JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

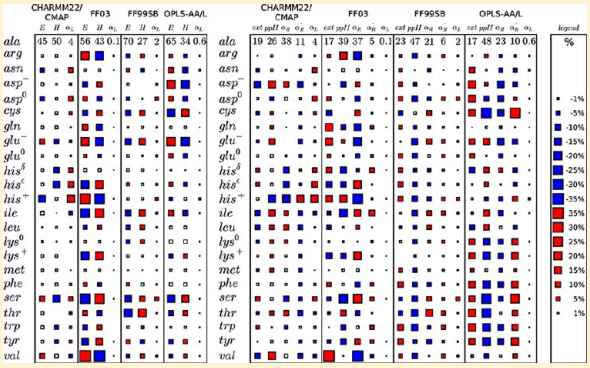
Critical Assessment of Current Force Fields. Short Peptide Test Case

Jiří Vymětal and Jiří Vondrášek*

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, 166 10 Prague 6, Czech Republic

Supporting Information

ABSTRACT: The applicability of molecular dynamics simulations for studies of protein folding or intrinsically disordered proteins critically depends on quality of energetic functions—force fields. The four popular force fields for biomolecular simulations, CHARMM22/CMAP, AMBER FF03, AMBER FF99SB, and OPLS-AA/L, were compared in prediction of conformational propensities of all common proteinogenic amino acids. The minimalistic model of terminally block amino acids (dipeptides) was chosen for assessment of side chain effects on backbone propensities. The precise metadynamics simulations revealed striking inconsistency of trends in conformational preferences as manifested by investigated force fields for both backbone and side chains. To trace this disapproval between force fields, the two related AMBER force fields were studied more closely. In the cases of FF99SB and FF03, we uncovered that the distinct tends were driven by different charge models. Additionally, the effects of recent correction for side chain torsion (FF99SB-ILDN) were examined on affected amino acids and exposed significant coupling between free energy profiles and propensities of backbone and side chain conformers. These findings have important consequences for further force field development.



INTRODUCTION

Atomistic molecular dynamics simulations provide invaluable insight in various biological phenomena but the center of interest is still a process of protein folding.[1,2] Due to the enormous progress in computational performance in the field of computer simulations nowadays, the highly resolved in silico experiments can be performed routinely in microsecond time scales and even the millisecond frontier was reached recently using specialized hardware.[3] These time scales allow the studying of a folding mechanism of fast folding protein domains[4] or the detailed characterization of unfolded/denatured states.[5] However, the quality of used energetic functions (force fields) might be the limiting factor for the utilization of simulations as a complement to experimental studies.[6]

It is well-known that the precise balance between propensities for helical and extended structures is an essential property of any force field intended for simultaneous description of folding α-helical, β-sheet, or mixed protein architectures,[6] as well as their unfolded states. As was documented in the not so distant past, the simulations of short alanine-based peptides has revealed overrated helical tendencies of current force fields.[7] The another evidence for imperfections in the parametrization followed from inability to predict correct kinetics and thermodynamics of α-helix formation.[8−10] These and further shortcomings resulted in the design of simple corrections to the latest generation of force fields. The backbone torsions of FF03 and FF99SB were reparametrized (FF03* and FF99SB*) for correct description of helix folding.[9] Later, the FF03w[11] correction was redesigned by the same authors for simulations with superior water model. The distinct corrections to FF99SB force field were elaborated recently for better reproduction of conformational behavior of small peptides (FF99SBϕ')[12] or proteins in solution (FF99SB$\phi\psi$)[13] based on NMR experimental data. The CHARMM22*[6] correction was designed to improve the helix−coil transition and to remove bias in folding simulations for CHARMM22 force field. All these previously mentioned corrections modified only a small amount of parameters in the initial force field, namely, the torsion parameters describing energetics of peptide backbone. Still, these changes brought improvement in reproduction of NMR observables of protein and peptides in comparison with the original parametrizations.[14,15]

In contrast to the backbone φ and ψ torsions, substantially less attention was paid to the rotamers of side chains. Parameters for these torsions were mostly overtaken from small model compounds and not accommodated for utilization in peptides or proteins. The exception: OPLS-AA/L involves reparametrized side chain torsions based on ab initio quantum calculation on small peptides.[16] The need for correction of several side chain torsions was also recognized in FF99SB force fields, resulting in the FF99SB-ILDN[17] variant. However, it had been shown that distributions of rotamer conformers predicted on model dipeptides and tetrapeptides by OPLS-AA/L and FF03 differ and also deviate from distributions observed in protein crystal structures in Protein Data Bank (PDB).[18]

Received: September 11, 2012

   dx.doi.org/10.1021/ct300794a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

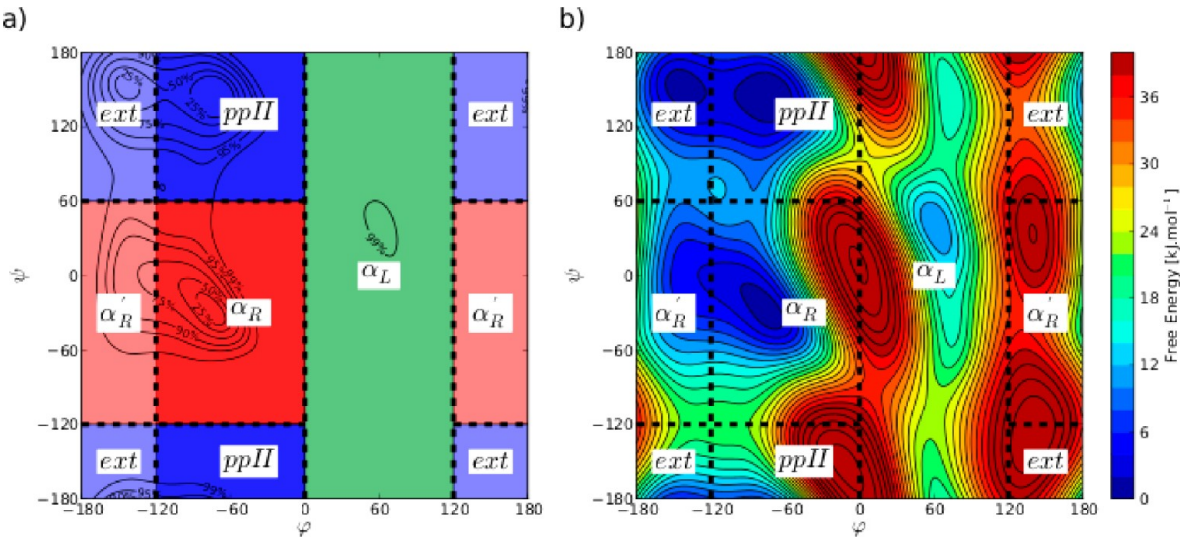


Figure 1. Definition of backbone conformers. The definition of backbone conformers were chosen according to the free energy profiles of dipeptides in terms of φ and ψ torsions. The 5-state models involves regular rectangular regions ext, ppII, α_R, and α_R'. Parts a and b show these regions as Ramachandran-like plot of "a typical dipeptide" with contours describing probability of occurrence (a) or in terms of free energy (b). The profile of the typical dipeptide represents an average profile of all dipeptides (except Gly and Pro) obtained in CHARMM22/CMAP, FF03, FF99SB, and OPLS-AA/L force fields. The 5-state model could not be optimal in same cases because individual force fields and dipeptides differ in features of highly populated regions. The border between ext and ppII (as well as α_R and α_R') is ambiguous among dipeptides and sometimes not differentiated. Therefore, the more general 3-state model with regions E (ext+ppII, in blue), H (α_R+α_R', in red), and α_L (in green) were also utilized.

The presence of a side chain induces specific structural propensity for each amino acid to secondary structure elements in proteins and even in short peptides.[19,20] The intrinsic propensities are still present in the smallest conceivable model systems—amino acid dipeptides—as were manifested by NMR and vibrational spectroscopy experiments.[21−23] These delicate conformational preferences may play important role in course of protein folding or in structural properties of denatured/unfolded state of proteins. The ability of force fields to reproduce intrinsic propensities of different amino acids and peptides or protein fragments of different composition is critical for their successful application for previously mentioned phenomena.

Despite of popularity of alanine dipeptide as the model compound in molecular modeling the detailed computational studies for other amino acids are rare as well as systematic studies of intrinsic propensity scales. Propensities of all amino acids were investigated in GGXGG host peptide using in lucem molecular mechanics (ilmm) and compared against distributions in PDB and long simulations of different proteins.[24] The conformational preferences of terminally blocked amino acids (Ace-X-Nme) were evaluated for CHARMM force field[25] and also OPLS-AA/L and FF03[18] by means of classical molecular dynamics. The different computational approach were utilized recently for assessment of propensities of amino acids in OPLS-AA/L force field. Instead of direct evaluation of frequency of different conformers in simulation the same quantities were derived from free energy profiles obtained by extensive umbrella sampling.[26]

As follows from the above-mentioned facts and references the critical assessment of the force fields is absolutely necessary condition in mapping realistic conformational space of peptides, their structural preferences and barriers between their minima. The resulting free energy profiles are in such case reliable measure not only for the structural preferences, but more importantly, they can provide information about mutual coupling in peptide torsion space. The critical point then is the obtaining of such profile in a reasonable time and with reasonable accuracy.

It is well-known that the free energy profiles can be achieved by more efficient biased free energy calculation method such as metadynamics.[27] In one of our papers, we had already shown that in the case of alanine dipeptide, metadynamics is viable tool for mapping the conformational and free energy space of peptides.[28] Utilizing the experience of our previous work, we continue in our effort to study the intrinsic propensities of all 20 natural amino acids and compare a performance of 4 different force fields largely used in simulations of biomolecules. Most importantly, we address a question of whether the investigated force fields are able to predict coherently any trends in propensities of terminally blocked amino acids.

METHODS

Model Systems. The model amino acid dipeptides (amino acid residues terminated by acetyl and N-methylamide group) were built by the tleap program from AMBERtools package.[29] The following amino acids and their protonation states were investigated in this study: alanine (ala), charged arginine (arg), asparagine (asn), charged aspartate (asp^-), neutral aspartate (asp^0), cysteine (cys), glutamine (gln), charged glutamate (glu^-), neutral glutamate (glu^0), glycine (gly), δ-protonated histidine (his^δ), ε-protonated histidine (his^ε), charged histidine (his^+), isoleucine (ile), leucine (leu), charged lysine (lys^+), neutral lysine (lys^0), methionine (met), phenylalanine (phe), proline (pro), serine (ser), threonine (thr), tryptophan (trp), tyrosine (tyr), and valine (val). Free energy profiles and propensities of this 25 model dipeptides were investigated in four different force fields: CHARMM22/CMAP,[30] AMBER FF03,[31] AMBER FF99SB,[32] and OPLS-AA/L.[16] The topology and force field parameters were assigned by pdb2gmx from parameter database provided and contributed in Gromacs package.[8,33,34] Model dipeptides were solvated by approximately 880 TIP3P water molecules in cubic boxes of single size 30 × 30 × 30 A. The net charge of

dx.doi.org/10.1021/ct300794a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

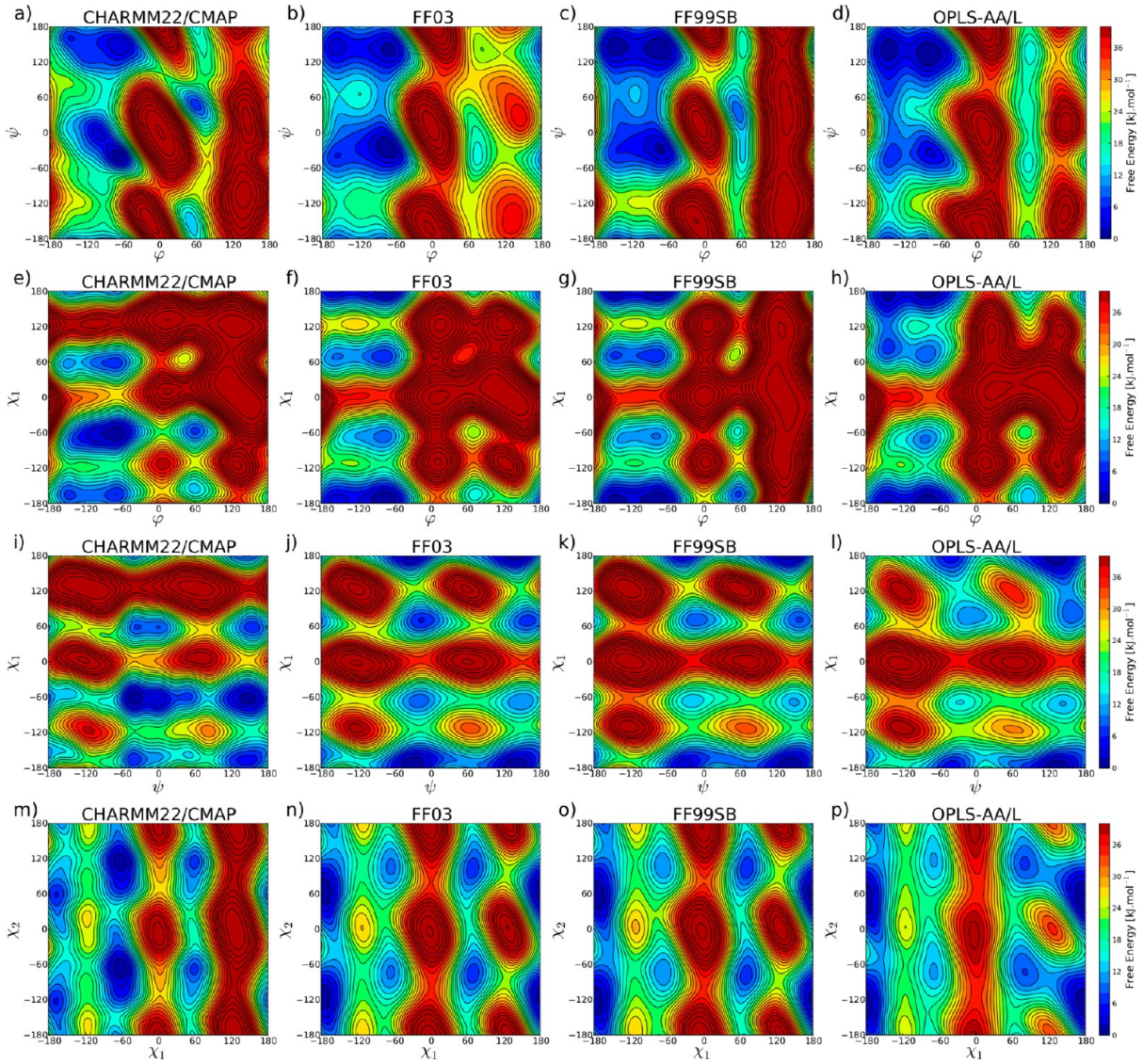


Figure 2. Free energy profiles of asp$^-$. The two-dimensional free energy profiles of asp$^-$ as obtained by metadynamics for CHARMM22/CMAP, FF03, FF99SB, and OPLS-AA/L force field in terms of φ/ψ (a–d), φ/χ_1 (e–h), ψ/χ_1 (i–l), and χ_1/χ_2 (m–p).

systems containing charged amino acids were neutralized by the addition of chlorine or sodium counterions. The geometries of solvated systems were subsequently optimized by L-BFGS minimizer. The position and velocities of water molecules were refined in short 0.5 ns equilibration simulation for temperature of 300 K and pressure of 1 bar.

Metadynamics and Collective Variables. All simulations were performed by means of Gromacs molecular dynamics package (version 4.5.3)[33] enhanced by Plumed plug-in[35] for metadynamics (version 1.2.2). Well-tempered version of metadynamics[36] together with bias-exchange algorithm[37] were used for extending conformational sampling and improving convergence.

The peptide backbone torsions (φ and ψ) and side chain torsions (χ_1 and χ_2) were chosen as collective variables (CV) for metadynamics. Because amino acid possess the side chains of different size and complexity, the side chain torsions χ_1 or χ_2 are not defined for all amino acids. Therefore, three bias-exchange protocols distinctive in used CV were designed for three classes of amino acids. The definition of torsions and CV used in each protocol are summarized in the Supporting Information (Tables S1.1 and S1.2). Each metadynamics protocol involves both one-dimensional and two-dimensional CVs, which represent individual torsion or couples of torsions.

All replicas in bias-exchange metadynamics simulation were conducted for 60 ns. The exchange between replicas was attempted each 4 ps. Bias potential was updated each 2 ps by Gaussian function with width (σ parameter) 0.2 and 0.3 rad for one-dimensional and two-dimensional CVs, respectively. The initial height of Gaussians 0.2 kJ·mol^{-1} was gradually decreased according to the well-tempered schema for bias factor $(T + \Delta T)/T = 10$. The values of all CVs in all replicas were deposited each 0.2 ps for further analysis.

Simulation Details. The equations of motions were integrated with time step of 2 fs. All bond lengths were constrained by the LINCS algorithm. The geometry of water molecules was constrained by the SETTLE algorithm. The temperature of 300 K was maintained by v-rescale thermostat with coupling constant 1.0 ps^{-1}. Berendsen's weak coupling barostat was utilized for pressure control with coupling constant 1.0 ps^{-1}. Electrostatic interactions were treated by Particle Mesh Ewald algorithm with real-space cutoff 10 Å and FFT spacing 10

dx.doi.org/10.1021/ct300794a | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

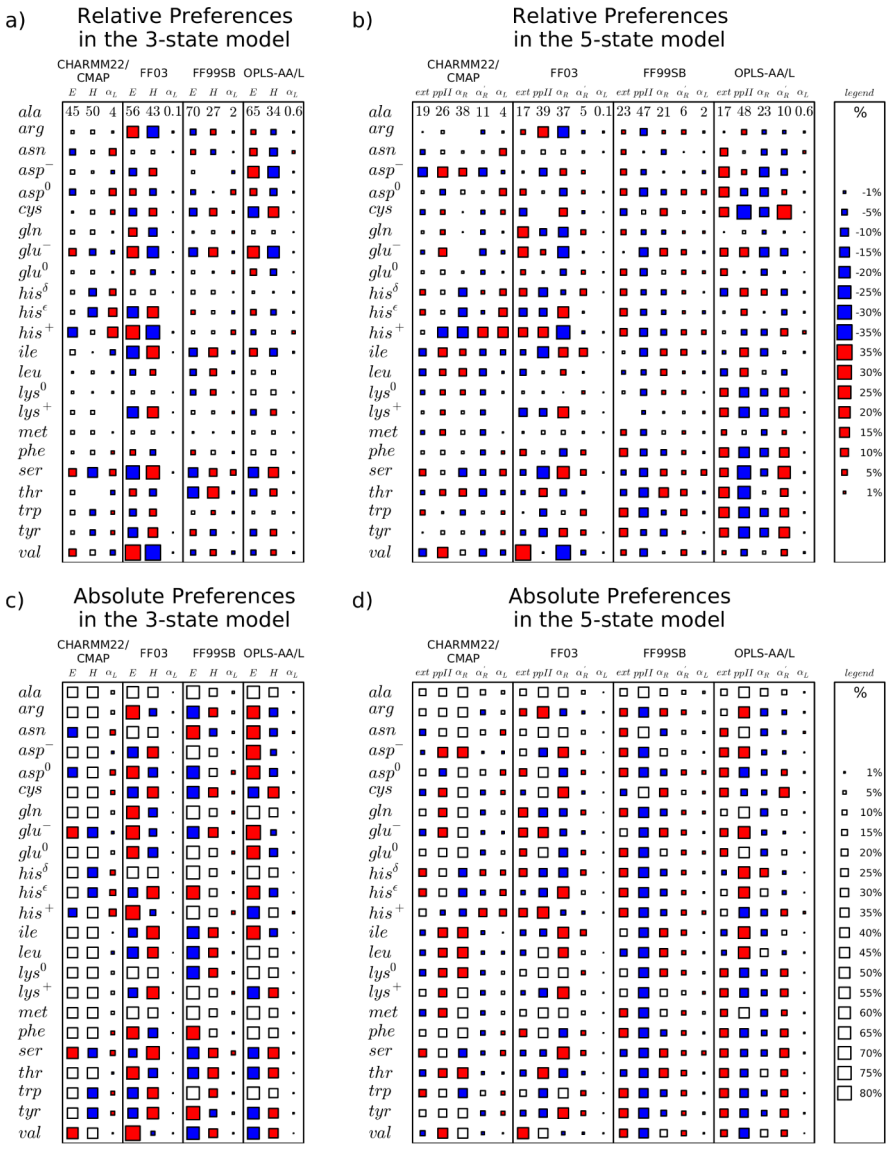


Figure 3. Relative population of backbone conformers. Relative and absolute backbone propensities of dipeptides are expressed by Hinton plots. Relative propensities were calculated as difference between populations (in %) of corresponding backbone conformers of given dipeptide and alanine dipeptide. Absolute populations of alanine dipeptide are listed in the first row of the tables. For other dipeptides, the difference from alanine dipeptide is represented by size of square and its color (less than alanine dipeptide in blue, more in red). Colorless squares denote statistically insignificant differences (the differences smaller than 2σ). Tables a and b report relative propensities for 3-state and 5-state model, respectively. Tables c and d report absolute propensities for 3-state and 5-state model, respectively.

Å . The nonbonded Lennard-Jones interactions were evaluated with simple cutoff of 14 Å accompanied by correction to energy and pressure tensor.

Definition of Regions. A strict definition of discrete backbone and side chain conformers was necessary for analysis of conformational preferences. Two models of backbone conformers were proposed based on φ and ψ torsions resulting in the 5-state and 3-state model. The comprehensive definition of regions on Ramachandran plot is provided in Figure 1.

Three rotamer states were discriminated by side chain torsion χ_1 and additional three or two states by torsion χ_2. The distinguished rotamers are listed in the Supporting Information (Table S1.3) with proper definition in terms of corresponding torsions.

Data Postprocessing and Error Analysis. The bias potential constructed in course of the metadynamics was recorded continuously as a list of parameters for Gaussians being summed to the total bias potential. The free energy profiles in individual CVs were reconstructed from the stored bias potential by sumhills utility provided in Plumed package using grids with resolution 500 points or 500 × 500 points for one- and two-dimensional CVs, respectively. The free energy profiles were analyzed after each addition of 1000 Gaussians to the bias potential. The average free energy profiles corresponding to the last 40 ns of simulations were reported as the final free energy profiles.

The population of the individual regions were calculated using the histogram reweighting method for well-tempered metadynamics.[38] This approach allows estimation of unbiased histograms of all investigated torsions from all replicas regardless which CV was biased in individual replica. The populations of backbone and side chain conformers were evaluated employing algorithm from ref 38 with bin size of 4° for all histograms. The calculations were performed for each replica providing

dx.doi.org/10.1021/ct300794a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Correlation of Propensities for Backbone Regions between Force Fields[a]

conformer	CHARMM22/CMAP vs FF03	CHARMM22/CMAP vs FF99SB	CHARMM22/CMAP vs OPLS-AA/L	FF03 vs FF99SB	FF03 vs OPLS-AA/L	FF99SB vs OPLS-AA/L
E	−0.16 ± 0.10	−0.26 ± 0.09	−0.03 ± 0.10	−0.07 ± 0.04	0.14 ± 0.04	0.16 ± 0.05
H	−0.13 ± 0.12	0.18 ± 0.10	−0.11 ± 0.11	0.03 ± 0.05	0.16 ± 0.04	0.13 ± 0.05
Ex	0.03 ± 0.06	0.37 ± 0.06	−0.12 ± 0.05	0.39 ± 0.04	−0.10 ± 0.04	0.08 ± 0.05
ppII	−0.04 ± 0.07	−0.09 ± 0.07	0.13 ± 0.05	−0.16 ± 0.06	0.03 ± 0.04	0.14 ± 0.05
α_R	0.00 ± 0.07	0.52 ± 0.06	−0.03 ± 0.07	0.18 ± 0.04	−0.16 ± 0.03	0.01 ± 0.04
α_R'	0.08 ± 0.04	−0.11 ± 0.04	0.01 ± 0.04	−0.00 ± 0.05	0.06 ± 0.03	0.20 ± 0.05
α_L	0.33 ± 0.08	0.54 ± 0.03	0.53 ± 0.04	0.49 ± 0.08	0.44 ± 0.10	0.67 ± 0.06

[a]Correlation between propensities of dipeptides for individual region measured by Spearman correlation coefficient. The uncertainty of reported values is expressed as a standard deviation.

reweighted histograms of φ, ψ, χ_1, and χ_2 torsions. The differences in histograms obtained from all replicas were used for estimation of uncertainty in final results.

The assumption that all replicas provided independent estimates of histograms were utilized for estimation of uncertainty in other derived values. The uncertainties in differences of conformational populations (relative propensities) were determined by simple permutation test. For the comparison of two populations, the difference were calculated for every possible combination of values coming from distinct replicas. The corresponding arithmetic average and standard deviation were reported as final relative propensity and its uncertainty.

The estimates of Spearman correlation coefficients were performed in the same fashion as relative propensities. Because of the astronomical number of permutations involving all replicas for all amino acids, only 10 000 permutations were picked randomly and evaluated. The values of the Spearman correlation coefficients were calculated by Scientific Tools for Python (SciPy.stats module).

All plots and other graphics were elaborated by means of the matplotlib library.

RESULTS

To illustrate the typical output of the metadynamic simulation (free energy profile), we used the aspartate dipeptide (asp^-)—a typical representant of a complex amino acid. The free energy profiles are shown in Figure 2. The two-dimensional free energy profiles in φ/ψ, φ/χ_1, ψ/χ_1, and χ_1/χ_2 (in rows) are presented for all four investigated force fields in different columns. The results of simulations of individual dipeptides are found in the Supporting Information (S2). We use the results in the following analysis.

The φ/ψ free energy profiles of different dipeptides (except Pro and Gly) share large amount of similarity within all four individual force fields. In case of the φ/ψ profiles (see Figure 2a−d), each force field manifests typical features present for all dipeptides with side chains. CHARMM22/CMAP exhibits 2 minima in α_R region, whereas the other force fields (FF03, FF99SB,OPLS-AA/L) differentiated more or less one minimum in α_R and the other in α_R' region. Another variation was observed in α_L region, which involves the canonical α_L conformer and 2 or 3 other distinctive minima. However, these regions are high in free energy and therefore populated rarely in classical molecular dynamics of dipeptides.

The qualitatively similar features were found for the free energy profiles in terms of φ/χ_1 and ψ/χ_1 torsions. All force fields and all dipeptides possessed by side-chains provided profiles resembling closely those in Figure 2e−h and i−l. The differences between individual amino acids and force fields was observed mainly in relative free energies of corresponding minima and hence in diverse population of conformers.

The free energy profiles in χ_1/χ_2 torsions reflected the hybridization state of the γ-atom in side chain of dipeptide. Nine rotamers were distinguished clearly for sp^3-hybridized C_γ atom as a result of steric hindrances with sp^3-hybridized C_β. On the other hand, six low energy regions were formed dipeptides with sp^2-hybridized C_γ, as shown in Figure 2m−p.

The obtained free energy profiles provided comprehensive information about position, shape, and depth of minima, as well as barriers. However, for sake of demonstrative comparison of different amino acids and force fields, it is necessary to move from individual profiles to quantified properties. This study focused on population of backbone conformers and rotamers expressing propensities of individual dipeptides rather then characteristics of the free energy profiles. The resulting propensities of dipeptides in different force fields represented by Hinton diagrams are shown in Figure 3. Our standard to which all other propensities are related is the alanine dipeptide. In addition to comparison of absolute affinities for given regions, the relative trends in propensities between given amino acid dipeptides were assessed. The inconsistency of all investigated force fields in predicting trends across dipeptides is obvious from all Hinton plots in Figure 3. The propensities related to alanine dipeptide for conformers in 3-state model (Figure 3a) as well as 5-state model (Figure 3b) did not reveal any consensus between studied force fields. Moreover, two related force fields from the AMBER family (FF99SB and FF03) differed strikingly in prediction of conformational preferences. The corresponding values of absolute and relative population of individual conformers in 3- and 5-state models together with estimated uncertainties are available in the Supporting Information (S3).

The lack of any common trends in calculated propensities was also quantified by Spearman correlation coefficient, which is able to detect nonlinear relationships in data. The resulting Spearman coefficients between propensities of dipeptides (except proline and glycine dipeptide) for individual regions as manifested by force fields are listed in Table 1. In the most cases, the values of correlation coefficient close to zero were calculated suggesting no correlation between propensities. The highest found value of correlation coefficient ($R \sim 0.5$) still admits only very poor correlation. The deterioration of correlation by outliers was excluded by detailed mutual scatter plots analysis of dipeptide populations in individual regions and force fields (see Supporting Information S5).

The preferences of dipeptides for χ_1 torsion, one of the most important parameters describing conformation of side chain, are reported in Figure 4. The predicted strong discrimination in favor of one rotamer was noticeable for asparagine (asn) and

dx.doi.org/10.1021/ct300794a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

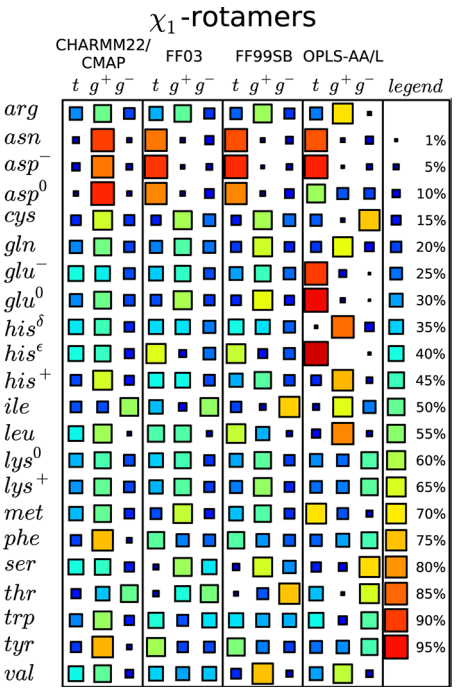


Figure 4. Propensities for χ_1 torsion. Propensities for individual χ_1 rotamers, trans (t), gauche$^+$ (g^+), and gauche$^-$ (g^-), are depicted in Hinton plot. The area of the square is proportional to the population of the given rotamer. The differences are also emphasized by colors. See the Supporting Information for values with estimated uncertainties.

aspartate (asp^-) dipeptide in all force fields. However, CHARMM22/CMAP prefers gauche$^+$ rotamers above the others, unlike FF03, FF99SB, and OPLS-AA/L populating trans rotamer. The very strong preferences for one particular χ_1 rotamer were observed further for OPLS-AA/L force field (glu^-, glu^0, his^δ, and his^ε). Most often, the free energy differences between the three rotamer states made them all (or at least two of them) feasible for population. The propensities for individual rotamers were found to be greatly force field dependent, as shown in Table 2 by Spearman correlation coefficients. The only significant correlation in propensities for χ_1 rotamers was observed for 2 AMBER force fields: FF03 and FF99SB. The lack of correlation between other force fields was proven to be not caused by few outliers (see Supporting Information S5 for detailed scatter plots).

Because of discrepancies in population of rotamers preferred by FF99SB and those found in experimental structures of selected proteins, the torsion correction in side chains were elaborated recently for certain amino acids (ile, leu, asp, and asn) as FF99SB-ILDN force field.[17]

The four above-mentioned dipeptides derived from amino acids affected by ILDN correction were added to our study and compared with their previous parametrization. The comparison is presented for sake of better readability on one-dimensional free energy profiles for asp^- dipeptide as an example (Figure 5). The correction dramatically influenced free energy shape of χ_1 and χ_2 torsions but more important is the fact that the dependence of ψ torsion was also affected by this change. The interference with free energy in terms of φ torsion was not proved confidently for aspartate dipeptide but statistically significant deviations on free energy profile were found for leucine dipeptide. The influence of ILDN correction on population of backbone conformers is summarized in Table 3 for the 3-state model. The 5-state model provided very similar picture (see Supporting Information S3). The population of φ/ψ conformers were changed remarkably in case of asp^- and ile dipeptide after applying ILDN correction. The effect on propensity of the two other dipeptides (asn, leu) was observed several-times smaller. The performance of the ILDN correction for χ_1 and χ_2, for which was intended, is listed in Table 4 for completeness. The ILDN correction completely swapped the ranking of rotamers. Basically, the propensities for trans χ_1 rotamer were weakened in favor of gauche$^+$ or gauche$^-$. The population of χ_2 rotamers were also changed dramatically for ile and leu dipeptides although the FF99SB-ILDN corrected only χ_1 torsional parameters for these amino acids.

The striking inconsistency in relative backbone propensities for individual dipeptides between FF03 and FF99SB force fields motivated us to question its source. Despite of sharing large amount of bonded and nonbonded parameters, FF03 and FF99SB differ in backbone torsion parameters and partial charge model.[31,39] Addressing the effect of different partial charges, the hybrid force field (HYB) was created. Parameters of the HYB force field originated completely from FF99SB except the partial charges, which were adopted from FF03. The metadynamics simulation of all dipeptides in hybrid force field were performed and analyzed in the same manner as for FF03, FF99SB, and the others.

The absolute population of backbone regions for HYB were found distinct from the parent FF99SB and FF03 force fields (see Supporting Information S3). However, the analysis of the trends in the relative propensities revealed strong resemblance between those manifested by FF03 and HYB. On the other hand, no common trends were found between FF99SB and HYB. The conclusive agreement in trends between FF03 and HYB is evident from mutual scatter plots presented in Figure 6 and Spearman correlation coefficients in Table 5. The correlation coefficients as well as scatter plots did not suggest any overall correlation of backbone propensities between HYB and FF99SB, unlike FF03.

DISCUSSION

In this study, we present conformational preferences of all standard amino acids in 4 different force fields in the simplest possible peptide model—an amino acid dipeptide. This model eliminates the various neighbors effects between amino acids in

Table 2. Correlation of Propensities for χ_1 Rotamers between Different Force Fields[a]

conformer	CHARMM22/CMAP vs FF03	CHARMM22/CMAP vs FF99SB	CHARMM22/CMAP vs OPLS-AA/L	FF03 vs FF99SB	FF03 vs OPLS-AA/L	FF99SB vs OPLS-AA/L
trans	−0.27 ± 0.04	−0.21 ± 0.04	−0.13 ± 0.04	0.91 ± 0.02	0.25 ± 0.04	0.27 ± 0.03
gauche$^+$	−0.31 ± 0.03	−0.27 ± 0.02	−0.01 ± 0.04	0.80 ± 0.02	0.14 ± 0.03	0.27 ± 0.02
gauche$^-$	0.42 ± 0.03	0.50 ± 0.03	0.10 ± 0.04	0.75 ± 0.02	0.45 ± 0.05	0.48 ± 0.03

[a]Correlation of propensities for individual χ_1 rotamers measured by Spearman correlation coefficient. The uncertainties expressed by standard deviation as estimated by permutations tests.

dx.doi.org/10.1021/ct300794a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

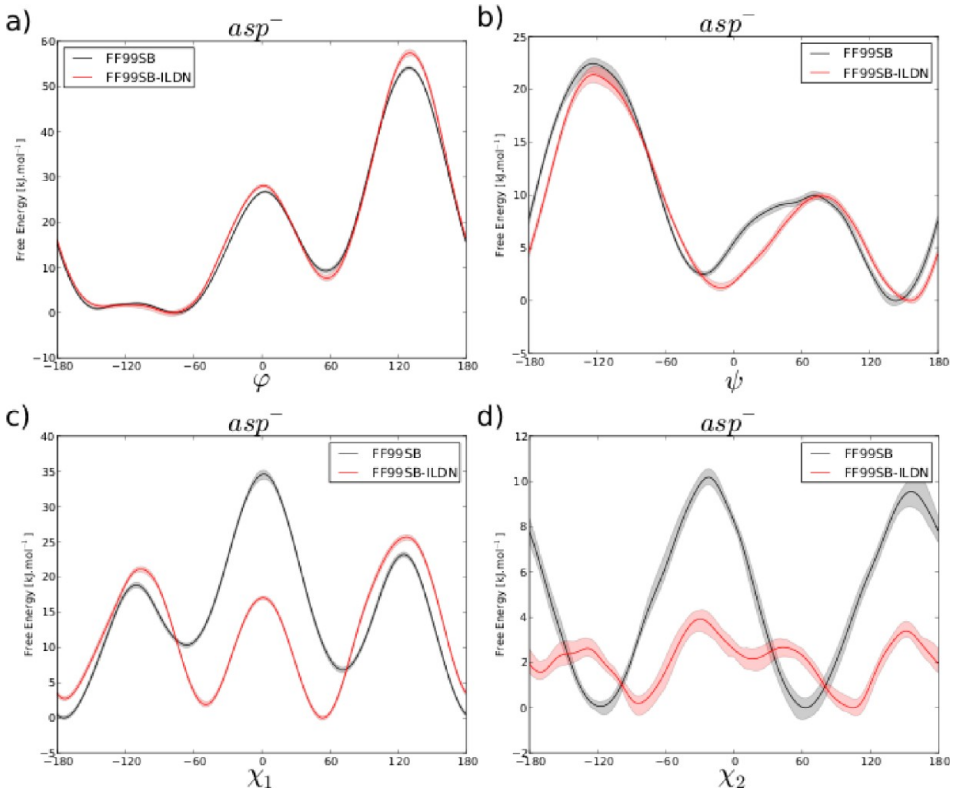


Figure 5. One-dimensional free energy profiles of aspartate dipeptide (asp^-) modeled by FF99SB and FF99SB-ILDN force field. Parts a, b, c, and d depict free energy profiles in terms of φ, ψ, χ_1, and χ_2 torsions, respectively. The width of the lines deliminates the uncertainty (90% confidence interval) as estimated from metadynamics.

Table 3. Effect of ILDN Correction on Population of Backbone Conformers[a]

	FF99SB			FF99SB-ILDN		
dipeptide	E	H	α_L	E	H	α_L
asn	74.1 ± 1.1	23.8 ± 1.0	2.0 ± 0.3	67.5 ± 0.8	29.9 ± 0.8	2.6 ± 0.2
asp^-	71.8 ± 1.3	27.4 ± 1.3	0.8 ± 0.1	55.6 ± 0.8	42.6 ± 0.9	1.8 ± 0.3
ile	60.3 ± 1.4	39.1 ± 1.3	0.7 ± 0.1	73.1 ± 0.5	26.0 ± 0.6	0.9 ± 0.1
leu	64.4 ± 0.9	34.2 ± 0.9	1.4 ± 0.1	66.3 ± 0.9	31.5 ± 1.0	2.2 ± 0.2

[a]Uncertainties reported as standard deviations.

peptide sequence, such as mutual side chain interactions and effect of the backbone atoms interactions. Therefore, it reflects the sequence independent intrinsic propensity of every amino acid for a local structure. The physical origin of this distinct propensities remains still unclear, and it is questionable whether it can be investigated by computer simulations using the current methods and models. We focused on the ability of molecular dynamics utilizing additive force fields to provide a consistent evaluation of the backbone and side chain intrinsic propensities.

Simulation Protocol. To obtain reliable and robust estimates of conformational preferences, we designed a protocol based on bias-exchange metadynamics accompanied by estimations of uncertainties. The proposed protocol were verified on test cases of the three amino acids dipeptides, the positively charged aspartate, hydrophobic isoleucine, and polar and bulky tryptophane dipeptides. The four independent simulations for each mentioned dipeptide and force field in question provided very similar results and confirmed the legitimacy of the estimated uncertainties (see the Supporting Information, Table S3.13). Our motivation for introducing this rather complicated protocol were grounded on the inability of classical equilibration simulations or simple metadynamics to provide reliable estimates of conformational preferences in dedicated computational time.

Water Models. All simulations were performed using simple TIP3P water model, which is a rather crude model and is overperformed by many others. However, AMBER and CHARMM force fields were originally parametrized just for application of TIP3P. On the other hand, authors of OPLS force field prioritized TIP4P model. We had decided to use TIP3P in all production simulations because of two reasons. First, to provide the same conditions for all simulation and to avoid any eventual bias introduced by utilizing different water model. Second, too high self-diffusion constant of TIP3P water reflecting its higher dynamics. In case of metadynamics as an artificial nonequilibrium simulation, the realistic kinetics cannot be achieved, and unrealistic fast dynamics of TIP3P does not pose an obstacle. Moreover, we expected that fast water molecules might facilitate faster conformational sampling. To assess the eventual effects of different water models on conformational preferences of dipeptides, we performed another test simulations for asp^-, ile, and trp dipeptides. The four

dx.doi.org/10.1021/ct300794a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 4. Effect of ILDN Correction on Population of Rotamers

	FF99SB			FF99SB-ILDN			FF99SB			FF99SB-ILDN		
	χ_1						χ_2					
dipeptides	t	g^+	g^-	t	g^+	g^-	t	g^+	g^-	t	g^+	g^-
asn	88.2 ± 0.7	3.0 ± 0.2	8.8 ± 0.5	42.0 ± 1.3	53.7 ± 1.3	4.3 ± 0.4		54.5 ± 0.9	45.5 ± 0.9		49.2 ± 1.0	50.8 ± 1.0
asp^-	92.4 ± 1.0	1.7 ± 0.3	5.9 ± 0.7	21.9 ± 0.9	25.5 ± 0.6	52.6 ± 0.8		49.8 ± 1.6	50.2 ± 1.6		49.6 ± 0.6	50.4 ± 0.6
ile	20.0 ± 1.1	6.6 ± 0.5	73.4 ± 1.4	11.0 ± 1.1	51.4 ± 1.7	37.6 ± 1.0	74.9 ± 1.5	6.5 ± 0.4	18.7 ± 1.3	59.3 ± 1.8	28.8 ± 1.6	11.9 ± 0.7
leu	62.7 ± 1.5	31.9 ± 1.0	5.4 ± 0.6	17.0 ± 1.2	81.6 ± 1.1	1.4 ± 0.1	34.3 ± 1.3	3.1 ± 0.5	62.6 ± 1.3	76.0 ± 1.1	2.4 ± 0.3	21.7 ± 1.2

independent simulations were conducted with additional SPC/E and TIP4P/Ew waters for each of the three test dipeptides. We found in our previous study[28] that this effect on alanine dipeptide were negligible, although it is known that water models differ in solvation of peptides[40] and influence significantly their stability.[41] Our trial simulations of the three aforementioned dipeptides revealed that there is a tendency of TIP3P water to slightly destabilize ppII region in favor of α_R regions compared with SPC/E and TIP4P/Ew. The largest magnitude of this effect were found for asp^- dipeptide, where the difference in population of the ppII region between water models were approximately 5% in each force field. We expect that the effect of water models for other dipeptides is comparable to uncertainties in the reported values.

Definition of Conformers. The definition of backbone conformers slightly differs between authors and makes any comparison rather complicated. We defined backbone conformers in terms of their φ and ψ torsions. It must be noted that there is no justification for using precise φ/ψ regions derived from protein structures for characterization of simple model as dipeptides, which is lacking the stabilization network of hydrogen bonds. The proposed ranges of torsions were accommodated to the obtained two-dimensional free energy profiles of dipeptides in all four force fields. However, the energy barriers between ppII and ext conformer is low in general, and therefore, the transition region is also populated significantly (see Figure 1a). The same was observed for α_R' region. The border between α_R and α_R' were even more difficult to assess, because some combinations of amino acid dipeptides and force fields did not manifest clear α_R' minimum but rather shallow low energy extension of α_R region. Therefore, any common dividing line between ppII and ext as well as α_R and α_R' is arbitrary. To overcome these difficulties, we also utilized a more general model involving only three states (E, H, and α_L) in similar fashion, as proposed Feig.[25] The three-state model do not suffer so much from ambiguity in borders between regions at the expense of lower resolution. Unlike the five-state model, all dividing lines in the three-state model pass through very low populated regions high in free energy. Both models do not have ambitions to describe precisely the system of free energy basins on the right side of Ramachandran plot. There are usually two or three minima present depending on force field used but with very low population. However, this regions are critical for stability of turns and preferences of amino acids for forming them in polypeptides.

Comparison of Intrinsic Propensities. The comparison of intrinsic propensities of dipeptides was evaluated in relative scales with alanine dipeptide as a baseline for each force field. The conformational preferences of alanine residue can be easily tuned by changing backbone torsion parameters, as was done in various force field corrections.[6,9,11−13,32,42] All these proposed corrections applied the same correction term also to the other amino acids residues (except glycine). Because of additivity of the potential energy terms in this class of force fields the conformational free energy is affected, in the monotonous although nonlinear way. Therefore, the trends in intrinsic propensities are preserved also in presence of general backbone corrections and only their magnitudes can differ. Hence, the trends observed for FF03 and FF99SB remain valid (except for numeric values) even for their corrected versions (FF03*, FF99SB*, FF03w, FF99SBφ', FF99SB$\varphi\psi$). On the other hand, as was demonstrated on case of ILDN parameters, ad hoc changes of torsional terms in side chains may shift significantly backbone propensities of given amino acid.

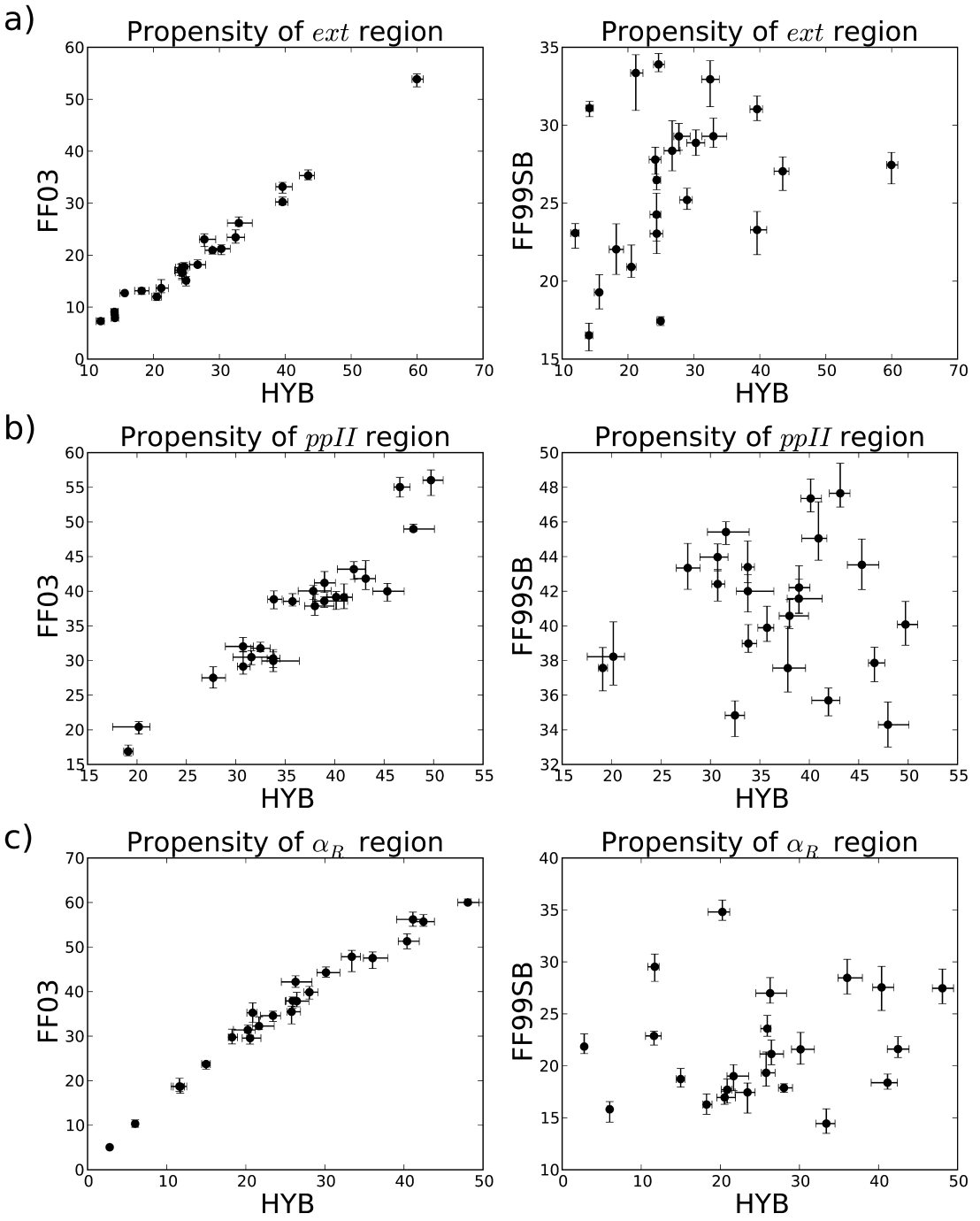


Figure 6. Trends in hybrid force field (HYB) and FF99SB and FF03. The population of the three most important region from 5-state model (ext (a), ppII (b), and α_R (c)) for all dipeptides (excluding pro and gly) as manifested by FF03, FF99SB, and HYB force field. The propensities correlate obviously between FF03 and HYB, contrary to the FF99SB. The error bars express the range of values obtained from different replicas.

Table 5. Correlation in Propensities for Backbone Regions Between HYB, FF03, and FF99SB[a]

conformer	ext	ppII	α_R	α_R'
HYB vs FF03	0.96 ± 0.01	0.91 ± 0.02	0.98 ± 0.01	0.94 ± 0.02
HYB vs FF99SB	0.36 ± 0.04	−0.04 ± 0.05	0.18 ± 0.04	−0.02 ± 0.05

[a]Spearman correlation coefficients measured between propensities for given backbone regions between HYB, FF03, FF99SB force fields for all dipeptides except pro and gly dipeptide. Uncertainties reported as standard deviation obtained by permutation test.

Basically, the distribution of side chain torsions is dependent on backbone torsions. This fact is well-known due to the statistical analysis of protein structures in the PDB[43] and is also the justification for development of various backbone dependent rotamer libraries.[44,45] Therefore, it could be more accurate to set up a complex definition of conformers involving both backbone and side chain torsions. We attempted to apply such approach, but it did not bring any additional insight or correlation between

dx.doi.org/10.1021/ct300794a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

force fields than presented in Tables 1 and 2. The correlation between populations of χ_1 (backbone independent) rotamers was found slightly higher than between different backbone conformers, and hence, the correlation coefficients for backbone dependent rotamers resembled closely values in Table 2. The correlation between populations of χ_1 rotamers in FF03 and FF99SB was the only significant correlation revealed in our study. However, this finding only makes more confusing the fact that no correlation were found between backbone conformers in these force fields.

The mutual disagreement in predicted trends was obtained for both the three-state and five-state backbone conformational models. Therefore, we can exclude that very low correlation was caused by any artificiality in definition of backbone conformers. The striking differences in calculated trends in intrinsic propensities for individual dipeptides across force fields cause doubts about the common parametrization strategy in force field development. On the other hand, the free energy differences between individual conformers, which reflect in relative populations, are usually very small (few kJ mol^{-1} or even less). Such precision in small energetic differences pose a challenge even for the most reliable quantum ab initio methods. Therefore, the force field cannot be expected to provide similar performance without careful parametrization. Predictions of trends in backbone conformational preferences introduced by different side chains seem to be very complicated task for force fields. We attempted to identify the terms that are critical in driving backbone preferences. The before mentioned and emphasized difference between two AMBER force fields suggested role of the electrostatics because both force fields differs in charge model. Although both FF99SB and FF03 use point charges provided by RESP procedure,[46] they differ in accounting of solvent effects. FF99SB has inherited the charges from its ancestor FF94 where the missing solvents effect were approximated by artificially high dipole moments obtained in gas phase SCF calculations in small basis set. Authors of FF03 parametrization tried to incorporate solvent effect in more physically sounded way using more realistic DFT calculations with implicit solvent model. Moreover, the charges on backbone atoms were not constrained but allowed to be specific for all amino acids in FF03. The hypothesis about role of charge model was tested by examination of hybrid model. Unfortunately, it cannot be decided easily whether the major contributions originate from scaled 1−4 electrostatic interactions or interactions between other more distant atoms. The reason for distinction of other force fields (CHARMM22/CMAP and OPLS-AA/L) is more uncertain. AMBER force fields, CHARMM22/CMAP, and OPLS-AA/L differ not only by charge model but also by parameters for Lennard-Jones interaction, most of the bonded interactions (bond, angles, torsions) and scaling factors for 1−4 nonbonded interactions. Each of the listed terms may be potential candidate responsible for disagreements in trends manifested by investigated force fields.

The importance of some more realistic force field parametrization seems to be more and more obvious as documented by growing number of publications on this topic. At the time we finished our study, two relevant papers appeared[47,48] on reparametrization of CHARMM force field, which we did not include into our analysis. There are two reasons we did not perform the calculations with CHARMM36 yet. The first reason is that at the time the papers were published there was no port to the GROMACS program package so we could not keep the uniformity of the method we applied for the other force fields. The second reason is that we could not find more test cases on CHARMM36 parametrization and its performance provided by comparison with existing benchmarks (as for example for noncovalent interactions validated by advanced ab initio methods).[49] We intend to continue in our effort to map systematically performance of commonly used force fields on the conformational behavior of peptides, so the suggested study with CHARMM36, as well as with other newly parametrized force fields, is our next task.

We are aware of the fact that valuable information provided by the presented study should be a comparison with experimental values and an assessment or classification of the tested force fields. It was also our initial intention to provide such data to demonstrate performances of individual force fields and recommend the most suitable one. Unfortunately, there were few reasons (as for example that only the matching experimental setup with the same model peptides in equivalent environment may be used as a reference) that we deliberately decided to not compare the obtained propensity scales to any scales derived from either experiments or PDB surveys. This condition disqualified both coil libraries and experimental studies of long peptides or peptides in presence of denaturing agents. The other factor influencing our decision is a dependence of the reported values on experimental technique and their interpretations. This was also discussed recently in the work of Elam et al.[50] It is apparent that the strong needs of the reliable empirical potential should stimulate experimentalists to provide more data on this subject to reach a reasonable level of confidence on this subject.

CONCLUSION

Our study brought important findings that we can summarize as follows:

The four commonly used force fields for simulation of biomolecules provide significantly different trends for propensities of amino acid torsions both for backbone and side chains.

It is not clear whether the found discrepancy is caused by insufficient parametrization or deficiency in the computational model due the simple forms of force fields potentials.

We explained that the major reason in discrepancy of the AMBER force fields (FF99SB and FF03) is caused by different charge models.

The reported inconsistency of the studied force-fields might influence interpretations of biomolecules simulations, namely, the process of protein folding or unfolding and influence of different environments on propensity of different thermodynamics states.

There is still a room for optimization. Although the application of residue specific parameters was avoided in course of force field development because of idea of transferability, the specific parameters may be necessary for fine-tuning of force fields toward experimental data and further work is necessary.

ASSOCIATED CONTENT

Supporting Information

Detailed description of collective coordinates and simulation protocol, free energy profiles for all amino acid dipeptides and all investigated force fields, as well as detailed tables of propensities and correlation scatter plots. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jiri.vondrasek@uochb.cas.cz.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by KONTAKT II programme LH11020 of Ministry of Education, Youth, and Sports of the Czech Republic. It was also a part of the research project No. Z40550506 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic. The access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005) is highly acknowledged.

■ REFERENCES

(1) Best, R. B. *Curr. Opin. Struct. Biol.* **2012**, *22*, 52−61.
(2) Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulten, K. *Nat. Phys.* **2010**, *6*, 751−758.
(3) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers. *Science* **2010**, *330*, 341−346.
(4) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517−520.
(5) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. *J. Am. Chem. Soc.* **2012**, *134*, 3787−3791.
(6) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Biophys. J.* **2011**, *100*, L47−49.
(7) Best, R. B.; Buchete, N.-V.; Hummer, G. *Biophys. J.* **2008**, *95*, L07−09.
(8) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472−2493.
(9) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.
(10) Gnanakaran, S.; García, A. E. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 773−782.
(11) Best, R. B.; Mittal, J. *J. Phys. Chem. B* **2010**, *114*, 14916−14923.
(12) Nerenberg, P. S.; Head-Gordon. *J. Chem. Theory Comput.* **2011**, *7*, 1220−1230.
(13) Li, D.-W.; Brüschweiler, R. *J. Chem. Theory Comput.* **2011**, *7*, 1773−1782.
(14) Beauchamp, K. a.; Lin, Y.-S.; Das, R.; Pande, V. S. *J. Chem. Theory Comput.* **2012**, *8*, 1409−1414.
(15) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *PloS ONE* **2012**, *7*, e32131.
(16) Kaminski, G. a.; Friesner, R. a.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(17) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.
(18) Jiang, F.; Han, W.; Wu, Y. *J. Phys. Chem. B* **2010**, *114*, 5840−5850.
(19) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13*, 211−222.
(20) Pace, C. N.; Scholtz, J. M. *Biophys. J.* **1998**, *75*, 422−427.
(21) Avbelj, F.; Grdadolnik, S. G.; Grdadolnik, J.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1272−1277.
(22) Grdadolnik, J.; Mohacek-Grosev, V.; Baldwin, R. L.; Avbelj, F. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 1794−1798.
(23) Schweitzer-Stenner, R.; Hagarman, A.; Measey, T. J.; Mathieu, D.; Schwalbe, H. *J. Am. Chem. Soc.* **2010**, *132*, 540−551.
(24) Beck, D. a C.; Alonso, D. O. V.; Inoyama, D.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 12259−12264.
(25) Feig, M. *J. Chem. Theory Comput.* **2008**, *4*, 1555−1564.
(26) Cruz, V. L.; Ramos, J.; Martinez-Salazar, J. *J. Phys. Chem. B* **2012**, *116*, 469−475.
(27) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(28) Vymetal, J.; Vondrásek, J. *J. Phys. Chem. B* **2010**, *114*, 5632−5642.
(29) Case, D. a; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(30) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(31) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(32) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(33) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(34) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. *J. Chem. Theory Comput.* **2010**, *6*, 459−466.
(35) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(36) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 1−4.
(37) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.
(38) Bonomi, M.; Barducci, A.; Parrinello, M. *J. Comput. Chem.* **2009**, *30*, 1615−1621.
(39) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(40) Florová, P.; Sklenovskỳ, P.; Banáš, P.; Otyepka, M. *J. Chem. Theory Comput.* **2010**, *6*, 3569−3579.
(41) Paschek, D.; Day, R.; García, A. E. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19840−19847.
(42) Vymětal, J.; Vondrášek. *J. Phys. Chem. Lett.* **2011**, *503*, 301−304.
(43) Chakrabarti, P.; Pal, D. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1−102.
(44) Shapovalov, M. V.; Dunbrack, R. L. *Structure* **2011**, *19*, 844−858.
(45) Lovell, S. C.; Davis, I. W.; Arendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 437−450.
(46) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(47) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.
(48) Best, R. B.; Mittal, J.; Feig, M.; Mackerell, A. D. *Biophys. J.* **2012**, *103*, 1045−1051.
(49) Řezáč, J.; Jurečka, P.; Riley, K. E.; Černý, J.; Valdes, H.; Pluháčková, K.; Berka, K.; Řezáč, T.; Pitoňák, M.; Vondrášek, J.; Hobza, P. *Coll. Czech CC* **2008**, *73*, 1261−1270.
(50) Elam, W. A.; Schrank, T. P.; Hilster, V. J. In *Protein and Peptide Folding, Misfolding, and Non-Folding*; Schweitzer-Stenner, R., Ed.; John Wiley & Sons: Hoboken, NJ, 2012; pp 159−185.