THE JOURNAL OF PHYSICAL CHEMISTRY B

Feature Article

pubs.acs.org/JPCB

# Residue-Specific Force Field Based on the Protein Coil Library. RSFF1: Modification of OPLS-AA/L

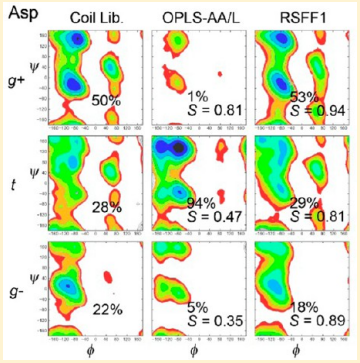
Fan Jiang,*[,†] Chen-Yang Zhou,[‡] and Yun-Dong Wu*[,†,‡]

[†]Laboratory of Computational Chemistry and Drug Design, Laboratory of Chemical Genomics, Peking University Shenzhen Graduate School, Shenzhen 518055, China

[‡]College of Chemistry, Peking University, Beijing 100871, China

Supporting Information

**ABSTRACT:** Traditional protein force fields use one set of parameters for most of the 20 amino acids (AAs), allowing transferability of the parameters. However, a significant shortcoming is the difficulty to fit the Ramachandran plots of all AA residues simultaneously, affecting the accuracy of the force field. In this Feature Article, we report a new strategy for protein force field parametrization. Backbone and side-chain conformational distributions of all 20 AA residues obtained from protein coil library were used as the target data. The dihedral angle (torsion) potentials and some local nonbonded (1-4/1-5/1-6) interactions in OPLS-AA/L force field were modified such that the target data can be excellently reproduced by molecular dynamics simulations of dipeptides (blocked AAs) in explicit water, resulting in a new force field with AA-specific parameters, RSFF1. An efficient free energy decomposition approach was developed to separate the corrections on $\phi$ and $\psi$ from the two-dimensional Ramachandran plots. RSFF1 is shown to reproduce the experimental NMR $^3J$-coupling constants of AA dipeptides better than other force fields. It has a good balance between $\alpha$-helical and $\beta$-sheet secondary structures. It can successfully fold a set of $\alpha$-helix proteins (Trp-cage and Homeodomain) and $\beta$-hairpins (Trpzip-2, GB1 hairpin), which cannot be consistently stabilized by other state-of-the-art force fields. Interestingly, the RSFF1 force field systematically overestimates the melting temperature (and the stability of native state) of these peptides/proteins. It has a potential application in the simulation of protein folding and protein structure refinement.




## 1. INTRODUCTION

In the last several decades, tremendous efforts have devoted to the bottom-up modeling of complex biomolecular systems, especially the atomistic molecular dynamics (MD) simulations.[1,2] Recently, using a powerful special-purpose computer, Shaw's group demonstrated that *ab initio* folding of a series of small peptides/proteins can be achieved, providing atomistic-level details of structures and dynamics.[3] Besides theoretical understandings of biologically relevant processes, biomolecular simulations have increasing applications such as structural refinement[4] and drug discovery.[5] However, their reliability and predictive power crucially depend on the accuracy of the force fields used to describe the interactions among atoms.[6]

Although there are important issues associated with the force field development, such as the solvent effect[7−9] and the electronic polarizability,[10−13] many recent efforts in improving classical protein force fields (such as AMBER,[14] CHARMM,[15] and OPLS-AA[16]) have been focused on the accurate description of backbone ($\phi$, $\psi$) and side-chain ($\chi$) conformational preferences (Scheme 1), owing to their essential roles in determining peptide and protein conformations. Figure 1 gives a brief summary. Early efforts included the fitting to gas-phase quantum mechanics (QM) $\phi$, $\psi$ energy surface of dipeptides (such as Ac-Ala-NHMe) at local-MP2 level (OPLS-AA/L,[17] CHARMM27[18,19]) or the fitting to dipeptide QM energy surface calculated with continuum solvent model (AMBER ff03[20]). Later, gas-phase QM conformational energies of tetrapeptides (such as Ac-$Ala_3$-NHMe) were used to fit the $\phi$, $\psi$ parameters (AMBER ff99SB).[21] Recently, side-chain $\chi$ potentials of Ile, Leu, Asp, and Asn in ff99SB have been improved by fitting to gas-phase QM (local-MP2 level) energies of dipeptides (ff99SB-ildn).[22]

Despite these efforts, previous peptide simulations indicated biased secondary structure preferences from various force fields,[23−31] which can result in failure of protein folding simulations. For example, the CHARMM27 cannot fold the all-$\beta$ protein WW domain due to overstabilization of $\alpha$-helical structures.[28] Some more recent efforts (AMBER ff99SB* and ff03* by Best et al.,[32] and CHARMM22* from Shaw's group[33]) aim to correct this problem by means of a minor adjustment to the backbone potential to reproduce the experimental $J$-couplings of $Ala_5$ and the $\alpha$-helical content of a poly-Ala-based peptide measured from NMR. These corrections can result in more balanced $\alpha$-helix and $\beta$-sheet preferences.[34] Very recently, Best et al. empirically optimized the backbone CMAP correction parameters for CHARMM force field on the basis

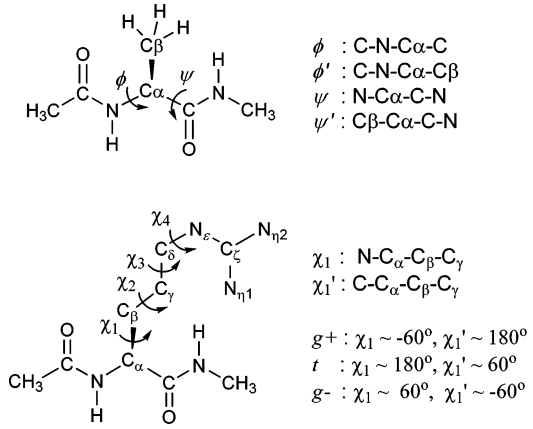
**Received:** February 18, 2014
**Revised:** May 2, 2014
**Published:** May 12, 2014





dx.doi.org/10.1021/jp5017449 | J. Phys. Chem. B 2014, 118, 6983−6998

**Scheme 1. Definitions of Backbone $\phi$, $\psi$ and Side-Chain $\chi_i$ Dihedral Angles**[a]




[a]Dipeptide (terminally blocked amino acid) models of alanine (top) and arginine (bottom) are used, with some hydrogen atoms omitted for clarity. The definition of three side-chain rotamers (g+/t/g-) is also given.

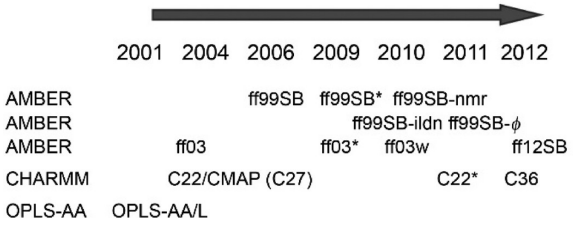



**Figure 1.** Recent developments of all-atom protein force fields. AMBER has more force field variants than CHARMM and OPLS-AA. "-ildn" is a modification for side-chain, whereas "*", "-NMR", and "-$\phi$" are modifications for backbone. Therefore, ff99SB*-ildn, ff99SB-ildn-NMR, and ff99SB-ildn-$\phi$ are also new variants of AMBER force fields.

of a similar strategy, together with new side-chain parameters from high-level QM calculations (CHARMM36).[35] Besides, a small modification of $\phi'$ potential in AMBER ff99SB was proposed to better reproduce *J*-couplings of short peptides (ff99SB-$\phi$).[36] There were also recent attempts to optimize backbone parameters on the basis of NMR chemical shifts on a few proteins (ff99SB-NMR).[37]

In these recent protein force fields, a single set of backbone parameters was used for all non-Gly/Pro amino acids (AAs), which were usually optimized on the Ala residue. However, recent evidence has indicated that this basic assumption can cause problems. For example, AMBER ff99SB force field gives significantly lower $\alpha$-helix content of poly-Ala-based peptides near 300 K compared with experimental results,[32] whereas it also underestimates the stability of $\beta$-hairpin peptide Trpzip-2.[38] Meanwhile, it actually well reproduces the experimental melting temperature ($T_m$) of Trp-cage mini-protein.[39] In the recent large-scale *ab initio* folding of 12 small proteins, another state-of-the-art force field CHARMM22* was found not to be able to stabilize the native state of Engrailed homeodomain.[3] Besides, it gives a quite low (260 K) $T_m$ for a thermal stable variant of Trp-cage. Very recently, it was found that current force fields have problems in describing the discrepancies in the local conformational preferences of different AA residues.[40] Therefore, there is still room for the improvement of these popular protein force fields.

For the development of an accurate protein force field, it is critical to obtain the local conformational free energy surface (Ramachandran plot) of each AA residue without secondary structure constraint. Recently, we carried out statistical analysis of conformational distributions of 20 AA residues from high-resolution protein crystal structures.[41,42] Inspired by earlier work of Swindells et al.,[43] we used the protein "coil library"[44−50] containing only the residues not in any secondary structure (i.e., without forming backbone hydrogen bond). The obtained conformational distributions of various residues are quite different from those obtained from the analysis of whole protein structures. For example, alanine significantly prefers $\alpha$-helical conformation in protein structures. However, in the coil library the polyproline-II ($P_{II}$) conformation is dominant for alanine, which agrees well with recent spectroscopic measurements of short peptides in aqueous solution.[51−56] In addition, conformational distributions of various residues from seven coil libraries with different restrictions are quite similar.[42] Thus, the coil library of the PDB allows us to obtain intrinsic conformational features of the 20 AA residues.

Another important feature is the side-chain $\chi_1$ rotamer distributions of residues other than Gly, Ala, and Pro. As shown in Scheme 1, each residue has three rotamers, g+, g−, and t, meaning the $C_\beta$−$C_\gamma$ bond is gauch+, gauch−, and trans with respect to the $C_\alpha$−N bond. Due to the interactions of the side chain with the backbone, these three rotamers may cause very different $\phi$, $\psi$ distributions (Ramachandran plots).[57,58] The statistical $\chi_1$ rotamer distributions from the protein coil library can be significantly different from those obtained from the whole protein structures, because in the latter case most residues are constrained in secondary structures.[41] We also found that current popular all-atom force fields do not reproduce the coil library rotamer distributions well.[41,42] On the other hand, our QM calculations of model molecules with the solvent effect of water reproduce the side-chain rotamer preferences and rotamer-dependent Ramachandran plots from coil library quite well.[42]

A second crucial issue is how to incorporate the conformational distributions of 20 AA residues from the coil library into a force field. The coil library statistics are related to free energy surfaces with solvent effect. It should be compared with similar free energies from MD simulation of solvated systems, such as dipeptides in water. However, free energy calculations require equilibrium MD simulations. Too many such calculations are impractical for optimizing parameters if a traditional trial-and-error approach is adapted. Furthermore, if different $\chi_1$ side-chain rotamers are considered separately, totally 55 coil-library Ramachandran plots (one for Gly/Ala, two for Pro, and three for each of the other 17 AAs) need to be fitted. Because these $\phi$, $\psi$ plots are fairly complicated and quite different, it is very difficult to develop one set of parameters to fit all of them simultaneously. A highly efficient parametrization strategy should be developed to avoid tremendous computational costs and human efforts.

In this paper, we present a new approach for the parametrization of a protein force field. Namely, we develop torsional parameters for each residue independently to fit the statistical conformational distributions derived from the protein coil library. Thus, the torsional parameters for backbone and side chain are residue specific. Although this approach can be applied to improve any available protein force field, we here use OPLS-AA/L[17] as an example. We show that the parametrization is quite simple, and the new force field, named


dx.doi.org/10.1021/jp5017449 | J. Phys. Chem. B 2014, 118, 6983−6998

RSFF1, gives much improved simulation results such as the $^3J$-coupling constants of dipeptides in solution, balance between $\alpha$ and $\beta$ secondary structures, as well as the reliability in folding of peptide and protein structures.

## 2. METHODS

**2.1. General Strategies.** The general energy expression (eq 1) of the RSFF1 force field is within the framework of a classical force field:

$$V_{total} = V_{bond} + V_{angle} + V_{torsion} + V_{local\text{-}LJ} + V_{LJ} + V_{Coulomb} \quad (1)$$

All bond stretching ($V_{bond}$) and angle bending ($V_{angle}$) potentials were adopted from the OPLS-AA/L without modification. The atomic $\sigma$ and $\varepsilon$ parameters of Lennard-Jones potential ($V_{LJ}$) and atomic charges of Coulomb potential ($V_{Coulomb}$) were also fully adopted from OPLS-AA/L. On the other hand, torsional potentials ($V_{torsion}$) for each AA residue were developed independently according to coil-library data. Besides, local Lennard-Jones potentials ($V_{local\text{-}LJ}$) between atoms separated by three covalent bonds (1-4 interactions) were treated differently from ordinary $V_{LJ}$, and some 1-5 and 1-6 interactions were also treated specially (included in the $V_{local\text{-}LJ}$). These are described in more detail below:

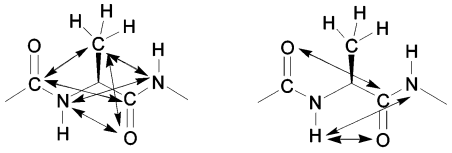
*Torsional Parameters.* Similar to potentials of most physical-based force fields, the torsion potentials in RSFF1 are Fourier expansions:

$$V(\theta) = \sum k_n \cos(n\theta) = \sum c_n \cos^n(\theta) \quad (2)$$

where $\theta$ is a given dihedral angle and the coefficients $k_n$ or $c_n$ are parameters. There are $3 \times 3 = 9$ coupled torsion terms for each $sp^3-sp^3$ bond rotation, and $3 \times 2 = 6$ terms for $sp^3-sp^2$ rotation. To reduce the number of adjustable parameters, the torsion terms involving hydrogen atoms were kept the same as in OPLS-AA/L, such as H−N−$C_\alpha$−C for $\phi$ and N−$C_\alpha$−$C_\beta$−$H_\beta$ for $\chi_1$. Because there are enough data from the coil library, AA-dependent torsion parameters were used for RSFF1. The torsion potential of peptide bond $V(\omega)$ and improper torsions were kept the same as OPLS-AA/L, because their purpose is to maintain planar geometry of conjugated systems.

*Local Lennard-Jones Parameters.* Besides the torsion terms, those 1-4 interactions shown in Scheme 2 also affect the bond rotation. Some of them were modified in RSFF1, because we found that the original values might give too strong effects. Furthermore, to optimize coupling between neighboring torsions, instead of using dihedral angle cross-terms or 2D grid-based CMAP-like corrections,[18] we choose to directly modify the related 1-5 or 1-6 $V_{local\text{-}LJ}$ parameters ($\varepsilon$, $\sigma$) whenever necessary. In most cases, we set $\varepsilon$ = 0.1 kJ/mol and only manually adjusted $\sigma$ for different strength of repulsion. Same $\varepsilon$ and $\sigma$ parameters are used for different AAs whenever possible, to reduce the efforts in manually adjusting them.

**Scheme 2. Modified Local Lennard-Jones Interactions ($V_{local\text{-}LJ}$) for Backbone $\phi$, $\psi$ Torsions in RSFF1: Left, 1-4 Interactions; Right, 1-5 Interactions**



$$V_{L\text{-}J}(r) = 4\varepsilon\left[(\sigma/r)^{12} - (\sigma/r)^{6}\right]$$

**2.2. Parameterization Flow.** All parametrizations were based on replica exchange molecular dynamics (REMD)[59] simulations of various AA dipeptides (Ac-X-NHMe) in water. Generally, initial parameters were assigned for a given AA residue and they were then updated by repeating the procedure shown in Figure 2, until the simulated results are difficult to be further improved. One cycle of parameter optimization can be regarded as two successive transformations.

*From Force Field Parameters to Probability Distributions.* After REMD simulation using current parameters, the obtained trajectory is analyzed to obtain various statistical distributions (free energy surfaces), including the Ramachandran plot $p(\phi,\psi)$, the $\chi_1$-rotamer-dependent backbone conformational preference, percentages of three $\chi_1$-rotamers, and potentials of mean force (PMFs) for $\chi$ torsions. They are then compared with corresponding data from coil library statistics. The details are described in sections 2.3 and 2.4.

*From Probability Distributions to Updated Parameters.* As shown in Figure 2, various parameters were updated in parallel. In essence, the torsion potential $V(\theta)$ is updated according to the difference between the coil library PMF and the simulated PMF:

$$\begin{aligned} \Delta G(\theta) &= G_{coil}(\theta) - G_{MD}(\theta) \\ &\rightarrow \Delta V(\theta) \\ &= V_{new}(\theta) - V_{old}(\theta) \end{aligned} \quad (3)$$

This strategy is similar to the iterative Boltzmann inversion (IBI) method.[60] The required changes of the related Fourier coefficients can be obtained by fitting $\Delta V(\theta)$ to discrete $\Delta G(\theta)$ values. Before applying eq 3, a decomposition method is applied to derive corrections on $\phi$ and $\psi$ potentials from the 2D $\phi$, $\psi$ distributions. The details are described in section 2.5 for backbone torsions and section 2.6 for side-chain torsions. Besides, local L-J parameters ($\varepsilon$, $\sigma$) were adjusted manually only when necessary.

**2.3. Molecular Dynamics Simulations.** When OPLS-AA/L or our new force field was used, each dipeptide molecule was solvated with 319−330 TIP4P/Ew[61] water molecules. For the simulations using AMBER and CHARMM force fields, similar numbers of TIP3P water molecules were used. The ionic Arg, Lys, Asp, and Glu side chains were neutralized with counterion ($Cl^-$ or $Na^+$). REMD simulations were performed using *Gromacs* version 4.5.4, with 12 replicas ranging from 298 to 451 K. The temperatures of intermediate replicas were calculated according to a recent study[62] to give uniform exchange rates of ~16%. Swaps between neighboring replicas were attempted every 0.45 ps. The velocity rescaling thermostat[63] with $\tau_T$ = 0.2 ps was used to maintain the *NVT* ensemble. The periodic box size was obtained from averaging last 1 ns of a 3 ns *NPT* preproduction run at 300 K and 1 atom. Electrostatics were treated using the particle-mesh Ewald (PME) method with a real-space cutoff of 0.9 nm. van der Waals interactions were cut off at 0.9 nm with the long-range dispersion correction for energy and pressure.

In all simulations, the mass of water oxygen atom was reduced from 16 to 2 amu to increase the sampling efficiency[64] without altering the thermodynamics equilibrium properties. All bond lengths involving hydrogen atoms were constrained by


dx.doi.org/10.1021/jp5017449 | J. Phys. Chem. B 2014, 118, 6983−6998

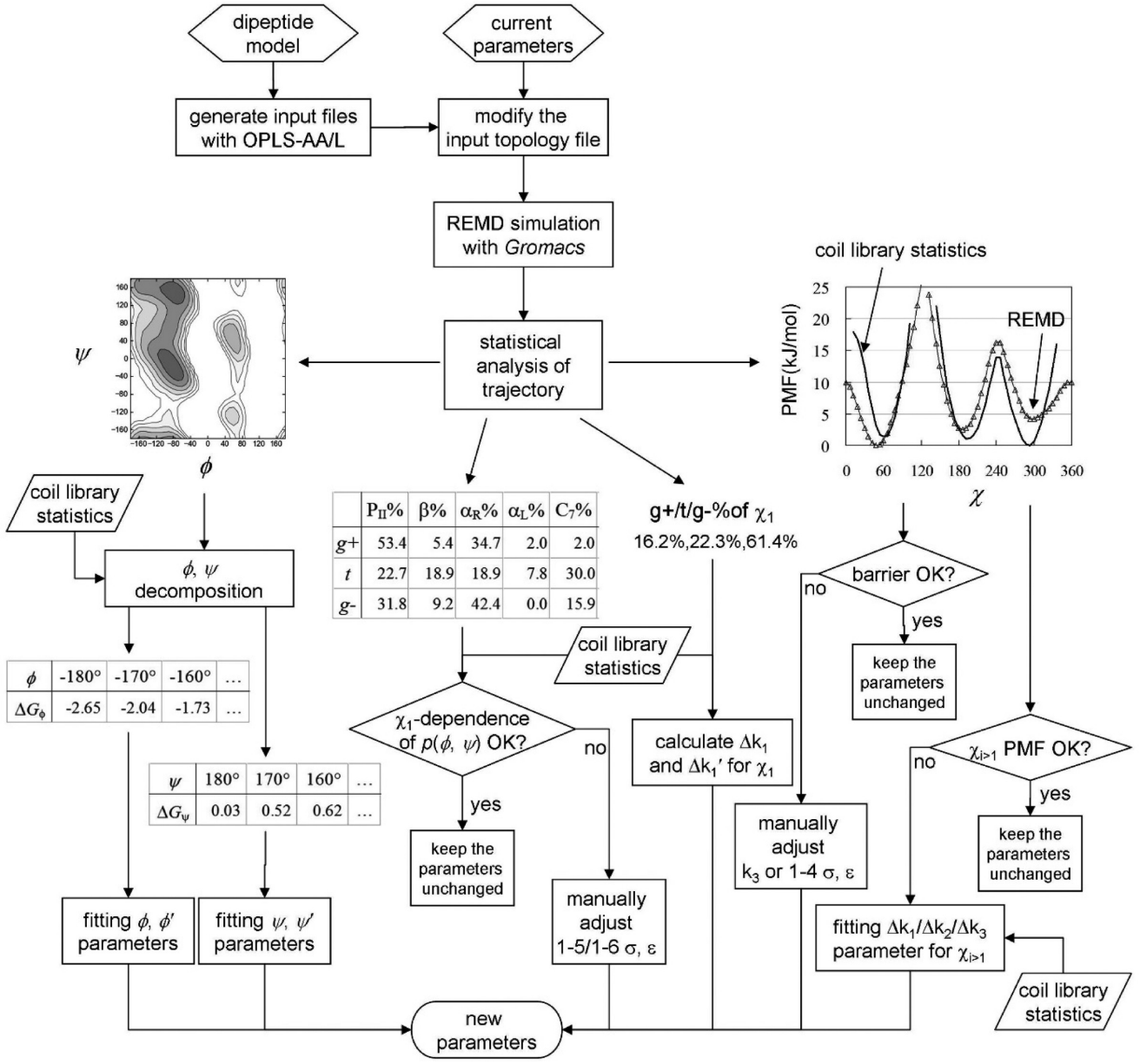



**Figure 2.** Flowchart of the overall procedure for one cycle of the parameter optimization. Details are explained in the text.

the LINCS,[65] allowing a time step of 3 fs. The REMD simulation of each dipeptide was carried out for ~90 ns per replica, and the structures were recorded every 0.6 ps. Trajectories from 298 K replica were used for statistical analysis, with the first 20 ns discarded. One such REMD simulation required ~24 h in real time on a 12-core 2.4 GHz Intel Xeon node.

**2.4. Protein Coil Library and Statistical Analysis.** Briefly, 6178 protein crystal structures with resolution <2.0 Å and *R* factor <0.2 were retrieved from the Protein Data Bank (PDB)[66] database with 50% sequence identity cutoff. The popular DSSP[48] program was used to assign secondary structures. Residues within any secondary structures—including the DSSP codes G($3_{10}$-helix), H($\alpha$-helix), I($\pi$-helix), B($\beta$-bridge), E($\beta$-sheet), and T(turn)—were all excluded from the coil library. Residues preceding proline or containing backbone atoms with *B* factor >35 were also excluded. Following our previous work,[41] residues with short polar side-chains (Asp, Asn, Ser, Thr) preceding any residue with $-60° < \psi < +60°$ were excluded, to avoid inter-residue H-bonding between their side-chain O atom and the successive backbone amide H atom.

Same as the previous work of Amir et al.[67] and ours, a 2D Gaussian kernel estimator was used to extract $\phi$, $\psi$ distributions, considering the periodicity of the dihedral angles:

$$n(\phi,\psi) = \sum_i w_i \exp[-\min(|\phi_i - \phi|, 360° - |\phi_i - \phi|)^2 + \min(|\psi_i - \psi|, 360° - |\psi_i - \psi|)^2/2\sigma^2] \quad (4)$$

Here $i$ counts for all residues of given type in the coil library, and $w_i$ is the $1/m$ weighting for $m$ identical chains in one PDB structure. $\phi_i$ and $\psi_i$ are the observed backbone dihedral angles for residue $i$. $10° \times 10°$ grids of $(\phi, \psi)$ and $\sigma = 10°$ were used. For AAs other than Ala and Gly, the statistics were carried out for three side-chain $\chi_1$ rotamers separately. The obtained distributions were shown in the Supporting Information Figure


dx.doi.org/10.1021/jp5017449 | J. Phys. Chem. B 2014, 118, 6983−6998

S1. The same approach was used to analyze the trajectories from REMD simulations with $w_i = 1$ for each structural frame. For side-chain $\chi$ distributions, a 1D Gaussian estimator was used with grid space of 6° and $\sigma = 7°$.

The similarity coefficient $S$ between $(\phi, \psi)$ distribution from coil library $n_{\mathrm{Coil}}(\phi, \psi)$ and that from simulation $n_{\mathrm{MD}}(\phi, \psi)$ can be calculated without normalization:

$$S = \frac{\sum n_{\mathrm{coil}}(\phi,\psi)\cdot n_{\mathrm{MD}}(\phi,\psi)}{\sqrt{\sum n_{\mathrm{Coil}}(\phi,\psi)^2}\cdot\sqrt{\sum n_{\mathrm{MD}}(\phi,\psi)^2}} \tag{5}$$

Only two identical distributions will give $S = 1$.

The probability distributions were obtained by normalizing the statistical counts $n(\phi,\psi)$:

$$p(\phi,\psi) = \frac{n(\phi,\psi) + \varepsilon}{\sum n(\phi,\psi)} \tag{6}$$

To avoid infinity free energy when $n = 0$, pseudocount $\varepsilon = 0.02$ is used with negligible changes in the allowed Ramachandran regions.

**2.5. Optimization of Backbone Dihedral Potentials.** The difference between 2D $\phi$, $\psi$ free energy surfaces from the coil library and REMD simulations is separated into the corrections for 1D $\phi$ component and $\psi$ component:

$$\Delta G_\phi(\phi) + \Delta G_\psi(\psi) = G_{\mathrm{Coil}}(\phi,\psi) - G_{\mathrm{MD}}(\phi,\psi) \tag{7}$$

To solve this equation, we can rewrite eq 7 into eq 8 by using hypothetical probability distributions for both $\Delta G_\phi$ and $\Delta G_\psi$:

$${}^{\delta}p_\phi(\phi)\cdot{}^{\delta}p_\psi(\psi) = p_{\mathrm{coil}}(\phi,\psi)/p_{\mathrm{MD}}(\phi,\psi) \tag{8}$$

Then ${}^{\delta}p_\phi$ and ${}^{\delta}p_\psi$ can be solved from the known $p_{\mathrm{Coil}}(\phi,\psi)$ and $p_{\mathrm{MD}}(\phi,\psi)$ by applying following two equations iteratively:

$${}^{\delta}p_\phi(\phi) = \frac{\sum_\psi p_{\mathrm{coil}}(\phi,\psi)}{\sum_\psi p_{\mathrm{MD}}(\phi,\psi)\cdot{}^{\delta}p_\psi(\psi)} \tag{9a}$$

$${}^{\delta}p_\psi(\psi) = \frac{\sum_\phi p_{\mathrm{coil}}(\phi,\psi)}{\sum_\phi p_{\mathrm{MD}}(\phi,\psi)\cdot{}^{\delta}p_\phi(\phi)} \tag{9b}$$

Uniform distribution of ${}^{\delta}p_\psi \equiv 1$ was used as the initial guess, and the convergence can always be achieved within 10 iterations. Then the ${}^{\delta}p_\phi$ and ${}^{\delta}p_\psi$ are converted to free energy scale separately:

$$\Delta G_\phi(\phi) = -RT\ln{}^{\delta}p_\phi(\phi) \tag{10a}$$

$$\Delta G_\psi(\psi) = -RT\ln{}^{\delta}p_\psi(\psi) \tag{10b}$$

The obtained $\Delta G_\phi$ and $\Delta G_\psi$ are discrete functions with 10° interval. They are fitted to analytical dihedral potentials in the force field:

$$\Delta V_\phi(\phi) + \Delta V_{\phi'}(\phi') = \sum_{n=0}^{5}\Delta c_{\phi,n}\cos^n(\phi) + \sum_{n=0}^{5}\Delta c_{\phi',n}\cos^n(\phi') \tag{11a}$$

$$\Delta V_\psi(\psi) + \Delta V_{\psi'}(\psi') = \sum_{n=0}^{5}\Delta c_{\psi,n}\cos^n(\psi) + \sum_{n=0}^{5}\Delta c_{\psi',n}\cos^n(\psi') \tag{11b}$$

where $\Delta c_{\phi,n}$, $\Delta c_{\phi',n}$, $\Delta c_{\psi,n}$, and $\Delta c_{\psi',n}$ are the changes of Fourier coefficients related to $V_\phi$, $V_{\phi'}$, $V_\psi$ and , $V_{\psi'}$ terms, respectively. Of course, there are no $V_{\phi'}$ and $V_{\psi'}$ terms for Gly. The zeroth-order ($n = 0$) terms are constant and do not affect the force field, but they are necessary as an offset to minimize the difference with target $\Delta G_\phi$ or $\Delta G_\psi$. Assuming the relationships $\phi' = \phi - 120°$ and $\psi' = \psi + 120°$, the parameters were fitted by minimizing the following penalty functions:

$$s_\phi = \sum w(\phi)[\Delta V_\phi(\phi) + \Delta V_{\phi'}(\phi - 120°) - \Delta G_\phi(\phi)]^2 \tag{12a}$$

$$s_\psi = \sum w(\psi)[\Delta V_\psi(\psi) + \Delta V_{\psi'}(\psi + 120°) - \Delta G_\psi(\psi)]^2 \tag{12b}$$

The $\phi$ and $\psi$ values with higher occurrences have higher weight $w$ in the fitting:

$$w(\phi) = \sqrt{\sum_\psi p_{\mathrm{coil}}(\phi,\psi)} \tag{13a}$$

$$w(\psi) = \sqrt{\sum_\phi p_{\mathrm{coil}}(\phi,\psi)} \tag{13b}$$

The square root of the probability corresponds to a Boltzmann weight at 600 K, similar to the 500 K previously used by Lindorff-Larson et al.[22] It is a compromise between equal weight (infinite $T$) that cannot ensure a high accuracy at the most probable conformations and the 300 K Boltzmann weight that will lead to large errors in the barrier regions. Besides, $w(\psi)$ is doubled at $\psi = -40°$ to achieve better fitting at the $\alpha$-helix conformation, which is highly populated in protein structures but much less favored in a coil library. A very simple version of self-adaptive evolution strategy was used in the parameter fitting. In each iteration, one parameter is randomly chosen for mutation, by adding a random value of normal distribution with standard deviation $\sigma$. Only a mutation that improves the fitness is accepted. Following the one-fifth success rule, if the acceptance rate of mutating a certain parameter is >20%, the corresponding $\sigma$ is increased to 1.5 $\sigma$; otherwise, the $\sigma$ is reduced to 0.6 $\sigma$. Excellent convergence can be achieved within $10^5$ iterations. Although Fourier expansion up to fifth-order was used for backbone dihedral angles, the fitting is well overdetermined due to 10° interval for $\Delta G_\phi$ and $\Delta G_\psi$. An actual example of the fitting was given in Figure S2 (Supporting Information). It is an important feature that our new force field places higher precision on describing the conformations with low free energies.

For AAs other than Ala/Gly/Pro, there are three side-chain $\chi_1$ rotamers, which can give quite different $\phi$, $\psi$ plots. Under the standard forms of current force fields, we cannot use different $\phi$, $\psi$ potentials for different rotamers. However, if we directly use eq 3 to obtain $\phi$, $\psi$ distribution regardless of the rotameric state, the most abundant rotamer will weight more for the optimized parameters and the Ramachandran plot of the least favored rotamer may not be well reproduced. To reduce this bias, we use the following to combine the three rotamer-dependent $\phi$, $\psi$ distributions:

$$n(\phi,\psi) = n_{g+}(\phi,\psi)/\sqrt{N_{g+}} + n_t(\phi,\psi)/\sqrt{N_t} + n_{g-}(\phi,\psi)/\sqrt{N_{g-}} \tag{14}$$


dx.doi.org/10.1021/jp5017449 | J. Phys. Chem. B 2014, 118, 6983−6998

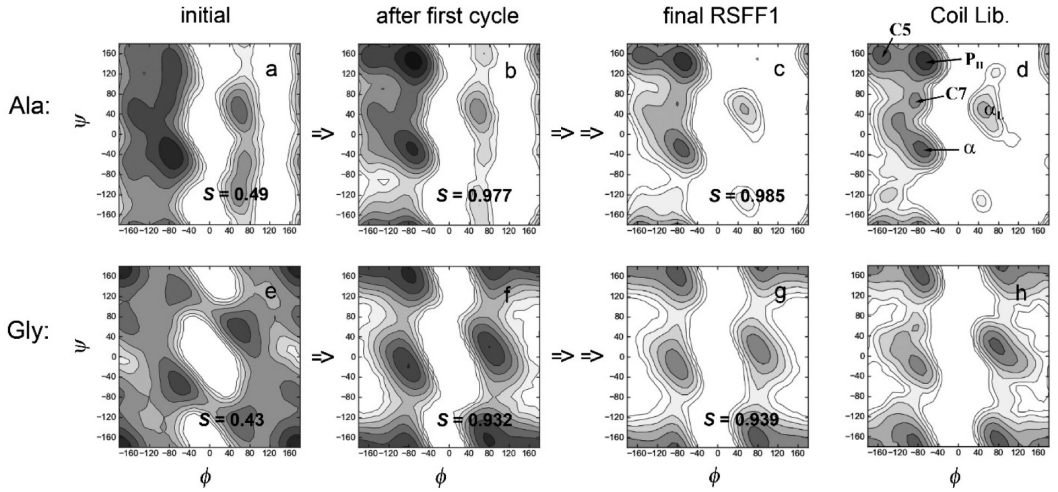



**Figure 3.** Simulated $\phi$, $\psi$ distributions during the development of RSFF1 force field, together with the target data from the PDB coil library. For both alanine (upper) and glycine (lower), $\phi$, $\psi$ distributions before (a, e) and after (b, f) the first cycle of the force field optimization and the results from the final optimized force field (c, g) are given. The similarity coefficient ($S$) with respect to the coil library plot is given for each simulated plot. Contours are drawn every $k_BT$ free energy difference. The same scale is used throughout the paper.

where $N_{g+}$, $N_t$, and $N_{g-}$ are the total numbers of the g+, t, and g− rotamers, respectively. The square root of total number ensures that the more abundant rotamer still weights more in the fitting. The obtained $n_{Coil}(\phi,\psi)$ and $n_{MD}(\phi,\psi)$ are then normalized to $p_{Coil}(\phi,\psi)$ and $p_{MD}(\phi,\psi)$ using eq 5. They are directly related to $\phi$, $\psi$ free energy surfaces on the basis of the Boltzmann distribution law.

**2.6. Optimization of Side-Chain Torsion Potentials.** Less Fourier terms (up to third order) were used for side-chain $\chi_i$ torsions. For the side-chain $\chi_1$ and $\chi_1'$ potentials, the new force field use fewer terms than OPLS-AA/L to reduce the number of parameters:

$$V(\chi_1) + V(\chi_1') = k_1 \cos(\chi_1) + k_3 \cos(3\chi_1) + k_1' \cos(\chi_1') \tag{15}$$

The update of parameters $k_1$ and $k_1'$ is based on comparing the simulated and target populations of g+/t/g− rotamers:

$$k_{1,new} - k_{1,old} = \Delta k_1 = \alpha RT \ln[(p_{g-,coil}/p_{t,coil})/(p_{g-,MD}/p_{t,MD})] \tag{16a}$$

$$k_{1,new}' - k_{1,old}' = \Delta k_1' = \alpha RT \ln[(p_{g-,coil}/p_{g+,coil})/(p_{g-,MD}/p_{g+,MD})] \tag{16b}$$

In practice, we found $\alpha = 0.6$ is a good choice. For $\beta$-branched Val and Ile, there are two N−$C_\alpha$−$C_\beta$−$C_\gamma$ dihedral angles and two C−$C_\alpha$−$C_\beta$−$C_\gamma$ dihedral angles on one $C_\alpha$−$C_\beta$ rotation. Thus, the $p_{g-}/p_t$ in eq 16a and $p_{g-}/p_{g+}$ in eq 16b were replaced by $p_{g-}/p_{g+}$ and $p_t/p_{g+}$, respectively. For Thr, The N−$C_\alpha$−$C_\beta$−$C_\gamma$ and C−$C_\alpha$−$C_\beta$−$C_\gamma$ dihedral potentials were set to zero, so eq 16a and eq 16b can be applied. Unlike $k_1$ and $k_1'$, $k_3$ is manually adjusted to reproduce the rotational barriers.

We use the functional form similar to OPLS-AA/L force field for $\chi_{i>1}$ torsion potentials:

$$V(\chi) = k_1 \cos(\chi) + k_2 \cos(2\chi) + k_3 \cos(3\chi) \tag{17}$$

In most cases, $k_1$ controls the trans/gauche preference, and $k_3$ controls the rotational barrier. Except for the $\chi_2$ of Asx and $\chi_3$ of Glx, we found $k_2 = 0$ can be used for all side-chain torsions. The rotation of terminal −CO−NH group ($\chi_2$ of Asn and $\chi_3$ of Gln) involves two coupled dihedral angles: C−C−C−O and C−C−C−N. The potential for the latter is set to zero to simplify the parametrization. The updates of Fourier coefficients $\Delta k_1$, $\Delta k_2$, $\Delta k_3$ were obtained from minimizing:

$$s = \Sigma\left[\Delta V(\chi) - RT \ln \frac{p_{coil}(\chi)}{p_{MD}(\chi)}\right]^2 \tag{18}$$

The $\chi$ values with relative free energy >20 kJ/mol from coil library are not included in the fitting. The fitting was also carried out using self-adaptive evolution strategy.

## 3. RESULTS AND DISCUSSION

**3.1. Alanine and Glycine.** We began our studies with Ala because most AAs are its derivatives. The reparameterization began with setting all four $\phi$, $\phi'$, $\psi$, $\psi'$ potentials to zero, with $\sigma$ = 0.270 nm and $\varepsilon$ = 0.1 kJ/mol for all the six 1-4 L-J interactions in Scheme 2. As shown in Figure 3a, the obtained $\phi$, $\psi$ distribution (a) was very different from the target coil library data ($S$ = 0.49). However, after only one cycle of optimization using our $\phi$, $\psi$ decomposition approach, the simulated $\phi$, $\psi$ plot (b) was significantly improved to $S$ > 0.97. This agreement is already close to the final RSFF1 force field ($S$ = 0.985). This highly efficient approach makes the optimization of torsion potentials no longer the bottleneck in our force field development.

As shown in Figure 3b, the densities for $\phi < -160°$ with $\psi$ in the range +40° to −80° are still higher than the coil library distributions whereas the $C_5$ basin is not deep enough. These can be improved by adding a weak repulsion between $H_i\cdots N_{i+1}$ to destabilize the $\alpha'$ conformation and a weak attraction between $H_i\cdots O_i$ to stabilize the $C_5$ conformation. We also reduce the repulsion between $O_{i-1}\cdots C_i$ to reduce the barrier at $\phi$ = 0°. When proper modifications shown in Table 1 are introduced, with several additional cycles of optimization, the $\phi$, $\psi$ distribution from the Ala dipeptide simulation agrees excellently with the coil library distributions.


dx.doi.org/10.1021/jp5017449 | J. Phys. Chem. B 2014, 118, 6983−6998

**Table 1. Parameters for All Modified 1-5 and 1-6 L-J Interactions**

| pair[a] | type | $\sigma$ (nm) | $\varepsilon$ (kJ/mol) | note |
|---|---|---|---|---|
| H···O | 1-5 | 0.180 | 5.0 | backbone |
| H···$N_{i+1}$ | 1-5 | 0.290 | 0.1 | backbone |
| $O_{i-1}$···C | 1-5 | 0.270 | 0.1 | backbone |
| $C_\gamma$···O | 1-5 | 0.230 | 5.0 | Asp |
| $C_\gamma$···O | 1-5 | 0.230 | 3.0 | Asn |
| $C_\gamma$···$O_{i-1}$ | 1-6 | 0.250 | 1.0 | Asx |
| $O_\delta$···C | 1-5 | 0.270 | 1.5 | Asx |
| $C_\gamma$···H | 1-5 | 0.290 | 0.1 | Asp |
| $C_\gamma$···H | 1-5 | 0.260 | 0.1 | Asn |
| $C_\gamma$···$C_{i-1}$ | 1-5 | 0.350 | 0.1 | Asx |
| $O_\delta$···N | 1-5 | 0.310 | 0.1 | Asx |
| $C_\gamma$···$N_{i+1}$ | 1-5 | 0.350 | 0.1 | Asx |
| $N_\delta$···O | 1-6 | 0.375 | 0.1 | Asn |
| $N_\delta$···N | 1-5 | 0.320 | 0.1 | Asx |
| $O_\gamma$···$C_{i-1}$ | 1-5 | 0.320 | 0.1 | Ser, Thr |
| $O_\gamma$···O | 1-5 | 0.330 | 0.1 | Ser, Thr |
| $O_\gamma$···$N_{i+1}$ | 1-5 | 0.345 | 0.1 | Ser, Thr |

[a]The subscript $i + 1$ or $i - 1$ indicates the atom is in the following or preceding residue.

Among 20 AAs, Gly has a special conformational flexibility due to its lack of a side chain. We then used Gly to examine the transferability of the $V_{\text{local-LJ}}$ parameters optimized on Ala. Interestingly, with optimized $\phi$, $\psi$ parameters, RSFF1 gives the $\phi$, $\psi$ distribution very similar to the coil library distribution (Figure 3g). Similar to the case of Ala, the $S$ value increases from 0.43 to 0.932 after just one cycle of optimization (Figure 3e,f). Because of the success on Gly residue, the same backbone $V_{\text{local-LJ}}$ parameters were used for all other AAs.

**3.2. Side-Chain $\chi$ Torsions.** As shown in Table 2, the first-order Fourier coefficients $k_1$ and $k_1'$ for the side-chain $\chi_1$ and

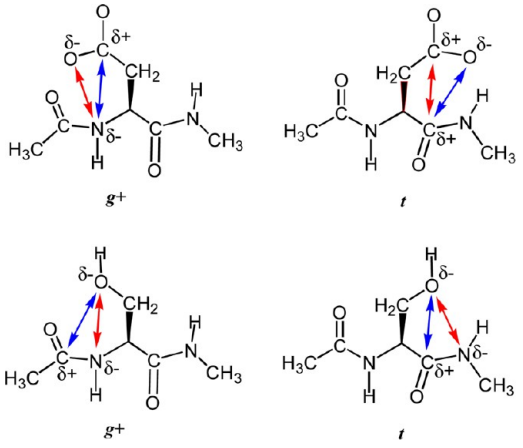
**Table 2. Side-Chain $\chi_1$ Fourier Coefficients (kJ/mol) of Some AA Residue from the OPLS-AA/L and the RSFF1 Force Fields**

| | OPLS-AA/L | | | | | RSFF1 | | |
|---|---|---|---|---|---|---|---|---|
| | $k_1$ | $k_2$ | $k_3$ | $k_1'$ | $k_2'$ | $k_1$ | $k_3$ | $k_1'$ |
| Glu | 10.4 | 0.5 | −0.5 | −3.4 | 1.2 | −1.9 | 1.0 | 1.0 |
| Gln | 2.4 | 1.1 | 1.7 | −5.6 | −1.7 | −1.3 | 1.0 | 0.9 |
| Lys | 1.3 | 0.4 | 0.8 | −4.7 | 1.1 | −1.1 | 1.0 | 1.2 |
| Ser | 11.4 | 1.8 | 2.2 | −11.8 | 1.8 | −3.6 | 0.8 | 0.6 |
| Thr | 11.4 | 1.8 | 2.2 | −11.8 | 1.8 | −3.7 | 1.0 | 0.1 |
| Asp | −9.4 | −2.0 | −0.6 | −4.3 | −6.5 | −0.6 | 0.8 | 5.5 |
| Asn | −7.3 | 1.4 | 3.1 | 2.2 | 3.4 | 1.1 | 0.8 | 4.2 |

$\chi_1'$ torsions (Scheme 1) in OPLS-AA/L have large deviations, with the ranges −9.4 to +11.4 and −11.8 to +2.2 kJ/mol, respectively. The large deviations might be partly resulted from using the gas-phase QM calculations for parametrization.[17] Especially, OPLS-AA/L gives a significant preference to the $t$ rotamer for Glu residue, due to a large $k_1$ of 10.4 kJ/mol. Indeed, in the gas phase the most stable conformation of Glu dipeptide forms a H-bond between side chain and backbone,[17] which is not favored in protein structures. From the coil library, Glu and Gln have very similar rotamer preferences,[41] indicating that the charged terminal group has limited effect.

For residues with short polar side-chains (Ser, Thr, Asp, and Asn), we noticed that the $\chi_1$ and $\chi_1'$ parameters in the OPLS-AA/L force field tend to compensate the unbalanced 1-4/1-5 electrostatic interactions. As shown in Scheme 3, the g+

**Scheme 3. Balance between 1-4 and 1-5 Electrostatic Interactions: Top, Asp; Bottom, Ser**




rotamer of Asp has attractive 1-4 interaction and repulsive 1-5 interaction, whereas the t rotamer of Asp has 1-4 repulsion and 1-5 attraction. In the original OPLS-AA/L force field, the 1-4 electrostatic interactions are scaled down by a factor of 0.5, which significantly favors the t rotamer and disfavors the g+ rotamer. In OPLS-AA/L, negative $k_1$ (−9.4 kJ/mol) for Asp still cannot fully compensate the strong t preference from the unbalanced electrostatic 1-4/1-5 interactions. A similar situation also occurred for Ser and Thr, in an opposite way to Asp (also shown in Scheme 3).

It might be more appropriate not to scale down the 1-4 electrostatic interactions. Indeed, Smith and Karplus found that reducing the 1-4 electrostatic interactions by 50% led to qualitatively incorrect trans−gauche energy for $n$-butane.[68] In developing ECEPP-05 force field, Scheraga et al. found that no scaling of 1-4 electrostatic interactions provided the best results for conformational energies of 1,3-propanediol.[69] In RSFF1 force field, we do not scale down the 1-4 electrostatics, resulting in more consistent torsion parameters. Within a few cycle of applying eq 16, these parameters were optimized to achieve excellent agreement with the coil library rotamer distributions (Figure S3, Supporting Information). The ranges of $k_1$ and $k_1'$ are reduced to −3.7 to +2.3 and −0.9 to +5.5 kJ/mol, respectively.

As shown in Figure 4, RSFF1 reproduces the whole $\chi_1$ free energy profiles (or PMFs) from the coil library quite well. A few rotational barriers from RSFF1 simulations are slightly lower than those from the coil library. The OPLS-AA/L only reproduces PMFs well for some hydrophobic residues (such as Leu, Phe, Tyr, Trp, Val, Ile), but it does not describe the rotamer distributions (relative free energies of the three minima) well in many cases. The problem is most serious for Glu, His, Cys, Ser, Thr, and Asp. In RSFF1, besides the $k_3$ parameter in eq 15, the 1-4 L-J parameters between non-hydrogen atoms are also adjusted to give a good description of the rotational barriers. These parameters are shared between different $\chi$ torsions and different AA types. These modified 1-4 L-J parameters resulted in much weaker interactions compared with the original ones in OPLS-AA (Table 3).

Besides $\chi_1$ rotation, all side-chain $\chi_{i>1}$ rotational free energy profiles were optimized to match coil library PMFs (Figure S4,


dx.doi.org/10.1021/jp5017449 | J. Phys. Chem. B 2014, 118, 6983−6998

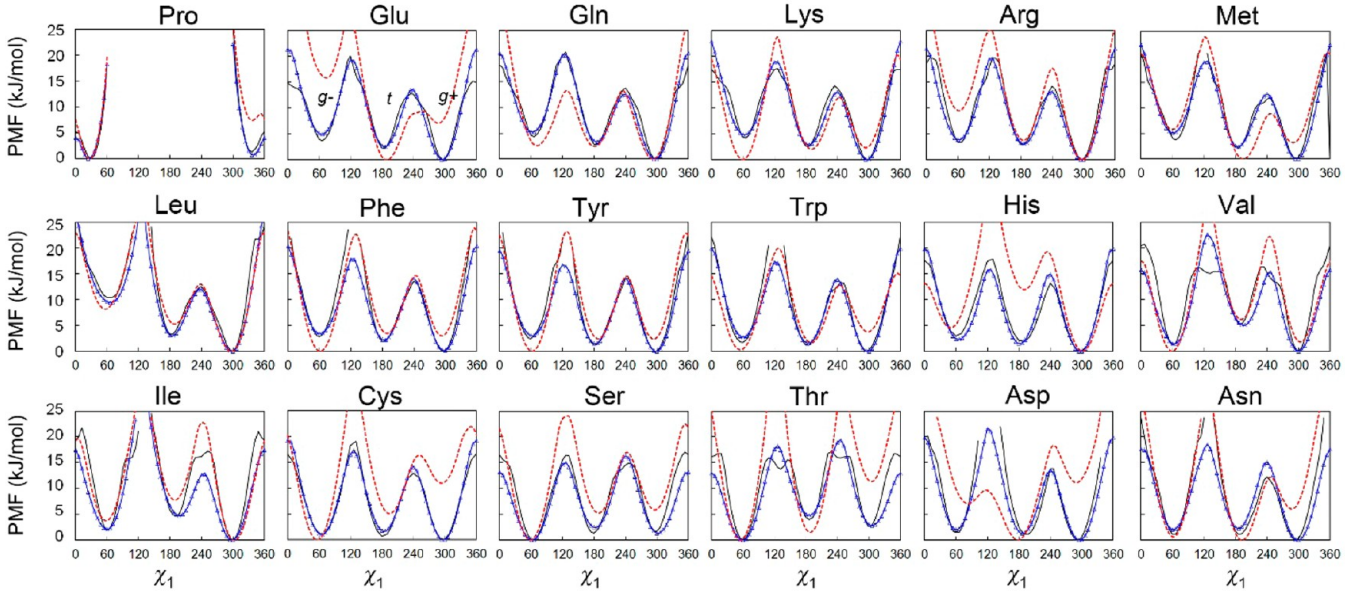



**Figure 4.** Free energy profiles of side-chain $\chi_1$ in 18 AA residues, from the coil library (black line) and from OPLS-AA/L (red dashed line) and RSFF1 (blue line with triangles) simulations.

**Table 3. Modified 1-4 L-J Parameters for Side-Chain Conformations**

| pair | $\sigma$ (nm) | $\varepsilon$ (kJ/mol) |
|---|---|---|
| $sp^3$-C···all-N | 0.280 | 0.1 |
| $sp^2$-C···all-N | 0.320 | 0.1 |
| $sp^3$-C···all-C | 0.290 | 0.1 |
| $sp^2$-C···$sp^2$-C | 0.310 | 0.1 |
| all-O···all-N | 0.300 | 0.1 |
| all-O···$sp^3$-C | 0.280 | 0.1 |
| all-O···$sp^2$-C | 0.300 | 0.1 |
| all-S···all-N | 0.330 | 0.1 |
| all-S···all-C | 0.330 | 0.1 |

Supporting Information). Using the fitting scheme described by eq 18 in section 2.6, the target PMF can be achieved by only one cycle of optimization (Figure S5, Supporting Information). As a result, fewer $\chi$ parameters are used in RSFF1 compared with OPLS-AA/L (Table S1, Supporting Information). The PMF of $\chi_2$ rotation in Asp is shown in Figure 5. From coil library, $\chi_2$ around 0° and 180° are mostly favored, which may be stabilized by n → $\pi^*$ interaction between Asp side-chain carboxylic O atom and backbone C atom, as indicated from QM calculations.[70] In the contrary, OPLS-AA/L and AMBER

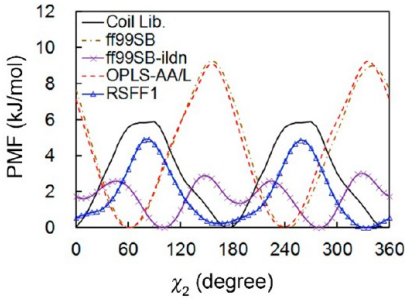



**Figure 5.** Free energy profiles for $\chi_2$ in Asp from the coil library and simulations with various force fields. Due to the symmetry of the $-COO^-$ group, $\chi_2$ and $\chi_2$ + 180° are identical.

ff99SB gave maximum free energies at $\chi_2$ near 165°. The $\chi_2$ PMF from ff03 force field (not shown) is very similar to that from ff99SB. Simulation using improved version ff99SB-ILDN still gives $\chi_2$ PMF different from the coil library PMF, although it is better than ff99SB. Indeed, the parametrization for Asp is most difficult because the side-chain has strong interactions with the backbone.

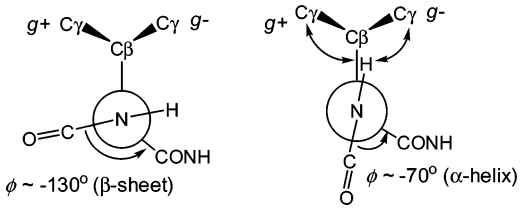
**3.3. Rotamer-Dependent Ramachandran Plots.** Because the high efficiency of our parametrization methods and sufficient data from coil library, it is very convenient to use different $\phi$, $\psi$ parameters for different AAs. Still, same set of $\phi$, $\psi$ parameters are used for AAs with very similar local conformational features: (1) Glu/Gln/Lys/Arg/Met/Leu with single $sp^3$ C$\gamma$ atom, (2) Phe/Tyr/Trp with single nonpolar $sp^2$ C$\gamma$ atom, and (3) Val/Ile with nonpolar $\beta$-branched side chains. Other AAs with more polar $\gamma$ atoms use their special $\phi$, $\psi$ parameters, including Ser, Thr, Cys, His, Asp, and Asn. The final $\phi$, $\psi$ torsion parameters are given in the Supporting Information (Table S2).

To account for the coupling between side-chain and backbone conformations, an additional 1-5 L-J interaction between the C$\gamma$ atom and the backbone amide H atom is added. This may be related to the fact that $\beta$-branched AAs (Val, Ile) with two nonpolar C$\gamma$ atoms have intrinsically highest $\beta$-sheet propensities and low $\alpha$-helix propensities (Scheme 4), as discussed earlier by Han et al.[71] In addition, Ala without the C$\gamma$ atom has the highest propensity for $\alpha$-helix formation. The steric repulsions between the polar H atom and nonpolar atoms can be missing if the van der Waals radius of amide H atom was ignored. Like most $V_{\text{local-LJ}}$ in RSFF1, we set $\varepsilon$ = 0.1 kJ/mol and adjusted the $\sigma$ value. Finally, $\sigma$ = 0.320 nm was used for the interactions with the $sp^3$ C$\gamma$ atom (most AAs) or S$\gamma$ atom (Cys), and slightly weaker repulsion of $\sigma$ = 0.310 nm was used for the interactions with $sp^2$ C$\gamma$ atom in aromatic side chains.

With this additional H···C$\gamma$/S$\gamma$ interaction and optimized backbone $\phi$, $\psi$ potentials, a significant increase of similarities ($S$) with coil library data can be achieved. As shown in Table 4, OPLS-AA/L simulations give $S$ < 0.8 for g− rotamers of most AAs, and $S$ < 0.9 for all t rotamers. On the other hand, RSFF1


dx.doi.org/10.1021/jp5017449 | J. Phys. Chem. B 2014, 118, 6983−6998

**Scheme 4. Newman Projections along the N−C$\alpha$ Bond ($\phi$ Torsion)[a]**




[a]The amide H atom is close to C$\beta$ and C$\gamma$ atoms when in $\alpha$-helical conformation. This H···C$\gamma$ repulsion also depends on the side-chain rotamer.

**Table 4. Similarity Coefficients ($S$) between Simulated Rotamer-Dependent $\phi$, $\psi$ Distributions and Coil Library Statistics**

| | OPLS-AA/L | | | | RSFF1 | | | |
|---|---|---|---|---|---|---|---|---|
| | g+ | t | g− | all | g+ | t | g− | all |
| **A** | | | | 0.83 | | | | 0.985 |
| **G** | | | | 0.35 | | | | 0.939 |
| **P** | 0.88 | | 0.94 | 0.86 | 0.99 | | 0.99 | 0.998 |
| **E** | 0.90 | 0.85 | 0.29 | 0.89 | 0.97 | 0.96 | 0.98 | 0.977 |
| **Q** | 0.93 | 0.91 | 0.72 | 0.94 | 0.97 | 0.97 | 0.95 | 0.980 |
| **K** | 0.92 | 0.85 | 0.68 | 0.77 | 0.97 | 0.96 | 0.96 | 0.976 |
| **R** | 0.92 | 0.88 | 0.64 | 0.90 | 0.97 | 0.96 | 0.93 | 0.973 |
| **M** | 0.95 | 0.85 | 0.56 | 0.90 | 0.96 | 0.96 | 0.94 | 0.965 |
| **L** | 0.95 | 0.86 | 0.71 | 0.94 | 0.98 | 0.94 | 0.97 | 0.981 |
| **F** | 0.90 | 0.81 | 0.61 | 0.70 | 0.97 | 0.94 | 0.97 | 0.980 |
| **Y** | 0.90 | 0.82 | 0.58 | 0.72 | 0.97 | 0.95 | 0.97 | 0.983 |
| **W** | 0.90 | 0.85 | 0.76 | 0.78 | 0.95 | 0.93 | 0.87 | 0.964 |
| **C** | 0.90 | 0.82 | 0.45 | 0.47 | 0.96 | 0.94 | 0.96 | 0.967 |
| **V** | 0.90 | 0.80 | 0.80 | 0.85 | 0.97 | 0.96 | 0.98 | 0.979 |
| **I** | 0.90 | 0.89 | 0.83 | 0.91 | 0.95 | 0.96 | 0.96 | 0.963 |
| **S** | 0.85 | 0.87 | 0.57 | 0.63 | 0.95 | 0.91 | 0.96 | 0.970 |
| **T** | 0.86 | 0.74 | 0.82 | 0.78 | 0.95 | 0.90 | 0.96 | 0.972 |
| **D** | 0.81 | 0.47 | 0.35 | 0.61 | 0.94 | 0.81 | 0.89 | 0.944 |
| **N** | 0.87 | 0.51 | 0.54 | 0.70 | 0.94 | 0.87 | 0.91 | 0.953 |

simulations give $S > 0.92$ for all except a few cases. As shown in Figure 6, there is an additional $\alpha'$ basin ($\phi, \psi \sim -140°, 30°$) in the $\phi$, $\psi$ plots of g− and t rotamers of Lys from OPLS-AA/L simulations, which is absent in coil library results. This additional $\alpha'$ basin agrees with the OPLS-AA/L simulation of Ala dipeptide, which was suppressed in RSFF1 by modified local interactions. All the $\chi_1$-dependent $\phi$, $\psi$ plots from RSFF1 simulations are given in the Supporting Information (Figure S6).

Also from Table 4, OPLS-AA/L gives especially low $S$ values (0.5 or less) for t and g− rotamers of Asx. From the coil library results (Figure 6), the t Asp favors the $C_7$-like conformation around $\phi, \psi \sim -80°, +80°$ and $\alpha_L$ conformation with $\phi > 0$. For g− Asp, the bottom of the $\alpha_R$ basin is shifted to $\phi, \psi \sim -110°, +10°$ with a higher population than extended conformations. These conformations are not well stabilized in OPLS-AA/L force field. Asn also has similar special features.

Modifications of some 1-5/1-6 L-J interactions for Asx (Table 1) were introduced to achieve significantly better agreement with coil library observations. All these modified interactions involve pairs between polar atoms and may function to compensate possible small inaccuracies in the water-mediated electrostatic interactions. Compared with the case for other AAs, full optimization of these parameters is rather difficult, which required most of the efforts in our RSFF1 parametrization. Interestingly, different from the majority situations of $\varepsilon$ = 0.1 kJ/mol, the modified L-J interactions between polar C and O atoms in Asx are attractive. The L-J potential with $\sigma$ = 0.230 nm and $\varepsilon$ = 5.0 kJ/mol gives an energy of −2.8 kJ/mol at the distance of 0.31 nm. There it can be stabilization from the n → $\pi^*$ interaction between the oxygen long pair and antibond $\pi$ orbital of the C═O group.[72] Still, RSFF1 gives the $\phi$, $\psi$ plot of t Asp not in exact agreement with the coil library ($S$ = 0.81) (Figure 6). Simple L-J potential may not fully account for the directionality of the n → $\pi^*$ interaction.

For Ser and Thr, our QM calculations indicated that their $O_\gamma$ and $C_{i-1}$ atoms can have short distance of <3.0 Å. This can lead to >4 kJ/mol repulsion when default OPLS-AA L-J parameters are used. We thus set $\varepsilon$ = 0.1 kJ/mol and $\sigma$ = 0.32 nm to give reduced repulsion at short C···O distance. This allows strong electrostatic attraction between the two oppositely charged atoms, which is sufficient to give satisfactory results.

**3.4. NMR *J* Couplings of Dipeptides.** To compare the performance of RSFF1 with that of other force fields, we calculate the NMR $^3J_{H_NH_\alpha}$ couplings of all 19 dipeptides (except Pro) from their $\phi$ torsions sampled in the simulations and compared them with experimental data reported by Avbelj et al.[50] The $^3J_{H_NH_\alpha}$ scalar coupling has been widely used in experimental characterization of conformations of short peptides in solution. It is sensitive to the distribution of backbone $\phi$ angle through the Karplus relationship:

$$^3J_{H_NH_\alpha} = A\ \cos^2(\phi - 60°) + B\cos(\phi - 60°) + C \quad (19)$$

Several different sets of empirical Karplus parameters ($A$, $B$, $C$ in eq 19) were reported, from different fittings of experimental $^3J_{H_NH_\alpha}$ values of different proteins to their X-ray or NMR structures.

From Table 5, for all Karplus parameter sets used, RSFF1 force field gives the lowest RMSD values, indicating better agreement with experimental $^3J_{H_NH_\alpha}$. Especially, when the 2007 parameter set is used, only RSFF1 gives the RMSD value (0.19 Hz) smaller than the estimated uncertainty ($\sigma$ = 0.36 Hz) in deriving the Karplus parameters. Unlike the RMSD values, different Karplus parameter sets give nearly the same correlation coefficients ($R$) between calculated and experimental $^3J_{H_NH_\alpha}$. The RSFF1 gives significantly higher $r$ values (>0.9) than other force fields.

As shown in Figure 7, experimental $^3J_{H_NH_\alpha}$ coupling of the Ala dipeptide is significantly smaller than its derivatives (non-Gly/Ala AAs) by 0.6−1.8 Hz. Different from experiments, A99sb*-ildn force field gives similar $^3J_{H_NH_\alpha}$ coupling (within 0.7 Hz) for Ala and its derivatives (except Val). The $^3J_{H_NH_\alpha}$ for Ala and some AAs such as Glu and Val are significantly overestimated. The overestimation of Ala $^3J_{H_NH_\alpha}$ is also observed in OPLS-AA/L simulation. Compared with A99sb*-ildn, the A99sb-ildn-NMR force field consistently reduced the $^3J_{H_NH_\alpha}$ of all AAs, resulting in good results for some AAs, but $^3J_{H_NH_\alpha}$ values for Gly, Cys, Asn, and His were considerably underestimated. Unlike other force fields, RSFF1 can reproduce the gap between $^3J_{H_NH_\alpha}$ of Ala and its derivatives. The excellent performance of RSFF1 agrees with the previous finding that


dx.doi.org/10.1021/jp5017449 | J. Phys. Chem. B 2014, 118, 6983−6998

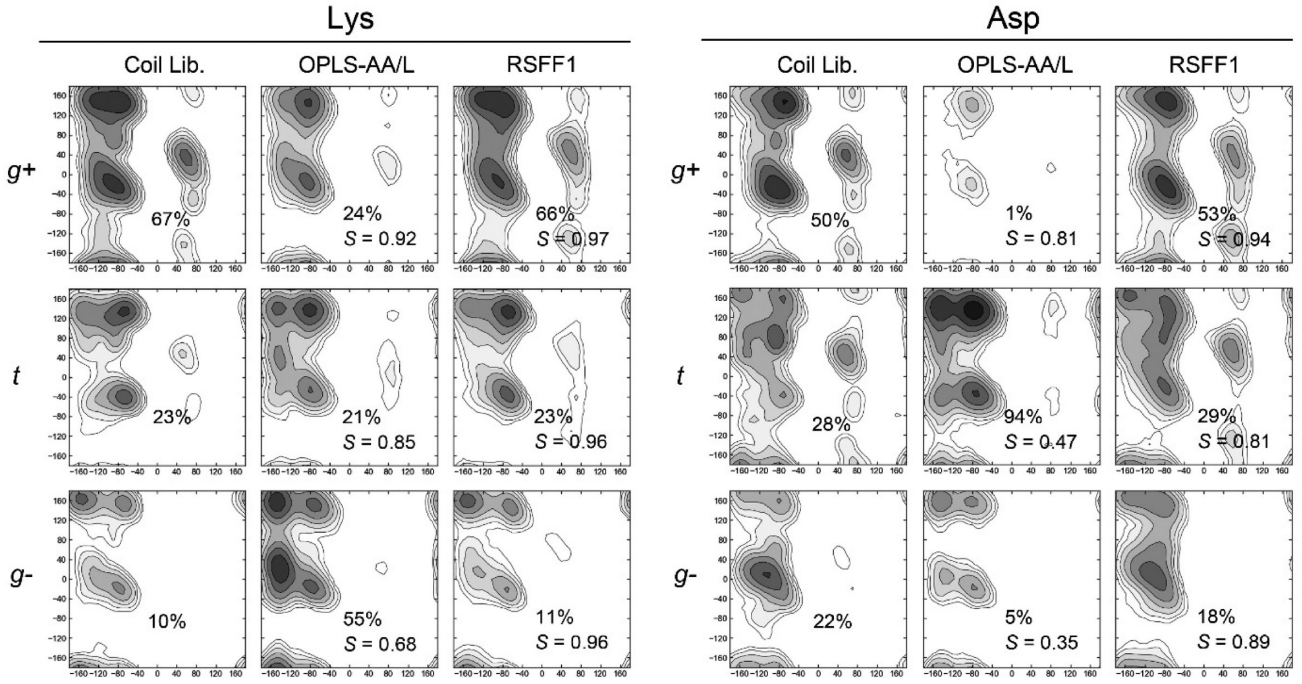



**Figure 6.** $\phi$, $\psi$ distributions for lysine (left) and aspartate (right) residues in different $\chi_1$ rotamers (g+/t/g−). Results are from the coil library statistics and those from dipeptide simulations with OPLS-AA/L and RSFF1 force fields.

**Table 5. Root Mean Square Deviation (RMSD) and the Correlation Coefficients ($R$) between Calculated and Experimental $^3J_{H_NH_\alpha}$ Couplings of the 19 Dipeptides (Pro Excluded)[a]**

| Karplus parameters | | | | 99SB*-ildn | | 99SB-ildn-NMR | | CHARMM22* | | OPLS-AA/L | | RSFF1 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| year | $A$ | $B$ | $C$ | RMSD | $R$ | RMSD | $R$ | RMSD | $R$ | RMSD | $R$ | RMSD | $R$ |
| 1984[73] | 6.40 | −1.40 | 1.90 | 0.66 | 0.68 | 0.38 | 0.62 | 0.37 | 0.71 | 0.62 | 0.78 | 0.31 | 0.92 |
| 1991[74] | 6.60 | −1.30 | 1.50 | 0.42 | 0.68 | 0.46 | 0.61 | 0.36 | 0.70 | 0.35 | 0.78 | 0.20 | 0.92 |
| 1993[75] | 6.51 | −1.76 | 1.60 | 0.72 | 0.67 | 0.41 | 0.62 | 0.39 | 0.72 | 0.68 | 0.78 | 0.34 | 0.91 |
| 1997[76] | 7.09 | −1.42 | 1.55 | 0.81 | 0.68 | 0.43 | 0.62 | 0.45 | 0.70 | 0.76 | 0.78 | 0.42 | 0.92 |
| 1999[77] | 7.90 | −1.05 | 0.65 | 0.37 | 0.69 | 0.59 | 0.60 | 0.46 | 0.65 | 0.30 | 0.78 | 0.29 | 0.92 |
| 2000[78] | 9.44 | −1.53 | −0.07 | 0.97 | 0.69 | 0.47 | 0.61 | 0.50 | 0.68 | 0.90 | 0.78 | 0.49 | 0.92 |
| 2007[79] | 8.40 | −1.36 | 0.33 | 0.55 | 0.69 | 0.46 | 0.61 | 0.35 | 0.68 | 0.47 | 0.78 | 0.19 | 0.92 |

[a]Results calculated using seven different Karplus parameter sets are listed; for each, the lowest RMSD value(s) among the four force fields are underlined.

the $^3J_{H_NH_\alpha}$ directly calculated from coil library $\phi$ distributions agree very well with those in dipeptides.[50]

To better understand these results, the $\phi$ distributions of two representative cases are shown in Figure 8. From the coil library, the $\phi$ distribution of Ala has a highest peak around −67° and a lower shoulder around −153°, which agree with previous work of Avbelj et al.[47] On the other hand, the coil library $\phi$ distribution of g+ Gln has a much higher population around −100° and only one peak. (Scheme 4). This agrees with much higher $^3J_{H_NH_\alpha}$ of Gln dipeptide. This large difference cannot be fully reproduced by current force fields. At $\phi$ = −120°, both OPLS-AA/L and CHARMM22* well reproduce the coil library value for Ala but cannot fully describe the increased population for g+ Gln. At $\phi$ = −65°, ff99SB-NMR agrees with coil library results for Ala but cannot fully follow the decrease in the population for g+ Gln. Indeed, the coil library $\phi$, $\psi$ distribution of Gln (ordinary AA) differ from that of Ala with $S$ = 0.86. In comparison, ff99SB variants and OPLS-AA/L give higher $S$ = 0.94−0.97. Our results imply that current force fields may underestimate the backbone conformational differences between Ala and its derivatives, suggesting that same parameters on all AAs may not be enough.

**3.5. Folding of Both $\alpha$-Helix Proteins and $\beta$-Hairpin Peptides.** The ability to fold peptides and small proteins is a stringent test of a force field, because even minor inaccuracies at single-residue level can lead to a significant perturbation of delicate balance among different structures. We carried out the folding simulations of Trp-cage[80] (Tc5b, a designed 20-residue $\alpha$-helix mini-protein), Trpzip-2[81] (a designed tryptophan zipper $\beta$-hairpin), and GB1 hairpin[82] (residues 41−56 of protein G B1 domain), using the RSFF1 and OPLS-AA/L, and the two state-of-the-art force fields[83] CHARMM22* and AMBER ff99SB*-ildn. We also carried out folding simulation of a much larger three-helix bundle protein Engrailed Homeodomain (1ENH)[84] using RSFF1. All folding simulations were carried out using REMD, initiated from unfolded structures.

The AMBER ff99SB*-ildn, CHARMM22*, and OPLS-AA/L cannot consistently stabilize the native structures of the four systems (Figure 9) as the dominant cluster. AMBER ff99SB*-ildn simulations of the two $\beta$-hairpins gave many quite different structures without a dominant cluster, although a small fraction


dx.doi.org/10.1021/jp5017449 | J. Phys. Chem. B 2014, 118, 6983−6998

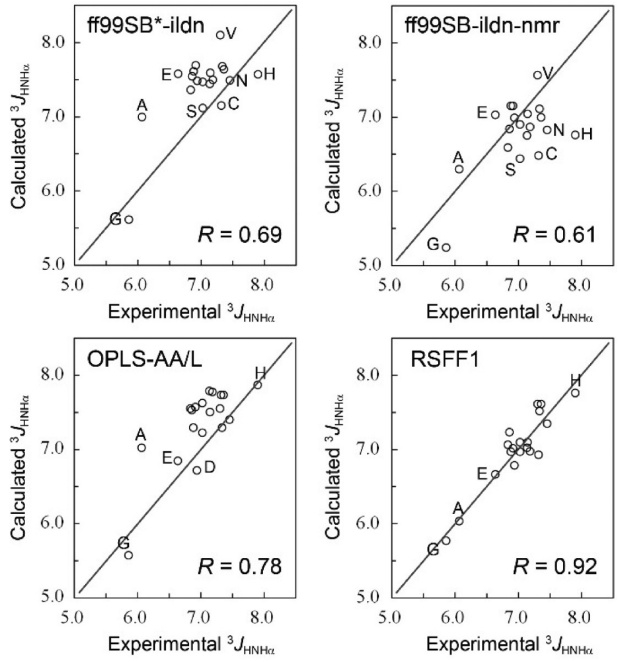



**Figure 7.** Calculated $^3J_{H_NH_\alpha}$ coupling constant in 19 dipeptides plotted against the corresponding experimental data. Each plot shows the results from simulations using each force fields, calculated using a recent (2007) Karplus parameter set.

of near-native structures was found for each. Indeed, even ff99SB with $\beta$-sheet propensity higher than ff99SB* still significantly underestimate the stability of Trpzip-2.[38] In a previous study, ff99SB simulations could not stabilize the folded structure of another $\beta$-hairpin peptide (Mbh12).[85] However, we noticed that previous folding simulations of GB1 hairpin using ff99SB* gave the $\beta$-hairpin structure as the most populated cluster (21%).[86] CHARMM22* can sample the native structure of the Trp-cage but did not give it as the dominant cluster in our simulation. Also, the native state of Engrailed Homeodomain is unstable in a previous MD simulation using CHARMM22*.[3] The OPLS-AA/L force field, which share same nonbonded parameters with RSFF1, cannot fold the Trp-cage. The simulation gave different structures without a dominant cluster, and structures from most populated clusters lack regular secondary structures.

The backbone torsion parameters for non-Gly/Pro AAs in current force fields were usually parametrized on Ala residue. Especially, the AMBER ff99SB*-ildn and CHARMM22* force fields were optimized to reproduce experimental data on Ala-based peptides, intended to achieve more balanced conformational preferences. For Figure 10a, they indeed give similarly $\alpha$-helicities of Ac-Ala$_{14}$-NHMe, reasonably agree with experiments for $T$ > 300 K. However, they give very different melting curves for Trp-cage and Trpzip-2 without any Ala residue. It seems that, at least for these systems, ff99SB*-ildn prefers $\alpha$-helix structure whereas CHARMM22* prefers $\beta$-hairpin structure. Therefore, current strategy of deriving backbone correction based on Ala-based peptides cannot fully solve the secondary structure biases existed in current protein force fields.

On the other hand, RSFF1 can successfully fold the two $\alpha$-helical proteins and two $\beta$-hairpins, each with the dominant cluster very similar to the experimental structure. The reliability of RSFF1 in stabilizing the native structures of various sequences may come from its ability to accurately describe different intrinsic conformational preferences of different AA residues by using the residue-specific parameters.

As shown in Figure 10, RSFF1 consistently overstabilizes the $\alpha$-helical structure of Ac-Ala$_{14}$-NHMe and the folded states of Trp-cage and the two $\beta$-hairpins. RSFF1 also overstabilizes the three-helix bundle Homeodomain, because a high population (80%) of folded structures is observed at temperature (330 K, the lowest replica) higher than its experimental $T_m$ (325 K). Interestingly, the $\phi$, $\psi$ distributions from coil library were originally used to model the denatured and intrinsically disordered peptides and proteins. However, because the underlying local interactions determining these intrinsic conformational features also exist in the folded states, the RSFF1 does not bias toward the unfolded state. On the contrary, it actually somehow overestimates the stability of the native state, which is better than uncertain secondary structure biases for some applications such as the structure prediction and refinement. The underline reason is still unknown, but it is possible that RSFF1 can be fine-tuned to achieve better agreement with experiments.

## 4. CONCLUSION AND OUTLOOK

In this work, we present our efforts in developing a new protein force field RSFF1, based on the $\phi$, $\psi$, and $\chi$ free energy surfaces (PMFs) of all 20 amino acids (AAs) from statistical analysis of protein coil library. A set of new methods has been established, by which excellent agreement can be achieved between PMFs from dipeptide simulations and the target PMFs. Especially, backbone torsion parameters, which are AA-dependent in RSFF1, can be easily optimized using our new $\phi$, $\psi$ decomposition approach. This work demonstrates that it is

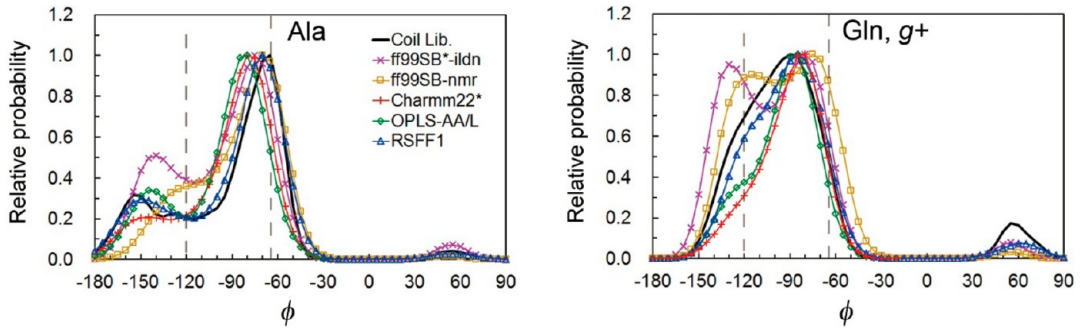



**Figure 8.** $\phi$ distributions of Ala (left) and g+ rotamer of Gln (right), from coil library statistics and force field simulations. The black dashed vertical lines indicate $\phi = -120°$ and $\phi = -65°$, corresponding to $\beta$-sheet and $\alpha$-helix/$P_{II}$ conformations, respectively.


dx.doi.org/10.1021/jp5017449 | J. Phys. Chem. B 2014, 118, 6983−6998

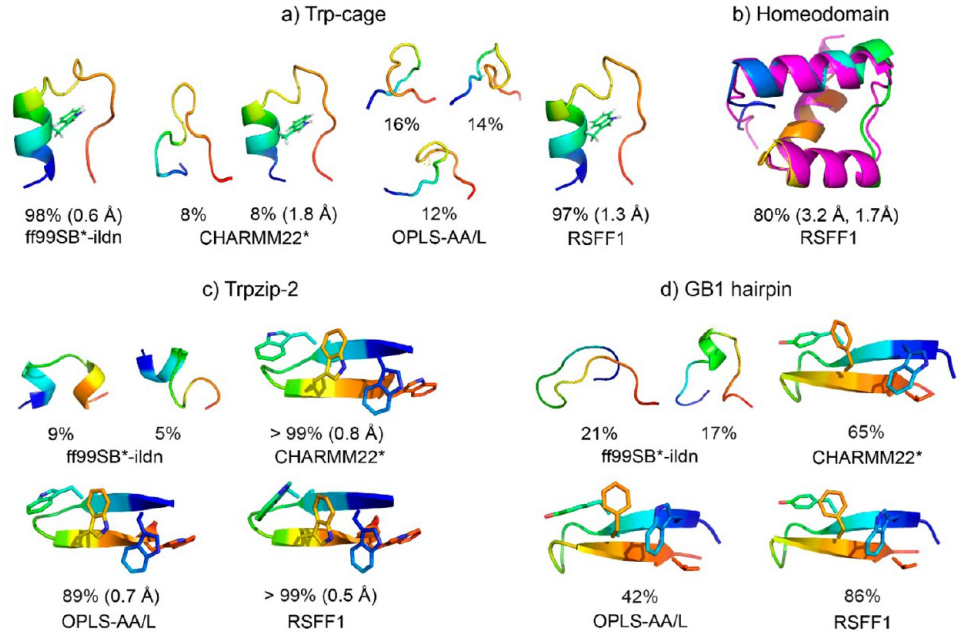



**Figure 9.** Representative structures of Trp-cage (a), engrailed homeodomain (b), Trpzip-2 (c), and GB1 hairpin (d) from REMD simulations using various force fields. The percentage of each cluster is given below its representative structure. For each structure similar to the corresponding experimental structure (a, 1L2Y; b, 1ENH; c, 1LE1), its backbone RMSD is also given in parentheses. For homeodomain, the predicted structure (rainbow) from RSFF1 simulation is superposed with its crystal structure (1ENH, magenta). The NMR study of the GB1 hairpin[82] did not give atomic coordinates, but the representative structures from CHARMM22*, OPLS-AA/L, and RSFF1 all have the same $\beta$-sheet H-bond pattern and aromatic side-chain packing as those derived from the experiments.

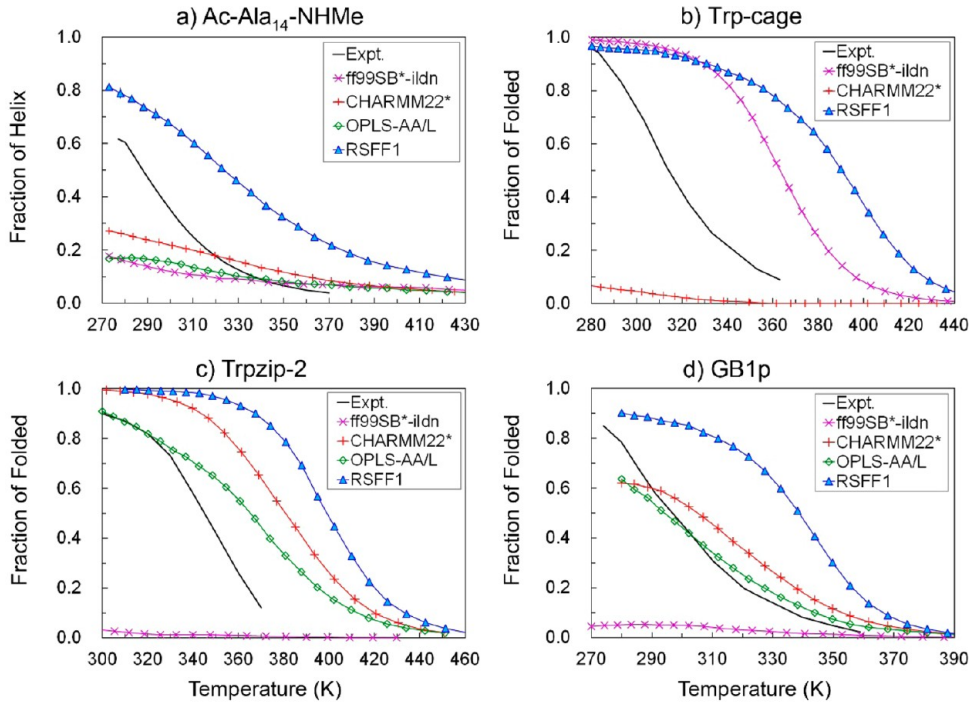



**Figure 10.** Melting curves of Ac-Ala$_{14}$-NHMe (a), Trp-cage (b), Trpzip-2 (c), and GB1 hairpin (d), from experiments and REMD simulations using various force field. The experimental melting curve of Ac-Ala$_{14}$-NHMe was calculated using AGADIR,[87] which is based on experimental helix nucleation and propagation parameters of Ala.

feasible to parametrize all rotable torsions in an all-atom force field based on free energy surfaces instead of potential energy surfaces.

During the parametrization, we found that not scaling 1-4 electrostatic interactions while significantly reducing 1-4 van der Waals (L-J) interaction is a good choice. Also, adding only three 1-5 L-J interactions ($H_i\cdots O_i$, $H_i\cdots N_{i+1}$, $O_{i-1}\cdots C_i$), which are the same for all AAs, is enough for the coupling between backbone $\phi$ and $\psi$ torsions. For Asp, Asn, Ser, and Thr, modifications of local polar interactions such as additional O···C attraction may also be necessary.


dx.doi.org/10.1021/jp5017449 | J. Phys. Chem. B 2014, 118, 6983−6998

We show that RSFF1 gives significantly improved simulation results for a variety of peptides and proteins. It well reproduces the NMR $J$ coupling constants of AA dipeptides, better than its parent OPLS-AA/L and some recent force fields (AMBER ff99B*-ildn, ff99SB-ildn-NMR, and CHARMM22*). The different intrinsic conformational preferences of various AA residues cannot be fully captured using a single set of backbone parameters. RSFF1 can also consistently fold a set of peptides and proteins including both $\alpha$-helix (Ac-Ala14-NHMe, Trp-cage, Homeodomain) and $\beta$-sheet (Trpzip-2, GB1 hairpin) ones, with similar overstabilization. In comparison, other force fields cannot correctly fold all of them simultaneously. This indicates that RSFF1 not only achieves a good balance between $\alpha$-helical and $\beta$-sheet structures but also is transferable among different sequences. Indeed, RSFF1 can successfully fold all of the 12 small fast-folding proteins recently studied by Lindorff-Larson et al.,[3] which will be reported elsewhere.

## ■ ASSOCIATED CONTENT

### Supporting Information

The detailed methods for the folding simulations, various free energy surfaces from the coil library, and RSFF1 simulations (Figure S1−S6), and all new torsion parameters in the RSFF1 force field (Tables S1, S2). This material is available free of charge via the Internet at http://pubs.acs.org. The implementation of RSFF1 in *Gromacs* will be provided upon request.

## ■ AUTHOR INFORMATION

### Corresponding Authors

*F. Jiang: e-mail, jiangfan@pku.edu.cn.
*Y.-D. Wu: e-mail, wuyd@pkusz.edu.cn.

### Notes

The authors declare no competing financial interest.

### Biographies

Fan Jiang received his B.Sc. and Ph.D. from Peking University (PKU) in 2003 and 2008, respectively. He worked in Yun-Dong Wu's lab both as a student at PKU and as a research associate (2008−2010) at HKUST. He is currently a research staff in PKU Shenzhen Graduate School. His current research focuses on developing peptide/protein simulation and structure prediction methods, especially on the efforts to merge physical-based and knowledge-based approaches.

Chen-Yang Zhou is a Ph.D. candidate in the College of Chemistry at Peking University working with Prof. Yun-Dong Wu. He received his B.Sc. in Chemistry from Peking University in 2010. His current research focuses on parametrization and validation of protein force fields.

Yun-Dong Wu is currently Chair Professor of Chemistry at Peking University. He is a member of the Chinese Academy of Science. He received his B.Sc. in chemistry from Lanzhou University in 1981, and his Ph.D. from University of Pittsburgh in 1986. After postdoctoral research with Prof. K. N. Houk at UCLA, he joined the faculty at the HongKong University of Science & Technology, becoming Chair Professor in 2007. His research focuses on understanding the mechanisms of organic reactions, molecular design with peptides, modeling of protein folding, and protein−protein interactions.

## ■ ACKNOWLEDGMENTS

We are grateful for the financial supports from the National Natural Science Foundation of China (Grant No. 21133002 for Y.-D.W. and 21203004 for F.J.), the Shenzhen Peacock Program (KQTD201103), and Peking University Shenzhen Graduate School.

## ■ REFERENCES

(1) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646−652.
(2) Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **2012**, *41*, 429−452.
(3) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517−520.
(4) Mirjalili, V.; Feig, M. Protein Structure Refinement through Structure Selection and Averaging from Molecular Dynamics Ensembles. *J. Chem. Theory Comput.* **2013**, *9*, 1294−1303.

(5) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9*, 71.
(6) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the Accuracy of Physical Models Used in Protein-Folding Simulations: Quantitative Evidence from Long Molecular Dynamics Simulations. *Curr. Opin. Struc. Biol.* **2014**, *24*, 98−105.
(7) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A Biomolecular Force Field Based on The Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.
(8) Jorgensen, W. L.; Tirado-Rives, J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6665−6670.
(9) Xu, Z.; Luo, H. H.; Tieleman, D. P. Modifying The OPLS-AA Force Field to Improve Hydration Free Energies for Several Amino Acid Side Chains Using New Atomic Charges and an Off-Plane Charge Model for Aromatic Residues. *J. Comput. Chem.* **2007**, *28*, 689−697.
(10) Tong, Y.; Mei, Y.; Li, Y. L.; Ji, C. G.; Zhang, J. Z. Electrostatic Polarization Makes a Substantial Contribution to the Free Energy of Avidin−Biotin Binding. *J. Am. Chem. Soc.* **2010**, *132*, 5137−5142.
(11) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. X-Pol Potential: An Electronic Structure-Based Force Field for Molecular Dynamics Simulation of a Solvated Protein in Water. *J. Chem. Theory. Comput.* **2009**, *5*, 459−467.
(12) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr.; et al. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.
(13) Wang, J.; Cieplak, P.; Li, J.; Wang, J.; Cai, Q.; Hsieh, M.; Lei, H.; Luo, R.; Duan, Y. Development of Polarizable Models for Molecular Mechanical Calculations II: Induced Dipole Models Significantly Improve Accuracy of Intermolecular Interaction Energies. *J. Phys. Chem. B* **2011**, *115*, 3100−3111.
(14) 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. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(15) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(16) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(17) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. 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.
(18) Mackerell, A. D.; Feig, M.; Brooks, C. L. 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.
(19) MacKerell, A. D.; Feig, M.; Brooks, C. L. Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(20) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(21) 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.
(22) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99sb Protein Force Field. *Proteins* **2010**, *78*, 1950−1958.
(23) Hu, H.; Elstner, M.; Hermans, J. Comparison Of A QM/MM Force Field and Molecular Mechanics Force Fields in Simulations of Alanine and Glycine "Dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in Water in Relation to the Problem of Modeling the Unfolded Peptide Backbone in Solution. *Proteins* **2003**, *50*, 451−463.
(24) Okur, A.; Strockbine, B.; Hornak, V.; Simmerling, C. Using PC Clusters to Evaluate the Transferability of Molecular Mechanics Force Fields for Proteins. *J. Comput. Chem.* **2003**, *24*, 21−31.
(25) Yoda, T.; Sugita, Y.; Okamoto, Y. Secondary-Structure Preferences of Force fields for Proteins Evaluated by Generalized-Ensemble Simulations. *Chem. Phys.* **2004**, *307*, 269−283.
(26) Wang, T.; Wade, R. C. Force Field Effects on a $\beta$-Sheet Protein Domain Structure in Thermal Unfolding Simulations. *J. Chem. Theory Comput.* **2006**, *2*, 140−148.
(27) Todorova, N.; Legge, F. S.; Treutlein, H.; Yarovsky, I. Systematic Comparison of Empirical Forcefields for Molecular Dynamic Simulation of Insulin. *J. Phys. Chem. B* **2008**, *112*, 11137−11146.
(28) Freddolino, P. L.; Park, S.; Roux, B.; Schulten, K. Force Field Bias in Protein Folding Simulations. *Biophys. J.* **2009**, *96*, 3772−3780.
(29) Matthes, D.; de Groot, B. L. Secondary Structure Propensities in Peptide Folding Simulations: A Systematic Comparison of Molecular Mechanics Interaction Schemes. *Biophys. J.* **2009**, *97*, 599−608.
(30) Vymětal, J.; Vondrášek, J. Metadynamics As a Tool for Mapping the Conformational and Free-Energy Space of Peptides - The Alanine Dipeptide Case Study. *J. Phys. Chem. B* **2010**, *114*, 5632−5642.
(31) Project, E.; Nachliel, E.; Gutman, M. Force Field-Dependant Structural Divergence Revealed During Long Time Simulations of Calbindin D9k. *J. Comput. Chem.* **2010**, *31*, 1864−1872.
(32) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix−Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.
(33) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47−L49.
(34) Mittal, J.; Best, R. B. Tackling Force-Field Bias in Protein Folding Simulations: Folding of Villin HP35 and Pin WW Domains in Explicit Water. *Biophys. J.* **2010**, *99*, L26−L28.
(35) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. J. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone $\phi$, $\psi$ and Side-Chain $\chi_1$ and $\chi_2$ Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.
(36) Nerenberg, P. S.; Head-Gordon, T. Optimizing Protein-Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *J. Chem. Theory Comput.* **2011**, *7*, 1220−1230.
(37) Li, D.; Brüschweiler, R. NMR-Based Protein Potentials. *Angew. Chem., Int. Ed.* **2010**, *49*, 6778−6780.
(38) Nymeyer, H. Energy Landscape of the Trpzip2 Peptide. *J. Phys. Chem. B* **2009**, *113*, 8288−8295.
(39) Day, R.; Paschek, D.; Garcia, A. E. Microsecond Simulations of the Folding/Unfolding Thermodynamics of the Trp-Cage Miniprotein. *Proteins* **2010**, *78*, 1889−1899.
(40) Vymetal, J.; Vondrasek, J. Critical Assessment of Current Force Fields. Short Peptide Test Case. *J. Chem. Theory Comput.* **2013**, *9*, 441−451.
(41) Jiang, F.; Han, W.; Wu, Y. D. Influence of Side Chain Conformations on Local Conformational Features of Amino Acids and Implication for Force Field Development. *J. Phys. Chem. B* **2010**, *114*, 5840−5850.
(42) Jiang, F.; Han, W.; Wu, Y. D. The Intrinsic Conformational Features of Amino Acids from a Protein Coil Library and Their Applications in Force Field Development. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3413−3428.

(43) Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. Intrinsic $\phi$, $\psi$ Propensities of Amino Acids, Derived from the Coil Regions of Known Structures. *Nat. Struct. Biol.* **1995**, *2*, 596−603.
(44) Serrano, L. Comparison between the $\phi$ Distribution of the Amino Acids in the Protein Database and NMR Data Indicates that Amino Acids have Various $\phi$ Propensities in the Random Coil Conformation. *J. Mol. Biol.* **1995**, *254*, 322−333.
(45) Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. Toward a Description of the Conformations of Denatured States of Proteins. Comparison of a Random Coil Model with NMR Measurements. *J. Phys. Chem.* **1996**, *100*, 2661−2666.
(46) Penkett, C. J.; Redfield, C.; Dodd, I.; Hubbard, J.; McBay, D. L.; Mossakowska, D. E.; Smith, R.; Dobson, C. M.; Smith, L. J. NMR Analysis of Main-Chain Conformational Preferences in an Unfolded Fibronectin-Binding Protein. *J. Mol. Biol.* **1997**, *274*, 152−159.
(47) Avbelj, F.; Baldwin, R. L. Role of Backbone Solvation and Electrostatics in Generating Preferred Peptide Backbone Conformations: Distributions of phi. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 5742−5747.
(48) Fitzkee, N. C.; Fleming, P. J.; Rose, G. D. The Protein Coil Library: A Structural Database of Nonhelix, Nonstrand Fragments Derived from the PDB. *Proteins* **2005**, *58*, 852−854.
(49) Jha, A. K.; Colubri, A.; Zaman, M. H.; Koide, S.; Sosnick, T. R.; Freed, K. F. Helix, Sheet, and Polyproline II Frequencies and Strong Nearest Neighbor Effects in a Restricted Coil Library. *Biochemistry* **2005**, *44*, 9691−9702.
(50) Avbelj, F.; Grdadolnik, S. G.; Grdadolnik, J.; Baldwin, R. L. Intrinsic Backbone Preferences Are Fully Present in Blocked Amino Acids. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 1272−1277.
(51) Poon, C.; Samulski, E. T.; Weise, C. F.; Weisshaar, J. C. Do Bridging Water Molecules Dictate the Structure of a Model Dipeptide in Aqueous Solution? *J. Am. Chem. Soc.* **2000**, *122*, 5642−5643.
(52) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. Polyproline II Structure in a Sequence of Seven Alanine Residues. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 9190−9195.
(53) Shi, Z.; Chen, K.; Liu, Z.; Kallenbach, N. R. Conformation of the Backbone in Unfolded Proteins. *Chem. Rev.* **2006**, *106*, 1877−1897.
(54) Grdadolnik, J.; Golič Grdadolnik, S.; Avbelj, F. Determination of Conformational Preferences of Dipeptides Using Vibrational Spectroscopy. *J. Phys. Chem. B* **2008**, *112*, 2712−2718.
(55) Schweitzer-Stenner, R. Distribution of Conformations Sampled by the Central Amino Acid Residue in Tripeptides Inferred from Amide I Band Profiles and NMR Scalar Coupling Constants. *J. Phys. Chem. B* **2009**, *113*, 2922−2932.
(56) Grdadolnik, J.; Mohacek-Grosev, V.; Baldwin, R. L.; Avbelj, F. Populations of the Three Major Backbone Conformations in 19 Amino Acid Dipeptides. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 1794−1798.
(57) McGregor, M. J.; Islam, S. A.; Sternberg, M. J. Analysis of the Relationship between Side-Chain Conformation and Secondary Structure In Globular Proteins. *J. Mol. Biol.* **1987**, *198*, 295−310.
(58) Chakrabarti, P.; Pal, D. The Interrelationships of Side-Chain and Main-Chain Conformations in Proteins. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1−102.
(59) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(60) Thomas, P. D.; Dill, K. A. An Iterative Method for Extracting Energy-Like Quantities from Protein Structures. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 11628−11633.
(61) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.
(62) Prakash, M. K.; Barducci, A.; Parrinello, M. Replica Temperatures for Uniform Exchange and Efficient Roundtrip Times in Explicit Solvent Parallel Tempering Simulations. *J. Chem. Theory Comput.* **2011**, *7*, 2025−2027.
(63) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 14101.
(64) Lin, I. C.; Tuckerman, M. E. Enhanced Conformational Sampling of Peptides via Reduced Side-Chain and Solvent Masses. *J. Phys. Chem. B* **2010**, *114*, 15935−15940.
(65) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(66) Berman, H.; Henrick, K.; Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10*, 980.
(67) Amir, E. D.; Kalisman, N.; Keasar, C. Differentiable, Multi-Dimensional, Knowledge-Based Energy Terms for Torsion Angle Probabilities and Propensities. *Proteins* **2008**, *72*, 62−73.
(68) Smith, J. C.; Karplus, M. Empirical Force Field Study of Geometries and Conformational Transitions of Some Organic Molecules. *J. Am. Chem. Soc.* **1992**, *114*, 801−812.
(69) Arnautova, Y. A.; Jagielska, A.; Scheraga, H. A. A New Force Field (ECEPP-05) for Peptides, Proteins, and Organic Molecules. *J. Phys. Chem. B* **2006**, *110*, 5025−5044.
(70) Pal, T. K.; Sankararamakrishnan, R. Quantum Chemical Investigations on Intraresidue Carbonyl-Carbonyl Contacts in Aspartates of High-Resolution Protein Structures. *J. Phys. Chem. B* **2010**, *114*, 1038−1049.
(71) Han, W.; Wu, Y.-D. Coarse-Grained Protein Model Coupled with a Coarse-Grained Water Model: Molecular Dynamics Study of Polyalanine-Based Peptides. *J. Chem. Theory Comput.* **2007**, *3*, 2146−2161.
(72) Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N. n→$\pi$* Interactions in Proteins. *Nat. Chem. Biol.* **2010**, *6*, 615−620.
(73) Pardi, A.; Billeter, M.; Wuthrich, K. Calibration of the Angular Dependence of the Amide Proton-C$^{\alpha}$ Proton Coupling Constants, $^{3}J_{HN\alpha}$, in a Globular Protein. Use of $^{3}J_{HN\alpha}$ for Identification of Helical Secondary Structure. *J. Mol. Biol.* **1984**, *180*, 741−751.
(74) Ludvigsen, S.; Andersen, K. V.; Poulsen, F. M. Accurate Measurements of Coupling Constants from Two-Dimensional Nuclear Magnetic Resonance Spectra of Proteins and Determination of $\phi$-Angles. *J. Mol. Biol.* **1991**, *217*, 731−736.
(75) Vuister, G. W.; Bax, A. Quantitative *J* Correlation: A New Approach for Measuring Homonuclear Three-Bond $J(H^{N}H^{\alpha})$ Coupling Constants in 15N-Enriched Proteins. *J. Am. Chem. Soc.* **1993**, *115*, 7772−7777.
(76) Hu, J.; Bax, A. Determination of $\phi$ and $\chi_1$ Angles in Proteins from $^{13}$C−$^{13}$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.
(77) Schmidt, J. M.; Blumel, M.; Lohr, F.; Ruterjans, H. Self-Consistent $^{3}J$ Coupling Analysis for the Joint Calibration of Karplus Coefficients and Evaluation of Torsion Angles. *J. Biomol. NMR* **1999**, *14*, 1−12.
(78) Case, D. A.; Scheurer, C.; Brüschweiler, R. Static and Dynamic Effects on Vicinal Scalar J Couplings in Proteins and Peptides: A MD/DFT Analysis. *J. Am. Chem. Soc.* **2000**, *122*, 10390−10397.
(79) Vogeli, B.; Ying, J.; Grishaev, A.; Bax, A. Limits on Variations in Protein Backbone Dynamics from Precise Measurements of Scalar Couplings. *J. Am. Chem. Soc.* **2007**, *129*, 9377−9385.
(80) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-Residue Protein. *Nat. Struct. Biol.* **2002**, *9*, 425−430.
(81) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Tryptophan Zippers: Stable, Monomeric $\beta$-Hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 5578−5583.
(82) Blanco, F. J.; Rivas, G.; Serrano, L. A Short Linear Peptide That Folds into a Native Stable $\beta$-Hairpin in Aqueous Solution. *Nat. Struct. Biol.* **1994**, *1*, 584−590.
(83) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS One* **2012**, *7*, e32131.
(84) Clarke, N. D.; Kissinger, C. R.; Desjarlais, J.; Gilliland, G. L.; Pabo, C. O. Structural Studies of the Engrailed Homeodomain. *Protein Sci.* **1994**, *3*, 1779−1787.

(85) Matthes, D.; de Groot, B. L. Secondary Structure Propensities in Peptide Folding Simulations: A Systematic Comparison of Molecular Mechanics Interaction Schemes. *Biophys. J.* **2009**, *97*, 599−608.
(86) Best, R. B.; Mittal, J. Free-Energy Landscape of the GB1 Hairpin in All-Atom Explicit Solvent Simulations with Different Force Fields: Similarities and Differences. *Proteins* **2011**, *79*, 1318−1328.
(87) Muñoz, V.; Serrano, L. Development of the Multiple Sequence Approximation within the AGADIR Model of $\alpha$-helix Formation: Comparison with Zimm-Bragg and Lifson-Roig Formalisms. *Biopolymers* **1997**, *41*, 495−509.