

# Improved side-chain torsion potentials for the Amber ff99SB protein force field

Kresten Lindorff-Larsen,<sup>1</sup> Stefano Piana,<sup>1</sup> Kim Palmo,<sup>1</sup> Paul Maragakis,<sup>1</sup> John L. Klepeis,<sup>1</sup> Ron O. Dror,<sup>1</sup> and David E. Shaw<sup>1,2\*</sup>

<sup>1</sup> D. E. Shaw Research, New York, New York 10036

<sup>2</sup> Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10032

## ABSTRACT

Recent advances in hardware and software have enabled increasingly long molecular dynamics (MD) simulations of biomolecules, exposing certain limitations in the accuracy of the force fields used for such simulations and spurring efforts to refine these force fields. Recent modifications to the Amber and CHARMM protein force fields, for example, have improved the backbone torsion potentials, remedying deficiencies in earlier versions. Here, we further advance simulation accuracy by improving the amino acid side-chain torsion potentials of the Amber ff99SB force field. First, we used simulations of model alpha-helical systems to identify the four residue types whose rotamer distribution differed the most from expectations based on Protein Data Bank statistics. Second, we optimized the side-chain torsion potentials of these residues to match new, high-level quantum-mechanical calculations. Finally, we used microsecond-timescale MD simulations in explicit solvent to validate the resulting force field against a large set of experimental NMR measurements that directly probe side-chain conformations. The new force field, which we have termed Amber ff99SB-ILDN, exhibits considerably better agreement with the NMR data.

Proteins 2010; 78:1950–1958.  
© 2010 Wiley-Liss, Inc.

**Key words:** molecular dynamics simulation; molecular mechanics; NMR; rotamer; side chain; protein dynamics; quantum mechanics; dihedral.

## INTRODUCTION

Molecular dynamics (MD) simulations with atomistic, physics-based force fields offer the potential to gain new insight into the functional mechanisms of biomolecules. The successful application of such simulations may, however, be limited by two major shortcomings. First, computational requirements restrict the amount of time that can be simulated and may thus prevent a simulation from exploring all relevant molecular conformations (the *sampling problem*). Second, inaccuracies in the potential energy function may bias the simulation toward incorrect conformations (the *force field problem*).<sup>1</sup> Although progress in overcoming the sampling problem could make new and important biological phenomena accessible to computational study for the first time, the success of such efforts is critically dependent on force field quality, because inaccuracies in the physical models used for molecular simulation may produce misleading results even in the absence of any computational limitations. Recent advances in both software and hardware have made possible the simulation of biologically relevant processes with atomistic accuracy on timescales well beyond the microsecond.<sup>1–3</sup> These developments, combined with continuous improvements to enhanced sampling techniques,<sup>4</sup> have placed ever greater demands on force field accuracy.

As exemplified by recent work on nucleic acids<sup>5</sup> and proteins,<sup>2</sup> long MD simulations have been used to highlight deficiencies in existing force fields, leading in turn to the development of new and improved versions.<sup>5</sup> Although much current development in this area is focused on the inclusion of polarization effects,<sup>6</sup> polarizable force fields are computationally more expensive than their fixed-charge counterparts, and recent studies suggest that there is still room for improvement of nonpolarizable force fields. Minor yet important modifications to the backbone torsion potentials incorporated in recent versions of the Amber and CHARMM force fields (Amber ff99SB<sup>7</sup> and the CMAP backbone energy correction to CHARMM22<sup>8</sup>) have led to improvements in accuracy compared with earlier releases, as demonstrated, for example, through the comparison of simulation results to experimental data.<sup>9</sup>

Here, we further refine the Amber ff99SB protein force field by optimizing the  $\chi_1$  torsion potentials for amino acid side chains. Among the torsional degrees of freedom in proteins, the  $\chi_1$  torsions are expected to be second only to the backbone torsions in importance for describing protein energetics,

Additional Supporting Information may be found in the online version of this article.  
Kresten Lindorff-Larsen and Stefano Piana contributed equally to this work.

\*Correspondence to: David E. Shaw, D. E. Shaw Research, New York, NY 10036. E-mail: David.Shaw@DEShawResearch.com

Received 11 December 2009; Revised 5 February 2010; Accepted 9 February 2010

Published online 9 March 2010 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22711

yet the relevant terms in the Amber force field have remained virtually identical for 25 years.<sup>10–12</sup> Because these  $\chi_1$  torsion potentials have not, to our knowledge, been systematically revised since their initial introduction, they seemed a natural target for improvement.

We here focused our efforts on those side chains that displayed the largest deviations from expected behavior and used a three-step procedure to refine the force field. First, we identified putatively problematic residue types by comparing the distribution of  $\chi_1$  dihedrals in simulations of short helical peptides with the corresponding statistics for residues in helices in the Protein Data Bank (PDB).<sup>13</sup> We found that four residue types (isoleucine, leucine, aspartate, and asparagine) exhibited particularly large deviations from the PDB distribution, suggesting that the ff99SB force field does not model these side chains well. Second, we obtained new  $\chi_1$  torsion potentials for these four residues by fitting force-field parameters to *ab initio* quantum level DF-LMP2<sup>14,15</sup> dihedral scans. Finally, we validated the refined force field using a large set of NMR data containing hundreds of measurements that directly probe the relevant side-chain conformations. We found substantial improvements for all four modified residues, as demonstrated by significantly closer agreement between the rotameric states observed in the simulations and those probed by NMR experiments.

## MATERIALS AND METHODS

### MD simulations of short polyalanine helices

We solvated terminally capped alanine-based helices with the sequence Ace-(Ala)<sub>4</sub>Xaa(Ala)<sub>4</sub>-NMA, where Xaa is any 1 of the 20 naturally occurring amino acids other than Gly, Ala, and Pro, in a cubic box with sides of  $\sim 27$  Å containing  $\sim 600$  water molecules. Protonation states were chosen to correspond to neutral pH. Because the goal was to compare the rotamer distributions observed in MD simulations of these peptides to the distributions observed in helices, we applied a weak restraint to both the  $\phi$  and  $\varphi$  torsion angles to ensure that the peptides stayed helical. These restraints were of the form:

$$V(\theta) = k_{\theta} \sum_{n=1}^4 \frac{(-1)^{n-1}}{n!} (1 + \cos(n(\theta - \theta_0))),$$

with reference values ( $\theta_0$ ) of  $122^\circ$  and  $133^\circ$  for  $\phi$  and  $\varphi$ , respectively, and a force constant ( $k_{\theta}$ ) of  $1 \text{ kcal mol}^{-1}$ . We note that although the reference values do not correspond to the helical region of the Ramachandran map, this cosine series acts as a restraint that ensures that the peptide remains in a helical conformation throughout the entire simulation without noticeably influencing the side-chain motion.

Each system was equilibrated at 300 K and 1 atm with 2.4 ns of MD simulation in the NPT ensemble. Then,

MD simulations were carried out in the NVT ensemble for 720 ns using the Nosé-Hoover thermostat with a relaxation time of 1 ps. All simulations were performed using the Desmond MD program<sup>16</sup> version 2.1.1.0 and either the Amber ff99SB<sup>7</sup> or the modified Amber ff99SB force field described herein, which we have termed ff99SB-ILDN. All bonds involving hydrogen atoms were constrained with the SHAKE algorithm.<sup>17</sup> A cutoff of 10 Å was used for the Lennard-Jones interaction and the short-range electrostatic interactions. The smooth particle mesh Ewald method<sup>18</sup> with a  $32 \times 32 \times 32$  grid and a fourth-order interpolation scheme was used to compute the long-range electrostatic interactions. The pairlists were updated every 10 fs with a cutoff of 10.5 Å. We used a multistep RESPA scheme<sup>19</sup> for the integration of the equations of motion with timesteps of 2.0, 2.0, and 6.0 fs for the bonded, short-range nonbonded, and long-range nonbonded interactions, respectively. To check for potential biases introduced by long-range interactions between peptides in periodic images, we repeated these simulations for four of the amino acids (Xaa: Ile, Leu, Asp, and Asn) using a larger box with side length 37 Å. We found that the results of these control simulations were within error of those using the smaller box sizes.

### MD simulations of small globular proteins

MD simulations of hen egg white lysozyme (HEWL), bovine pancreatic trypsin inhibitor (BPTI), ubiquitin (Ubq), and the B3 domain of Protein G (GB3) were performed using Desmond version 2.1.0.1 and the Amber ff99SB or the modified Amber ff99SB-ILDN force fields. The TIP3P water model<sup>20</sup> was used for simulations of HEWL, Ubq, and GB3, and the TIP4P-Ew water model<sup>21</sup> was used for simulations of BPTI. Simulation parameters were the same as in the simulations of small helical peptides, apart from the fact that a  $64^3$  PME grid was used for HEWL and a  $48^3$  grid was used for BPTI, Ubq, and GB3. Simulations of HEWL, BPTI, Ubq, and GB3 were initiated from PDB<sup>22</sup> entries 6LYT, 5PTI, 1UBQ, and 1P7E solvated in cubic water boxes containing 10,594, 4215, 6080, and 5156 water molecules, respectively. The net charge of the proteins was neutralized with sodium or chloride ions. Each system was initially subject to energy minimization, followed by 1.2 ns of MD simulation in the NPT ensemble during which the temperature was increased linearly from 10 to 300 K, and position restraints on the backbone atoms were annealed from 1.0 to  $0.0 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ . After this initial relaxation, each system was simulated for 6 ns in the NPT ensemble. The frame of this simulation with the volume closest to the average volume was selected as the starting conformation for a production run of 1.2  $\mu\text{s}$  in the NVT ensemble. The trajectories obtained from the NVT runs were used for subsequent data analysis.

## Calculation of NMR properties

For all four protein systems, experimentally measured  $^3J$  coupling constants for  $H^\alpha-C^\alpha-C^\beta-H^{\beta_{1,2}}$  dihedrals are available<sup>23–27</sup> (and Bax, personal communication). The experimental values were compared to those calculated using a Karplus relationship<sup>28</sup> from the torsion angles observed in the MD simulations. For BPTI, HEWL, and Ubq, stereo-specific assignments allow us to distinguish between couplings for  $H^{\beta_1}$  and  $H^{\beta_2}$ . In GB3, where stereospecific assignments are not available, we used the independently measured  $C^\beta-H^{\beta_{1,2}}$  residual dipolar couplings (RDCs) to determine the most likely assignment, as has been suggested previously.<sup>26</sup> In addition to calculating  $H^\alpha-C^\alpha-C^\beta-H^{\beta_{1,2}}$  couplings for all four proteins, we also calculated  $N-C^\alpha-C^\beta-C^\gamma$  and  $C'-C^\alpha-C^\beta-C^\gamma$  couplings in GB3 and Ubq<sup>29,30</sup> and  $C'-C^\alpha-C^\beta-H^{\beta_{1,2}}$  couplings in HEWL.<sup>24</sup> For the  $N-C^\alpha-C^\beta-C^\gamma$  and  $C'-C^\alpha-C^\beta-C^\gamma$  couplings in Ile, Val, and Thr, we used Karplus parameters from Chou *et al.*<sup>30</sup>; for all other couplings, we used amino acid-specific parameters from Pérez *et al.*<sup>31</sup>

Side-chain RDCs for GB3 and HEWL were calculated as ensemble averages, as described earlier.<sup>32</sup> For HEWL, the alignment tensor was first determined using a set of backbone HN RDCs, and the resulting alignment tensor was then used to calculate RDCs for Asn side-chain amides.<sup>33</sup> As the experiment reports only the sum of the two RDCs for the  $N^{\delta}-H^{\delta_{1,2}}$  bonds, we calculated the same sum from the simulations. In GB3, the same procedure was used to determine the alignment tensor from a set of backbone couplings, resulting in a calculation of  $C^\beta-H^{\beta_{1,2}}$  couplings.<sup>26</sup> In total, 390 scalar couplings and 50 RDCs were calculated from the MD simulations and compared to experimental values. The complete dataset, together with the values calculated using ff99SB and ff99SB-ILDN, is available in the Supporting Information for this article.

## Ab initio calculations

Quantum mechanical (QM) calculations were performed at the MP2 level of theory, using local and density-fitting approximations,<sup>14</sup> with an augmented triple-zeta basis set (aug-cc-pVTZ) via the MOLPRO program.<sup>15</sup> Full scans of the potential energy surface (PES) around the  $\chi_1$  bond were performed for Ace-Xaa-NMA peptides, where Xaa was either Ile, Leu, Asp, or Asn. For each point on the PES, the geometry of the system was fully relaxed with the  $\chi_1$ ,  $\chi_2$ ,  $\phi$ , and  $\varphi$  angles constrained. In the Ile and Leu calculations,  $\chi_1$  was varied between  $-180^\circ$  and  $180^\circ$  in  $15^\circ$  increments, and for each value of  $\chi_1$ ,  $\chi_2$  values of  $-60^\circ$ ,  $60^\circ$ , and  $180^\circ$  were considered. A total of 72 points were optimized for each of these two residues. For Asp and Asn, both  $\chi_1$  and  $\chi_2$  were varied between  $-180^\circ$  and  $180^\circ$  in  $30^\circ$  increments. A total of 72 and 144 points were optimized for Asp and Asn, respectively (note that the calculation of the Asp  $\chi_1/\chi_2$  torsion map required half the number of points because of the

symmetry of the  $\chi_2$  torsion in Asp). In all calculations,  $\phi$  and  $\varphi$  were kept fixed to the values of  $-135^\circ$  and  $135^\circ$ , respectively, corresponding to the extended conformation.

## Parameter fitting

A fit to the potential energy scans was performed by calculating the difference between the molecular mechanics energies and the *ab initio* energies for each point on the PES. The energy terms of the  $\chi_1$  torsion in Ile and Leu and the  $\chi_1$  and  $\chi_2$  torsions in Asp and Asn were then optimized to minimize the following function:

$$\Phi = \sum_{i=1}^N (E_i^{\text{QM}} - E_i^{\text{MM}})^2 e^{-\beta E_i^{\text{QM}}},$$

where  $E^{\text{MM}}$  and  $E^{\text{QM}}$  are the force-field and QM energies, respectively, and  $N$  is the number of conformations optimized at the QM level. The differences between  $E^{\text{MM}}$  and  $E^{\text{QM}}$  are weighted by a Boltzmann factor  $e^{-\beta E_i^{\text{QM}}}$ . We set the inverse temperature,  $\beta = 1/kT$ , to  $1.0 \text{ mol kcal}^{-1}$  so as to assign to each point a weight that is intermediate between a fit to the energies ( $\beta = 0.0 \text{ mol kcal}^{-1}$ , i.e., uniform weights) and a fit to the Boltzmann populations at room temperature ( $\beta \sim 1.7 \text{ mol kcal}^{-1}$ ). Our choice of  $\beta$ , corresponding to a temperature of 500 K ( $\beta = 1/kT$ ), ensures a high level of accuracy at the minima of the energy profile without giving rise to large errors in the barrier regions. The force-field energy,  $E^{\text{MM}}$ , is given by the Amber ff99SB energy,  $E^{\text{A99SB}}$ , plus a new torsion term, that replaces the existing Amber ff99SB torsion,  $V^{\text{A99SB}}(\theta)$ :

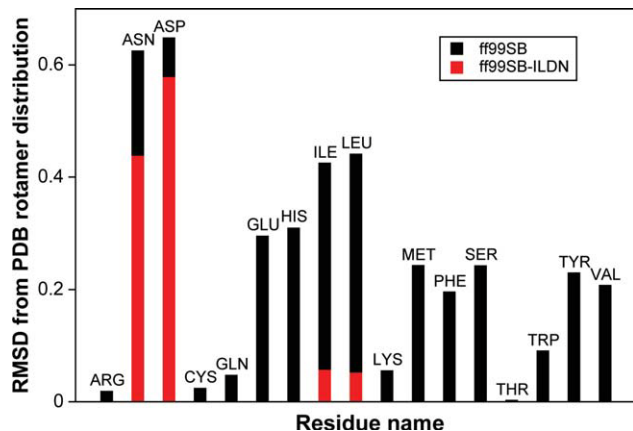
$$E^{\text{MM}} = E^{\text{A99SB}} - V^{\text{A99SB}}(\theta) + k_0 + \sum_{m=1}^M k_m (1 + \cos(m\theta - \theta_0)).$$

In this equation,  $k_0$  is a constant, the  $k_m$ s are the parameters of the fit and represent the force constants for the  $M$  terms in the cosine expansion, and  $\theta_0$  was fixed to  $0.0^\circ$ , consistent with the Amber force-field convention. Formulated in this way, the resulting parameters define a new torsion potential that is meant to replace the existing torsion term,  $V^{\text{A99SB}}(\theta)$ , in ff99SB. The number of terms,  $M$ , used in the cosine expansion was two for Ile, three for Leu, and six for both  $\chi_1$  and  $\chi_2$  in Asp and Asn. Allowing for a larger number of parameters in the Ile and Leu torsions did not result in any substantial improvement of the least-squares fit.

## RESULTS

### Comparison of rotamer distributions from MD simulations with the PDB statistics

Our approach to the refinement of the Amber ff99SB force field focused on improving the description of the side-chain  $\chi_1$  torsion angle. Previous studies have shown that the distribution of structures in the PDB may be a good approximation for the distributions expected to be



**Figure 1**

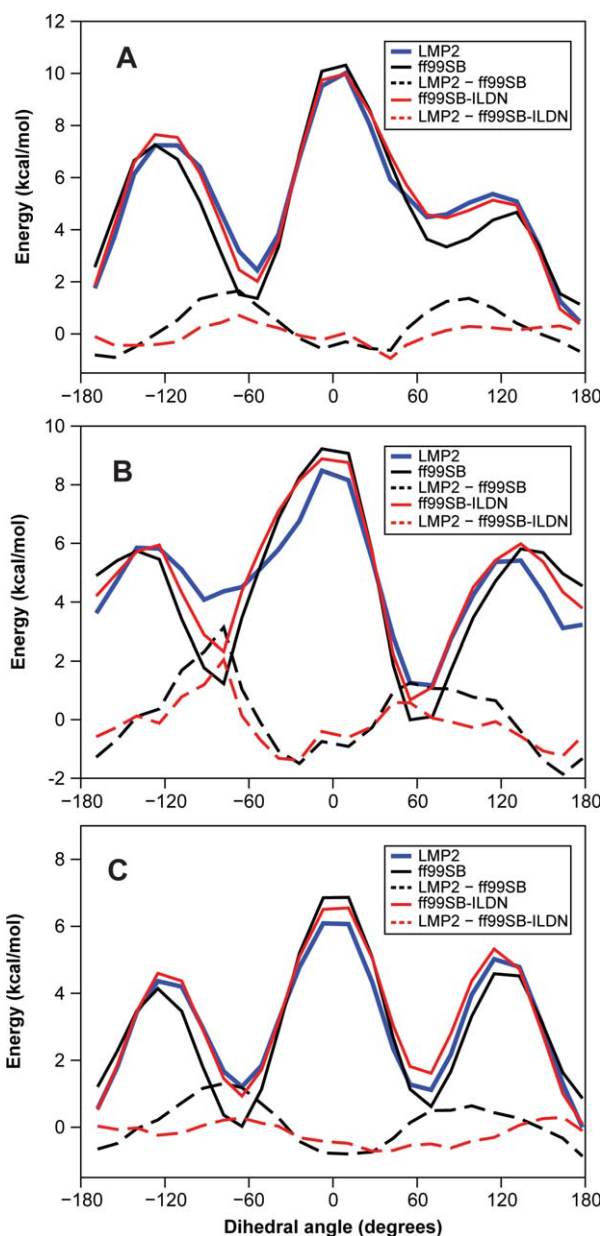
Simulation of small alpha-helical peptides with the Amber ff99SB and ff99SB-ILDN force fields. The plot shows the RMSD between the calculated rotamer distributions for each residue type and the distribution observed for the same residue in helices in the PDB. The values of the  $\chi_1$  dihedral observed in the simulations were partitioned into “plus,” “minus,” or “trans” rotamers as described previously,<sup>13</sup> and the RMSD was calculated over this three-state classification. The black bars show the results obtained using the ff99SB force field, and the red bars show the results for Ile, Leu, Asp, and Asn using the new side-chain torsion parameters (ff99SB-ILDN) described in this article.

observed in an MD simulation.<sup>8,34–36</sup> However, such agreements are not expected to be sufficiently quantitative to parameterize force fields, and we instead use comparisons between MD and PDB statistics solely to identify residues whose side chains may be inaccurately described by the force field. More specifically, we performed a series of MD simulations of short helical peptides with the sequences  $(\text{Ala})_4\text{Xaa}(\text{Ala})_4$ , where Xaa is any 1 of the 20 amino acids apart from Gly, Ala, and Pro. From these simulations, we calculated the relative populations of the plus (p), minus (m) and trans (t)  $\chi_1$  rotamers for each residue and compared them to the relative populations observed for the same residue in helices in the PDB<sup>13</sup> (Fig. 1; see also Supporting Information). This comparison shows clearly that the  $\chi_1$  distributions for four residues (Ile, Leu, Asp, and Asn) differ significantly from those found in the PDB. We find that this result is robust with respect to the similarity metric used to compare the distributions and the length of the simulations. On the basis of these observations, we selected these four residues for refinement of the  $\chi_1$  torsion parameters, as described below. We note here that subsequent comparisons between simulations of proteins and NMR measurements, as described further below, found the same four residues to deviate the most.

#### Fitting of torsion potentials to the QM-calculated energies

Quantum level *ab initio* calculations at the DF-LMP2 level of theory were used to calculate torsion energy pro-

files for the four amino acids (Fig. 2 and Supporting Information). As Asp and Asn display more complicated rotameric preferences for  $\chi_2$  than Ile and Leu, and



**Figure 2**

Torsion energy profiles for rotation around the  $\chi_1$  angle in isoleucine. Energy profiles were calculated for three different values of the  $\chi_2$  angle, namely (A)  $-60^\circ$ , (B)  $60^\circ$ , and (C)  $178^\circ$ . The dihedral angle shown here is defined as  $\text{N}-\text{C}^\alpha-\text{C}^\beta-\text{C}^\gamma$ . *Ab initio* energies calculated at the LMP2 level are reported in solid blue lines, whereas Amber ff99SB force field energies are reported in solid black lines. A modified torsion term for the  $\chi_1$  angle (see Table I) was introduced to maximize the agreement between the *ab initio* and the force-field energies. The resulting Amber ff99SB-ILDN energies are reported in solid red lines. The dashed lines show the differences between the QM and molecular mechanics energies (with ff99SB in black dashed lines and ff99SB-ILDN in red dashed lines).

because  $\chi_1$  and  $\chi_2$  torsions appeared to be more strongly coupled in Asp and Asn than in Ile and Leu, we calculated the full two-dimensional  $\chi_1/\chi_2$  energy profile for Asp and Asn and fitted the resulting QM data to new torsion profiles for both  $\chi_1$  and  $\chi_2$ . For Ile and Leu, we calculated QM torsion scans for  $\chi_1$  at three values of  $\chi_2$ . Although the underlying physical reason for the discrepancies between the QM and force-field energies is not clear, we decided to follow the approach used to modify the ff99SB backbone potential by modifying the torsion energy terms in the force field. The modification of bonded terms such as those for torsion angles will only directly influence a small number of atoms and thereby reduces the possibility of introducing unwanted side effects (when compared with, for example, the modification of nonbonded terms).

In previous studies, force-field torsion parameters have been optimized against quantum chemical calculations using a range of target functions.<sup>7,8,37</sup> The choice of target function may affect the resulting force field because it may implicitly weigh different regions of the energy surface differently. In our calculations, we found that it was not possible to obtain sub-kcal mol<sup>-1</sup> accuracy on a fit to the entire potential energy profile by simply introducing an additional torsion term. For this reason, the direct fit to the energy profile, while giving a good fit to the rotational barrier regions, produces unacceptable errors in the relative rotamer populations. On the other hand, a fit to the room temperature populations introduces substantial errors of up to several kcal mol<sup>-1</sup> in the barrier regions, as these have negligible Boltzmann factors at room temperature. As described in more detail in the Materials and Methods section, we have thus adopted an intermediate approach of least-square fitting to the Boltzmann population at 500 K. We found that this choice ensures, in practice, that the weights of the high-energy regions, although smaller, are not completely negligible. It also strikes a good balance between the need to obtain good equilibrium populations (residual errors in these regions are typically <0.5 kcal mol<sup>-1</sup>) and reasonable torsion barriers (errors in the barriers are typically between 0.5 and 2 kcal mol<sup>-1</sup>) (Fig. 2 and Supporting Information). The  $\chi_1$  torsion corrections can be introduced on several sets of atoms that would all, in the absence of fluctuations of bond lengths and angles, be related by rotational symmetries. As we, however, decided to follow the convention in the Amber force fields to fix the phase shift,  $\theta_0$ , to zero, this symmetry is broken in terms of the resulting torsion terms. For each  $\chi_1$  correction we considered, we therefore fitted to, one at a time, both the N-C<sup>α</sup>-C<sup>β</sup>-C<sup>γ</sup> and the C'-C<sup>α</sup>-C<sup>β</sup>-C<sup>γ</sup> dihedral angles and chose the one that gave rise to the best fit to the *ab initio* data. This turned out to be N-C<sup>α</sup>-C<sup>β</sup>-C<sup>γ</sup> for Ile, N-C<sup>α</sup>-C<sup>β</sup>-C<sup>γ</sup> for Asp, and C'-C<sup>α</sup>-C<sup>β</sup>-C<sup>γ</sup> for Leu and Asn. Adding corrections to more than one of the torsion angles that define  $\chi_1$ , or, equivalently, allowing

**Table I**

List of Modified Parameters for the  $\chi_1$  and  $\chi_2$  Torsion Potentials in Selected Amino Acids of the Amber ff99SB Force Field

Res.	Angle	$\theta_0$	$k_1$	$k_2$	$k_3$	$k_4$	$k_5$	$k_6$
Ile	N-C <sup>α</sup> -C <sup>β</sup> -C <sup>γ</sup> <sup>2</sup>	0.0	0.195	-0.846				
Leu	C-C <sup>α</sup> -C <sup>β</sup> -C <sup>γ</sup>	0.0	0.571	-0.358	0.135			
Asp	N-C <sup>α</sup> -C <sup>β</sup> -C <sup>γ</sup>	0.0	-2.635	-1.190	-0.007	0.423	0.232	-0.213
	C <sup>α</sup> -C <sup>β</sup> -C <sup>γ</sup> -O <sup>δ1,2</sup>	0.0	0.0	-0.443	0.0	-0.138	0.0	-0.013
Asn	C-C <sup>α</sup> -C <sup>β</sup> -C <sup>γ</sup>	0.0	0.571	-0.596	0.118	-0.417	0.104	-0.101
	C <sup>α</sup> -C <sup>β</sup> -C <sup>γ</sup> -N <sup>δ</sup>	0.0	-1.046	-0.181	-0.035	0.100	0.130	-0.106

The parameters are in kcal mol<sup>-1</sup> and correspond to the torsion potentials that are defined in the main text. Note that for the  $\chi_2$  torsion in Asp, the correction is applied to both side-chain oxygen atoms.

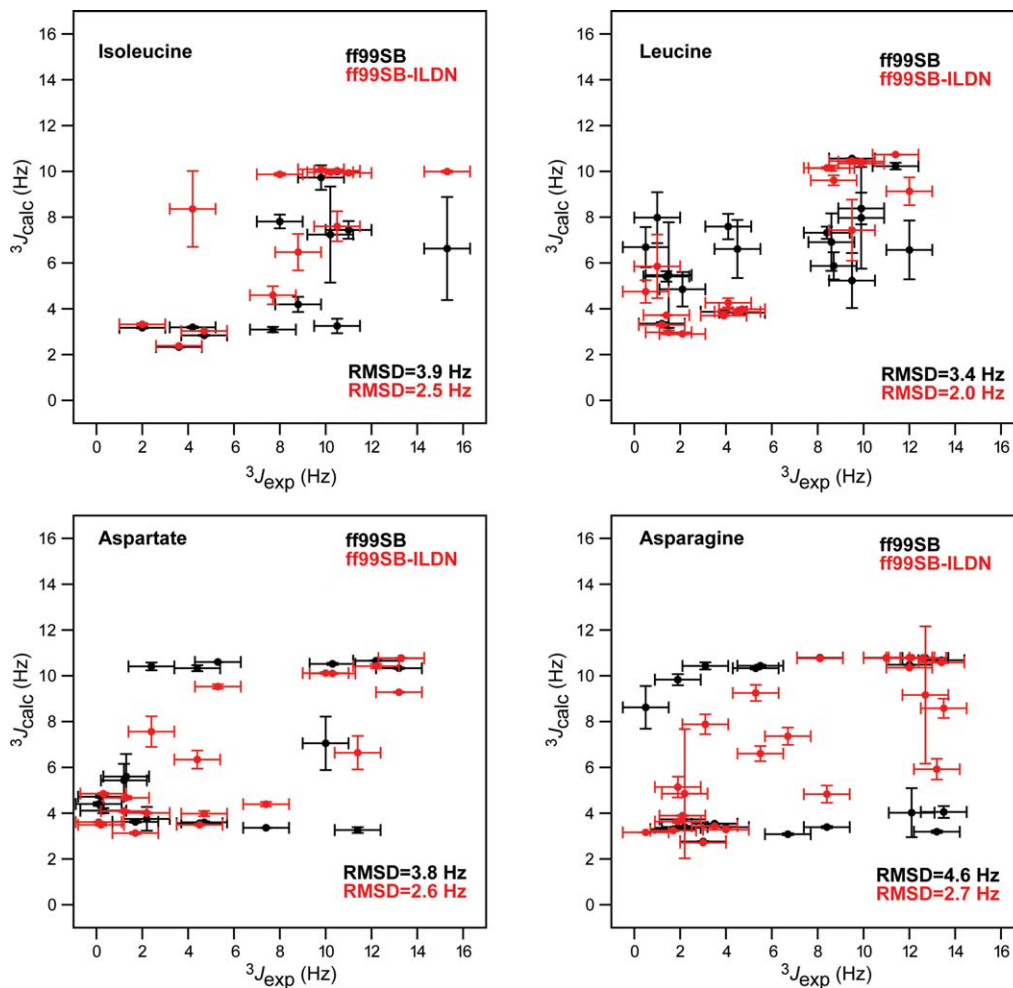
the phase to be nonzero, turned out in practice only to lead to a modest improvement in the quality of the fit, and thus the torsion term for only a single angle was modified. The corrections introduced with this procedure are reported in Table I and range from ~1 kcal mol<sup>-1</sup> for Leu up to ~5 kcal mol<sup>-1</sup> for Asp. We term the force field that results from replacing the original dihedral terms in ff99SB with these optimized parameters “ff99SB-ILDN” (ILDN being the one-letter code for the side chains whose potentials we modify).

### Rotamer distribution in the ff99SB-ILDN force field

As a first test, we repeated the simulations of the short helical peptides using the modified side-chain torsion potentials. For all four residues that we modified, we find that the ff99SB-ILDN force field improves the agreement with the PDB distribution (Fig. 1; see also Supporting Information). This result is encouraging as the information about the PDB distribution was not used in any way to modify the torsion parameters. We observe a substantial improvement for both Leu and Ile and a marginal one for Asp and Asn. The underlying reason for this difference is not clear. Although even a “perfect” force field would not necessarily reproduce exactly the PDB distribution, the deviations observed for Asp and Asn may be caused by errors in the parametrization of the nonbonded interactions. Such errors cannot be completely compensated for or corrected by introducing a modified torsion potential. As described in the following section, however, we observed substantial improvements for all four residue types when ff99SB and ff99SB-ILDN were evaluated using solution-state NMR measurements.

### Validation through comparison to NMR data

Matching the PDB rotamer distribution is not a direct control that can be used to evaluate the quality of a force field. A more important and stringent test is the ability of the force field to reproduce experimental quantities that directly probe the side-chain conformations of proteins in solution. A large amount of such experimental

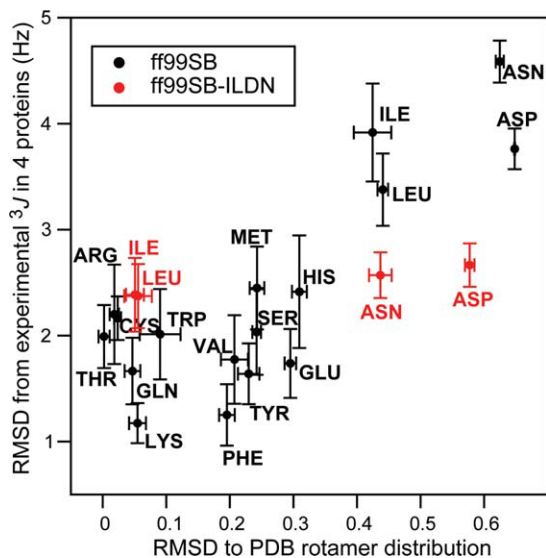


**Figure 3**

Comparison of experimental NMR  $^3J$  scalar couplings and corresponding values calculated from the MD simulations of HEWL, BPTI, Ubq, and GB3. The plots show  $\text{H}^{\alpha}\text{-C}^{\alpha}\text{-C}^{\beta}\text{-H}^{\beta 1,2}$  couplings that directly probe the side-chain  $\chi_1$  angles. Values before (black) and after (red) the side-chain torsion potential refinement are reported for the four residues (Ile, Leu, Asp, and Asn) whose side-chain potentials were modified. Each panel is labeled with the RMSD between the experimental scalar couplings, and the values calculated using the two force fields.

data is available in the form of NMR side-chain  $^3J$  scalar couplings and RDCs. We have performed MD simulations on the microsecond timescale for four globular proteins (BPTI, ubiquitin, GB3, and lysozyme) in which a large amount of high-quality NMR data probing the side-chain conformation is available. Relatively long simulations are required to achieve a strong convergence of the calculated NMR quantities, as rotameric changes occur on a broad range of timescales. Simulations were performed using both the standard ff99SB force field and the modified ff99SB-ILDN force field. For each of the four proteins, we found the native state to be very stable in the simulations with both ff99SB and ff99SB-ILDN as evidenced, for example, by low average backbone root-mean-square deviations (RMSD) from the experimentally determined structures ( $\leq 1$  Å in the simulations of BPTI, Ubq, and GB3 and  $< 2$  Å in the simulations of HEWL).

We calculated a large number of scalar couplings from these simulations and compared them to the experimental values; the results for Ile, Leu, Asp, and Asn are shown in Figure 3. It is clear that many outliers that are present in the simulations with ff99SB are not present in the simulations with ff99SB-ILDN. To quantify the agreement between experiment and simulation, we calculated the RMSD between the experimentally derived and simulation-derived scalar couplings on a per-residue-type basis. The results are shown in Figure 4, where these RMSD values are plotted against the results obtained from the analysis of the helical peptides described above. The results for ff99SB show clearly that the four residues that were selected for force-field refinement based on the comparison between the distribution of rotamers in the PDB and in MD simulations of helical peptides are also the residues that display the largest deviations between



**Figure 4**

Comparison of different metrics used to evaluate the quality of the side-chain description in force fields. The RMSD between the calculated rotamer distribution and the distribution observed in the PDB is plotted versus the RMSD between the calculated and experimental side-chain NMR  $^3J$  couplings ( $H^\alpha-C^\alpha-C^\beta-H^{\beta 1,2}$ ). The values calculated after refinement of the side-chain torsion potentials are reported in red.

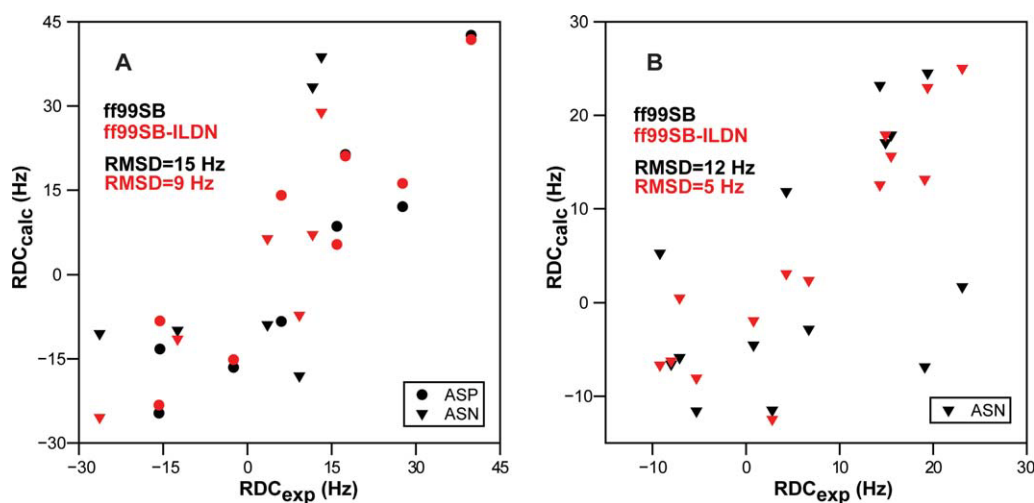
the calculated and experimental NMR data. This observation validates our approach of using the deviation from the PDB rotamer distribution in helices as a metric to

identify residues that require refinement of their side-chain torsion parameters. In agreement with the visual inspection of Figure 3, the results in Figure 4 show that the description of all four residues that were modified in ff99SB-ILDN improved significantly after the refinement. Notably, for Asp and Asn, the agreement with the NMR data improves much more than the agreement with the PDB rotamer distribution.

We also compared the simulations to experimentally measured RDCs that act as an alternative probe of the amino acid side-chain orientations. We calculated RDCs for  $C^\beta-H^{\beta 1,2}$  bonds in GB3 and compared them to the experimental values [Fig. 5(a)]. We observed improvements for both Asp and Asn residues, although these comparisons are complicated by the fact that stereospecific assignments are not available. For HEWL, we compared our simulations to the experimentally measured RDCs for the side-chain amide groups in Asn residues [Fig. 5(b)]. These values depend both on the  $\chi_1$  and  $\chi_2$  torsions and also show significant improvements in the ff99SB-ILDN force field.

## DISCUSSION

We propose a set of improved side-chain torsion parameters for the Amber ff99SB force field. The refinement was limited to the four residues (isoleucine, leucine, aspartate, and asparagine) that appear to be most problematic in ff99SB when comparing the rotamer distribution observed in MD simulation of short helices



**Figure 5**

Comparison of experimental residual dipolar couplings and values calculated from the MD simulations of (A) GB3 and (B) HEWL. In A, we show a comparison between experiment and simulation for  $C^\beta-H^{\beta 1,2}$  RDCs in GB3 for Asp (circles) and Asn (triangles) residues. In B, we show a comparison between experiment and simulation for  $N^\delta-H^{\delta 1,2}$  RDCs in HEWL for Asn (triangles) residues. The experimental values were reported as the sum of the couplings to the two side-chain amide protons, and so we calculated the same sum from the MD simulations. In both A and B, the results are shown for both simulations with ff99SB (black symbols) and with ff99SB-ILDN (red symbols). Each panel is labeled with the RMSD between the experimental couplings, and the values calculated using the two force fields.

with the rotamer distribution taken from helices in the PDB. The new parameters were obtained by fitting to new QM data and were validated against a large set of NMR data. The consistent improvements observed for all four of the residues that we modified suggest that our approach is robust and general. Nevertheless, we decided against modifying additional residues, as it would become increasingly difficult to demonstrate significant improvements for those residues. The corrections introduced here for Ile, Leu, Asp, and Asn range between 1 and 5 kcal mol<sup>-1</sup> and can thus have a noticeable impact on the stability of protein folds and flexible loops, particularly in long MD simulations that exceed the timescales of the rotations of buried or partially buried side chains. Because the new torsion potentials described here represent a clear improvement of those in the existing force field and do not appear to cause undesirable side effects, we recommend the usage of ff99SB-ILDN over ff99SB in MD simulations of proteins.

## ACKNOWLEDGMENTS

The authors thank all the experimentalists whose data they have used and without whom this study would not have been possible. In particular, they thank Ad Bax and John L. Marquardt for sharing unpublished data on ubiquitin and Christina Redfield and Lorna Smith for sharing data on lysozyme. They also thank Michael Eastwood for a careful reading of the manuscript and Rebecca Kastleman for editorial assistance.

## REFERENCES

- Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE. Long-timescale molecular dynamics simulations of protein structure and function. *Curr Opin Struct Biol* 2009;19:120–127.
- Freddolino PL, Liu F, Gruebele M, Schulten K. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys J* 2008;94:L75–L77.
- Shaw DE, Dror RO, Salmon JK, Grossman JP, Mackenzie KM, Bank JA, Young C, Deneroff MM, Batson B, Bowers KJ, Chow E, Eastwood MP, Ierardi DJ, Klepeis JL, Kuskin JS, Larson RH, Lindorff-Larsen K, Maragakis P, Moraes MA, Piana S, Shan Y, Towles B. Millisecond-scale molecular dynamics simulations on Anton. In: *Proceedings of the 2009 ACM/IEEE Conference on Supercomputing (SC09)*. Washington, DC: ACM Press; 2009.
- Liwo A, Czaplowski C, Oldziej S, Scheraga HA. Computational techniques for efficient conformational sampling of proteins. *Curr Opin Struct Biol* 2008;18:134–139.
- Pérez A, Marchán I, Svozil D, Spomer J, Cheatham TE, III, Laughton CA, Orozco M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J* 2007;92:3817–3829.
- Friesner RA. Modeling polarization in proteins and protein-ligand complexes: methods and preliminary results. *Adv Protein Chem* 2006;72:79–104.
- Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 2006;65:712–725.
- Mackerell AD, Jr, Feig M, Brooks CL, III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 2004;25:1400–1415.
- Best RB, Buchete NV, Hummer G. Are current molecular dynamics force fields too helical? *Biophys J* 2008;95:L07–L09.
- Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc* 1984;106:765–784.
- Weiner SJ, Kollman PA, Nguyen DT, Case DA. An all atom force field for simulations of proteins and nucleic acids. *J Comput Chem* 1986;7:230–252.
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 1995;117:5179–5197.
- Lovell SC, Word JM, Richardson JS, Richardson DC. The penultimate rotamer library. *Proteins* 2000;40:389–408.
- Werner H-J, Manby FR, Knowles PJ. Fast linear scaling second-order Moller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J Chem Phys* 2003;118:8149–8160.
- Werner HJ, Knowles PJ, Lindh R, Manby FR, Schuetz M, Celani P, Korona T, Rauhut G, Amos RD, Bernhardsson A, Berning A, Cooper DL, Deegan MJO, Dobbyn AJ, Eckert F, Hampel C, Hetzer G, Llyod AW, McNicholas SJ, Meyer W, Mura ME, Nicklass A, Palmieri P, Pitzer R, Schumann U, Stoll H, Stone AJ, Tarroni R, Thorsteinsson T. MOLPRO, version 2006.1, a package of *ab initio* programs. Available at: <http://www.molpro.net>.
- Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolosváry I, Moraes MA, Sacerdoti FD, Salmon JK, Shan Y, Shaw DE. Scalable algorithms for molecular dynamics simulations on commodity clusters. In: *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing (SC06)*. Washington, DC: ACM Press; 2006.
- Lippert RA, Bowers KJ, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolosváry I, Shaw DE. A common, avoidable source of error in molecular dynamics integrators. *J Chem Phys* 2007;126:046101.
- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys* 1995;103:8577.
- Tuckerman M, Berne BJ, Martyna GJ. Reversible multiple time scale molecular dynamics. *J Chem Phys* 1992;97:1990.
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–935.
- Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* 2004;120:9665–9678.
- Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, Westbrook J. The Protein Data Bank and the challenge of structural genomics. *Nat Struct Biol* 2000;7:957–959.
- Smith LJ, Sutcliffe MJ, Redfield C, Dobson CM. Analysis of phi and chi 1 torsion angles for hen lysozyme in solution from 1H NMR spin-spin coupling constants. *Biochemistry* 1991;30:986–996.
- Grimshaw SB. Novel approaches to characterizing native and denatured proteins by NMR, Doctoral Thesis, University of Oxford, 1999.
- Schwalbe H, Grimshaw SB, Spencer A, Buck M, Boyd J, Dobson CM, Redfield C, Smith LJ. A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Sci* 2001;10:677–688.
- Miclet E, Boisbouvier J, Bax A. Measurement of eight scalar and dipolar couplings for methine-methylene pairs in proteins and nucleic acids. *J Biomolecular NMR* 2005;31:201–216.
- Berndt KD, Güntert P, Orbons LP, Wüthrich K. Determination of a high-quality nuclear magnetic resonance solution structure of the



- bovine pancreatic trypsin inhibitor and comparison with three crystal structures. *J Mol Biol* 1992;227:757–775.
28. Karplus M. Contact electron-spin coupling of nuclear magnetic moments. *J Chem Phys* 1959;30:11–15.
  29. Hu J-S, Bax A. Determination of  $\phi$  and  $\chi_1$  angles in proteins from  $^{13}\text{C}$ - $^{13}\text{C}$  three-bond J couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond? *J Am Chem Soc* 1997;119:6360–6368.
  30. Chou JJ, Case DA, Bax A. Insights into the mobility of methyl-bearing side chains in proteins from  $^3J_{\text{CC}}$  and  $^3J_{\text{CN}}$  couplings. *J Am Chem Soc* 2003;125:8959–8966.
  31. Pérez C, Lohr F, Ruterjans H, Schmidt JM. Self-consistent Karplus parametrization of  $^3J$  couplings depending on the polypeptide side-chain torsion  $\chi_1$ . *J Am Chem Soc* 2001;123:7081–7093.
  32. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M. Simultaneous determination of protein structure and dynamics. *Nature* 2005;433:128–132.
  33. Higman VA, Boyd J, Smith LJ, Redfield C. Asparagine and glutamine side-chain conformation in solution and crystal: A comparison for hen egg-white lysozyme using residual dipolar couplings. *J Biomol NMR* 2004;30:327–346.
  34. Butterfoss GL, Hermans J. Boltzmann-type distribution of side-chain conformation in proteins. *Protein Sci* 2003;12:2719–2731.
  35. Morozov AV, Kortemme T, Tsemekhman K, Baker D. Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc Natl Acad Sci USA* 2004;101:6946–6951.
  36. Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M. Relation between native ensembles and experimental structures of proteins. *Proc Natl Acad Sci USA* 2006;103:10901–10906.
  37. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 2001;105:6474–6487.