Influence of Side Chain Conformations on Local Conformational Features of Amino Acids and Implication for Force Field Development

Fan Jiang,[†,‡] Wei Han,[‡] and Yun-Dong Wu*[,†,‡]

Laboratory of Chemical Genomics, Shenzhen Graduate School of Peking University, Shenzhen 518055, China, and Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

Received: September 21, 2009; Revised Manuscript Received: January 28, 2010

Statistical analysis of coil regions in protein structures has been used to obtain the local backbone ϕ, ψ preferences of amino acids, which agree well with the NMR experiments of unfolded peptides and proteins. We analyzed the conformational features of amino acid residues in a restricted coil library of 4220 high-resolution protein crystal structures. In addition to Gly, Ala, and Pro, the ϕ, ψ distribution (Ramachandran plot) of each amino acid is analyzed with respect to three side chain conformers: g+ ($\chi_1 \sim -60°$), g− ($\chi_1 \sim 60°$), and t ($\chi_1 \sim 180°$). The statistical study indicates that the effect of side chain conformations on ϕ, ψ distributions is even greater than the effect of amino acid types. On the basis of the χ_1, ϕ, ψ conformational preferences, the amino acids in addition to Gly, Pro, and Ala can be divided into five types: (1) ordinary amino acids, (2) Ser, (3) Asp and Asn, (4) Val and Ile, and (5) Thr, each with distinguished χ_1 rotamers. The α-helix, β-sheet, and type-I β-turn preferences of the different rotamers of various amino acid types can be captured by their intrinsic ϕ, ψ preferences from our coil library. Molecular dynamics simulations of dipeptide Ac−X−NHMe and tetrapeptide Ac−A−X−A−NHMe models give nearly the same side chain rotamer distributions. However, for many amino acids, both OPLS−AA/L and AMBER−FF03 force fields give very different χ_1 rotamer distributions from the coil library. This may partially explain why dipeptide models sometimes cannot reproduce those of protein structures well. The current coil library analysis may be valuable in improving the force field for protein simulations.

Introduction

The local backbone conformational preferences of α-amino acid residues (described by ϕ, ψ torsional angle-dependent free energy maps or Ramachandran plots) contribute significantly to protein structures and folding. There has been much effort committed to obtaining local backbone conformational preferences of the 20 natural amino acids, including their secondary structure propensities, from both statistical analysis of protein crystal structures[1−11] and spectroscopic (CD, NMR) studies[12−21] of peptides or proteins in aqueous solution. Despite the tremendous efforts, the topic of local conformational features of peptides remains an active research subject.[22−26]

The Ramachandran plots were originally derived from the conformational analysis of simple dipeptide models.[27,28] About a decade ago, Street and Mayo used similar small peptide models to derive the intrinsic β-sheet propensities of 17 amino acids through simulations, which are in good correlation with experimental results.[29] However, recent molecular dynamics (MD) simulations of small peptides (GGXGG[25] or dipeptides[26]) gave ϕ, ψ preferences of 20 amino acids significantly different from statistical results of protein crystal structures.

In recent years, more attention has been paid to the analysis of the coil library (the residues outside the α-helices and β-sheets) from known protein structures, with the hope of obtaining intrinsic conformational features of amino acids. Good correlations between coil library ϕ, ψ distributions of amino acids and their secondary structure propensities have been obtained.[7,11] The statistical results from the coil library also agree very well with the NMR experiments of designed unfolded peptides in solution, in which the local interactions dominate.[17−20,30] The statistical ϕ, ψ distributions obtained from the coil library have also been successfully used in the simulations of unfolded peptides and proteins,[31−33] and in de novo protein structure prediction.[34]

These coil library studies indicate that the secondary structure propensities of different amino acids do largely come from their intrinsic ϕ, ψ preferences. In force field development, backbone torsional ϕ, ψ parameters are usually optimized on Gly and Ala dipeptides or tetrapeptides to reproduce the results from quantum mechanics (QM) calculations or protein crystal structures. These parameters are usually used for other amino acids without modification.[35−44] The inconsistency between recent MD simulations of small peptides and observed local conformational preferences may indicate the need to consider the effect of the side chains.

It was found that both the secondary structure and amino acid type can affect the χ_1 (the rotation around Cα−Cβ bond) distributions.[45] The backbone-dependent side chain rotamer library has been already established.[46−48] It was also found that side chain conformation χ_1 significantly affects the ϕ, ψ preferences of various amino acids.[49,50] However, there is still a lack of quantitative evaluation of the intrinsic influence of different side chain rotamers on backbone conformations, which needs high-resolution χ_1-dependent ϕ, ψ plots based on a reliable coil library.

The side chain conformation itself is also very important because the interactions between side chains contribute greatly

* To whom correspondence should be addressed. Phone: 00852-2358-7391. Fax: 00852-2357-1594. E-mail: chydwu@ust.hk.
† Shenzhen Graduate School of Peking University.
‡ The Hong Kong University of Science and Technology.

10.1021/jp909088e
Published on Web 04/14/2010

to protein folding and binding. Previous NMR studies have indicated that the local conformational preferences of side chains can be extracted from protein coil residues.[51,52] It will be interesting to see whether the χ_1 distributions of various amino acids from the coil library can be reproduced by MD simulations with commonly used force fields.

Here, we report the construction of a new restricted coil library and the establishment of side chain-dependent local conformational preferences or (ϕ, ψ, χ_1) Ramachandran plots. We also investigated the intrinsic conformational features obtained from our coil library and compared them with the side chain preferences in various secondary structures. We also report that the side chain conformational preferences from simulations of commonly used OPLS-AA/L and AMBER-FF03 force fields can be very different from those obtained from the coil library.

Methods

Statistical Analysis of PDB Structures. Protein X-ray crystal structures of resolution <2.0 Å and R factor <0.2 were selected and downloaded from the Protein Data Bank (PDB),[53] with 50% sequence identity cutoff. There were totally 4220 protein structures and 2.0×10^6 residues. When there were m identical chains in a protein structure, each chain was counted with a $1/m$ statistical weight. The widely used DSSP[54] program was used to assign secondary structures. Coil residues were defined as the residues that are not in any secondary structures, including 3_{10}-helix (DSSP code G), α-helix (H), π-helix (I), β-sheet (E), β-bridge (B), and various types of turns (T). In most previous coil libraries,[7,9,10,55−57] residues in turn structures (most of them being β-turns) were not excluded, but here, we used more restricted criteria, following the work of Jha et al.[11] Also following the studies by Jha et al., all the residues adjacent to any secondary structures were excluded from our restricted coil library (Coil-R) to eliminate the biases against the conformations of related secondary structures. The elimination of biases is confirmed by obtaining Ramanchandran plots from only mainly-α proteins or mainly-β proteins (Supporting Information Figure S1). The residues preceding proline[56] or having any backbone atom with temperature factor $B > 36$ were also excluded.

Different from all previous coil libraries, there is a further restriction for our Coil-R. That is to exclude residues preceding any residue in α-conformation ($-60° < \psi < 60°$, both α_L and α_R). Previous studies[58−62] have indicated that the side chains of some amino acids, such as Ser, Asp, and Asn, can form an H-bond with the backbone of successive residues in α-conformer. The strong local inter-residue side chain/backbone H-bonding can significantly alter the rotameric preference of the involved side chain. This effect can be avoided by allowing only the following residue to be in extended conformations.

Because backbone dihedral angles ϕ, ψ are continuous real variables, the probability for any given exact ϕ, ψ values are zero. To obtain the preference for a certain backbone conformation (ϕ, ψ), the Gauss function was used to sample the conformational space:

$$n_{X,\chi}(\phi,\psi) = \sum_i^{X,\chi} w_i \exp\left[-\frac{(\phi_i-\phi)^2+(\psi_i-\psi)^2}{2\sigma^2}\right] \quad (1)$$

Here, i counts for all residues of given amino acid type X in side chain rotamer χ in our coil library, and w_i is the $1/m$ weighting for m identical chains. For the Ramachandran plots in Figure 1, 10° × 10° grids of (ϕ, ψ) and standard deviation σ of 7° were used. The general features of the obtained ϕ, ψ plots remained the same when σ was increased to 10°.

The similarity coefficient, S, between two distributions $\mathbf{X} = \{x_i\}$, $\mathbf{Y} = \{y_i\}$ were simply calculated as

$$S_{XY} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\cdot\sqrt{\sum y_i^2}} = \cos\alpha_{XY} \quad (2)$$

S_{XY} can be regarded as the cosine of the angle α between two vectors $\mathbf{X}$ and $\mathbf{Y}$. Here, x_i and y_i obtained from eq 1 are all nonnegative values, so the S_{XY} must between 0 and 1. When and only when the two distributions are actually the same after normalization does $S_{XY} = 1$. We also calculated the Jensen−Shannon divergences among all the distributions, which anticorrelate well with the similarity coefficients. (Tables S7−S14 in the Supporting Information)

To calculate the percentages in Figure 4, the counts of a certain amino acid type at a certain side chain rotamer (g+/t/g−) for given ϕ, ψ region were divided by the total number of same amino acid in *all* side chain rotamers in the same library. Therefore, both side chain and backbone conformational preferences were included.

All-Atom Molecular Dynamics Simulations. MD simulations of Ac−X−NHMe (dipeptides) and Ac−Ala−X−Ala−NHMe (tetrapeptides) were carried out with the Gromacs[63] program (version 3.3.1) using the OPLS-AA/L[64,38] all-atom force field for the 17 amino acids having side chain rotamers. Each peptide was solvated by 340−370 TIP4P[65] waters in a periodic cubic box. Berendsen temperature and pressure coupling with $\tau_T = 0.2$ ps and $\tau_P = 0.5$ ps were used to maintain $T = 300$ K and $P = 1$ atm. A cutoff of 10 Å was used for nonbonded interactions, and the reaction field method ($\varepsilon_r = 80$ for water) was used to treat long-range electrostatic interactions. For each peptide other than Asn and Thr, two different trajectories of 100−700 ns (1 fs time step) were generated. For the peptides of Asn and Thr, trajectories of much longer time (>2 μs) were collected to reach adequate sampling of their side chain rotamers, and the results were conformed by replica-exchange molecular dynamics (~300−419 K, 15 replicas). A summary of results is given in Supporting Information Table S15.

MD simulations of Ac−X−NHMe were also carried out with the Amber[66,67] program using the AMBER-FF03[39] all-atom force field. Each peptide was solvated by 520−610 TIP3P[68] waters in a truncated octahedron box. SHAKE[69] was used to constrain all bonds involving hydrogen. Langevin dynamics with the collision frequency 2.0 ps^{-1} was used to maintain $T = 300$ K. Berendsen coupling with $\tau_P = 2.0$ ps was used to maintain $P = 1$ atm. Nonbonded interactions were treated with the particle mesh Ewald (PME) method with a 9 Å cutoff. For each peptide, 1−3 trajectories of 100−400 ns (2 fs time step) were generated. We also implemented the AMBER-FF03 force field with Gromacs 3.3.1 program and obtained quite similar results. A detailed description and summary of the results are given in Supporting Information Table S16.

In both OPLS-AA/L and AMBER-FF03 simulations, to mimic pH = 7 conditions, Arg/Lys side chains are positively charged, and Asp/Glu side chains are negatively charged. The whole systems were then neutralized with counterions (Cl^- or Na^+). In all simulations, structures were recorded every 1.0 ps. The first 10 ns of each trajectory was discarded for preequilibrium except for the OPLS-AA/L simulations of Asn and Thr,

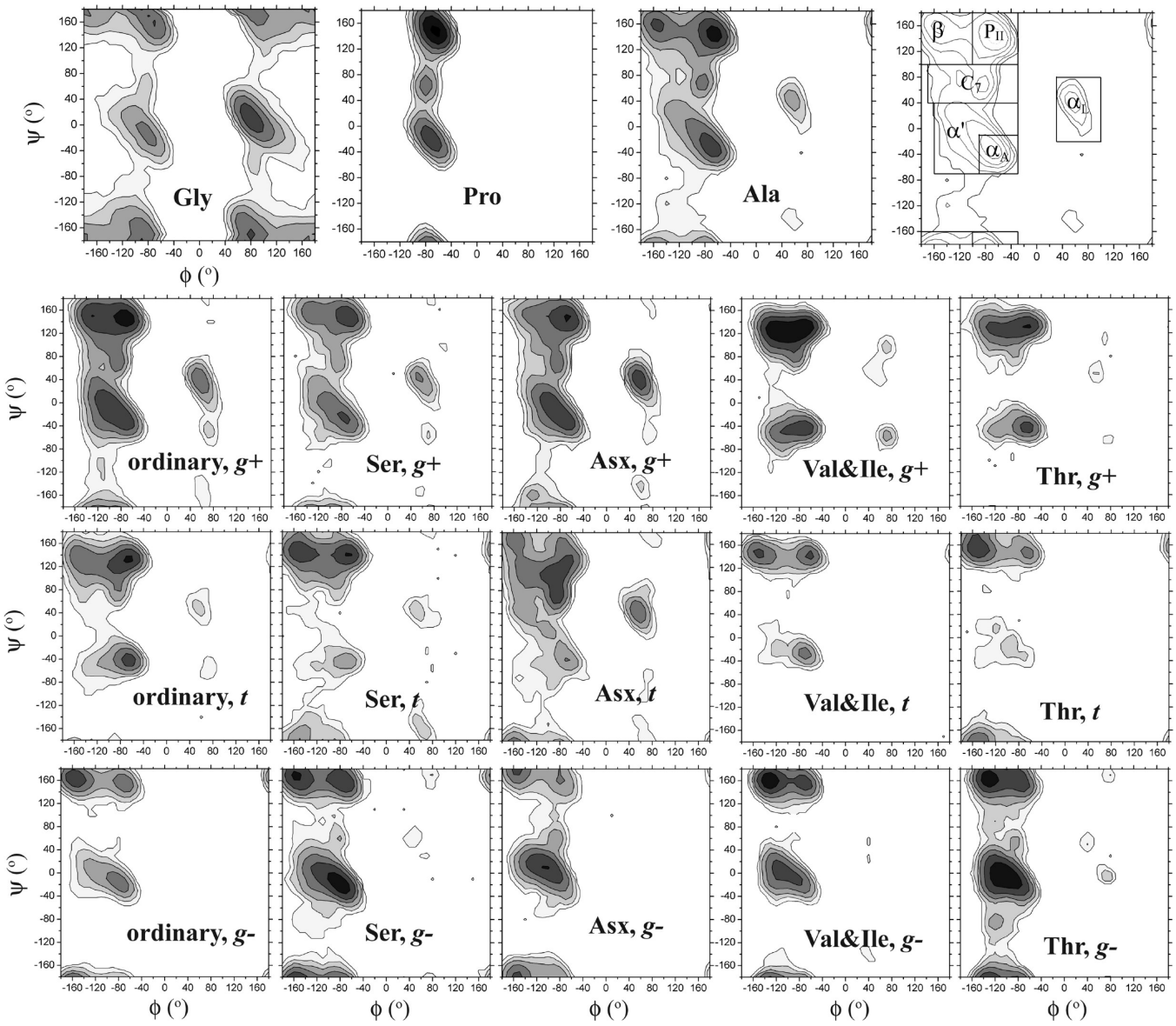


Figure 1. ϕ, ψ distributions of different side chain χ_1 rotamers (g+, t, g−) of five amino acid types, obtained from our restricted coil library statistics. Darker region indicates higher density, and two neighboring contour lines correspond to $e = 2.7$ times change of probability. The plots of different rotamers of the same amino acid type are drawn at the same scale such that they can be directly compared.

in which the first 100 ns was discarded for each trajectory. For each peptide, if not specified, the presented results are from the summation of all its trajectories.

Results and Discussion

Overall Distributions. Table 1 summarizes the side chain conformational preferences of all 17 amino acids with three side chain χ_1 rotamers (Gly, Ala, Pro have no rotamers). Eleven of them have similar behavior, and they are grouped as ordinary amino acids. They have the following order of preferences: g+ > t > g− in nearly all situations, except for Phe, Tyr, and Trp in the α-helix and Trp, His, Cys in turns. As found previously, Leu significantly disfavors the g− side chain rotamer (<3%) in all situations. Cys is also a little special for its high g+ preference for the α-helix (75%) and relatively low g+ (45%) and high g− (23%) in coil. Among them, the five amino acids (Glu, Gln, Lys, Arg, Met) with unbranched sp^3 Cγ in their side chains have very similar rotamer preferences. In the coil region, they have roughly 62−68% g+, 22−26% t, and 10−13% g−. The percentages do not change much when compared with results from our restricted coil library (Coil-R), which are 60−65% g+, 23−26% t, and 11−17% g−, respectively. This is interesting, since their side chains are terminated with very different groups, including highly nonpolar, neutral polar, positively charged, and negatively charged groups. Their backbone ϕ, ψ distributions are also very similar to recent studies.[24] This indicates that only the part of side chains close to the backbone is important in determining local conformational behaviors.

Different from the ordinary amino acids, several amino acids have much different side chain conformational preferences. Ser has a β−OH group. From our restricted coil library, Coil-R, there is a high preference for the OH group to take the g− conformation (53%) rather than g+ (29%) and t (17%) conformations. This preference for g− becomes even more significant (62%) in the turns. The g− rotamer of Ser is also significant in the α-helix (36%) and β-sheet (38%) structures when compared with ordinary amino acids. Thr also has a β−OH, but it has an additional β-methyl group. Therefore, its conformational preference is about the combination of those of

TABLE 1: The Percentages of Three Side Chain χ_1 Rotamers for Each Amino Acid Type in Various Secondary Structures or Coil Region

	α-helix			β-sheet			turns			coil[a]			Coil-R[b]		
	g+ %	t %	g− %	g+ %	t %	g− %	g+ %	t %	g− %	g+ %	t %	g− %	g+ %	t %	g− %
							Ordinary								
Glu	58.5	37.8	3.7	45.1	43.7	11.2	65.2	20.3	14.5	61.9	26.3	11.8	60.1	26.0	13.9
Gln	59.6	37.8	2.6	46.6	41.7	11.7	74.3	15.2	10.5	67.9	21.7	10.4	64.0	24.6	11.4
Lys	52.5	45.0	2.6	49.9	40.6	9.5	69.6	19.4	10.9	68.3	22.2	9.5	64.9	24.6	10.6
Arg	52.1	44.7	3.2	49.2	38.6	12.1	70.7	18.1	11.1	65.1	21.5	13.4	60.6	22.8	16.6
Met	67.8	30.2	1.9	46.7	40.0	13.3	77.0	14.7	8.4	61.7	25.5	12.7	61.3	24.8	13.9
Leu	63.8	35.9	0.3	53.9	43.1	3.0	85.8	13.2	1.0	78.6	19.7	1.6	73.1	25.1	1.8
Phe	41.0	57.4	1.6	56.1	23.9	20.0	72.4	17.4	10.2	61.5	25.3	13.2	51.6	29.5	19.0
Tyr	40.9	56.4	2.8	55.7	26.0	18.3	69.9	17.7	12.5	59.1	27.2	13.7	48.4	32.8	18.8
Trp	39.4	52.9	7.7	55.8	26.2	18.0	52.4	17.3	30.3	53.2	30.7	16.1	44.7	32.0	23.3
His	50.0	45.2	4.7	48.2	36.3	15.6	70.6	13.3	16.2	54.3	31.0	14.7	55.6	25.5	18.9
Cys	74.6	20.4	4.9	43.3	35.6	21.1	63.3	9.3	27.4	45.2	31.9	22.9	45.8	25.0	29.2
							Special								
Ser	41.2	22.5	36.3	27.5	34.3	38.2	29.7	8.5	61.7	22.1	27.4	50.5	29.4	17.2	53.4
Asp	77.8	17.6	4.6	35.7	54.8	9.4	54.1	14.5	31.4	32.6	45.2	22.1	48.2	28.9	22.8
Asn	76.7	20.8	2.5	48.2	42.0	9.8	64.9	18.8	16.3	39.1	38.7	22.2	55.1	26.0	18.9
Val	87.3	5.1	7.6	75.5	7.8	16.7	51.0	11.6	37.4	60.6	7.5	31.9	57.4	9.3	33.3
Ile	87.9	7.2	4.8	79.6	8.6	11.8	44.4	16.9	38.7	65.2	9.3	25.5	61.2	12.1	26.7
Thr	73.5	1.3	25.1	52.1	13.7	34.2	18.9	2.4	78.7	24.2	10.2	65.6	27.3	9.0	63.7

[a] All coil residues outside α-helix, 3_{10}-helix, β-bridge, β-sheet, and various turns. [b] For residues in our restricted coil library.

Ser and ordinary amino acids. Thus, its g− rotamer with β-CH_3 at the g+ position has an even higher preference than Ser in coil (66%) and turns (79%). In addition, the t rotamer of Thr is much less preferred in all secondary structures due to its β-CH_3 in the unfavorable g− position.

Asp and Asn (Asx) have $-COO^-$ and $-CONH_2$ substitutions at the β-carbon, respectively. Their conformational preferences are significantly different from those of Glu and Gln (Glx), which have γ-COO^- and γ-$CONH_2$ groups. In the Coil-R, Asx have a lower g+ preference than Glx but higher g− preference. This feature is also true in the turn structures. Asx, on the other hand, has a considerably higher g+ preference (77%) than the Glx (59%) in α-helix. It should also be noticed that for Ser and Asx, their t percentages decrease significantly by more than 1/3 from the entire coil region to our restricted Coil-R. The reason is that their t rotamers are more favored than g+ and g− in interacting with the backbone of successive residues of turnlike conformations (including α-helix capping).[58,60,62]

Both Val and Ile have two β-alkyl groups, and they disfavor the t rotamer. In particular, they have a highest preference for g+ conformation in α-helix and β-sheet structures, so the two β-alkyl groups can take the g+ and t positions to avoid the g− position. On the basis of these results, the amino acids having side chain rotamers can be divided into five types: types I (ordinary), II (Ser), III (Asx), IV (Val and Ile), and V (Thr). This classification is similar to that reported about two decades ago by McGregor et al.,[45] except that we separate Ser and Thr as two types.

Side-Chain-Dependent ϕ, ψ Plots. The above classification of amino acids other than Gly, Ala, and Pro is supported by similarity analysis of their χ_1-dependent Ramachandran plots. As shown in the Supporting Information (Tables S8−S10), except for a few cases, the similarity coefficients (S) among the 11 ordinary amino acids are all above 0.9 for g+, t, and g− rotamers. The same is true for the Asp/Asn and Val/Ile groups. Therefore, we present only 18 Ramachandran plots in Figure 1: Gly, Pro, Ala, and g+, t, g− rotamers for each of the five groups.

As shown in Figure 1, the obtained Ramachandran plots of Gly, Pro, and Ala are in good agreement with previous results. The ϕ, ψ distribution of Ala from Coil-R also agrees well with previous ab initio MP2 calculations of Ala dipeptide with solvent effect.[70] There are five major conformational basins in the ϕ, ψ plot of Ala, as indicated in the top right of Figure 1. Among them, β and P_{II} conformers exist in all non-Gly/Pro types and can be merged into one large basin in some cases (g+ rotamers of Asx, Val and Ile, Thr). The α basin (α′ and α_A, the later for "authentic" α-helical region) also exists in all types, but with quite variable preferences. C_7 and α_L are less favored and can be absent in some types. For Ala and ordinary amino acids (when combining the distributions of g+, t, g− rotamers together), the P_{II} conformer has the highest probability. This agrees with recent experimental studies of unfolded small peptides in aqueous solution.[71−75]

The most significant result from Figure 1 is that for a particular amino acid, the three rotamers (g+, t, g−) can have quite different ϕ, ψ distributions. Table 2 gives the similarity analysis among different rotamers. It clearly indicates that the effect of different side chain rotamers can be greater than different amino acid types. The g+ rotamers of the ordinary amino acids Ser and Asx have quite similar ϕ, ψ distributions (similarity coefficient $S > 0.9$). In contrast, for ordinary amino acids, $S = 0.63$ between g+ and t rotamers, $S = 0.47$ between g+ and g−, and $S = 0.23$ between t and g−. We also obtained the similarity matrix for g+, t, and g− rotamers of individual amino acids (Supporting Information Tables S7−S10). The ordinary amino acids at the same side chain rotamer are quite similar, but for most amino acids, the S values among different side chain rotamers are less than 0.6.

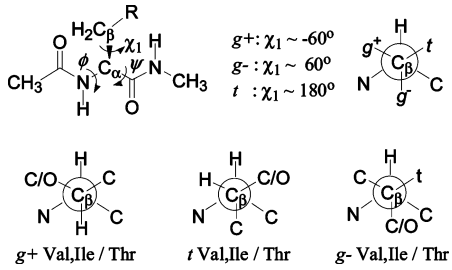
The ordinary amino acids Ser and Asx are all monosubstituents of Ala. When compared to Ala, only g+ rotamers of them have considerable ($S > 0.8$) similarities in ϕ, ψ distributions. On the other hand, their t and g− rotamers are quite different from Ala ($S = 0.34-0.75$), as shown in Figure 1. This implies that the side chains in the g+ rotamer have less interaction with backbone atoms. Since g+ is also the most favored rotamer for ordinary amino acids and Asx, we may guess that the local interactions are mainly repulsive, such as the *syn*-pentane effect previously discussed by Dunbrack and Karplus.[47]

TABLE 2: Similarity Matrix of the ϕ, ψ Plots of 18 Types in Figure 1

					g+					t					g−				
		Gly	Pro	Ala	ord[a]	Ser	Asx	V, I	Thr	ord	Ser	Asx	V, I	Thr	ord	Ser	Asx	V, I	Thr
	Gly	1.00[b]	0.33	0.36	0.35	0.39	0.42	0.10	0.14	0.18	0.26	0.26	0.22	0.21	0.26	0.37	0.26	0.23	0.27
	Pro	0.33	1.00	0.88	0.62	0.78	0.71	0.27	0.44	0.55	0.43	0.31	0.62	0.17	0.31	0.45	0.19	0.29	0.24
	Ala	0.36	0.88	1.00	0.83	0.93	0.86	0.47	0.64	0.75	0.68	0.48	0.79	0.40	0.53	0.65	0.34	0.43	0.41
g+	ord	0.35	0.62	0.83	1.00	0.95	0.91	0.57	0.61	0.63	0.67	0.47	0.67	0.37	0.47	0.71	0.49	0.67	0.70
	Ser	0.39	0.78	0.93	0.95	1.00	0.94	0.49	0.60	0.66	0.64	0.48	0.70	0.34	0.48	0.73	0.44	0.60	0.61
	Asx	0.42	0.71	0.86	0.91	0.94	1.00	0.47	0.56	0.62	0.58	0.53	0.60	0.26	0.42	0.74	0.55	0.51	0.65
	V, I	0.10	0.27	0.47	0.57	0.49	0.47	1.00	0.92	0.82	0.75	0.36	0.46	0.14	0.12	0.15	0.08	0.15	0.13
	Thr	0.14	0.44	0.64	0.61	0.60	0.56	0.92	1.00	0.95	0.76	0.35	0.57	0.17	0.16	0.21	0.09	0.17	0.15
t	ord	0.18	0.55	0.75	0.63	0.66	0.62	0.82	0.95	1.00	0.81	0.45	0.68	0.22	0.23	0.27	0.12	0.19	0.16
	Ser	0.26	0.43	0.68	0.67	0.64	0.58	0.75	0.76	0.81	1.00	0.54	0.78	0.57	0.49	0.40	0.18	0.38	0.27
	Asx	0.26	0.31	0.48	0.47	0.48	0.53	0.36	0.35	0.45	0.54	1.00	0.31	0.34	0.36	0.34	0.27	0.23	0.22
	V, I	0.22	0.62	0.79	0.67	0.70	0.60	0.46	0.57	0.68	0.78	0.31	1.00	0.52	0.55	0.51	0.17	0.42	0.30
	Thr	0.21	0.17	0.40	0.37	0.34	0.26	0.14	0.17	0.22	0.57	0.34	0.52	1.00	0.85	0.55	0.23	0.53	0.34
g−	ord	0.26	0.31	0.53	0.47	0.48	0.42	0.12	0.16	0.23	0.49	0.36	0.55	0.85	1.00	0.75	0.35	0.57	0.46
	Ser	0.37	0.45	0.65	0.71	0.73	0.74	0.15	0.21	0.27	0.40	0.34	0.51	0.55	0.75	1.00	0.65	0.64	0.80
	Asx	0.26	0.19	0.34	0.49	0.44	0.55	0.08	0.09	0.12	0.18	0.27	0.17	0.23	0.35	0.65	1.00	0.43	0.74
	V, I	0.23	0.29	0.43	0.67	0.60	0.51	0.15	0.17	0.19	0.38	0.23	0.42	0.53	0.57	0.64	0.43	1.00	0.77
	Thr	0.27	0.24	0.41	0.70	0.61	0.65	0.13	0.15	0.16	0.27	0.22	0.30	0.34	0.46	0.80	0.74	0.77	1.00

[a] Ordinary amino acids. [b] Underlined values are of $S \geq 0.8$.

CHART 1: Definition of Side Chain χ_1 Rotamers: β-Substituted Ala (upper), and β-Branched Amino Acids (bottom)



As for the β-branched amino acids (Val and Ile, Thr), they have two substitutions on C_β (See Chart 1). Their g+ rotamer means C_β substitutions in both g+ and t positions. Similarly, the t rotamer has t and g− C_β substitutions and the g− has g− and g+ C_β substitutions. Thus, there are considerable similarities ($S > 0.8$) between their g+ rotamer and the t rotamer of ordinary amino acids and between the t rotamer of Thr and the g− rotamer of ordinary amino acids.

Intrinsic Side-Chain Rotamer Preferences. Figure 2 shows the distributions of χ_1 of 16 AAs in four different backbone conformational regions (β, P_{II}, α' and α_A). In every case, χ_1 is distributed in narrow regions near 60°, 185°, and 295° for the g−, t, and g+ rotamers, respectively. The distributions are somewhat more broad for Phe, Tyr, and Trp. Therefore, it is reasonable to simplify χ_1 distributions as the percentages of the three rotamers. The results are summarized in Table 3. For all amino acids, the t rotamer is significantly suppressed when backbone conformation is in α' region, and the g− rotamer is suppressed in the P_{II} region. For Asx and β-branched amino acids, their g− rotamers are especially favored in the α' region but not in the α_A region.

In our coil library, we do not restrict the identity of neighboring residues. Previous studies of coil library and unfolded peptides[11,76−78] indicate that a residue's local conformational preference can be significantly affected by both amino acid type and backbone conformation of the residue's nearest neighbors (NN). It is necessary to see whether the χ_1 preferences obtained from our coil library are biased due to the NN effect. Since Ser, Asx, and β-branched and aromatic amino acids were found to have a larger NN effect than others, a more strictly restricted coil library Coil-R+ was obtained, in which the residues adjacent to any of these mentioned amino acids were excluded. We also obtained a much more strictly restricted coil library (Coil-GA) in which the residues can be adjacent only to Gly or Ala.

As shown in Figure 3, χ_1 preferences of all 17 amino acids from Coil-R+ agree excellently with those from Coil-R, even though the size of Coil-R+ is only 1/4 of Coil-R. There is also a good correlation between Coil-R and Coil-GA. This indicates that the intrinsic side chain conformational preferences can be obtained from our Coil-R without significant influence of the NN effect. To further compare the Coil-R and Coil-GA, the numbers of occurrences of 17 × 3 side chain rotamers in four backbone ϕ, ψ regions (β, P_{II}, α, α') were obtained. Despite small numbers ($n < 10$) for most data points, significant correlations ($r = 0.84-0.98$) can also be obtained (Supporting Information Figure S3).

It is interesting to compare the obtained side chain preferences of certain ϕ, ψ with the preferences for related secondary structures. From Figure 4, it is clear that when the amino acid types with side chain rotamers are combined, good correlations can be obtained for α-helix, β-sheet, and also β-turns. This implies that the rotamer preferences for the secondary structures are largely contributed from a local effect. The g− rotamer of amino acids other than Ser and Thr is found to be avoided in α-helix, which was previously attributed to the steric clash between the side chain and the carbonyl group of the *i*-3 residue.[45] However, without the α-helical structure, the g− rotamer is also strongly disfavored at the same backbone conformation (−64°, −42°).

The local structures in an α-helix and type-I β-turn belong to an α conformational basin. However, they have quite different preferences. In general, the g− rotamers are less disfavored in type-I β-turns than in α-helices, and t rotamers are avoided in the second position of type-I β-turns but have percentages similar to g+ in α-helices. In addition, in the first position of β-turns, the preference is in the order Ala > g− of Ser > g− of Thr, but changes to g− of Thr > g− of Ser > Ala in the second position of the turns. This agrees with previous finding that the χ_1 distribution can change dramatically due to a small change in ϕ, ψ. [48]

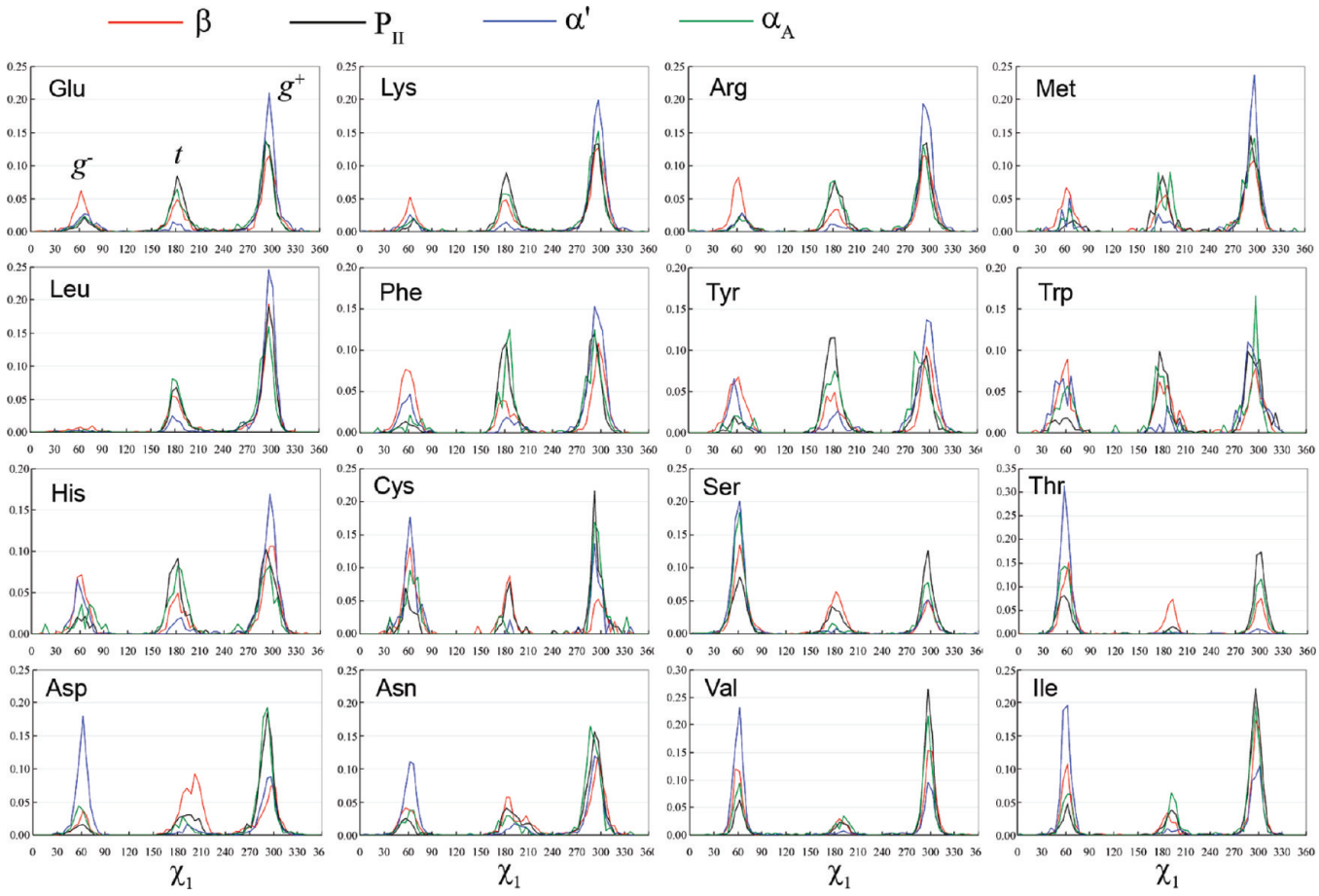


Figure 2. The χ_1 distributions (in degrees) for 16 amino acids in four different ϕ, ψ regions defined in Figure 1: β, P_{II}, α', and α_A. Gln not shown here is very similar to Glu and Lys.

TABLE 3: Backbone-Dependent χ_1 Rotamer Preferences from Coil-R Statistics[a]

	β			P_{II}			α'			α_A		
	g+ %	t %	g− %	g+ %	t %	g− %	g+ %	t %	g− %	g+ %	t %	g− %
E	52	23	25	55	35	9	79	6	15	59	30	11
Q	57	22	21	54	37	9	87	7	6	55	33	12
K	60	22	19	56	38	7	83	6	10	62	29	9
R	52	17	31	56	33	11	83	6	11	53	37	10
M	48	25	27	57	35	8	79	9	12	59	31	10
L	69	26	5	72	28	1	90	9	1	66	33	1
F	43	20	37	50	43	7	72	9	20	50	40	9
Y	41	23	36	41	52	7	68	11	21	50	40	10
W	32	33	35	52	38	11	54	8	37	40	33	26
H	50	21	29	51	40	9	68	8	24	41	40	20
C	22	29	49	58	21	21	38	2	60	54	6	40
S	19	29	52	43	20	37	22	4	75	29	6	65
D	36	52	12	73	20	7	38	7	56	73	10	16
N	51	28	21	66	24	10	49	9	42	69	16	15
V	51	11	38	72	10	18	31	3	66	59	14	27
I	57	12	31	73	14	13	36	4	59	60	18	22
T	25	23	52	60	6	34	4	2	94	40	2	58

[a] The definitions of the β, P_{II}, α', α_A regions are illustrated in Figure 1.

Comparison with MD Simulations. The side chain rotamer distributions from molecular dynamics simulations are given in Table 4. First, results of dipeptide and tetrapeptide models using the same OPLS-AA/L force field are compared in Figure 5a. The two peptide models have very similar side chain rotamer preferences, as indicated by the correlation coefficient of $r = 0.99$. This indicates that the dipeptide model may be good enough for studying the intrinsic side chain conformational features. We also tested longer octapeptide models of various sequences. In most cases, the neighboring residues do not significantly affect the obtained rotamer distributions (Supporting Information Table S17).

As shown in Figure 5b, the MD simulations with the OPLS-AA/L do not reproduce the observed g+/t/g− distributions from the coil library Coil-R well. Several amino acids (Glu (E), Asp (D), Asn (N), Met (M), and Thr (T)) have their t rotamer

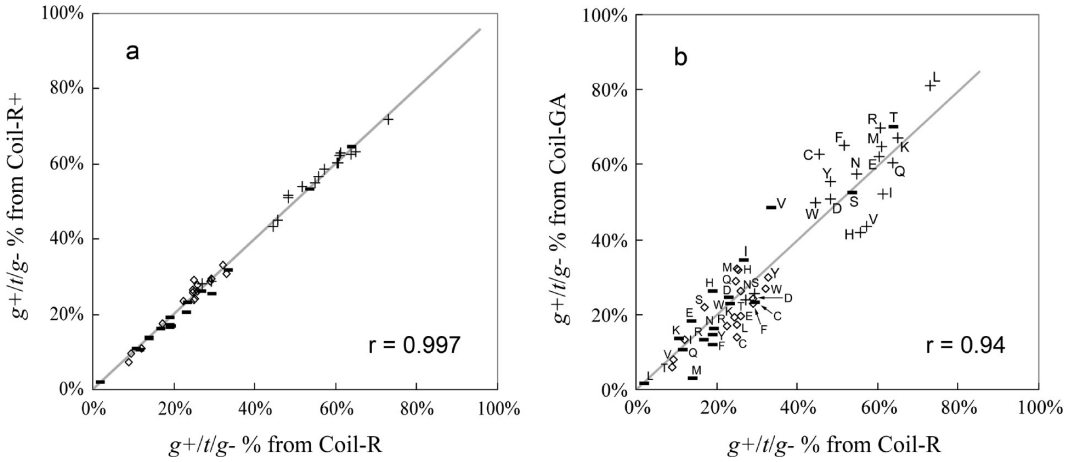


Figure 3. Correlations of the side chain rotamer preferences (in percentages) obtained from Coil-R with (a) Coil-R+. The neighboring residues cannot be Ser, Asx, β-branched, or aromatic amino acid. (b) Coil-GA. The neighboring residues can be only Gly or Ala. Plus sign (+) is for g+ rotamers; horizontal bar (−), for g− rotamers; and open diamond, for t rotamers. The same symbols for g+, t, g− are also applied in other figures.

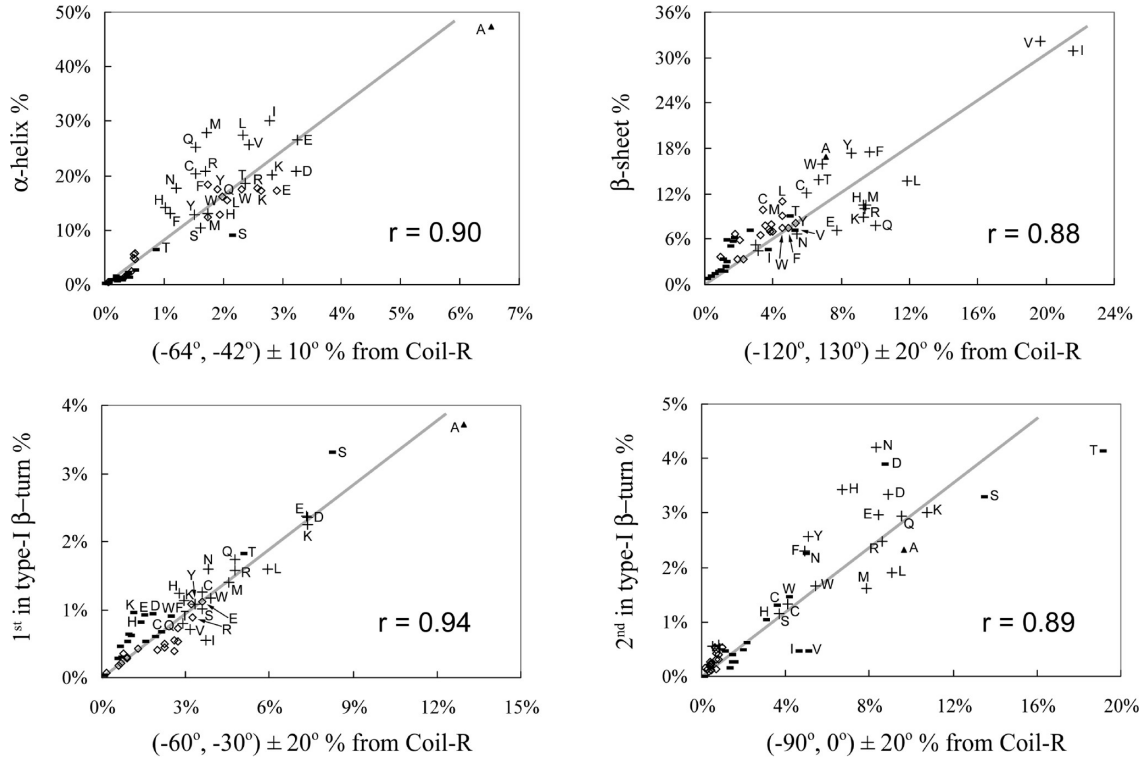


Figure 4. Correlations between the percentages of certain secondary structures and the local preferences of corresponding ϕ, ψ, for all non-Gly/Pro amino acids. We use eq 1 (see Methods) to obtain the local conformational preferences, and the ±10° and ±20° indicate the σ values used for the sampling.

preferences overestimated, and Cys (C), Lys (K), Val (V), Trp (W), Phe (F), and Tyr (Y) have their g− rotamer preferences overestimated.

The AMBER-FF03 force field was also used to carry out simulations of the dipeptide models. It gives distributions quite different from those of the OPLS-AA/L force field. As in Figure 6a, there is no correlation in side chain rotamer distributions between the OPLS-AA/L and AMBER-FF03 force fields. Furthermore, the statistical rotamer preferences also cannot be reproduced by the AMBER-AA03 simulations as shown in Figure 6b. Comparing the AMBER-FF03 simulations with the Coil-R results, the amino acids that have significant different χ_1 preferences include Asp (D), Asn (N), Ser (S), Cys (C), Leu (L), Ile (I), Val (V), Phe (F), Tyr (Y), Trp (W), and Lys (K).

Most amino acids in the Coil-R library have intrinsic rotamer preferences of g+ in the range of 45−65%, t in the range of 23−33%, and g− in the range of 11−30%. However, both OPLS-AA/L and AMBER-FF03 force fields give much more scattered distributions ranging from 0 to 90%. This may imply that the molecular force field may exaggerate the interactions between side chain and local backbone atoms. The different protonation states of His are treated differently with molecular force fields, which can hardly be distinguished from crystal structures. From Table 4, two forms of neutral His give out very different rotamer distributions using either the OPLS-AA/L or AMBER-FF03 force field, indicating significant interactions between side chain N_δ or N_ε groups and the backbone in these force fields. However, the observed rotamer distribution of His

TABLE 4: Percentages of Three Side Chain Rotamers of Each Amino Acid, Obtained from MD Simulations

	OPLS-AA/L						AMBER-FF03		
	Ac−X−NHMe[a]			Ac−A−X−A−NHMe[b]			Ac−X−NHMe		
	g+ %	t %	g− %	g+ %	t %	g− %	g+ %	t %	g− %
Glu (E)	6	93	1	8	91	1	49	28	23
Gln (Q)	64	21	15	63	20	17	48	22	30
Lys (K)	21	17	63	18	24	58	43	30	27
Arg (R)	79	19	2	67	30	3	52	35	13
Met (M)	23	70	7	22	72	6	64	18	18
Leu (L)	84	13	4	82	13	5	48	47	5
Phe (F)	29	23	48	17	22	60	27	40	34
Tyr (Y)	25	24	52	18	21	61	17	60	23
Trp (W)	11	35	54	9	40	51	16	37	47
His^{δ} (H^{δ})	81	1	17	72	2	26	40	31	30
His^{ε} (H^{ε})	0	98	1	0	99	1	8	63	30
Cys (C)	2	17	81	3	22	75	63	15	22
Ser (S)	12	15	73	12	15	72	57	4	39
Asp (D)	2	90	8	3	90	7	3	92	5
Asn (N)	7	58	35	6	55	39	2	80	17
Val (V)	27	7	67	26	7	67	32	30	38
Ile (I)	70	5	24	74	3	23	9	28	63
Thr (T)	1	41	58	2	31	67	36	6	58

[a] Dipeptide model of amino acid X. [b] Tetrapeptide model with A = Ala adjacent to X. His^{δ} has hydrogen on N_{δ} but not N_{ε}; His^{ε} has hydrogen on N_{ε} but not N_{δ}.

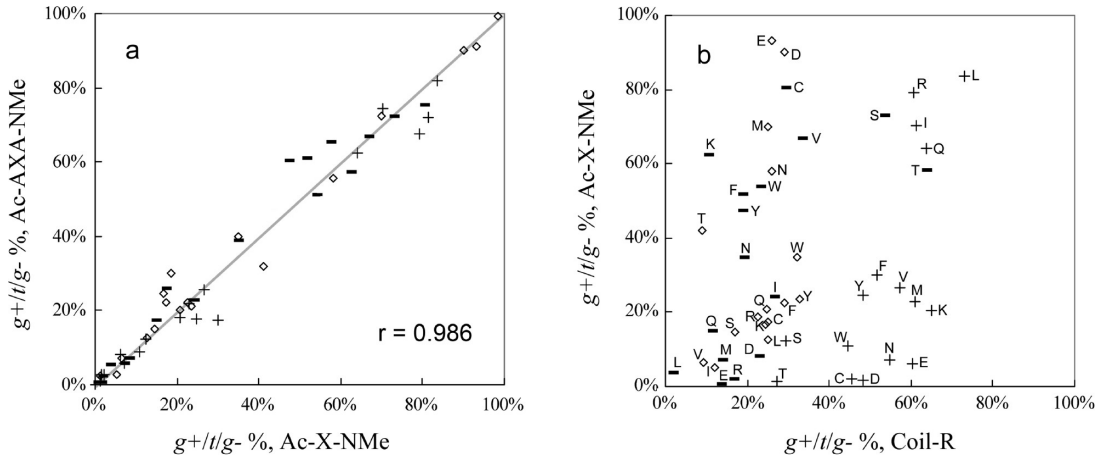


Figure 5. The rotamer preferences obtained from OPLS-AA/L simulations. (a) The correlation between dipeptides (Ac−X−NHMe) results and tetrapeptides (Ac−AXA−NHMe) results; (b) dipeptides simulations compared with our restricted coil library statistics. The symbols have the same meanings as in Figure 3.

is quite similar to nonpolar Phe, and they also have very similar ϕ, ψ distributions.[24] The actual interactions between side chain polar groups and the backbone might be much smaller than those given by the force fields.

Our studies indicate a potential defect in current molecular force fields and strategy for force field development. It has been found that the MD simulations of Gly and Ala dipeptides using various MM force fields cannot well reproduce the coil library ϕ, ψ distributions, which can be better reproduced by QM/MM simulations.[79] Here, the samplings of χ_1 for other amino acids are also questionable. Actually, for nonpolar side chains of Val, Ile, and Leu, recent gas-phase QM calculations of dipeptide models can reproduce the observed rotamer distributions better than the MM method.[80]

Similar to Figure 1, the χ_1-dependent Ramachandran plots of all amino acids from both OPLS-AA/L (Supporting Information Figure S4) and AMBER-FF03 (Supporting Information Figure S5) simulations are obtained. The plots of four of the amino acids obtained by the OPLS-AA/L simulations are shown in Figure 7. As for the backbone ϕ, ψ distributions, the MD results do capture some important features of Coil-R observations, except for an additional basin of the region $\phi < -120°$ and $\psi < 80°$. Some interrelationship between the side chain and backbone conformations can be reproduced by both force fields, such as the different positions of β, P_{II}, and α basins for different rotamers, the relatively disfavored α' region for t rotamers, and the disfavored P_{II} for the g− rotamers. However, both force fields fail to reproduce the highly populated C_7 and α_L basins (Figure 1) for the t rotamer of Asx.

Some differences in the Ramachandran plots between the coil library and simulations are also obvious, quite different from those shown in Figure 1. The statistics of the coil-R library indicate that for a given rotamer, for example, g+, different amino acids usually have quite similar ϕ, ψ distributions (Supporting Information Table S8, Figure S2), but they can have very different χ_1 preferences. When comparing Table 3 with Tables 5 and 6 of the backbone-dependent rotamer preferences, it is clear that the g+/t/g− percent differences between PDB statistics and MD simulations are consistent among different backbone conformations. From OPLS-AA/L simulations, Glu/Met/Asp/Asn give a much higher t percent and Cys/Lys/Phe/Trp/Val give much higher g− percent consistently for all four

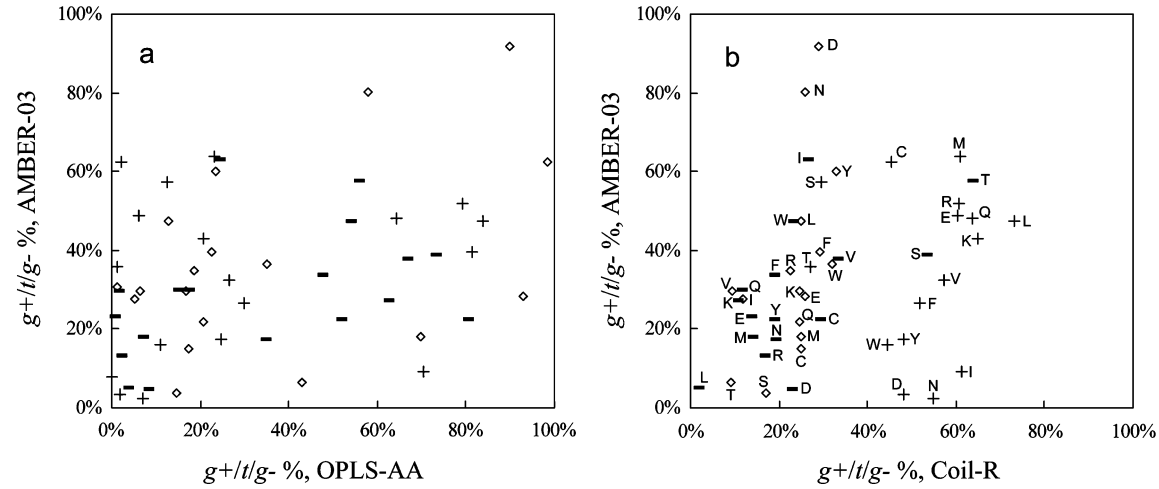


Figure 6. The rotamer preferences obtained from AMBER-FF03 simulations of dipeptides (a) compared with the results from OPLS-AA/L simulations and (b) compared with the results from our restricted coil library. The symbols have the same meanings as in Figure 2.

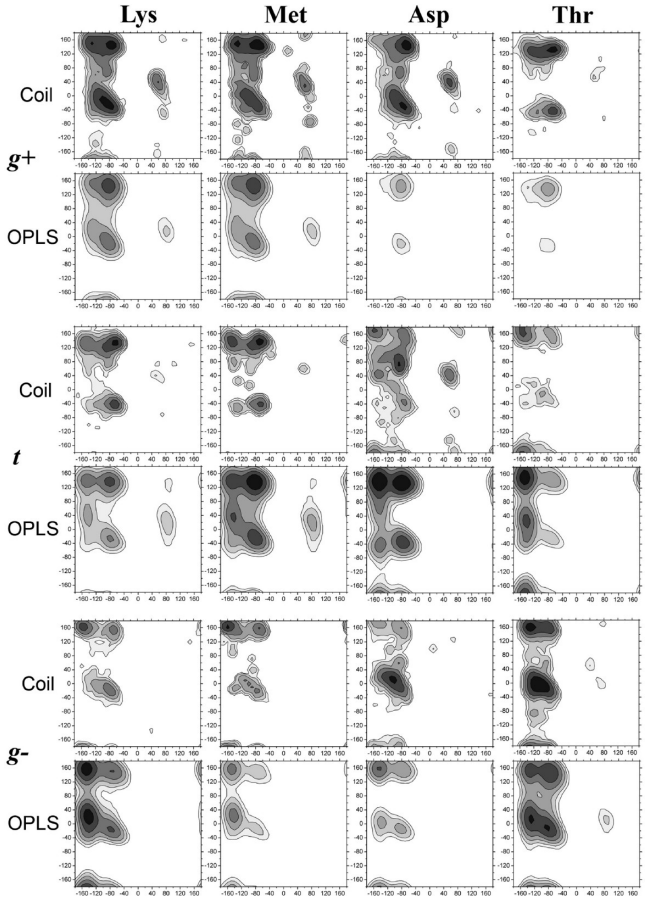


Figure 7. Comparison between χ_1-dependent ϕ, ψ distributions from the Coil-R coil library and those from dipeptide OPLS-AA/L simulations. Only four amino acids are shown here; others are given in the Supporting Information. The density scale used is the same as Figure 1.

TABLE 5: Backbone-Dependent χ_1 Rotamer Preferences from Dipeptide Simulations with OPLS-AA/L Force Field

	β			P_{II}			α'			α		
	g+	t	g−	g+	t	g−	g+	t	g−	g+	t	g−
E	5	94	1	7	93	0	8	92	0	8	92	0
Q	50	25	25	73	22	5	61	11	27	67	23	10
K	12	14	73	43	30	27	14	7	80	29	22	49
R	72	24	4	82	18	0	85	12	3	78	22	1
M	23	66	11	25	73	2	24	61	15	19	78	3
L	75	18	7	87	12	1	83	7	10	86	13	1
F	16	26	58	51	38	11	21	5	74	46	21	33
Y	14	26	60	43	45	12	18	5	78	40	22	38
W	7	29	64	19	64	17	11	14	75	17	48	35
H^{δ}	67	2	31	89	1	10	74	1	25	86	1	13
H^{ϵ}	0	98	2	0	99	1	0	96	3	0	99	1
C	1	21	78	6	40	53	1	6	93	4	16	80
S	7	20	73	29	22	49	7	5	87	19	6	75
D	1	89	10	3	92	5	2	78	20	2	91	6
N	5	54	41	9	82	10	7	23	70	10	68	23
V	23	12	64	45	4	51	7	4	89	21	3	75
I	66	11	23	88	2	11	24	9	67	67	5	28
T	1	73	27	4	13	83	0	24	76	2	7	91

TABLE 6: Backbone-Dependent χ_1 Rotamer Preferences from Dipeptide Simulations with AMBER-FF03 Force Field

	β			P_{II}			α'			α		
	g+	t	g−	g+	t	g−	g+	t	g−	g+	t	g−
E	38	33	29	62	29	9	56	15	29	55	26	19
Q	39	20	40	61	27	12	59	17	23	52	35	14
K	41	18	41	50	25	25	44	18	38	36	39	25
R	46	33	21	56	34	9	61	24	15	47	44	9
M	51	23	26	73	18	10	81	8	12	76	17	8
L	46	39	14	56	41	3	56	41	3	36	63	1
F	18	35	47	29	51	20	37	25	38	27	51	22
Y	12	41	47	17	66	17	25	43	32	13	72	15
W	9	33	57	15	51	34	22	21	56	16	39	45
H^{δ}	27	29	44	46	42	12	51	17	32	39	41	20
H^{ϵ}	4	68	27	7	78	16	12	41	46	8	68	23
C	37	30	32	65	17	18	66	6	28	67	11	22
S	43	16	41	76	6	18	55	1	44	59	1	40
D	1	95	4	2	95	2	6	87	7	3	93	4
N	3	77	20	3	86	10	3	64	34	1	87	12
V	24	37	39	61	16	23	12	18	70	30	34	37
I	11	27	61	29	14	57	4	21	75	11	37	51
T	28	21	51	49	4	48	14	2	84	36	1	63

ϕ, ψ regions. A similar phenomenon is also found from AMBER-FF03 simulations, such as the underestimated percent of t for Ser. There may be inherent problems for force fields to sample χ_1 angle, regardless of ϕ, ψ.

More studies are still needed to understand why there are significant differences between observations and simulations. The commonly used force fields for biomolecular simulations are carefully parametrized to reproduce the thermodynamics and other experimental properties of liquids.[81] This indicates that their intermolecular/nonlocal nonbonded parameters are quite reliable. However, the local nonbonded interactions may have quite different behaviors, such as the 1−4 van der Waals interactions and maybe also the 1−5 nonbonded interactions. In addition, the desolvation effect of the side chain on the backbone polar atoms can also contribute significantly to the

local conformational preferences of amino acids.[10,82,83] There is still room for further improving force field parameters of these local interactions.

Conclusions

In this work, we constructed a restricted coil library to study the intrinsic local conformational behaviors of residues in proteins. This coil library gives conformational features that correlate well with those in more strictly constrained coil libraries in which the neighboring effect is eliminated. Therefore, this library appears to allow us to derive good intrinsic local conformational features of amino acids.

It is found that the χ_1 dihedral angle distributes in a narrow range for every rotamer of every amino acid. The Ramachandran plots of the three rotamers of a given amino acid can be significantly different. Therefore, it is important to incorporate this side chain rotamer distribution in future protein force field development.

Molecular dynamics simulations using the widely applied OPLS-AA/L and AMBER-FF03 force fields result in side chain rotamer distributions that are quite different from those of the statistics of the coil library. The problem is most serious for those residues with polar side chains or with aromatic rings, suggesting the need for improvement of the force fields.

Acknowledgment. We are grateful to the Research Grants Council of Hong Kong (CA06/07.SC05, 663509) for financial support of this research.

Supporting Information Available: Actual statistical counts for Table 1, local conformational preferences of the 54 types (20 amino acids with three χ_1 states) from both Coil-R and Coil-GA coil libraries, the 54 × 54 similarity matrix and divergence matrix, and the results from MD simulations of longer octapeptides. The original data of the 54 ϕ, ψ plots will be provided upon request. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13*, 211–222.
(2) Levitt, M. *Biochemistry* **1978**, *17*, 4277–4285.
(3) Niefind, K.; Schomburg, D. *J. Mol. Biol.* **1991**, *219*, 481–497.
(4) Rooman, M. J.; Kocher, J.-P. A.; Wodak, S. J. *J. Mol. Biol.* **1991**, *221*, 961–979.
(5) Kang, H. S.; Kurochkina, N. A.; Lee, B. *J. Mol. Biol.* **1993**, *229*, 448–460.
(6) Muñoz, V.; Serrano, L. *Proteins* **1994**, *20*, 301–311.
(7) Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. *Nat. Struct. Biol.* **1995**, *2*, 596–603.
(8) Walther, D.; Cohen, F. E. *Acta Crystallogr., D* **1999**, *55*, 506–517.
(9) Novmöller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr., D* **2002**, *58*, 768–776.
(10) Avbelj, F.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5742–5747.
(11) Jha, A. K.; Colubri, A.; Zaman, M. H.; Koide, S.; Sosnick, T. R.; Freed, K. F. *Biochemistry* **2005**, *44*, 9691–9702.
(12) Sueki, M.; Lee, S.; Powers, S. P.; Denton, J. B.; Konishi, Y.; Scheraga, H. A. *Macromolecules* **1984**, *17*, 148–155.
(13) O'Neil, K.; DeGrado, W. *Science* **1990**, *250*, 646–651.
(14) Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. *Protein Sci.* **1994**, *3*, 843–852.
(15) Kim, C. A.; Berg, J. M. *Nature* **1993**, *362*, 267–270.
(16) Smith, C. K.; Withka, J. M.; Regan, L. *Biochemistry* **1994**, *33*, 5510–5517.
(17) Serrano, L. *J. Mol. Biol.* **1995**, *254*, 322–333.
(18) Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. *J. Phys. Chem.* **1996**, *100*, 2661–2666.
(19) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494–506.
(20) Penkett, C. J.; Redfield, C.; Dodd, I.; Hubbard, J.; McBay, D. L.; Mossakowska, D. E.; Smith, R. A. G.; Dobson, C. M.; Smith, L. J. *J. Mol. Biol.* **1997**, *274*, 152–159.
(21) Peti, W.; Hennig, M.; Smith, L. J.; Schwalbe, H. *J. Am. Chem. Soc.* **2000**, *122*, 12017–12018.
(22) Tosatto, S. C.; Battistutta, R. *BMC Bioinformatics* **2007**, *8*, 155.
(23) Malkov, S. N.; Živković, M. V.; Beljanski, M. V.; Hall, M. B.; Zarić, S. D. *J. Mol. Model.* **2008**, *14*, 769–775.
(24) Dahl, D. B.; Bohannan, Z.; Mo, Q.; Vannucci, M.; Tsai, J. *J. Mol. Biol.* **2008**, *378*, 749–758.
(25) Beck, D. A. C.; Alonso, D. O. V.; Inoyama, D.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 12259–12264.
(26) Feig, M. *J. Chem. Theory Comput.* **2008**, *4*, 1555–1564.
(27) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95–99.
(28) Ramachandran, G. N.; Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 283–438.
(29) Street, A. G.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9074–9076.
(30) Avbelj, F.; Grdadolnik, S. G.; Grdadolnik, J.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1272–1277.
(31) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099–13104.
(32) Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002–17007.
(33) Betancourt, M. R. *J. Phys. Chem. B* **2008**, *112*, 5058–5069.
(34) Fujitsuka, Y.; Chikenji, G.; Takada, S. *Proteins* **2006**, *62*, 381–398.
(35) 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.
(36) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908–5920.
(37) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, J. D.; Evanseck, M. J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(38) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
(39) 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. A. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
(40) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III *J. Am. Chem. Soc.* **2004**, *126*, 698–699.
(41) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III *J. Comput. Chem.* **2004**, *25*, 1400–1415.
(42) MacKerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584–1604.
(43) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.
(44) Fujitani, H.; Matsuura, A.; Sakai, S.; Sato, H.; Tanida, Y. *J. Chem. Theory Comput.* **2009**, *5*, 1155–1165.
(45) McGregor, M. J.; Islam, S. A.; Sternberg, M. J. E. *J. Mol. Biol.* **1987**, *198*, 295–310.
(46) Dunbrack, R. L., Jr.; Karplus, M. *J. Mol. Biol.* **1993**, *230*, 543–574.
(47) Dunbrack, R. L., Jr.; Karplus, M. *Nat. Struct. Biol.* **1994**, *1*, 334–340.
(48) Dunbrack, R. L., Jr.; Cohen, F. E. *Protein Sci.* **1997**, *6*, 1661–1681.
(49) Chakrabarti, P.; Pal, D. *Protein Eng.* **1998**, *11*, 631–647.
(50) Chakrabarti, P.; Pal, D. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1–102.
(51) West, N. J.; Smith, L. J. *J. Mol. Biol.* **1998**, *280*, 867–877.
(52) Hennig, M.; Bermel, W.; Spencer, A.; Dobson, C. M.; Smith, L. J.; Schwalbe, H. *J. Mol. Biol.* **1999**, *288*, 705–723.
(53) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(54) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.
(55) O'Connell, T. M.; Wang, L.; Tropsha, A.; Hermans, J. *Proteins* **1999**, *36*, 407–418.
(56) Lovell, S. C.; Davis, I. W.; Arendall, W., III; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins* **2003**, *50*, 437–450.
(57) Fitzkee, N. C.; Fleming, P. J.; Rose, G. D. *Proteins* **2005**, *58*, 852–854.
(58) Vijayakumar, M.; Qian, H.; Zhou, H.-X. *Proteins* **1999**, *34*, 497–507.

(59) Wan, W.-Y.; Milner-White, E. J. *J. Mol. Biol.* **1999**, *286*, 1633–1649.
(60) Penel, S.; Hughes, E.; Doig, A. J. *J. Mol. Biol.* **1999**, *287*, 127–143.
(61) Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *96*, 14258–14263.
(62) Jiang, F.; Wu, Y.-D. *Chem. J. Chin. U.* **2008**, *29*, 2371–2376.
(63) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
(64) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
(65) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(66) Case, D. A.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
(67) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*, University of California: San Francisco, 2008.
(68) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.
(69) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
(70) Wang, Z.-X.; Duan, Y. *J. Comput. Chem.* **2004**, *25*, 1699–1716.
(71) Poon, C.; Samulski, E. T.; Weise, C. F.; Weisshaar, J. C. *J. Am. Chem. Soc.* **2000**, *122*, 5642–5643.
(72) Weise, C. F.; Weisshaar, J. C. *J. Phys. Chem. B* **2003**, *107*, 3265–3277.
(73) Eker, F.; Cao, X.; Nafie, L.; Huang, Q.; Schweitzer-Stenner, R. *J. Phys. Chem. B* **2003**, *107*, 358–365.
(74) Eker, F.; Griebenow, K.; Cao, X.; Nafie, L.; Schweitzer-Stenner, R. *Biochemistry* **2004**, *43*, 613–621.
(75) Shi, Z.; Chen, K.; Liu, Z.; Sosnick, T. R.; Kallenbach, N. R. *Proteins* **2006**, *63*, 312–321.
(76) Pappu, R. V.; Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12565–12570.
(77) Zaman, M. H.; Shen, M. Y.; Berry, R. S.; Freed, K. F.; Sosnick, T. R. *J. Mol. Biol.* **2003**, *331*, 693–711.
(78) Keskin, O.; Yuret, D.; Gursoy, A.; Turkay, M.; Erman, B. *Proteins* **2004**, *55*, 992–998.
(79) Hu, H.; Elstner, M.; Hermans, J. *Proteins* **2003**, *50*, 451–463.
(80) Renfrew, P. D.; Butterfoss, G. L.; Kuhlman, B. *Proteins* **2008**, *71*, 1637–1646.
(81) Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665–6670.
(82) Avbelj, F. *J. Mol. Biol.* **2000**, *300*, 1335–1359.
(83) Avbelj, F.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1309–1313.

JP909088E