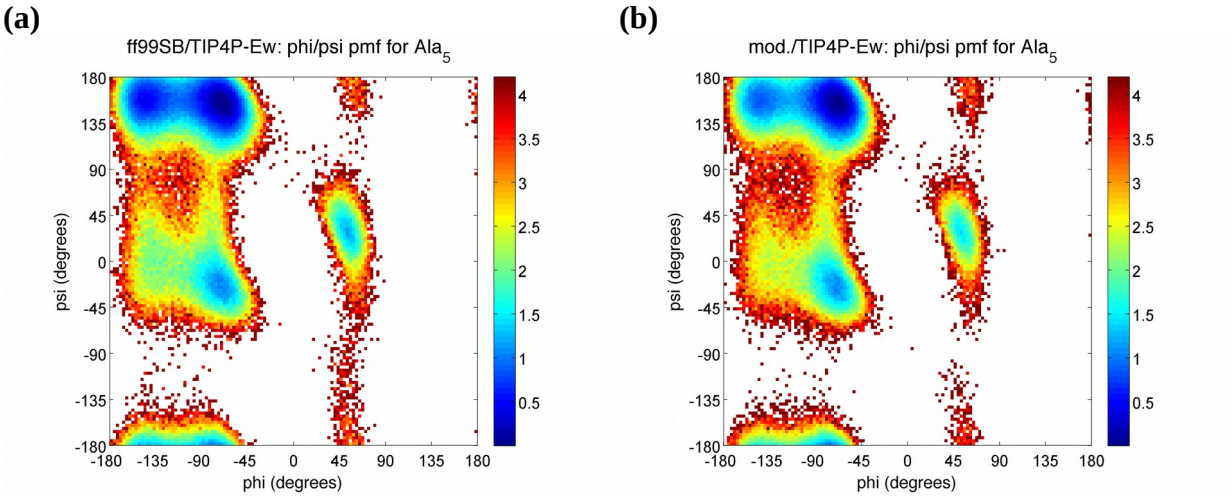

Figure 2. Average conformational preferences of residues 2-4 in the Ala_5 peptide using (a) the unmodified ff99SB force field or (b) the ff99SB force field with optimized ϕ' potential. Conformational preferences are represented as a potential of mean force (pmf), $W(\phi,\psi) = -RT \log p(\phi,\psi)$, with relative free energies given in kcal/mol.

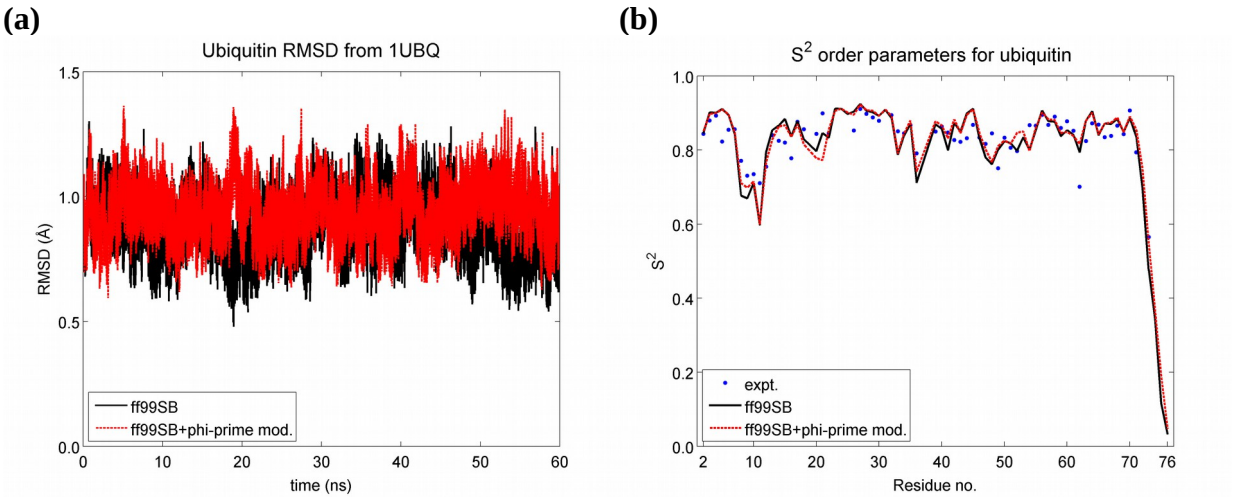

Figure 3. Native state stability and dynamics of unmodified ff99SB (black) and ff99SB with optimized ϕ' potential (red) for ubiquitin. (a) Root mean square distance (RMSD) to crystal structure 1UBQ[49] over 60 ns of simulation. The mean RMSDs are 0.88 Å for unmodified ff99SB and 0.94 Å for ff99SB with the optimized potential. (b) Lipari-Szabo order parameters (S^2) for ubiquitin at 300 K, pH 4.7, with experimentally derived isotropic values shown in blue dots[50]. The root mean square errors relative to the experimental S^2 values are 0.044 for both unmodified ff99SB and ff99SB with the optimized ϕ' potential.

Table 1. χ^2 values for calculated scalar couplings of Ala_3 at (a) 350 K, (b) 325 K, (c) 300 K, and (d) 275 K for various force field and water model combinations. Values are given as the means over four independent simulations, with the standard errors of the means given in parentheses. For RREMD simulations, probabilities are given as means over two independent simulations, with the differences between the two simulations given in parentheses.

(a) 350 K

	Orig.	**DFT1**	**DFT2**
ff99SB (TIP3P)	1.82 (0.09)	1.30 (0.03)	1.40 (0.09)
ff99SB (TIP4P-Ew)	1.60 (0.06)	1.18 (0.01)	1.16 (0.06)
ff99SB* (TIP4P-Ew)	1.54 (0.05)	1.20 (0.02)	1.11 (0.05)

(b) 325 K

	Orig.	**DFT1**	**DFT2**
ff99SB (TIP3P)	1.63 (0.07)	0.97 (0.05)	1.21 (0.07)
ff99SB (TIP4P-Ew)	1.39 (0.04)	0.87 (0.04)	0.96 (0.04)

(c) 300 K

	Orig.	**DFT1**	**DFT2**
ff99SB (TIP3P)	2.38 (0.21)	1.21 (0.13)	1.80 (0.21)
ff99SB (TIP4P-Ew)	1.90 (0.20)	0.99 (0.09)	1.31 (0.19)

(d) 275 K

	Orig.	**DFT1**	**DFT2**
ff99SB (TIP3P)	2.92 (0.16)	1.97 (0.10)	2.51 (0.15)
ff99SB (TIP3P, RREMD)	2.76 (0.54)	1.91 (0.35)	2.36 (0.54)
ff99SB (TIP4P-Ew)	1.95 (0.33)	1.53 (0.13)	1.59 (0.29)
ff99SB (TIP4P-Ew, RREMD)	2.14 (0.07)	1.49 (0.06)	1.72 (0.08)
ff99SB* (TIP4P-Ew, RREMD)	2.71 (0.13)	1.90 (0.06)	2.30 (0.11)

Table 2. Conformational preferences of the central residue of Ala_3 at (a) 350 K, (b) 325 K, (c) 300 K, and (d) 275 K for various force field and water model combinations. Values are given as the means over four independent simulations, with the standard errors of the means given in parentheses. For RREMD simulations, probabilities are given as means over two independent simulations, with the differences between the two simulations given in parentheses.

(a) 350 K

	α	**β**	**PPII**	**other**
ff99SB (TIP3P)	0.169 (0.008)	0.390 (0.002)	0.392 (0.002)	0.048 (0.006)
ff99SB (TIP4P-Ew)	0.118 (0.006)	0.420 (0.005)	0.426 (0.006)	0.036 (0.012)
ff99SB* (TIP4P-Ew)	0.139 (0.007)	0.412 (0.006)	0.413 (0.004)	0.036 (0.012)

(b) 325 K

	α	**β**	**PPII**	**other**
ff99SB (TIP3P)	0.142 (0.004)	0.391 (0.006)	0.422 (0.002)	0.044 (0.014)
ff99SB (TIP4P-Ew)	0.100 (0.006)	0.416 (0.004)	0.456 (0.006)	0.028 (0.008)

(c) 300 K

	α	**β**	**PPII**	**other**
ff99SB (TIP3P)	0.121 (0.004)	0.396 (0.004)	0.450 (0.002)	0.032 (0.005)
ff99SB (TIP4P-Ew)	0.081 (0.002)	0.411 (0.004)	0.488 (0.005)	0.021 (0.006)

(d) 275 K

	α	**β**	**PPII**	**other**
ff99SB (TIP3P)	0.102 (0.009)	0.390 (0.004)	0.476 (0.006)	0.032 (0.012)
ff99SB (TIP3P, RREMD)	0.106 (0.004)	0.388 (0.017)	0.479 (0.007)	0.027 (0.020)
ff99SB (TIP4P-Ew)	0.068 (0.005)	0.402 (0.002)	0.519 (0.005)	0.011 (0.007)
ff99SB (TIP4P-Ew, RREMD)	0.069 (0.015)	0.399 (0.007)	0.511 (0.004)	0.021 (0.003)
ff99SB* (TIP4P-Ew, RREMD)	0.063 (0.011)	0.402 (0.008)	0.526 (0.003)	0.009 (0.005)

Table 3. χ^2 values for calculated scalar couplings of Gly_3 at 300 K for various force field and water model combinations. Values are given as the means over four independent simulations, with the standard errors of the means given in parentheses.

	All couplings			**No $^2J(N',C_\alpha)$ or $^3J(C,C')$ coupling**		
	Orig.	**DFT1**	**DFT2**	**Orig.**	**DFT1**	**DFT2**
ff99SB (TIP3P)	3.21 (0.03)	3.45 (0.05)	3.26 (0.04)	0.57 (0.02)	1.14 (0.04)	0.73 (0.03)
ff99SB (TIP4P-Ew)	2.93 (0.04)	3.11 (0.08)	2.96 (0.05)	0.47 (0.05)	0.98 (0.11)	0.62 (0.06)
ff99SB* (TIP4P-Ew)	2.92 (0.02)	3.08 (0.06)	2.93 (0.03)	0.46 (0.02)	0.96 (0.06)	0.60 (0.03)

Table 4. χ^2 values for calculated scalar couplings of Val_3 at 300 K for various force field and water model combinations. Values are given as the means over four independent simulations, with the standard errors of the means given in parentheses. For RREMD simulations, probabilities are given as means over two independent simulations, with the differences between the two simulations given in parentheses.

	All couplings			**No $^3J(C,C')$ coupling**		
	Orig.	**DFT1**	**DFT2**	**Orig.**	**DFT1**	**DFT2**
ff99SB (TIP3P)	2.00 (0.17)	2.99 (0.06)	2.31 (0.18)	1.22 (0.16)	2.50 (0.03)	1.64 (0.16)
ff99SB (TIP3P, RREMD)	1.88 (0.29)	3.13 (0.06)	2.18 (0.30)	1.11 (0.29)	2.67 (0.03)	1.54 (0.30)
ff99SB (TIP4P-Ew)	1.91 (0.25)	3.33 (0.17)	2.24 (0.29)	1.21 (0.29)	2.97 (0.20)	1.68 (0.33)
ff99SB	1.73	3.24	2.05	0.97	2.81	1.42

(TIP4P-Ew, RREMD)	(0.15)	(0.09)	(0.16)	(0.18)	(0.08)	(0.18)
ff99SB* (TIP4P-Ew, RREMD)	1.69 (0.14)	3.09 (0.12)	1.98 (0.15)	0.92 (0.13)	2.64 (0.10)	1.33 (0.14)

Table 5. χ^2 values for calculated scalar couplings of Val_3 at 300 K for unmodified ff99SB and ff99SB with the optimized ϕ' backbone dihedral angle potential, both with TIP4P-Ew water. Values are given as the means over two independent RREMD simulations, with the differences between the two simulations given in parentheses. Only the Orig. Karplus equation parameters are used for these calculations.

	All couplings	**No ^{3}J(C,C') coupling**
ff99SB (unmodified)	1.73 (0.15)	0.97 (0.18)
Opt. ϕ' dihedral potential	1.53 (0.02)	1.10 (0.00)

Table 6. χ^2 values for calculated scalar couplings of Ala_5 at 300 K for unmodified ff99SB and ff99SB with the optimized ϕ' backbone dihedral angle potential, both with TIP4P-Ew water. Values are given as the means over two independent RREMD simulations, with the differences between the two simulations given in parentheses.

	All couplings			**No ^{3}J(C,C') coupling**		
	Orig.	**DFT1**	**DFT2**	**Orig.**	**DFT1**	**DFT2**
ff99SB (unmodified)	2.44 (0.10)	1.87 (0.06)	2.14 (0.11)	1.73 (0.09)	1.20 (0.03)	1.37 (0.08)
Opt. ϕ' dihedral potential	1.33 (0.04)	2.13 (0.05)	1.26 (0.02)	0.86 (0.05)	1.85 (0.05)	0.86 (0.02)